Processos de Ordem Infinita Estocasticamente Perturbados

Lucas Moreira

Tese apresentada
ao
Instituto de Matemática, Estatística e Computação Científica
da
Universidade Estadual de Campinas
para
obtenção do título
de
Doutor em Estatística

Programa: Estatística

Orientadora: Nancy Lopes Garcia

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da FAPESP

Campinas, 16 fevereiro de 2012

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA

Lucas Moreira

Processos de Ordem Infinita Estocasticamente Perturbados

TESE DE DOUTORADO
APRESENTADA AO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA DA UNICAMP PARA OBTENÇÃO DO TÍTULO DE DOUTOR EM ESTATÍSTICA

ORIENTADORA: NANCY LOPES GARCIA

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO LUCAS MOREIRA, E ORIENTADA PELA PROFA. DRA. NANCY LOPES GARCIA

Nancy L. Garcia

Nancy Lopes Garcia
Orientadora

CAMPINAS, 2012

FICHA CATALOGRÁFICA ELABORADA POR ANA REGINA MACHADO – CRB8/5467
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA – UNICAMP

M813p	Moreira, Lucas, 1984- Processos de ordem infinita estocasticamente perturbados / Lucas Moreira. – Campinas, SP : [s.n.], 2012. Orientador: Nancy Lopes Garcia. Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica. 1. Processo estocástico. 2. Estatística robusta. I. Garcia, Nancy Lopes, 1964-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em inglês: Processes of infinite order stochastically perturbed
Palavras-chave em inglês:
Stochastic processes
Robust statistics
Área de concentração: Estatística
Titulação: Doutor em Estatística
Banca examinadora:
Nancy Lopes Garcia [Orientador]
Jesus Enrique Garcia
Alexsandro Giacomo Grimbert Gallo
Florencia Graciela Leonardi
Miguel Natalio Abadi
Data da defesa: 16-02-2012
Programa de Pós-Graduação: Estatística

Tese de Doutorado defendida em 16 de fevereiro de 2012 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

____

Prof(a). Dr(a). NANCY LOPES GARCIA

____

Prof(a). Dr(a). JESUS ENRIQUE GARCIA

____

Prof(a). Dr(a). ALEXSANDRO GIACOMO GRIMBERT GALLO

____

Prof(a). Dr(a). FLORENCIA GRACIELA LEONARDI

____

Prof(a). Dr(a). MIGUEL NATALIO ABADI

Dedico este trabalho à minha esposa Luciene, pela compreensão dos momentos ausentes e pela colaboração constante nas profícuas opiniões.

Aos meus pais, tradução mais simples e sólida do amor puro e incondicional.

Agradecimentos

Á Deus, por ter guardado meus passos e por ter me dado sabedoria necessária em minhas decisões.

À Profa. Dra. Nancy Lopes Garcia, pela participação ativa e direta neste passo gigantesco a caminho do nosso engrandecimento profissional, me ensinando a conciliar os momentos de austeridade e ternura, fatores primordiais na realização de um trabalho científico, meu eterno agradecimento.

Ao Prof. Dr. Serguei Popov e sua esposa, Profa. Dra. Marina Vachkovskaia por confiarem em minha capacidade acadêmica e pelos ricos ensinamentos.

À Profa. Dra. Denise Duarte pela amizade e apoio em toda a minha carreira acadêmca.

À Prof. Dr. Miguel Abadi pela amizade, respeito e pelos ricos ensinamentos.

Aos meus amigos Rodrigo Lambert, Jozé Evangelista, Renan Chaves e Márcio Valk, pelo companheirismo e principalmente pelo apoio.

Ao Instituto de Matemática, Estatística e Computação Científica onde tive a oportunidade de dar um importante rumo ao crescimento científico e profissional.

Aos professores do Departamento de Estatística da Unicamp, pela disponibilidade e atenção.

Aos demais colegas do IMECC e às demais pessoas que me ajudaram durante esta jornada.

À FAPESP, pelo suporte financeiro.

Resumo

Inspirados em Collet, Galves e Leonardi (2008), a motivação original deste texto é responder a seguinte questão: é possível recuperar a árvore de contextos de uma cadeia de alcance variável através de uma amostra perturbada da cadeia? Inicialmente, consideramos cadeias binárias de ordem infinita nas quais um dos símbolos pode ser modificado com uma probabilidade pequena e fixada. Provamos que as probabilidades de transição da cadeia perturbada estão uniformemente próximas das probabilidades de transição correspondentes da cadeia original se a probabilidade de contaminação é suficientemente pequena. Por meio deste resultado, fomos capazes de responder afirmativamente à pergunta inicial deste trabalho, ou seja, é possível recuperar a árvore de contextos do processo original mesmo utilizando uma amostra contamina no procedimento de estimação. Com isso, mostramos que o estimador da árvore de contextos utilizado é robusto. Em seguida, consideramos o seguinte modelo: dadas duas cadeias de alcance variável, tomando valores num mesmo alfabeto finito, a cada instante do tempo, o novo processo escolhe aleatoriamente um dos dois processos originais com uma probabilidade grande e fixa. A cadeia obtida dessa maneira pode então ser vista como uma perturbação estocástica da cadeia que está sendo escolhida com probabilidade maior. Para esse modelo, obtivemos resultados semelhantes aos obtidos para o modelo inicial.

Palavras-chave: Processo estocástico, Estatística robusta.

Abstract

Inspired by Collet, Galves and Leonardi (2008), the original motivation of this paper is to answer the following question: Is it possible to recover the context tree of a length variable chain range through a disturbed sample of chain? Initially consider binary chains of infinite order in which one of the symbols can be modified with a small and fixed probability. We prove that the transition probabilities of the perturbed chain are uniformly close to the corresponding transition probabilities of the original chain if the probability of contamination is small enough. Through this result, we were able to answer affirmatively to the initial question of this work, i.e., it is possible to recover the context tree of the original process using a sample contaminates the estimation procedure. With this, we show that the estimator of the context tree used is robust. Next, consider the following model: given two length variable chains, taking values in the same finite alphabet, at each instant of time, the new process randomly chooses one of the two processes with a large and fixed probability . The chain obtained with greater probability can be seen as a stochastic disturbance of the original chain. For this model, we obtained similar results to the those obtained for the initial model.

Keywords: Stochastic process, Robust statistics.

Sumário

Lista de Figuras

Capítulo 1

Introdução

Inspirados em Collet, Galves e Leonardi (2008), a motivação original deste texto é responder a seguinte questão: é possível recuperar a árvore de contextos de uma cadeia de alcance variável através de uma amostra perturbada da cadeia? Recordamos que em uma cadeia de alcance variável a probabilidade condicional do próximo símbolo, dado o passado, depende de uma porção do passado cujo comprimento é uma função do próprio passado. Em Rissanen (1983) foi introduzida esta classe de modelos, denominada *fontes de memória finita* ou *máquinas de árvores*. Na literatura estatística recente, estes modelos tornaram-se populares e são chamadas *cadeias de alcance variável*.

A extensão de um modelo com memória variável para uma situação não-Markoviana, em que os contextos são ainda finitos, porém com comprimento ilimitado, ocorre naturalmente. Com a leitura de Galves e Löcherbach (2008) é possível fazer um levantamento recente acerca do tema. Assim, podemos considerar não apenas cadeias de alcance variável perturbadas aleatoriamente, mas também cadeias estocásticas de ordem infinita aleatoriamente perturbadas.

Para tornar a apresentação de nossos resultados mais clara, lembramos o conceito de *contexto*. Rissanen utilizou a palavra contexto para designar o sufixo minimal de uma sequência de símbolos passados, que é suficiente para definir a probabilidade do próximo símbolo. Ele também notou que o conjunto de todos os contextos satisfaz a propriedade do sufixo, que significa que nenhum contexto é sufixo próprio de outro contexto. Esta propriedade permite representar o conjunto de todos os contextos como o conjunto de folhas de uma árvore enraizada e rotulada, chamada de *árvore de contextos* do processo. Com essa representação, a cadeia pode ser descrita pela árvore de todos os contextos e por uma família associada de probabilidades de transição sobre o conjunto de símbolos. Dado um contexto, sua medida de probabilidade associada fornece a probabilidade do próximo símbolo dado qualquer passado possuindo este contexto como sufixo.

Em seu artigo original, Rissanen não só introduziu a classe de modelos de memória variável como também propôs um algoritmo para estimar a árvore de contextos, o chamado algoritmo Contexto. Desde então, várias versões do algoritmo Contexto surgiram na literatura. Em todas as versões a decisão de podar um ramo é tomada considerando uma função *custo*. Um ramo é podado se a função custo assume um valor menor do que um determinado valor limite. A árvore de contextos estimada é a menor árvore satisfazendo esta condição.

Neste texto, utilizamos uma versão do algoritmo Contexto introduzido em a Galves, Maume-Deschamps e Schmitt (2006) para árvores finitas e estendido para árvores ilimitadas em Galves e Leonardi (2008). Nesta versão, a decisão de podar um ramo é tomada considerando a diferença entre as probabilidades condicionais estimadas do ramo original e as correspondentes do ramo podado, por meio de um valor limite adequado. Através de desigualdades exponenciais para as probabilidades de transição estimadas associadas aos contextos candidatos, esses artigos mostram não só a consistência desta versão do algoritmo Contexto, mas também proporcionam uma cota superior exponencial para a taxa de convergência.

Em nosso primeiro modelo, consideramos cadeias binárias de ordem infinita nas quais um dos símbolos pode ser modificado com uma probabilidade pequena e fixada. Nosso primeiro resultado mostra que as probabilidades condicionais da cadeia de memória variável original e as do processo perturbado estão uniformemente próximas se, a probabilidade de contaminação é pequena o suficiente.

Utilizando o estimador da árvore de contextos apresentado em Galves e Leonardi (2008) e nosso primeiro resultado fomos capazes de recuperar a árvore de contextos do processo original através de uma amostra contaminada, respondendo afirmativamente à pergunta original deste trabalho. Este resultado também nos diz que o estimador da árvore de contextos utilizado é robusto, pois, mesmo tendo como base uma amostra contaminada este ainda consegue recuperar a árvore de contextos do cadeia original.

Em seguida, consideramos o seguinte modelo: dadas duas cadeias de alcance variável, tomando valores num mesmo alfabeto finito, a cada instante do tempo, o novo processo escolhe aleatoriamente um dos dois processos originais com uma probabilidade grande e fixa. A cadeia obtida dessa maneira pode então ser vista como uma pertubação estocástica da cadeia que está sendo escolhida com probabilidade maior. Para esse modelo, obtivemos resultados semelhantes aos do problema inicial.

Nosso trabalho está organizado da seguinte forma: no Capítulo 2 apresentamos as definições básicas e nossos resultados principais e o Capítulo 3 é dedicado à prova desses resultados. No Capítulo 4 comparamos os resultados obtidos nesse trabalho com os correspondentes apresentados em Collet, Galves e Leonardi (2008).

Capítulo 2

Processos Estocasticamente Perturbados

2.1 Definições

Seja $\mathcal{A}$ o alfabeto $\{0, 1, \ldots, N-1\}$, com tamanho $|\mathcal{A}| = N$. Dados dois inteiros $m \leq n$ denotamos por ω_m^n a sequência de símbolos $\omega_m, \ldots, \omega_n$ de $\mathcal{A}$ e $\mathcal{A}_m^n$ o conjunto de tais sequências. O comprimento da sequência ω_m^n é denotado por $l\left(\omega_m^n\right)$ e é definido por $l\left(\omega_m^n\right) = n - m + 1$. Qualquer sequência ω_m^n com $m > n$ representa uma sequência vazia. A mesma notação é estendida para o caso $m = -\infty$.

Dadas duas sequências ω e v, com $l(\omega) < \infty$, denotamos por $v\omega$ a sequência de comprimento $l(u) + l(\omega)$ obtida pela concatenação das duas sequências. Dizemos que uma sequência s é um *sufixo* da sequência ω se existe uma sequência u, com $l(u) \geq 1$, tal que $\omega = us$. Neste caso escrevemos $s \prec \omega$. Quando $s \prec \omega$ ou $s = \omega$ escrevemos $s \preceq \omega$. Dada uma sequência finita ω denotamos por $suf(\omega)$ o maior sufixo de ω.

Ao longo deste texto, consideraremos $(X_t)_{t\in\mathbb{Z}}$ e $(Y_t)_{t\in\mathbb{Z}}$ processos estocásticos, estacionários e ergódicos sobre o alfabeto finito $\mathcal{A} = \{0, 1, \ldots, N-1\}$, com $|\mathcal{A}| = N$. Dadas duas seqências $\omega,\ v \in \mathcal{A}_{-\infty}^{-1}$ e símbolos $a,\ b \in \mathcal{A}$, denotamos por

$$p_X(a \,|\, \omega) = \mathbb{P}(X_0 = a \,|\, X_{-1} = \omega_{-1}, X_{-2} = \omega_{-2}, \ldots),$$

$$p_Y(b \,|\, v) = \mathbb{P}(Y_0 = b \,|\, Y_{-1} = v_{-1}, Y_{-2} = v_{-2}, \ldots),$$

as versões regulares das probabilidades condicionais dos processos. Dadas duas sequências finitas ω, $v \in \mathcal{A}_{-j}^{-1}$ denotamos por

$$p_X(\omega) = \mathbb{P}\left(X_{-j}^{-1} = \omega\right),$$

$$p_Y(v) = \mathbb{P}\left(Y_{-j}^{-1} = v\right),$$

as probabilidades estacionárias dos cilindros definidos pelas sequências ω e v, respectivamente.

Assumiremos que o processo (X_t) satisfaz as seguintes condições

1 *Não nulidade*, ou seja

$$\alpha_X := \inf\left\{p_X(a \mid \omega) : a \in \mathcal{A},\ \omega \in \mathcal{A}_{-\infty}^{-1}\right\} > 0,$$

2 *Taxa de continuidade somável*, ou seja

$$\beta_X := \sum_{k\in\mathbb{N}} \beta_{k,X}$$

onde a sequência $\{\beta_{k,X}\}_{k\in\mathbb{N}}$ é definida por

$$\beta_{k,X} := \sup\left\{\left|1 - \frac{p_X(a \mid \omega)}{p_X(a \mid v)}\right| : a \in \mathcal{A},\ v,\ \omega \in \mathcal{A}_{-\infty}^{-1}\ \ \textit{com}\ \ \omega \overset{k}{=} v\right\}. \tag{2.1}$$

Aqui, $\omega \overset{k}{=} v$ significa que existe uma sequência u, com $l(u) = k$, tal que $u \prec \omega$ e $u \prec v$. A sequência $\{\beta_{k,X}\}_{k\in\mathbb{N}}$ é chamada *taxa de continuidade* do processo (X_t).

Iremos supor que o processo (Y_t) também satisfaz as condições de não nulidade e taxa de continuidade somável com constantes α_Y e β_Y.

Seja (Z_t) um processo de ordem infinita sobre o alfabeto $\mathcal{A} = \{0, 1, \ldots, N-1\}$. Denotamos por $p_Z(\omega_{-j}^{-1})$ a probabilidade do cilindro

$$\mathbb{P}\left(Z_{-j}^{-i} = \omega_{-j}^{-1}\right).$$

Para toda sequência $\omega_{-\infty}^{-1}$ denotamos por

$$p_Z(a \mid \omega) = \mathbb{P}(Z_0 = a \mid Z_{-1} = \omega_{-1}, Z_{-2} = \omega_{-2}, \ldots)$$

as probabilidades de transição regulares correspondendo ao processo (Z_t).

Definição 2.1. *Uma sequência $\omega \in \mathcal{A}_{-j}^{-1}$ é um contexto para o processo (X_t) se esta satisfaz*

1 Para toda sequência semi-infinita $x_{-\infty}^{-1}$ tendo ω como um sufixo,

$$\mathbb{P}\left(X_0 = a \mid X_{-\infty}^{-1} = x_{-\infty}^{-1}\right) = p_X(a \mid \omega),\ \textit{para todo } a \in \mathcal{A}.$$

2 Nenhum sufixo de ω satisfaz 1.

Um contexto infinito é uma sequência semi-infinita $\omega_{-\infty}^{-1}$ tal que nenhum dos seus sufixos ω_{-j}^{-1}, $j = 1, 2, \ldots$ é um contexto.

A Definição 2.1 implica que o conjunto de todos contextos (sendo finito ou infinito) pode ser representado como uma árvore com raiz e rótulos (ver a Observação 2.2). Esta árvore é chamada *árvore de contextos* do processo (X_t) e será denotada por $\mathcal{T}_X$. A hipótese de não nulidade implica que a árvore de contextos do processo (X_t) é completa, isto é, qualquer sequência em $\mathcal{A}_{-\infty}^{-1}$ pertence a $\mathcal{T}_X$ ou têm sufixo que pertence a $\mathcal{T}_X$. Dizemos que a árvore de contextos $\mathcal{T}_X$ é *limitada* se esta possui um número finito de sequências. No caso infinito dizemos que $\mathcal{T}_X$ é *ilimitada*.

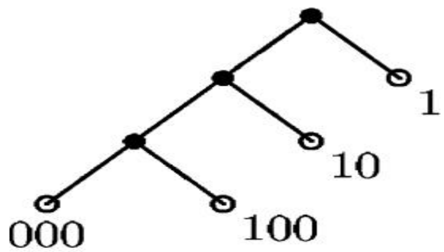

Figura 2.1: Árvore de contextos de um processo de renovação ($k_0 = 3$).

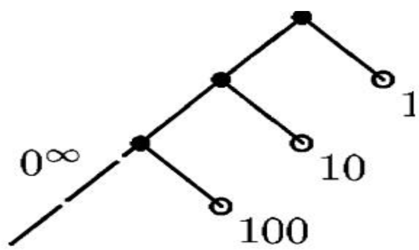

Figura 2.2: Árvore de contextos de um processo de renovação ($k_0 = \infty$).

Observação 2.2. *Cada sequência $s = a_1^k \in \mathcal{T}_X$ é vista como um caminho de uma folha para a raiz (elaborado com a raiz no topo), consistindo de k nós rotulados pelos símbolos $a_1, \ldots, a_k$. Uma sequência semi-infinita $a_{-\infty}^{-1} \in \mathcal{T}_X$ é vista como um caminho infinito para a raiz, conforme as Figuras 2.1 e 2.2. As sequências $s \in \mathcal{T}_X$ são também identificadas com as folhas da árvore $\mathcal{T}_X$. A folha s é identificada com o caminho ligando s com a raiz, como nas Figuras 2.1 e 2.2. Analogamente, os nós da árvore $\mathcal{T}_X$ são identificados com os sufixos finitos de todos $s \in \mathcal{T}_X$ (finito ou infinito), sendo a raiz identificada com a sequência vazia. Os filhos de um nó s são aquelas sequências as, com $a \in \mathcal{A}$, que são nós próprios, isto é, sufixos de algum $s' \in \mathcal{T}_X$.*

Denotamos por $d(\mathcal{T}_X)$ ou $|\mathcal{T}_X|$ a profundidade da árvore $\mathcal{T}_X$, ou seja,

$$d(\mathcal{T}_X) = \max\{l(s) : s \in \mathcal{T}_X\}.$$

Quando a árvore de contextos possui profundidade $d(\mathcal{T}_X) = k_0 < \infty$ o processo (X_t) é uma cadeia de Markov de ordem k_0. Neste caso, a árvore de contextos fornece uma descrição parcimoniosa do processo, pois, a coleção das $(|\mathcal{A}| - 1)|\mathcal{T}_X|$ probabilidades de transição é suficiente para descrever o *processo, ao invés de* $(|\mathcal{A}|-1)\,|\mathcal{A}|^{k_0}$ *probabilidades de transição. Note que a árvore de contextos de um processo independente e identicamente distribuídos processo (i.i.d.) consiste apenas da raiz* $\emptyset$*, assim,* $|\mathcal{T}_X| = 1$*.*

Nosso próximo exemplo foi retirado de Csiszár e Talata (2006).

Exemplo 2.3. *(Processo de Renovação). Seja* $\mathcal{A} = \{0,1\}$ *e suponhamos que a distância entre as ocorrências de* $1's$ *são i.i.d. Denotamos por* P_j *a probabilidade de que esta distância seja* j*, ou seja,* $P_j = p_X\left(10^{j-1}1\right)/p_X(1)$*, e seja*

$$q_k = \sum_{j=k}^{\infty} P_j,$$

$k \geq 1$*. Então, para* $k \geq 1$*, temos*

$$p_X(1|10^{k-1}) = \frac{P_k}{q_k} \triangleq P_{k,X}$$

($P_{k,X}$ *é indefinido se* $q_k = 0$*). Fazendo* $P_{0,X} = p_X(1)$ *e denotando por* k_0 *o menor inteiro tal que, para todo* $k \geq k_0$*,* $p_{X,k}$ *é constante ou indefinido (*$k_0 = \infty$ *caso não exista tal inteiro). Então, os contextos são as sequências* 10^{i-1}*,* $i \leq k_0$*, e a sequência* 0^{k_0} *(se* $k_0 < \infty$*) ou a sequência semi-infinita* 0^{∞} *(se* $k_0 = \infty$*), ver Figuras 2.1 e 2.2.*

Dado um inteiro K denotamos por $\mathcal{T}_X\,|_K$ a árvore $\mathcal{T}_X$ truncada no nível K, isto é

$$\mathcal{T}_X\,|_K = \{\omega \in \mathcal{T}_X : l(\omega) \leq K\} \cup \{\omega : l(\omega) = k \text{ e } \omega \prec u, \text{para algum } u \in \mathcal{T}_X\}\,.$$

Seja $Z_1, Z_2, \ldots, Z_n$ uma amostra aleatória do processo (Z_t). Para toda sequência finita ω, com $l(\omega) \leq n$, denotaremos por $N_n(\omega)$ o número de ocorrências da sequência na amostra, isto é

$$N_n(\omega) = \sum_{t=0}^{n-l(\omega)} \mathbf{1}_{\left\{Z_{t+1}^{t+l(\omega)} = \omega\right\}}.$$

Para todo elemento $a \in \mathcal{A}$ e para toda sequência finita ω, a probabilidade de transição empírica $\hat{p_{Z_n}}(a \mid \omega)$ é definida por

$$\hat{p_{Z_n}}(a \mid \omega) = \frac{N_n(\omega a) + 1}{N_n(\omega .) + |\mathcal{A}|},$$

onde

$$N_n(\omega .) = \sum_{b \in \mathcal{A}} N_n(\omega b).$$

Uma modificação do estimador da árvore de contextos de Rissanen proposta em Galves e Leonardi (2008) será apresentada em seguida. Em primeiro lugar, definamos, para qualquer sequência finita ω, o operador

$$\Delta_n(\omega) = \max_{a\in\mathcal{A}} \left|\hat{p_{Z_n}}(a \mid \omega) - \hat{p_{Z_n}}(a \mid suf(\omega))\right|.$$

O operador $\Delta_n(\omega)$ calcula a distância máxima entre as probabilidades de transição empírica associadas à sequência ω e aquela associada à sequência $suf(\omega)$.

Definição 2.4. *Para todo $\delta > 0$ e $d < n$, o estimador da árvore de contextos $\hat{\mathcal{T}}_n^{\delta,d}$ é o conjunto contendo todos os $\omega \in \mathcal{A}_{-d}^{-1}$, tais que $\Delta_n(a\ suf(\omega)) > \delta$ para algum $a \in \mathcal{A}$ e $\Delta_n(u\omega) \leq \delta$ para todo $u \in \mathcal{A}_{-d}^{-l(\omega)}$.*

Dado um inteiro $k \geq 1$, definamos

$$\mathcal{C}_k = \{u \in \mathcal{T}_X\,|_k : p_X(a \mid u) \neq p_X(a \mid suf(u)), \text{para algum a} \in \mathcal{A}\}$$

e

$$D_k = \min_{u\in C_k} \max_{a\in\mathcal{A}} \{|p_X(a \mid u) - p_X(a \mid suf(u))|\}.$$

Pela definição, podemos ver que $D_k > 0$ para todo $k \geq 1$.

2.2 Resultados

2.2.1 Primeiro Modelo

Inicialmente, consideramos (X_t) um processo como na Seção 2.1, mas tomando valores no alfabeto binário $\mathcal{A} = \{0, 1\}$. Estamos interessados no efeito de um ruído Bernoulli lançado no processo (X_t), independente dos símbolos sucessivos do processo (X_t), ou seja, consideramos $(\xi_t)_{t\in\mathbb{Z}}$ uma sequência de variáveis aleatórias i.i.d. tomando valores em $\{0, 1\}$, independente do processo (X_t), com

$$\mathbb{P}(\xi_t = 1) = 1 - \varepsilon,$$

onde ε é um parâmetro de ruído fixado em $(0, 1)$.

Definimos o processo estocasticamente perturbado $(Z_t)_{t\in\mathbb{Z}}$ por

$$Z_t = X_t.\xi_t. \tag{2.2}$$

Genericamente (Z_t) é um processo de ordem infinita.

Nosso primeiro resultado prova que as probabilidades de transição da cadeia perturbada estão uniformemente próximas das probabilidades de transição correspondentes da cadeia original se a probabilidade de mudança é pequena o suficiente.

Teorema 2.5. *Para (X_t) e (Z_t) como acima e para todo $\varepsilon \in (0,1)$, temos*

$$\sup_{k} \sup_{a,\ \omega_{-k}^{-1}} \left| p_Z\left(a|\omega_{-k}^{-1}\right) - p_X\left(a|\omega_{-k}^{-1}\right) \right| \leq \varepsilon \left[1 + \frac{4\beta_X}{\min\left(1, (1+\varepsilon)\alpha_X \beta_X^*\right)} \right], \tag{2.3}$$

onde $k \geq 0$, $a \in \mathcal{A}$, $\omega_{-k}^{-1} \in \mathcal{A}_{-k}^{-1}$ e $\beta_X^* = \prod_{k=0}^{+\infty}(1-\beta_{k,X}) > 0$.

Observação 2.6. *Aqui e em todo o texto consideramos eventos condicionais definidos por sequências vazias. Por exemplo, aparece um destes eventos quando $k = 0$ no Teorema 2.5. Nestes casos, a convenção é que esses eventos são removidos das expressões condicionais.*

Com o intuito de recuperar a árvore de contextos truncada do processo (X_t) através de uma amostra do processo perturbado (Z_t), estabelecemos o segundo resultado desta seção.

Teorema 2.7. *Seja K um inteiro e considere $Z_1, Z_2, \ldots, Z_n$ uma amostra aleatória do processo perturbado (Z_t). Então, existem uma constante c_1 e um inteiro d dependendo do processo (X_t) tais que para todo $\epsilon \in (0, D_d/2c_1)$, para todo $\delta \in (c_1\epsilon, D_d - c_1\epsilon)$, existe $n_0(\delta)$ tal que para todo $n > n_0$ temos*

$$\mathbb{P}\left(\hat{\mathcal{T}}_n^{\delta,d}|_K \neq \mathcal{T}_X|_K \right) \leq c_2 exp\left\{-c_3(n-d)\right\}.$$

As constantes são explícitas e dadas por:

1. $c_1 = 2\left[1 + \frac{4\beta_X}{min(\alpha_X\beta_X^*, 1)}\right]$,
2. $d = max_{u \notin \mathcal{T}_X,\ l(u)<K} min\left\{k : existe \quad \omega \in \mathcal{C}_k \quad com \quad u \prec \omega\right\}$,
3. $n_0 = \frac{6}{\left\{D_d - \delta - 2\varepsilon\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta_X^*)}\right]\right\}\alpha_X^d(1-\varepsilon)^d} + d$,
4. $c_2 = 2^d 12 e^{\frac{1}{e}}$,
5. $c_3 = \frac{\left[\min(D_d-\delta,\delta) - 2\bar{k}\right]^2 \alpha_X^{2d}(1-\varepsilon)^{3d+1}}{256e(d+1)\left(1+\frac{\beta_X}{\alpha_X}\right)}$,
6. $\bar{k} = \varepsilon\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta^*)}\right] + \frac{3}{(n-d)\alpha_X^d(1-\varepsilon)^d}$.

Como uma consequência do Teorema 2.7, obtivemos o seguinte resultado de consistência forte que nos diz que, para todo inteiro K, podemos recuperar a árvore de contextos truncada do processo (X_t) para quase toda amostra infinita do processo perturbado (Z_t).

Corolário 2.8. *Para todo inteiro K e para quase toda amostra infinita $Z_1, Z_2, \ldots$ existe um $\overline{n}$ tal que, para todo $n \geq \overline{n}$ temos*

$$\hat{\mathcal{T}}_n^{\delta,d}|_K = \mathcal{T}_X|_K,$$

onde d e δ são escolhidas como no Teorema 2.7.

Observação 2.9. *Tanto o Teorema 2.7 como o Corolário 2.8 nos dizem que o estimador* $\hat{\mathcal{T}}_n^{\delta,d}\big|_K$ *é robusto, no sentido que, mesmo que o processo de estimação tenha como base uma amostra aleatória do processo perturbado* (Z_t)*, o estimador* $\hat{\mathcal{T}}_n^{\delta,d}\big|_K$ *consegue recuperar a árvore de contextos truncada* $\mathcal{T}_X\big|_K$.

2.2.2 Segundo Modelo

Para os próximos resultados definiremos um novo processo de ordem infinita, que também será denotado por (Z_t). O novo modelo pode ser resumido do seguinte modo: dadas duas cadeias de Markov de alcance variável, tomando valores num mesmo alfabeto finito, a cada instante do tempo, o novo processo escolhe aleatoriamente um dos dois processos originais com uma probabilidade grande e fixa. A cadeia obtida dessa maneira pode então ser vista como uma pertubação estocástica da cadeia que está sendo escolhida com probabilidade maior. Formalmente, consideramos $\mathcal{A}$ o alfabeto finito $\{0, 1, \ldots, N-1\}$, com $|\mathcal{A}| = N$. Agora, sejam (X_t) e (Y_t) processos tomando valores em $\mathcal{A}$ e satisfazendo as condições de não nulidade e taxa de continuidade somável, como na Seção 2.1. Seja $(\xi_t)_{t\in\mathbb{Z}}$ uma sequência de variáveis aleatórias i.i.d. tomando valores em $\{0, 1\}$, independente dos processos (X_t) e (Y_t), com

$$\mathbb{P}(\xi_t = 1) = 1 - \varepsilon,$$

onde ε é um parâmetro fixado em $(0, 1)$. Assumimos que os processos (X_t) e (Y_t) são independentes. Definimos o novo processo $(Z_t)_{t\in\mathbb{Z}}$ por

$$Z_t = \begin{cases} X_t, & \text{se } \xi_t = 1, \\ Y_t, & \text{se } \xi_t = 0. \end{cases} \tag{2.4}$$

Genericamente (Z_t) é um processo de ordem infinita e pode ser interpretado como uma pertubação estocástica do processo (X_t). Para esse modelo, obtivemos resultados semelhantes aos do problema inicial, que serão apresentados em seguida.

O próximo teorema diz que a diferença entre as probabilidades condicionais do próximo símbolo, dado um passado finito de qualquer comprimento fixado, é uniformemente limitada superiormente pela probabilidade de mudança multiplicada por uma constante fixa.

Teorema 2.10. *Para* (X_t)*,* (Y_t) *e* (Z_t) *como acima e para todo* $\varepsilon \in (0, 1)$*, temos*

$$\sup_k \sup_{a,\ \omega_{-k}^{-1}} \left| p_Z\left(a|\omega_{-k}^{-1}\right) - p_X\left(a|\omega_{-k}^{-1}\right)\right| \leq \varepsilon \left[2 + \frac{4(N-1)\beta_X}{\min\left(1, \alpha\beta^*_{min}\right)}\right],$$

onde $k \geq 0$*,* $a \in \mathcal{A}$*,* $\omega_{-k}^{-1} \in \mathcal{A}_{-k}^{-1}$*,* $\beta^*_X = \prod_{k=0}^{+\infty}(1 - \beta_{k,X}) > 0$ *e* $\alpha\beta^*_{min} = \min\{\alpha_X\beta^*_X, \alpha_Y\}$.

Com o objetivo de recuperar a árvore de contextos truncada do processo (X_t) através de uma amostra do processo perturbado (Z_t), estabelecemos o seguinte resultado.

Teorema 2.11. *Seja K um inteiro e seja $Z_1, Z_2, \ldots, Z_n$ uma amostra aleatória do processo (Z_t). Então, existem uma constante c_1 dependendo dos processos (X_t) e (Y_t) e um inteiro d dependendo do processo (X_t) tais que para todo $\epsilon \in (0, D_d/2c_1)$, para todo $\delta \in (c_1\epsilon, D_d - c_1\epsilon)$, existe $n_0(\delta)$ tal que para todo $n > n_0$ temos*

$$\mathbb{P}\left(\hat{\mathcal{T}}_n^{\delta,d}\,|_K \neq \mathcal{T}_X\,|_K\right) \leq c_2 exp\left\{-c_3(n-d)\right\}.$$

As constantes são explícitas e dadas por:

1. $c_1 = 4\left[1 + \frac{2(N-1)\beta_X}{min((\alpha\beta^*)_{min},1)}\right]$,
2. $d = max_{u \notin \mathcal{T}_X,\ l(u)<K} min\left\{k : existe \quad \omega \in \mathcal{C}_k \quad com \quad u \prec \omega\right\}$,
3. $n_0 = \frac{2(N+1)}{\left\{D_d - \delta - 4\varepsilon\left[1+\frac{2(N-1)\beta_X}{\min(1,(\alpha\beta^*)_{min})}\right]\right\}\alpha_{min}^d} + d$,
4. $\alpha_{min} = \min\{\alpha_X, \alpha_Y\}$,
5. $c_2 = 48N^d(N+1)e^{\frac{1}{e}}$,
6. $c_3 = \frac{\left[\min(D_d-\delta,\delta)-2\bar{k}\right]^2\alpha^{2d}}{128N^2e(d+1)\left(1+\frac{\beta}{\alpha}\right)_{min}}$,
7. $\left(1+\frac{\beta}{\alpha}\right)_{min} = \min\left\{\left(1+\frac{\beta_X}{\alpha_X}\right), \left(1+\frac{\beta_Y}{\alpha_Y}\right)\right\}$,
8. $\bar{k} = \frac{\varepsilon c_1}{2} + \frac{N+1}{(n-d)\alpha_{min}^d}$.

O seguinte corolário nos diz que, para todo inteiro K, podemos recuperar a árvore de contextos truncada do processo (X_t) para quase toda amostra infinita do processo perturbado (Z_t).

Corolário 2.12. *Para todo inteiro K e para quase toda amostra infinita $Z_1, Z_2, \ldots$ existe um $\overline{n}$ tal que, para todo $n \geq \overline{n}$ temos*

$$\hat{\mathcal{T}}_n^{\delta,d}\,|_K = \mathcal{T}_X\,|_K\,,$$

onde d e δ são escolhidas como no Teorema 2.11.

Observação 2.13. *Tanto o Teorema 2.11 como o Corolário 2.12 nos dizem que o estimador $\hat{\mathcal{T}}_n^{\delta,d}\,|_K$ é robusto, no sentido que, mesmo que o processo de estimação tenha como base uma amostra aleatória do processo perturbado (Z_t), o estimador $\hat{\mathcal{T}}_n^{\delta,d}\,|_K$ consegue recuperar a árvore de contextos truncada $\mathcal{T}_X\,|_K$.*

Capítulo 3

Demonstrações

3.1 Demonstração do Teorema 2.5

Lembramos que nesta e na próxima seção o alfabeto $\mathcal{A}$ é o conjunto binário $\{0,1\}$ e o processo (Z_t) é definido por

$$Z_t = X_t.\xi_t.$$

Para este modelo podemos verificar as seguintes equivalências

$$\{Z_{-j}=0\} = \{X_{-j}=0, \xi_{-j}=0\} \bigcup \{X_{-j}=0, \xi_{-j}=1\} \bigcup \{X_{-j}=1, \xi_{-j}=0\}, \tag{3.1}$$

$$\{Z_{-j}=1\} = \{X_{-j}=1, \xi_{-j}=1\}, \tag{3.2}$$

para todo $j \in \mathbb{Z}$. Além disso, para todo $\omega_{-j} \in \mathcal{A}$, podemos escrever

$$\{X_{-j}=1, Z_{-j}=\omega_{-j}\} = \{X_{-j}=1, \xi_{-j}=\omega_{-j}\}. \tag{3.3}$$

Pela inclusão $\{X_{-j}=0\} \subset \{Z_{-j}=0\}$ e por (3.3), obtemos

$$\left\{X_{-j}^0 = u_{-j}^0, Z_{-j}^0 = \omega_{-j}^0\right\} \quad = \quad \left\{X_{-j}^0 = u_{-j}^0, \bigcap_{0\le l\le j:\ u_{-l}\neq 0} \xi_{-l} = w_{-l}\right\}, \tag{3.4}$$

para todo $\omega_{-j}^0 \in \mathcal{A}_{-j}^0$. Enfatizamos que as equivalências acima serão utilizadas nas demonstrações dos próximos resultados.

Iniciamos a prova do Teorema 2.5 com três lemas preparatórios.

Lema 3.1. *Para todo $\varepsilon \in (0,1)$, qualquer $\omega_{-\infty}^0$, quaisquer $k > j \ge 0$ e quaisquer $a, b \in \mathcal{A}$, temos*

$$\left|\mathbb{P}\left(X_0 = \omega_0 | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) - p_X(\omega_0|\omega_{-\infty}^{-1})\right| \le \beta_{j,X}.$$

Demonstração. Observamos, para todo $j \ge 0$, que

$$\mathbb{P}\left(X_0 = \omega_0 | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) = \frac{p_1}{p_2}, \tag{3.5}$$

sendo p_1 e p_2 definidos por

$$p_1 = \sum_{u_{-k}^{-j-2}} p_X\left(u_{-k}^{-j-2} a \omega_{-j}^{-1} \omega_0\right) \mathbb{P}\left(Z_{-k}^{-j-1} = \omega_{-k}^{-j-2} b | X_{-k}^{-j-1} = u_{-k}^{-j-2} a\right),$$

$$p_2 = \sum_{u_{-k}^{-j-2}} p_X\left(u_{-k}^{-j-2} a \omega_{-j}^{-1}\right) \mathbb{P}\left(Z_{-k}^{-j-1} = \omega_{-k}^{-j-2} b | X_{-k}^{-j-1} = u_{-k}^{-j-2} a\right),$$

onde os últimos dois somatórios são sobre todas as sequências $u_{-k}^{-j-2} \in \mathcal{A}_{-k}^{-j-2}$ tais que

$$\mathbb{P}\left(X_{-k}^{-j-1} = u_{-k}^{-j-2} a, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1} b\right) \neq 0.$$

Para atestar a validade de (3.5), temos, pela definição de probabilidade condicional, que

$$p_1 \;=\; \sum_{u_{-k}^{-j-2}} p_X\left(u_{-k}^{-j-2} a \omega_{-j}^{-1} \omega_0\right) \frac{\mathbb{P}\left(X_{-k}^{-j-1} = u_{-k}^{-j-2} a, Z_{-k}^{-j-1} = \omega_{-k}^{-j-2} b\right)}{\mathbb{P}\left(X_{-k}^{-j-1} = u_{-k}^{-j-2} a\right)}.$$

Considerando a equivalência entre eventos (3.4), fazendo $b = \omega_{-j-1}$ e utilizando a independência das mudanças (ξ_t), tem-se

$$\begin{aligned} p_1 &= \sum_{u_{-k}^{-j-2}} \mathbb{P}\left(X_{-k}^{0} = u_{-k}^{-j-2} a \omega_{-j}^{-1} \omega_0, \bigcap_{j+1 \le l \le k:\; u_{-l} \neq 0} \xi_{-l} = w_{-l}\right) \\ &= \sum_{u_{-k}^{-j-2}} \mathbb{P}\left(X_{-k}^{-j-2} = u_{-k}^{-j-2}, X_{-j-1} = a, X_{-j}^{-1} = \omega_{-j}^{-1}, X_0 = \omega_0, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \\ &= \mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, X_0 = \omega_0, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right). \end{aligned}$$

Analogamente, obtem-se

$$p_2 \;=\; \mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right).$$

Assim, podemos escrever

$$\begin{aligned} \frac{p_1}{p_2} &= \frac{\mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, X_0 = \omega_0, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}{\mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)} \\ &= \mathbb{P}\left(X_0 = \omega_0 | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \end{aligned}$$

e, dessa forma, verificamos (3.5). Agora, como em Fernández e Galves (2002), temos

$$\inf_{v_{-\infty}^{-j-1}} p_X\left(\omega_0 | v_{-\infty}^{-j-1} \omega_{-j}^{-1}\right) \le p_X\left(\omega_0 | u_{-k}^{-j-2} a \omega_{-j}^{-1}\right) \le \sup_{v_{-\infty}^{-j-1}} p_X\left(\omega_0 | v_{-\infty}^{-j-1} \omega_{-j}^{-1}\right). \tag{3.6}$$

Por (3.6) e pela continuidade do processo (X_t) segue que

$$p_X\left(\omega_0|\omega_{-\infty}^{-1}\right) - \beta_{j,X} \leq p_X\left(\omega_0|u_{-k}^{-j-2}a\omega_{-j}^{-1}\right) \leq p_X\left(\omega_0|\omega_{-\infty}^{-1}\right) + \beta_{j,X}. \tag{3.7}$$

Para comprovar (3.7) recomendamos ao leitor Collet, Galves e Leonardi (2008). Assim, por (3.5) e (3.7), obtemos

$$p_X\left(\omega_0|\omega_{-\infty}^{-1}\right) - \beta_{j,X} \leq \frac{p_1}{p_2} \leq p_X\left(\omega_0|\omega_{-\infty}^{-1}\right) + \beta_{j,X},$$

ou, equivalentemente,

$$\left|\mathbb{P}\left(X_0 = \omega_0|X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) - p_X\left(\omega_0|\omega_{-\infty}^{-1}\right)\right| \leq \beta_{j,X},$$

como queríamos. □

Lema 3.2. *Para todo $\varepsilon \in (0,1)$, qualquer $k \geq 0$ e qualquer ω_{-k}^0, temos*

$$p_Z\left(\omega_0|\omega_{-k}^{-1}\right) \geq (1-\varepsilon)\alpha_X \tag{3.8}$$

e

$$\mathbb{P}\left(X_0 = \omega_0|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \geq \alpha_X. \tag{3.9}$$

Além disso, para todo $0 \leq j \leq k$ temos

$$\mathbb{P}\left(X_{-j-1} = \omega_{-j-1}|X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \geq \alpha_X\beta_X^*. \tag{3.10}$$

Demonstração. Inicialmente, vamos verificar a seguinte afirmação: se vale a desigualdade (3.9) então também vale (3.8). Para ver isto, vamos, primeiramente, analisar o caso $\omega_0 = 0$. Considerando (3.1) e a independência das mudanças, segue que

$$\begin{aligned} p_Z\left(\omega_0|\omega_{-k}^{-1}\right) &= \mathbb{P}\left(X_0 = \omega_0, \xi_0 = 0|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) + \mathbb{P}\left(X_0 = \omega_0, \xi_0 = 1|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) + \\ &+ \mathbb{P}\left(X_0 = \omega_0', \xi_0 = 0|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \\ &= \varepsilon\mathbb{P}\left(X_0 = \omega_0|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) + (1-\varepsilon)\mathbb{P}\left(X_0 = \omega_0|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) + \\ &+ \varepsilon\mathbb{P}\left(X_0 = {\omega'}_0|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \end{aligned}$$

onde $\omega_0' \neq \omega_0$. Assim,

$$p_Z\left(\omega_0|\omega_{-k}^{-1}\right) \geq (1+\varepsilon)\alpha_X.$$

Analogamente, para o caso em que $\omega_0 \neq 0$, podemos escrever

$$p_Z\left(\omega_0|\omega_{-k}^{-1}\right) = \mathbb{P}\left(X_0 = \omega_0, \xi_0 = 1|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \geq (1-\varepsilon)\alpha_X.$$

Portanto, para todo $\omega_0 \in \mathcal{A}$, temos

$$p_Z\left(\omega_0|\omega_{-k}^{-1}\right) \geq (1-\varepsilon)\alpha_X.$$

Para demonstrarmos a segunda asserção deste lema, afirmamos, por (3.6), que vale da seguinte identidade

$$\mathbb{P}\left(X_0 = \omega_0|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) = \lim_{l\to+\infty} \frac{p_1}{p_2}, \tag{3.11}$$

onde

$$p_1 \;:=\; \sum_{u_{-l}^{-1}} p_X\left(\omega_0|\omega_{-\infty}^{-l-1}u_{-l}^{-1}\right)\mathbb{P}\left(X_{-l}^{-1} = u_{-l}^{-1}|X_{-\infty}^{-l-1} = \omega_{-\infty}^{-l-1}\right)\mathbb{P}(\bigcap_{-k\leq -t\leq -1:\; u_{-t}\neq 0} \xi_{-t} = w_{-t}),$$

$$p_2 \;:=\; \sum_{u_{-l}^{-1}} \mathbb{P}\left(X_{-l}^{-1} = u_{-l}^{-1}|X_{-\infty}^{-l-1} = \omega_{-\infty}^{l-1}\right)\mathbb{P}(\bigcap_{-k\leq -t\leq -1:\; u_{-t}\neq 0} \xi_{-t} = w_{-t}),$$

e os somatórios acima são sobre todas as sequências $u_{-l}^{-1} \in \mathcal{A}_{-l}^{-1}$ tais que

$$\mathbb{P}\left(X_{-l}^{-1} = u_{-l}^{-1}, Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \neq 0.$$

Pela hipótese de não nulidade, temos

$$p_X\left(\omega_0|\omega_{-\infty}^{-l-1}u_{-l}^{-1}\right) \geq \alpha_X.$$

Agora, utilizando a última expressão e a identidade (3.11), temos, para cada l, que

$$\frac{p_1}{p_2} \geq \alpha_X.$$

Consequentemente,

$$\lim_{l\to+\infty} \frac{p_1}{p_2} = \mathbb{P}\left(X_0 = \omega_0|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \geq \alpha_X,$$

o que prova (3.9). Para demonstrarmos a validade de (3.10), note que

$$\mathbb{P}\left(X_{-j-1} = \omega_{-j-1}|X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) = \frac{p_5}{p_6}, \tag{3.12}$$

onde

$$p_5 := \sum_{x_{-k}^{-j-2}} \mathbb{P}\left(Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}|X_{-k}^{-j-2} = x_{-k}^{-j-2}\right)\mathbb{P}\left(X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, X_{-k}^{-j-2} = x_{-k}^{-j-2}\right),$$

$$p_6 := \sum_{x_{-k}^{-j-2}} \mathbb{P}\left(Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}|X_{-k}^{-j-2} = x_{-k}^{-j-2}\right)\mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-k}^{-j-2} = x_{-k}^{-j-2}\right)$$

e os somatórios acima são sobre todas as sequências $x_{-k}^{-j-2} \in \mathcal{A}_{-k}^{-j-2}$ tais que

$$\mathbb{P}\left(X_{-k}^{-j-2} = x_{-k}^{-j-2}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \neq 0.$$

Para verificar (3.12), observamos, pela independência das mudanças e por (3.4), que são válidas as seguintes identidades

$$\begin{aligned}
p_5 &= \sum_{x_{-k}^{-j-2}} \frac{\mathbb{P}\left(Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}, X_{-k}^{-j-2} = x_{-k}^{-j-2}\right)}{\mathbb{P}\left(X_{-k}^{-j-2} = x_{-k}^{-j-2}\right)} \mathbb{P}\left(X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, X_{-k}^{-j-2} = x_{-k}^{-j-2}\right) \\
&= \sum_{x_{-k}^{-j-2}} \mathbb{P}\left(\bigcap_{j+2\leq l\leq k:\ x_l\neq 0} \xi_{-l} = w_{-l}\right) \mathbb{P}\left(X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, X_{-k}^{-j-2} = x_{-k}^{-j-2}\right) \\
&= \sum_{x_{-k}^{-j-2}} \mathbb{P}\left(X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, X_{-k}^{-j-2} = x_{-k}^{-j-2}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \\
&= \mathbb{P}\left(X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right).
\end{aligned}$$

Da mesma forma, podemos mostrar que

$$p_6 = \mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)$$

e deste modo concluímos a validade da identidade (3.12), pois,

$$\mathbb{P}\left(X_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) = \frac{\mathbb{P}\left(X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}{\mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}.$$

Afirmamos que vale a seguinte identidade

$$\begin{aligned}
p_7 &:= \frac{\mathbb{P}\left(X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, X_{-k}^{-j-2} = x_{-k}^{-j-2}\right)}{\mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-k}^{-j-2} = x_{-k}^{-j-2}\right)} \\
&= \frac{\prod_{l=1}^{j+1} p_X\left(\omega_{-l} | x_{-k}^{-j-2} \omega_{-j-1}^{-l-1}\right) \prod_{l=j+2}^{k} p_X\left(x_{-l} | x_{-k}^{-l-1}\right)}{\prod_{l=1}^{j} \mathbb{P}\left(X_{-l} = \omega_{-l} | X_{-k}^{-j-2} = x_{-k}^{-j-2}, X_{-j}^{-l-1} = \omega_{-j}^{-l-1}\right) \prod_{l=j+2}^{k} p_X\left(x_{-l} | x_{-k}^{-l-1}\right)},
\end{aligned}$$

Para comprovar a última afirmação, recomendamos a leitura de Collet, Galves e Leonardi (2008).

Pela última igualdade, segue que

$$
\begin{aligned}
p_7 &= \frac{\displaystyle\prod_{l=1}^{j+1} p_X\left(\omega_{-l}|x_{-k}^{-j-2}\omega_{-j-l}^{-l-1}\right)}{\displaystyle\prod_{l=1}^{j} \mathbb{P}\left(X_{-l}=\omega_{-l}|X_{-k}^{-j-2}=x_{-k}^{-j-2}, X_{-j}^{-l-1}=\omega_{-j}^{-l-1}\right)} \\
&= p_X\left(\omega_{-j-1}|x_{-k}^{-j-2}\right)\prod_{l=1}^{j}\frac{p_X\left(\omega_{-l}|x_{-k}^{-j-2}\omega_{-j-1}^{-l-1}\right)}{\mathbb{P}\left(X_{-l}=\omega_{-l}|X_{-k}^{-j-2}=x_{-k}^{-j-2}, X_{-j}^{-l-1}=\omega_{-j}^{-l-1}\right)} \\
&\geq \alpha_X\prod_{l=1}^{j}(1-\beta_{j-l,X}) \\
&\geq \alpha_X\beta_X^*.
\end{aligned}
$$

Consequentemente, temos

$$
\frac{\mathbb{P}\left(X_{-j-1}^{-1}=\omega_{-j-1}^{-1}, X_{-k}^{-j-2}=x_{-k}^{-j-2}\right)}{\mathbb{P}\left(X_{-j}^{-1}=\omega_{-j}^{-1}, X_{-k}^{-j-2}=x_{-k}^{-j-2}\right)} \geq \alpha_X\beta_X^*.
$$

A demonstração da terceira afirmação deste lema segue da última desigualdade e da identidade (3.12). □

Lema 3.3. *Para todo $\varepsilon\in(0,1)$, quaisquer $k>j\geq 0$ e qualquer ω_{-k}^{0}, temos*

$$
\mathbb{P}\left(X_{-j-1}={\omega'}_{-j-1}|X_{-j}^{-1}=\omega_{-j}^{-1}, Z_{-k}^{-j-1}=\omega_{-k}^{-j-1}\right) \leq \frac{\varepsilon}{(1+\varepsilon)\alpha_X\beta_X^*},
$$

onde ${\omega'}_{-j-1}\neq\omega_{-j-1}$.

Demonstração. Pela definição de probabilidade condicional, escrevemos

$$\begin{aligned}
p &:= \mathbb{P}\left(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}\right) \\
&= \frac{\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1}, Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}{\mathbb{P}\left(Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)} \\
&= \frac{\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1}, X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-j-1} = \omega_{-j-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}{\mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-j-1} = \omega_{-j-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)} \\
&= \frac{\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1}, X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-j-1} = \omega_{-j-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}{\mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)} \times \\
&\times \frac{\mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}{\mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-j-1} = \omega_{-j-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)},
\end{aligned}$$

onde $\omega'_{-j-1} \neq \omega_{-j-1}$. Assim, podemos reescrever p como segue

$$p = \frac{\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1}, Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}{\mathbb{P}\left(Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}. \tag{3.13}$$

Lembramos, para $\omega'_{-j-1} \neq 0$, que vale a seguinte equivalência entre eventos

$$\left\{X_{-j-1} = \omega'_{-j-1}, Z_{-j-1} = \omega_{-j-1}\right\} = \left\{X_{-j-1} = \omega'_{-j-1}, \xi_{-j-1} = \omega_{-j-1}\right\}.$$

Considerando a identidade acima, a equação (3.13) e a independência das mudanças, podemos, neste caso, escrever

$$p = \frac{\varepsilon \mathbb{P}\left(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}{\mathbb{P}\left(Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}. \tag{3.14}$$

Para o caso $\omega'_{-j-1} = 0$, temos

$$\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1}, Z_{-j-1} = \omega_{-j-1}\right) = 0.$$

Utilizando a última equivalência entre eventos e a equação (3.13), temos, neste caso, que

$$p = 0.$$

Logo, em todos os casos ($\omega'_{-j-1} \neq 0$ ou $\omega'_{-j-1} = 0$), podemos limitar p como segue

$$p \leq \frac{\varepsilon \mathbb{P}\left(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}{\mathbb{P}\left(Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}. \tag{3.15}$$

Note, pela independência das mudanças e por (3.1), que

$$\begin{aligned}
\bar{p} &:= \mathbb{P}\left(Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \\
&= \varepsilon \mathbb{P}\left(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \\
&+ \varepsilon \mathbb{P}\left(X_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) + \\
&+ (1-\varepsilon) \mathbb{P}\left(X_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right),
\end{aligned}$$

se $\omega_{-j-1} = 0$. Pelo Lema 3.2 e pela última igualdade, vale a seguinte desigualdade

$$\bar{p} \geq (1+\varepsilon) \alpha_X \beta_X^*. \tag{3.16}$$

Para o caso $\omega_{-j-1} \neq 0$, temos que

$$\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}\right) = 0, \tag{3.17}$$

pois, dado que o evento $\{Z_{-j-1} = 1\}$ ocorreu então, por (2.2), o evento $\{X_{-j-1} = 0\}$ não pode ocorrer. Portanto, em todos os casos, considerando (3.13), (3.15), (3.16) e (3.17), obtemos

$$\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}\right) \leq \frac{\varepsilon}{(1+\varepsilon)\alpha_X \beta_X^*}.$$

□

Por meio dos três lemas anteriores demonstraremos o Teorema 2.5.

Demonstração do Teorema 2.5. *Afirmamos, para todo $a \in \mathcal{A}$ e para toda sequência $\omega_{-k}^{-1} \in \mathcal{A}_{-k}^{-1}$, que*

$$\left| p_Z\left(a | \omega_{-k}^{-1}\right) - \mathbb{P}\left(X_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \right| \leq \varepsilon. \tag{3.18}$$

Vamos atestar a validade da desigualdade (3.18), analisando, separadamente, os casos $a = 0$ e $a \neq 0$. Para o caso $a = 0$, consideramos (3.1) para obter

$$\begin{aligned}
\mathbb{P}\left(Z_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}\right) &= (1-\varepsilon)\mathbb{P}\left(X_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}\right) + \varepsilon \mathbb{P}\left(X_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \\
&+ \varepsilon \mathbb{P}\left(X_0 = a' | Z_{-k}^{-1} = \omega_{-k}^{-1}\right),
\end{aligned}$$

onde $a' \neq 0$. Considerando a última igualdade, segue

$$\left|p_Z\left(a|\omega_{-k}^{-1}\right) - \mathbb{P}\left(X_0 = a|Z_{-k}^{-1} = \omega_{-k}^{-1}\right)\right| \leq \varepsilon.$$

Utilizando (3.2) obtemos, para o caso $a \neq 0$, que

$$p_Z\left(a|\omega_{-k}^{-1}\right) = (1-\varepsilon)\mathbb{P}\left(X_0 = a|Z_{-k}^{-1} = \omega_{-k}^{-1}\right).$$

Da última identidade, tem-se

$$\left|p_Z\left(a|\omega_{-k}^{-1}\right) - \mathbb{P}\left(X_0 = a|Z_{-k}^{-1} = \omega_{-k}^{-1}\right)\right| \leq \varepsilon.$$

Logo, para todo $a \in \mathcal{A}$, podemos escrever a desigualdade (3.18). Agora, procederemos como na demonstração do Teorema 1 de Collet, Galves e Leonardi (2008). Inicialmente, observamos que para o caso $k = 0$ a asserção do Teorema 2.5 é válida trivialmente. Então, resta-nos provar a desigualdade (2.3) para $k \geq 1$. Neste caso, podemos escrever

$$\begin{aligned} s &:= \mathbb{P}(X_0 = a|Z_{-k}^{-1} = \omega_{-k}^{-1}) - \mathbb{P}(X_0 = a|X_{-k}^{-1} = \omega_{-k}^{-1}) \\ &= \sum_{j=0}^{k-1}\Big[\mathbb{P}(X_0 = a|X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ &- \mathbb{P}(X_0 = a|X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2})\Big]. \end{aligned}$$

Cada parcela do somatório acima pode ser reescrita da seguinte maneira

$$\begin{aligned} p_j &:= \mathbb{P}(X_0 = a|X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ &- \mathbb{P}(X_0 = a|X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}) \\ &= \sum_{b\in\{0,1\}}\Big[\mathbb{P}(X_0 = a|X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = b, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ &- \mathbb{P}(X_0 = a|X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2})\Big] \\ &\times \mathbb{P}(X_{-j-1} = b|X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}). \end{aligned}$$

Para cada j, $0 \leq j \leq k-1$, o somatório p_j possui duas parcelas. A primeira, correspondendo a $b = \omega'_{-j-1}$, com $\omega'_{-j-1} \neq \omega_{-j-1}$, pode ser limitada superiormente por

$$\begin{aligned} k_{1,j} &:= \Big|\mathbb{P}(X_0 = a|X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = \bar{\omega}_{-j-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ &- \mathbb{P}(X_0 = a|X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2})\Big| \times \\ &\times \mathbb{P}(X_{-j-1} = \omega'_{-j-1}|X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}). \end{aligned}$$

Afirmamos que

$$k_{1,j} \leq 2\beta_{j,X} \frac{\varepsilon}{(1+\varepsilon)\alpha_X \beta_X^*}. \tag{3.19}$$

De fato, por meio do Lema 3.3, podemos escrever

$$\Big|\mathbb{P}(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1})\Big| \leq \frac{\varepsilon}{(1+\varepsilon)\alpha_X \beta_X^*}. \tag{3.20}$$

Agora, definimos a constante $k_{2,j}$ por

$$\begin{aligned} k_{2,j} \;&:=\; \Big|\mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = \bar{\omega}_{-j-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) + \\ &+\; \mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2})\Big|. \end{aligned}$$

Então, pela desigualdade triangular e pelo Lema 3.1, segue que

$$\begin{aligned} k_{2,j} \;&\leq\; \Big|\mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = \bar{\omega}_{-j-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) - p_X\left(a|\omega_{-\infty}^{-1}\right)\Big| + \\ &+\; \Big|\mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}) - p_X\left(a|\omega_{-\infty}^{-1}\right)\Big| \leq 2\beta_{j,X}. \end{aligned} \tag{3.21}$$

Logo, por (3.20) e (3.21) segue a desigualdade (3.19). Agora, para cada j, vamos limitar a parcela correspondendo a $b = \omega_{-j-1}$ de p_j. Pelas propriedades da função módulo, essa parcela pode ser limitada superiormente por $c_{3,j}$, onde

$$\begin{aligned} c_{3,j} \;&:=\; \Big|\mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = \omega_{-j-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) - \\ &-\; \mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2})\Big| \times \\ &\times\; \mathbb{P}(X_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}). \end{aligned}$$

Esta última expressão pode ser mojorada $c_{4,j}$, onde

$$\begin{aligned} c_{4,j} \;&:=\; \sum_{c \in \{0,1\}} \Big|\mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = \omega_{-j-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) - \\ &-\; \mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}, Z_{-j-1} = c)\Big| \times \\ &\times\; \mathbb{P}(X_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \times \\ &\times\; \mathbb{P}(Z_{-j-1} = c | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}). \end{aligned}$$

A parcela de $c_{4,j}$ correspondendo a $c = \omega_{-j-1}$ é nula e quando $c = \omega'_{-j-1}$, com $\omega'_{-j-1} \neq \omega_{-j-1}$, a correspondente parcela de $c_{4,j}$ pode ser limitada superiormente por $2\beta_{j,X}\varepsilon$. Com isso

$$|p_j| \leq 2\beta_{j,X} \frac{\varepsilon}{(1+\varepsilon)\alpha_X \beta_X^*} + 2\beta_{j,X}\varepsilon. \tag{3.22}$$

Agora, pela desigualdade triangular, temos

$$\begin{aligned}\left|p_Z\left(a|\omega_{-k}^{-1}\right)-p_X\left(a|\omega_{-k}^{-1}\right)\right| &\leq \left|p_Z\left(a|\omega_{-k}^{-1}\right)-\mathbb{P}\left(X_0=a|Z_{-k}^{-1}=\omega_{-k}^{-1}\right)\right|+\\ &+\ \left|\mathbb{P}\left(X_0=a|X_{-k}^{-1}=\omega_{-k}^{-1}\right)-\mathbb{P}\left(X_0=a|Z_{-k}^{-1}=\omega_{-k}^{-1}\right)\right|.\end{aligned}$$

Para majorar a primeira parcela no segundo membro da última desigualdade, utilizamos a expressão (3.18). Para a segunda, utilizamos as desigualdades (3.19) e (3.22). Assim,

$$\left|p_Z\left(a|\omega_{-k}^{-1}\right)-p_X\left(a|\omega_{-k}^{-1}\right)\right|\leq\varepsilon+\sum_{j=0}^{k-1}\left(2\beta_{j,X}\frac{\varepsilon}{(1+\varepsilon)\alpha_X\beta_X^*}+2\varepsilon\beta_{j,X}\right).$$

Portanto,

$$\sup_{a\in\mathcal{A},\ \omega_{-k}^{-1}\in\mathcal{A}_{-k}^{-1}}\left|p_Z\left(a|\omega_{-k}^{-1}\right)-p_X\left(a|\omega_{-k}^{-1}\right)\right|\leq\varepsilon\left[1+\frac{4\beta_X}{\min\left(1,(1+\varepsilon)\alpha_X\beta_X^*\right)}\right],$$

como queríamos.

3.2 Demonstração do Teorema 2.7

A prova do nosso segundo resultado baseia-se em cinco lemas. O primeiro deles é o Lema 3.4 de Galves e Leonardi (2008). Recordemos, por conveniência, este resultado.

Lema 3.4. *(Galves, Leonardi). Seja (X_t) um processo estocástico estacionário satisfazendo as hipóteses de não nulidade e taxa de continuidade somável. Então, existe uma sequência somável $(\rho_{l,X})_{l\in\mathbb{N}}$, satisfazendo*

$$\sum_{l\geq 1}\rho_{l,X}\leq 1+\frac{2\beta_X}{\alpha_X}$$

tal que para todo $i\geq 1$, para todo $k>i$, para todo $j\geq 1$ e para toda sequência finita ω_1^j, valem as seguintes desigualdades

$$\sup_{x_1^i\in\mathcal{A}^i}\left|\mathbb{P}\left({X_k}^{k+j-1}=\omega_1^j\mid X_1^i=x_1^i\right)-p_X\left(\omega_1^j\right)\right|\leq\sum_{l=0}^{j-1}\rho_{k-i+l,X}.$$

As constantes α_X e β_X aparecendo no resultado do último lema são as mesmas que foram definidas na Seção 2.1.

O Lema 3.4 será usado na prova do seguinte resultado envolvendo as mesmas quantidades α_X, β_X e $(\rho_{l,X})_{l\in\mathbb{N}}$.

Lema 3.5. *Existe uma sequência somável $(\rho_{l,X})_{l\in\mathbb{N}}$, satisfazendo*

$$\sum_{l\in\mathbb{N}}\rho_{l,X}\leq 2\left(1+\frac{\beta_X}{\alpha_X}\right),$$

tal que para todo $i \geq 1$, todo $k \geq i$, todo $j \geq 1$ e toda sequência finita ω_1^j, vale a seguinte desigualdade

$$\sup_{x_1^i, \theta_1^i \in \mathcal{A}^i} \left| \mathbb{P}\left(Z_k^{k+j-1} = \omega_1^j \mid X_1^i = x_1^i, \xi_1^i = \theta_1^i \right) - p_Z\left(\omega_1^j\right) \right| \leq \frac{\sum_{l=0}^{j-1} \rho_{k-i+l,X}}{(1-\varepsilon)^j}.$$

Demonstração. Das propriedades de probabilidade condicional, temos, para quaisquer $x_1^i, \theta_1^i \in \mathcal{A}_1^i$, que

$$\begin{aligned} k_1 &:= \left| \mathbb{P}\left(Z_k{}^{k+j-1} = \omega_1^j \mid X_1^i = x_1^i, \xi_1^i = \theta_1^i \right) - p_Z\left(\omega_1^j\right) \right| \\ &= \left| \sum_{x_k^{k+j-1}} \mathbb{P}\left(X_k^{k+j-1} = x_k^{k+j-1}, Z_k^{k+j-1} = \omega_1^j | X_1^i = x_1^i, \xi_1^i = \theta_1^i \right) - p_Z\left(\omega_1^j\right) \right|, \end{aligned} \tag{3.23}$$

onde o último somatório é sobre todas as sequências $x_k^{k+j-1} \in \mathcal{A}_1^j$ satisfazendo

$$\left\{ X_k^{k+j-1} = x_k^{k+j-1}, Z_k^{k+j-1} = \omega_1^j \right\} \neq \emptyset. \tag{3.24}$$

Por meio da identidade (3.23), afirmamos que

$$k_1 = \left| \sum_{x_k^{k+j-1}} p_X\left(x_k^{k+j-1} | x_1^i \right) \mathbb{P}\left(Z_k^{k+j-1} = \omega_1^j | X_k^{k+j-1} = x_k^{k+j-1} \right) - p_Z\left(\omega_1^j\right) \right|. \tag{3.25}$$

Para atestar (3.25) note, pela definição de probabilidade condicional e pela independência dos ruídos, que

$$\begin{aligned} k_1 &= \left| \sum_{x_k^{k+j-1}} \mathbb{P}\left(X_k^{k+j-1} = x_k^{k+j-1}, Z_k^{k+j-1} = \omega_1^j | X_1^i = x_1^i, \xi_1^i = \theta_1^i \right) - p_Z\left(\omega_1^j\right) \right| \\ &= \left| \sum_{x_k^{k+j-1}} \frac{\mathbb{P}\left(X_k^{k+j-1} = x_k^{k+j-1}, Z_k^{k+j-1} = \omega_1^j, X_1^i = x_1^i \right) \mathbb{P}\left(\xi_1^i = \theta_1^i \right)}{\mathbb{P}\left(X_1^i = x_1^i \right) \mathbb{P}\left(\xi_1^i = \theta_1^i \right)} - p_Z\left(\omega_1^j\right) \right| \\ &= \left| \sum_{x_k^{k+j-1}} \frac{\mathbb{P}\left(X_k^{k+j-1} = x_k^{k+j-1}, Z_k^{k+j-1} = \omega_1^j, X_1^i = x_1^i \right) p_X\left(x_k^{k+j-1} \right)}{p_X\left(x_1^i \right) p_X\left(x_k^{k+j-1} \right)} - p_Z\left(\omega_1^j\right) \right| \\ &= \left| \sum_{x_k^{k+j-1}} \frac{p_X\left(x_1^i x_k^{k+j-1} \right) \mathbb{P}\left(\bigcap_{k \leq l \leq k+j:\ x_l \neq 0} \xi_l = w_l \right) p_X\left(x_k^{k+j-1} \right)}{p_X\left(x_1^i \right) p_X\left(x_k^{k+j-1} \right) p_Z} - p_Z\left(\omega_1^j\right) \right|. \end{aligned}$$

Assim, temos

$$\begin{aligned} k_1 &= \left| \sum_{x_k^{k+j-1}} \frac{p_X\left(x_1^i x_k^{k+j-1}\right) \mathbb{P}\left(Z_k^{k+j-1} = \omega_1^j, X_k^{k+j-1} = x_k^{k+j-1}\right)}{p_X\left(x_1^i\right) p_X\left(x_k^{k+j-1}\right)} - p_Z\left(\omega_1^j\right) \right| \\ &= \left| \sum_{x_k^{k+j-1}} p_X\left(x_k^{k+j-1} | x_1^i\right) \mathbb{P}\left(Z_k^{k+j-1} = \omega_1^j | X_k^{k+j-1} = x_k^{k+j-1}\right) - p_Z\left(\omega_1^j\right) \right|. \end{aligned}$$

Deste modo, comprovamos a validade de (3.25). Agora, considerando a independência das mudanças, tem-se

$$\begin{aligned} k_3 &:= \sum_{x_k^{k+j-1} \in \mathcal{A}_1^j} \mathbb{P}\left(Z_k^{k+j-1} = \omega_1^j | X_k^{k+j-1} = x_k^{k+j-1}\right) \mathbb{P}\left(X_k^{k+j-1} = x_k^{k+j-1}\right) \\ &= \sum_{y_k^{k+j-1}} \mathbb{P}\left(\bigcap_{k \le l \le k+j-1:\ y_l \ne 0} \xi_l = \omega_l\right) \mathbb{P}\left(X_k^{k+j-1} = y_k^{k+j-1}\right) \\ &= \sum_{x_k^{k+j-1} \in \mathcal{A}_1^j} \mathbb{P}\left(X_k^{k+j-1} = x_k^{k+j-1}, Z_k^{k+j-1} = \omega_1^j\right) \\ &= p_Z\left(\omega_1^j\right), \end{aligned}$$

onde, na segunda igualdade, o somatório é sobre todas as sequências $y_k^{k+j-1} \in \mathcal{A}_1^j$ satisfazendo (3.24). Desta forma,

$$p_Z\left(\omega_1^j\right) = \sum_{x_k^{k+j-1} \in \mathcal{A}_1^j} \mathbb{P}\left(Z_k^{k+j-1} = \omega_1^j | X_k^{k+j-1} = x_k^{k+j-1}\right) \mathbb{P}\left(X_k^{k+j-1} = x_k^{k+j-1}\right). \tag{3.26}$$

Através das identidades (3.25) e (3.26) limitamos superiormente k_1 do seguinte modo

$$\begin{aligned} k_1 &\le \sum_{x_k^{k+j-1} \in \mathcal{A}_1^j} \left| \mathbb{P}\left(Z_k^{k+j-1} = \omega_1^j | X_k^{k+j-1} = x_k^{k+j-1}\right) \left[\mathbb{P}\left(X_k^{k+j-1} = x_k^{k+j-1} | X_1^i = x_1^i\right) \right.\right. \\ &- \left.\left. \mathbb{P}\left(X_k^{k+j-1} = x_k^{k+j-1}\right) \right] \right|. \end{aligned} \tag{3.27}$$

Pela inequação (3.27) e pelo Lema 3.4, temos

$$\begin{aligned} k_1 &\le \left| \sum_{x_k^{k+j-1} \in \mathcal{A}_1^j} \mathbb{P}\left(Z_k^{k+j-1} = \omega_1^j | X_k^{k+j-1} = x_k^{k+j-1}\right) \right| \sum_{l=0}^{j-1} \rho_{k-i+l,X} \\ &= \left| \sum_{y_k^{k+j-1}} \mathbb{P}\left(\bigcap_{k \le l \le k+j:\ y_l \ne 0} \xi_l = w_l\right) \right| \sum_{l=0}^{j-1} \rho_{k-i+l,X}, \end{aligned} \tag{3.28}$$

onde o último somatório é sobre todas as sequências $y_k^{k+j-1} \in \mathcal{A}_1^j$ satisfazendo (3.24). Agora, podemos mostrar que

$$\sum_{y_k^{k+j-1}} \mathbb{P}\left(\bigcap_{k \leq l \leq k+j:\ y_l \neq 0} \xi_l = w_l \right) \quad \leq \quad \frac{1}{(1-\varepsilon)^j}. \tag{3.29}$$

Portanto, através de (3.23), (3.28) e (3.29), obtemos

$$\left| \mathbb{P}\left(Z_k{}^{k+j-1} = \omega_1^j \mid X_1^i = x_1^i, \xi_1^i = \theta_1^i \right) - p_Z\left(\omega_1^j\right) \right| \leq \frac{\sum_{l=0}^{j-1} \rho_{k-i+l}}{(1-\xi)^j},$$

como queríamos. □

Enfatizamos que a técnica utilizada na demonstração dos próximos lemas desta seção foi inteiramente baseada nas demonstrações dos Lemas 10, 11 e 12 de Collet, Galves e Leonardi (2008). Por conta disto, omitiremos a prova de algumas identidades e inclusões que aparecem no restante desta seção.

A prova do próximo lema é uma consequência da Proposição 4 de Dedecker e Doukhan (2003). Para a conveniência do leitor, recordamos este resultado.

Proposição 3.6. *Seja $(Y_t)_{t \in \mathbb{N}}$ uma sequência de variáveis aleatórias centradas, quadrado integráveis e seja $\mathcal{M}_i$ a σ-algebra gerada por $Y_0, ..., Y_i$. Definimos $S_n = Y_0 + ... + Y_n$ e*

$$b_{i,n} = \max_{i \leq l \leq n} \left\| Y_i \sum_{k=i}^{l} \mathbb{E}(Y_k | \mathcal{M}_i) \right\|_{p/2}.$$

Então, para qualquer $p \geq 2$,

$$\|S_n\|_p \leq \left(2p \sum_{i=1}^{n} b_{i,n} \right)^{1/2}.$$

Lema 3.7. *Para toda sequência finita ω e para todo $t > 0$, temos*

$$\mathbb{P}\left(|N_n(\omega) - (n - l(\omega) + 1)p_Z(\omega)| > t \right) \leq e^{\frac{1}{e}} \exp\left[- \frac{-t^2(1-\varepsilon)^{l(\omega)}}{4e[n - l(\omega) + 1]l(\omega)\left(1 + \frac{\beta_X}{\alpha_X}\right)} \right].$$

Além disso, para todo $a \in \mathcal{A}$ e qualquer $n > \frac{|\mathcal{A}|+1}{tq(\omega)} + l(\omega)$, obtemos

$$\mathbb{P}\left(|\hat{p_{Z_n}}(a|\omega) - p_Z(a|\omega)| > t \right) \leq 3e^{\frac{1}{e}} \exp\left[-(n - l(\omega)) \frac{\left[t - \frac{3}{(n-l(\omega))p_Z(\omega)} \right]^2 p_Z(\omega)^2 (1-\varepsilon)^{l(\omega a)}}{64el(\omega a)\left(1 + \frac{\beta_X}{\alpha_X}\right)} \right].$$

Demonstração. Para toda sequência finita $\omega_1^j \in \mathcal{A}_1^j$, tem-se

$$N_n(\omega_1^j)$$

$$= \sum_{t=0}^{n-j} \prod_{1\le i\le j:\ \omega_i\neq 0} \left[\mathbf{1}_{\{X_{t+i}=\omega_i\}}\mathbf{1}_{\{\xi_{t+i}=1\}}\right] \prod_{1\le s\le j:\ \omega_i=0} \left[\mathbf{1}_{\{X_{t+s}=\omega_s\}} + \mathbf{1}_{\left\{X_{t+s}=\omega_i'\right\}}\mathbf{1}_{\{\xi_{s+t}=0\}}\right],$$

onde, como na Seção 2.1, $N_n(\omega_1^j)$ é o número de ocorrências da sequência ω_1^j numa amostra aleatória $Z_1, Z_2, \ldots, Z_n$ do processo (Z_t). Definimos o processo $\{U_t :\ t \in \mathbb{Z}\}$ por

$$U_t = \prod_{1\le i\le j:\ \omega_i\neq 0} \left[\mathbf{1}_{\{X_{t+i}=\omega_i\}}\mathbf{1}_{\{\xi_{t+i}=1\}}\right] \prod_{1\le s\le j:\ \omega_s=0} \left[\mathbf{1}_{\{X_{t+s}=\omega_s\}} + \mathbf{1}_{\{X_{t+s}=\omega_s'\}}\mathbf{1}_{\{\xi_{s+t}=0\}}\right] - p_Z(\omega_1^j),$$

onde $\omega_s' \neq \omega_s$, para $1 \le s \le j$. Denotamos por $\mathcal{M}_i$ a σ-algebra gerada por $U_0, ..., U_i$. Note que

$$\mathbb{E}\left(U_t\right) = \mathbb{E}\left(\mathbf{1}_{\left\{Z_{t+j}^{t+1}=\omega_j^1\right\}} - p_Z(\omega_1^j)\right) = 0,$$

ou seja, as variáveis U_t são centradas. Além disso, $\|U_t\|_{\frac{r}{2}} \le 1$. Agora, aplicando a Proposição 3.6, obtemos

$$\left\|N_n(\omega_1^j) - (n-j+1)p_Z(\omega_1^j)\right\|_r \le \left(2r\sum_{t=0}^{n-j} \max_{t\le l\le n-j} \left\|U_t \sum_{k=t}^{l} \mathbb{E}(U_k|\mathcal{M}_t)\right\|_{r/2}\right)^{1/2}, \tag{3.30}$$

para todo $r \ge 2$. Utilizando as desigualdades de Cauchy-Schwarz e triangular para a métrica $\|.\|_{r/2}$, temos

$$\begin{aligned} \left\|U_t \sum_{k=i}^{l} \mathbb{E}(U_k|\mathcal{M}_t)\right\|_{r/2} &\le \|U_t\|_{r/2} \left\|\sum_{k=t}^{l} \mathbb{E}(U_k|\mathcal{M}_t)\right\|_{r/2} \\ &\le \sum_{k=t}^{l} \|\mathbb{E}(U_k|\mathcal{M}_t)\|_{\infty}, \end{aligned}$$

pois, $\|.\|_p \le \|.\|_{pz} \le \ldots \le \|.\|_\infty$, para quaisquer $1 \le p \le q$, com $p,\ q \in \mathbb{N}$. Com isso, escrevemos

$$\left\|N_n(\omega_1^j) - (n-j+1)p_Z(\omega_1^j)\right\|_r \le \left(2r\sum_{t=0}^{n-j} \|U_t\|_{r/2} \sum_{k=t}^{n-j} \|\mathbb{E}(U_k|\mathcal{M}_t)\|_\infty\right)^{1/2}. \tag{3.31}$$

Segue, pela definição $\|.\|_\infty$, que

$$\begin{aligned}\|\mathbb{E}(U_k|\mathcal{M}_t)\|_\infty &= \sup_{\sigma_0^t\in\mathcal{A}_0^t}\left|\mathbb{E}(U_k|U_0^t=\sigma_0^t)\right| \\ &= \sup_{x_1^{t+j},\theta_1^{t+j}\in\mathcal{A}_1^{t+j}}\left|\mathbb{P}(Z_{k+j}^{k+1}=\omega_j^1|X_1^{t+j}=x_1^{t+j},\xi_1^{t+j}=\theta_1^{t+j})-p_Z(\omega_1^j)\right|.\end{aligned}$$

Da última igualdade, do Lema 3.5 e da desigualdade (3.31), temos

$$\left\|N_n(\omega_1^j)-(n-j+1)p_Z(\omega_1^j)\right\|_r \le \left[\frac{4r}{(1-\varepsilon)^j}(n-j+1)j\left(1+\frac{\beta_X}{\alpha_X}\right)\right]^{1/2}. \tag{3.32}$$

Agora, seja

$$B=\frac{4}{(1-\varepsilon)^j}(n-j+1)j\left(1+\frac{\beta_X}{\alpha_X}\right).$$

Como em Dedecker e Prieur (2005) obtemos, para todo $t>0$, que

$$\begin{aligned}\mathbb{P}\left(\left|N_n(\omega_1^j)-(n-j+1)p_Z(\omega_1^j)\right|>t\right) &\le \min\left(1,\frac{\mathbb{E}\left(\left|N_n(\omega_1^j)-(n-j+1)p_Z(\omega_1^j)\right|^r\right)}{t^r}\right) \\ &\le \min\left(1,\left[\frac{rB}{t^2}\right]^{\frac{r}{2}}\right).\end{aligned} \tag{3.33}$$

A função $r\to(cr)^{\frac{r}{2}}$, $c=\frac{B}{t^2}$, possui um mínimo em $r_0=\frac{1}{ec}$. Como em Galves e Leonard (2008), comparando o valor desta função com 1 e r_0 com 2 podemos inferir que

$$\min\left(1,\left[\frac{rB}{t^2}\right]^{\frac{r}{2}}\right)\le\exp\left\{-\frac{t^2}{eB}+e^{-1}\right\}. \tag{3.34}$$

Das desigualdades (3.33) e (3.34), segue que

$$\mathbb{P}\left(|N_n(\omega)-(n-l(\omega)+1)p_Z(\omega)|>t\right)\le e^{\frac{1}{e}}\exp\left\{-\frac{t^2(1-\varepsilon)^{l(\omega)}}{4e(n-l(\omega)+1)l(\omega)\left(1+\frac{\beta_X}{\alpha_X}\right)}\right\}.$$

Dessa forma, concluímos a primeira asserção deste lema. Para provar a segunda afirmação, observamos que

$$\left|p_Z(a|\omega)-\frac{(n-l(\omega))p_Z(\omega a)+1}{(n-l(\omega))p_Z(\omega)+|\mathcal{A}|}\right|\le\frac{|\mathcal{A}|+1}{(n-l(\omega))p_Z(\omega)}. \tag{3.35}$$

Agora, da desigualdade triagular, tem-se

$$|p_Z(a|\omega)-\hat{p_Z}(a|\omega)|\le\left|p_Z(a|\omega)-\frac{(n-l(\omega))p_Z(\omega a)+1}{(n-l(\omega))p_Z(\omega)+|\mathcal{A}|}\right|+\left|\hat{p_Z}(a|\omega)-\frac{(n-l(\omega))p_Z(\omega a)+1}{(n-l(\omega))p_Z(\omega)+|\mathcal{A}|}\right|,$$

onde,

$$\hat{p_{Z_n}}(a \mid \omega) = \frac{N_n(\omega a) + 1}{N_n(\omega .) + |\mathcal{A}|}.$$

Utilizando (3.35), escrevemos

$$|p_Z(a|\omega) - \hat{p_Z}(a|\omega)| \leq \frac{|\mathcal{A}| + 1}{(n - l(\omega))p_Z(\omega)} + \left| \hat{p_Z}(a|\omega) - \frac{(n - l(\omega))p_Z(\omega a) + 1}{(n - l(\omega))p_Z(\omega) + |\mathcal{A}|} \right|.$$

Pela última desigualdade, obtemos a inclusão entre eventos

$$\{|p_Z(a|\omega) - \hat{p_Z}(a|\omega)| > t\} \subset \left\{ \left| \hat{p_Z}(a|\omega) - \frac{(n - l(\omega))p_Z(\omega a) + 1}{(n - l(\omega))p_Z(\omega) + |\mathcal{A}|} \right| > t - \frac{|\mathcal{A}| + 1}{(n - l(\omega))p_Z(\omega)} \right\},$$

sempre que

$$t > \frac{|\mathcal{A}| + 1}{(n - l(\omega))p_Z(\omega)}.$$

Pela última inclusão, podemos escrever

$$\mathbb{P}\left(|p_Z(a|\omega) - \hat{p_Z}(a|\omega)| > t\right)$$

$$\leq \mathbb{P}\left(\left| \hat{p_Z}(a|\omega) - \frac{(n - l(\omega))p_Z(\omega a) + 1}{(n - l(\omega))p_Z(\omega) + |\mathcal{A}|} \right| > t - \frac{|\mathcal{A}| + 1}{(n - l(\omega))p_Z(\omega)} \right),$$

para todo $n \geq \frac{|\mathcal{A}|+1}{tq(\omega)} + l(\omega)$. Fazendo $t' = t - \frac{|\mathcal{A}|+1}{(n-l(\omega))p_Z(\omega)}$, vemos que

$$\mathbb{P}\left(\left| \hat{p_Z}(a|\omega) - \frac{(n - l(\omega))p_Z(\omega a) + 1}{(n - l(\omega))p_Z(\omega) + |\mathcal{A}|} \right| > t' \right)$$

$$\begin{aligned} \leq \quad & \mathbb{P}\left(|N_n(\omega a) - (n - l(\omega))p_Z(\omega a)| > \frac{t'}{2}\left[(n - l(\omega))p_Z(\omega) + |\mathcal{A}|\right] \right) \\ + \quad & \sum_{b \in \mathcal{A}} \mathbb{P}\left(|N_n(\omega b) - (n - l(\omega))p_Z(\omega b)| > \frac{t'}{2|\mathcal{A}|}\left[(n - l(\omega))p_Z(\omega) + |\mathcal{A}|\right] \right). \end{aligned}$$

Aplicando a primeira asserção deste lema, limitamos cada uma das parcelas no segundo membro da última inequação. Portanto,

$$\mathbb{P}\left(|p_Z(a|\omega) - \hat{p_Z}(a|\omega)| > t\right) \leq 3e^{\frac{1}{e}} \exp\left\{ -\left[t - \frac{3}{(n - l(\omega))p_Z(\omega)} \right]^2 \frac{[p_Z(\omega)]^2 (n - l(\omega))(1 - \varepsilon)^{l(\omega)+1}}{64e(l(\omega) + 1)\left(1 + \frac{\beta}{\alpha}\right)} \right\},$$

que é a segunda afirmação deste lema. $\square$

Lema 3.8. *Para todo* $\delta > 2\left[1+\frac{4\beta_X}{\min(\alpha_X\beta^*_X,1)}\right]\varepsilon$, *todo* $\omega \in \mathcal{T}_X$, $u\omega \in \hat{\mathcal{T}}_n^{\delta,d}$ *e*

$$n > \frac{6}{\left\{\delta - 2\varepsilon\left[1+\frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta^*_X)}\right]\right\}\alpha_X^d(1-\varepsilon)^d} + d$$

temos

$$\mathbb{P}(\Delta_n(u\omega) > \delta) \leq 12e^{\frac{1}{e}}\exp\left\{-(n-d)\frac{\left[\frac{\delta}{2}-\bar{k}\right]^2\alpha_X^{2d}(1-\varepsilon)^{3d+1}}{64e(d+1)\left(1+\frac{\beta_X}{\alpha_X}\right)}\right\},$$

onde

$$\bar{k} = \varepsilon\left[1+\frac{4\beta}{\min(1,(1+\varepsilon)\alpha_X\beta^*_X)}\right] + \frac{3}{(n-d)\alpha_X^d(1-\varepsilon)^d}.$$

Demonstração. Lembramos que

$$\Delta_n(u\omega) = \max_{a\in\mathcal{A}}\left|\hat{p_{Z_n}}(a|u\omega) - \hat{p_{Z_n}}(a|suf(u\omega))\right|.$$

Note, pela propriedade do sufixo (item 2 da Definição 2.1), que se $\omega \in \mathcal{T}_X$ então

$$p_X(a|u\omega) = p_X(a|suf(u\omega)),$$

para qualquer sequência finita u e qualquer símbolo $a \in \mathcal{A}$. Logo, pela desigualdade triangular, tem-se

$$\begin{aligned}|\hat{p_{Z_n}}(a|u\omega) - \hat{p_{Z_n}}(a|suf(u\omega))| &\leq |\hat{p_{Z_n}}(a|u\omega) - p_Z(a|u\omega)| + |p_X(a|u\omega) - p_Z(a|u\omega)| \\ &+ |p_X(a|suf(u\omega)) - p_Z(a|suf(u\omega))| \\ &+ |p_Z(a|suf(u\omega)) - \hat{p_Z}(a|suf(u\omega))|.\end{aligned}$$

Pelo Teorema 2.5, temos

$$|\hat{p_{Z_n}}(a|u\omega) - p_Z(a|u\omega)| + |p_X(a|u\omega) - p_Z(a|u\omega)|$$

$$\leq |\hat{p_{Z_n}}(a|u\omega) - p_Z(a|u\omega)| + \varepsilon\left[1+\frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta^*_X)}\right]$$

e

$$|p_Z(a|suf(u\omega)) - \hat{p_Z}(a|suf(u\omega))| + |p_X(a|suf(u\omega)) - p_Z(a|suf(u\omega))|$$

$$\leq |p_Z(a|suf(u\omega)) - \hat{p_Z}(a|suf(u\omega))| + \varepsilon\left[1+\frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta^*_X)}\right].$$

Logo, vale a seguinte desigualdade

$$\mathbb{P}\left(\max_{a\in\mathcal{A}}|\hat{p_{Z_n}}(a|u\omega) - \hat{p_{Z_n}}(a|suf(u\omega))| > \delta\right)$$

$$\leq \sum_{a\in\mathcal{A}} \left[\mathbb{P}\left(|\hat{p_{Z_n}}(a|u\omega) - p_Z(a|suf(u\omega))| > \frac{\delta}{2} - \varepsilon\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta_X^*)}\right]\right) \right.$$
$$\left. + \ \mathbb{P}\left(|p_Z(a|suf(u\omega)) - \hat{p_Z}(a|suf(u\omega))| > \frac{\delta}{2} - \varepsilon\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta_X^*)}\right]\right)\right],$$

para todo $\delta > 2\varepsilon\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta_X^*)}\right]$. Note que se $u\omega \in \hat{\mathcal{T}}_n^{\delta,d}$ então, pela Definição 2.4, temos

$$l(u\omega) = d \text{ e } l(suf(u\omega)) \leq d. \tag{3.36}$$

Considerando o Lema 3.2, segue que

$$p_Z(u\omega) \geq \alpha_X^d(1-\varepsilon)^d \text{ e } p_Z(suf(u\omega)) \geq \alpha_X^d(1-\varepsilon)^d.$$

Agora, utilizando o Lema 3.7 com $t = \frac{\delta}{2} - \varepsilon\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta_X^*)}\right]$, obtemos

$$\mathbb{P}\left(|\hat{p_{Z_n}}(a|u\omega) - p_Z(a|u\omega)| > \frac{\delta}{2} - \varepsilon\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta_X^*)}\right]\right)$$
$$\leq 3e^{\frac{1}{e}} \exp\left\{ -(n-d)\frac{\left[t - \frac{3}{(n-d)\alpha_X^d(1-\varepsilon)^d}\right]^2 \alpha_X^{2d}(1-\varepsilon)^{3d+1}}{64e(d+1)\left(1+\frac{\beta_X}{\alpha_X}\right)} \right\}, \tag{3.37}$$

e

$$\mathbb{P}\left(|p_Z(a|suf(u\omega)) - \hat{p_Z}(a|suf(u\omega))| > \frac{\delta}{2} - \varepsilon\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta_X^*)}\right]\right)$$
$$\leq 3e^{\frac{1}{e}} \exp\left\{ -(n-d)\frac{\left[t - \frac{3}{(n-d)\alpha_X^d(1-\varepsilon)^d}\right]^2 \alpha_X^{2d}(1-\varepsilon)^{3d+1}}{64e(d+1)\left(1+\frac{\beta_X}{\alpha_X}\right)} \right\}, \tag{3.38}$$

para

$$n > \frac{6}{\left\{\delta - 2\varepsilon\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta_X^*)}\right]\right\}(1-\varepsilon)^d\alpha_X^d} + d.$$

Considerando as cotas superiores para as probabilidades estabelecidas em (3.37) e (3.38), obtem-se

$$\mathbb{P}\left(\max_{a\in\mathcal{A}} |\hat{p_{Z_n}}(a|u\omega) - \hat{p_{Z_n}}(a|suf(u\omega))| > \delta \right)$$
$$\leq 12e^{\frac{1}{e}} \exp\left\{ -(n-d)\frac{\left[\frac{\delta}{2} - \bar{k}\right]^2 \alpha_X^{2d}(1-\varepsilon)^{3d+1}}{64e(d+1)\left(1+\frac{\beta_X}{\alpha_X}\right)} \right\},$$

onde,

$$\bar{k} = \varepsilon\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta_X^*)}\right] + \frac{3}{(n-d)\alpha_X^d(1-\varepsilon)^d},$$

como queríamos. $\square$

Lema 3.9. *Existe d tal que para todo*

$$\delta < D_d - 2\varepsilon \left[1 + \frac{4\beta_X}{\min\left(1, (1+\varepsilon)\alpha_X \beta_X^*\right)}\right],$$

todo $\omega \in \hat{\mathcal{T}}_n^{\delta,d}$, com $l(\omega) < K$, $\omega \notin \mathcal{T}_X$, e todo

$$n > \frac{6}{\left\{D_d - \delta - 2\varepsilon \left[1 + \frac{4\beta_X}{\min\left(1,(1+\varepsilon)\alpha_X\beta_X^*\right)}\right]\right\} \alpha_X^d (1-\varepsilon)^d} + d$$

temos

$$\mathbb{P}\left(\bigcap_{u\omega \in \mathcal{T}_X|_d} \{\Delta_n(u\omega) \leq \delta\}\right) \leq 6e^{\frac{1}{e}} \exp\left[-(n-d)\frac{\left[\frac{D_d-\delta}{2} - \bar{k}\right]^2 \alpha_X^{2d}(1-\varepsilon)^{3d+1}}{64e(d+1)\left(1+\frac{\beta_X}{\alpha_X}\right)}\right]$$

onde, como no Lema 3.8, temos

$$\bar{k} = \varepsilon\left[1 + \frac{4\beta_X}{\min\left(1, (1+\varepsilon)\alpha\beta^*\right)}\right] + \frac{3}{(n-d)\alpha_X^d(1-\varepsilon)^d}.$$

Demonstração. Como em Collet, Galves e Leonardi (2008), seja

$$d = \max_{u \notin \mathcal{T}_X,\ l(u)<K} \min\{k : \text{ existe } \omega \in C_k \text{ com } \omega \succ u\}.$$

Então, pelas definições de C_d e da constante d, existe $\bar{u\omega} \in \mathcal{T}_X|_d$ tal que

$$p_X(a|\bar{u\omega}) \neq p_X(a|suf(\bar{u\omega}))$$

para algum $a \in \mathcal{A}$. Sabemos que

$$\mathbb{P}\left(\bigcap_{u\omega \in \mathcal{T}_X|_d} \{\Delta_n(u\omega) \leq \delta\}\right) \leq \mathbb{P}\left(\Delta_n(\bar{u\omega}) \leq \delta\right). \tag{3.39}$$

Podemos ver que

$$\begin{aligned}
|\hat{p_{Z_n}}(a|\bar{u\omega}) - \hat{p_{Z_n}}(a|suf(\bar{u\omega}))| \;&\geq\; |p_X(a|\bar{u\omega}) - p_X(a|suf(\bar{u\omega}))| \\
&-\; |\hat{p_{Z_n}}(a|suf(\bar{u\omega})) - p_Z(a|suf(\bar{u\omega}))| \\
&-\; |p_Z(a|\bar{u\omega}) - p_X(a|\bar{u\omega})| - |\hat{p_{Z_n}}(a|\bar{u\omega}) - p_Z(a|\bar{u\omega})| \\
&-\; |p_Z(a|suf(\bar{u\omega})) - p_X(a|suf(\bar{u\omega}))|
\end{aligned}$$

para todo $a \in \mathcal{A}$. Lembramos que

$$D_d = \min_{u \in C_d} \max_{a \in \mathcal{A}} \{|p_X(a \mid u) - p_X(a \mid suf(u))|\}.$$

Assim, se $u\bar{\omega} \in C_d$, temos

$$\max_{a \in \mathcal{A}} |p_X(a|u\bar{\omega}) - p_X(a|suf(u\bar{\omega}))| \geq D_d.$$

Utilizando o Teorema 2.5, podemos escrever

$$\max_{a \in \mathcal{A}} |p_Z(a|u\bar{\omega}) - p_X(a|u\bar{\omega})| \leq \varepsilon \left[1 + \frac{4\beta_X}{\min(1, (1+\varepsilon)\alpha_X \beta_X^*)}\right]$$

e

$$\max_{a \in \mathcal{A}} |p_Z(a|suf(u\bar{\omega})) - p_X(a|suf(u\bar{\omega}))| \leq \varepsilon \left[1 + \frac{4\beta_X}{\min(1, (1+\varepsilon)\alpha_X \beta_X^*)}\right].$$

Logo,

$$\begin{aligned} \Delta_n(u\bar{\omega}) \;\; &\geq \;\; D_d - 2\varepsilon \left[1 + \frac{4\beta_X}{\min(1, (1+\varepsilon)\alpha_X \beta_X^*)}\right] \\ &- \;\; \max_{a \in \mathcal{A}} |\hat{p_Z}_n(a|u\bar{\omega}) - p_Z(a|u\bar{\omega})| - \max_{a \in \mathcal{A}} |\hat{p_Z}_n(a|suf(u\bar{\omega})) - p_Z(a|suf(u\bar{\omega}))|. \end{aligned}$$

Pela última desigualdade, tem-se

$$\begin{aligned} \mathbb{P}(\Delta_n(u\bar{\omega}) \leq \delta) \;\; &\leq \;\; \mathbb{P}(\bigcap_{a \in \mathcal{A}} \{|\hat{p_Z}_n(a|u\bar{\omega}) - p_Z(a|u\bar{\omega})| \geq \frac{D_d - \delta - 2\varepsilon \left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X \beta_X^*)}\right]}{2}\}) \\ &+ \;\; \mathbb{P}(\bigcap_{a \in \mathcal{A}} \{|\hat{p_Z}_n(a|u\bar{\omega}) - p_Z(a|u\bar{\omega})| \geq \frac{D_d - \delta - 2\varepsilon \left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X \beta_X^*)}\right]}{2}\}), \end{aligned}$$

sempre que

$$\delta < D_d - 2\varepsilon \left[1 + \frac{4\beta_X}{\min(1, (1+\varepsilon)\alpha_X \beta_X^*)}\right].$$

Finalmente, considerando a última cota superior obtida para $\mathbb{P}(\Delta_n(u\bar{\omega}) \leq \delta)$, a equação (3.39) e fazendo

$$t = \frac{D_d - \delta - 2\varepsilon \left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X \beta_X^*)}\right]}{2}$$

no Lema 3.7, temos

$$\mathbb{P}\left(\bigcap_{u\omega \in \mathcal{T}_X|_d} \{\Delta_n(u\omega) \leq \delta\}\right) \leq 6e^{\frac{1}{e}} \exp\left[-(n-d)\frac{c^2 \alpha_X^{2d}(1-\varepsilon)^{3d+1}}{64e(d+1)\left(1 + \frac{\beta_X}{\alpha_X}\right)}\right],$$

onde

$$c = \frac{D_d - \delta}{2} - \varepsilon\left[1 + \frac{4\beta_X}{\min\left(1, (1+\varepsilon)\alpha_X\beta_X^*\right)}\right] - \frac{3}{(n-d)\alpha_X^d(1-\varepsilon)^d}$$

e

$$n > \frac{6}{\left\{D_d - \delta - 2\varepsilon\left[1 + \frac{4\beta_X}{\min\left(1,(1+\varepsilon)\alpha_X\beta_X^*\right)}\right]\right\}\alpha_X^d(1-\varepsilon)^d} + d.$$

Com isso, concluímos a asserção deste lema. □

Demonstração do Teorema 2.7. *Como em Collet, Galves e Leonardi (2008), definimos*

$$O_{n,\delta}^{K,d} = \bigcup_{\substack{\omega \in \mathcal{T}_X \\ l(\omega)<K}} \bigcup_{u\omega \in \hat{\mathcal{T}}_n^{\delta,d}} \left\{\Delta_n(u\omega) > \delta\right\},$$

e

$$U_{n,\delta}^{K,d} = \bigcup_{\substack{\omega \in \hat{\mathcal{T}}_n^{\delta,d} \\ l(\omega)<K}} \bigcap_{u\omega \in \mathcal{T}_X|_d} \left\{\Delta_n(u\omega) \le \delta\right\}.$$

Então, se $d < n$, podemos ver que

$$\left\{\hat{\mathcal{T}}_n^{\delta,d}|_K \neq \mathcal{T}_X|_K\right\} \subseteq O_{n,\delta}^{K,d} \cup U_{n,\delta}^{K,d}.$$

Logo,

$$\mathbb{P}\left(\hat{\mathcal{T}}_n^{\delta,d}|_K \neq \mathcal{T}_X|_K\right) \le \sum_{\substack{\omega \in \mathcal{T}_X \\ l(\omega)<K}} \sum_{u\omega \in \hat{\mathcal{T}}_n^{\delta,d}} \mathbb{P}\left(\Delta_n(u\omega) > \delta\right) + \sum_{\substack{\omega \in \hat{\mathcal{T}}_n^{\delta,d} \\ l(\omega)<K}} \mathbb{P}\left(\bigcap_{u\omega \in \mathcal{T}_X|_d} \left\{\Delta_n(u\omega) \le \delta\right\}\right).$$

Observamos, pela Definição 2.4, que a soma

$$s_1 := \sum_{\substack{\omega \in \mathcal{T}_X \\ l(\omega)<K}} \sum_{u\omega \in \hat{\mathcal{T}}_n^{\delta,d}} \mathbb{P}\left(\Delta_n(u\omega) > \delta\right)$$

possui no máximo 2^{d-1} parcelas. Agora, pelo Lema 3.8, tem-se

$$s_1 \le 2^{d-1} 12 e^{\frac{1}{e}} \exp\left\{-(n-d)\frac{\left[\frac{\delta}{2} - \bar{k}\right]^2 \alpha_X^{2d}(1-\varepsilon)^{3d+1}}{64e(d+1)\left(1 + \frac{\beta_X}{\alpha_X}\right)}\right\},$$

onde

$$\bar{k} = \varepsilon\left[1 + \frac{4\beta_X}{\min\left(1, (1+\varepsilon)\alpha_X\beta_X^*\right)}\right] + \frac{3}{(n-d)\alpha_X^d(1-\varepsilon)^d},$$

$$n > \frac{3}{\left\{\delta - 2\varepsilon\left[1 + \frac{4\beta_X}{\min\left(1,(1+\varepsilon)\alpha_X\beta_X^*\right)}\right]\right\}\alpha_X^d(1-\varepsilon)^d} + d$$

e

$$\delta > 2\left(1 + \frac{4\beta_X}{\min(\alpha_X \beta_X^*, 1)}\right)\varepsilon.$$

Por outro lado, a soma

$$s_2 := \sum_{\substack{\omega \in \hat{\mathcal{T}}_n^{\delta,d} \\ l(\omega) < K}} \mathbb{P}\left(\bigcap_{u\omega \in \mathcal{T}_X|_d} \{\Delta_n(u\omega) \leq \delta\}\right)$$

possui no máximo 2^d *parcelas. Utilizando o Lema 3.9, temos*

$$s_2 \leq 2^d 6 e^{\frac{1}{e}} \exp\left[-(n-d)\frac{\left[\frac{D_d - \delta}{2} - \bar{k}\right]^2 \alpha_X^{2d}(1-\varepsilon)^{3d+1}}{64e(d+1)\left(1 + \frac{\beta_X}{\alpha_X}\right)}\right]$$

onde

$$n > \frac{6}{\left\{D_d - \delta - 2\varepsilon\left[1 + \frac{4\beta_X}{\min(1, (1+\varepsilon)\alpha_X\beta_X^*)}\right]\right\}\alpha_X^d(1-\varepsilon)^d} + d$$

e

$$\delta < D_d - 2\varepsilon\left[1 + \frac{4\beta_X}{\min(1, (1+\varepsilon)\alpha_X\beta_X^*)}\right].$$

Agora, sejam $c_1 = 2\left[1 + \frac{4\beta_X}{\min(1,(1+\varepsilon)\alpha_X\beta_X^*)}\right]$ *e*

$$n_0 = \frac{6}{\left\{D_d - \delta - 2\varepsilon\left[1 + \frac{4\beta_X}{\min(1, (1+\varepsilon)\alpha_X\beta_X^*)}\right]\right\}\alpha_X^d(1-\varepsilon)^d} + d.$$

Assim, para todo $\epsilon \in (0, D_d/2c_1)$, *para todo* $\delta \in (c_1\varepsilon, D_d - c_1\varepsilon)$ *e* $n > n_0$ *temos, considerando as cotas superiores para* s_1 *e* s_2 *obtidas acima, que*

$$\mathbb{P}\left(\hat{\mathcal{T}}_n^{\delta,d}|_K \neq \mathcal{T}_X|_K\right) \leq 2^d 12 e^{\frac{1}{e}} \exp\left[-(n-d)\frac{\left[\min(D_d - \delta, \delta) - 2\bar{k}\right]^2 \alpha_X^{2d}(1-\varepsilon)^{3d+1}}{256e(d+1)\left(1 + \frac{\beta_X}{\alpha_X}\right)}\right],$$

onde $\bar{k}$ *é como acima. Definindo as contantes* c_2 *e* c_3 *por*

$$c_2 = 2^d 12 e^{\frac{1}{e}},$$

$$c_3 = \frac{\left[\min(D_d - \delta, \delta) - 2\bar{k}\right]^2 \alpha_X^{2d}(1-\varepsilon)^{3d+1}}{256e(d+1)\left(1 + \frac{\beta_X}{\alpha_X}\right)}.$$

Portanto,

$$\mathbb{P}\left(\hat{\mathcal{T}}_n^{\delta,d}|_K \neq \mathcal{T}_X|_K\right) \leq c_2 \exp\left[-(n-d)c_3\right],$$

o que conclui a prova do Teorema 2.7.

Demonstração do Corolário 2.8. *Pelo Teorema 2.7, temos*

$$\sum_n \mathbb{P}\left(\hat{\mathcal{T}}_n^{\delta,d}\mid_K \neq \mathcal{T}_X\mid_K\right) \leq \sum_n c_2 e^{[-(n-d)c_3]} < \infty,$$

pois, as cotas para a probabilidade de erro na estimação da árvore de contextos truncada são somáveis em n para apropriadas escolhas de d e δ. Assim, pelo Lema Borel-Cantelli, obtemos

$$\mathbb{P}\left(\left[\hat{\mathcal{T}}_n^{\delta,d}\mid_K \neq \mathcal{T}_X\mid_K i.v.\right]\right) = 0,$$

ou, equivalentemente, para todo inteiro K e para quase toda amostra infinita $Z_1, Z_2, \ldots$ existe um $\overline{n}$ tal que, para todo $n > \overline{n}$ temos

$$\hat{\mathcal{T}}_n^{\delta,d}\mid_K = \mathcal{T}_X\mid_K.$$

3.3 Demonstração do Teorema 2.10

Recordamos que nesta e na próxima seção, o alfabeto $\mathcal{A}$ é o conjunto finito $\{0, 1, \ldots, N-1\}$ e o processo (Z_t) é definido por

$$Z_t = \begin{cases} X_t, & \text{se } \xi_t = 1, \\ Y_t, & \text{se } \xi_t = 0. \end{cases}$$

Para esse modelo, valem as seguintes equivalências

$$\{X_{-j} = \omega_{-j}, Y_{-j} = \omega_{-j}\} \subset \{Z_{-j} = \omega_{-j}\}, \tag{3.40}$$

$$\{X^0_{-j} = x^0_{-j}, Y^0_{-j} = y^0_{-j}, Z^0_{-j} = \omega^0_{-j}\} \tag{3.41}$$

$$= \left\{X^0_{-j} = x^0_{-j}, Y^0_{-j} = y^0_{-j}, \bigcap_{0\leq l\leq j:\ x_{-l}\neq y_{-l},\ \omega_{-l}=x_{-l}} \xi_{-l} = 1, \bigcap_{0\leq i\leq j:\ x_{-i}\neq y_{-i},\ \omega_{-i}=y_{-i}} \xi_{-i} = 0\right\},$$

$$\{X_{-j} = x_{-j}, Z_{-j} = x_{-j}\} = \{X_{-j} = x_{-j}, \xi_{-j} = 1\} \bigcup \{X_{-j} = x_{-j}, Y_{-j} = x_{-j}, \xi_{-j} = 0\}, \tag{3.42}$$

$$\{Y_{-j} = y_{-j}, Z_{-j} = y_{-j}\} = \{Y_{-j} = y_{-j}, \xi_{-j} = 0\} \bigcup \{Y_{-j} = y_{-j}, X_{-j} = y_{-j}, \xi_{-j} = 1\}, \tag{3.43}$$

$$\{Z_{-j} = z_{-j}\} = \{X_{-j} =_{-j}, \xi_{-j} = 1\} \bigcup \{Y_{-j} = z_{-j}, \xi_{-j} = 0\}, \tag{3.44}$$

para todo $j \in \mathbb{Z}$ e quaisquer sequências x^0_{-j}, y^0_{-j}, ω^0_{-j} $\mathcal{A}^0_{-j}$. Além disso, temos

$$\{X_{-j} = x_{-j}, Z_{-j} = z_{-j}\} = \{X_{-j} = x_{-j}, Y_{-j} = z_{-j}, \xi_{-j} = 0\}, \tag{3.45}$$

se $x_{-j} \neq z_{-j}$ e

$$\{Y_{-j} = y_{-j}, Z_{-j} = z_{-j}\} = \{Y_{-j} = y_{-j}, \xi_{-j} = 1, X_{-j} = z_{-j}\}, \tag{3.46}$$

se $y_{-j} \neq z_{-j}$. Essas equivalências serão utilizadas nas demonstrações dos próximos resultados.

A prova do Teorema 2.10 baseia-se nos três próximos lemas preparatórios.

Lema 3.10. *Para todo $\varepsilon \in (0,1)$, quaisquer $k > j \geq 0$, todo $\omega_{-\infty}^{0}$ e quaisquer $a, b \in \mathcal{A}$, temos*

$$\left|\mathbb{P}\left(X_0 = \omega_0 | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) - p_X(\omega_0|\omega_{-\infty}^{-1})\right| \leq \beta_{j,X}.$$

Demonstração. Para todo $j \geq 0$, condicionando nos valores de X_{-k}^{-j-2} e considerando a independência entre os processos (ξ_t), (Y_t) e (X_t), tem-se

$$\mathbb{P}\left(X_0 = \omega_0 | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) = \frac{p_1}{p_2}, \tag{3.47}$$

onde

$$p_1 := \sum_{u_{-k}^{-j-2}} p_X\left(u_{-k}^{-j-2} a \omega_{-j}^{-1} \omega_0\right) \mathbb{P}\left(Z_{-k}^{-j-1} = \omega_{-k}^{-j-2} b | X_{-k}^{-j-1} = u_{-k}^{-j-2} a\right), \tag{3.48}$$

$$p_2 := \sum_{u_{-k}^{-j-2}} p_X\left(u_{-k}^{-j-2} a \omega_{-j}^{-1}\right) \mathbb{P}\left(Z_{-k}^{-j-1} = \omega_{-k}^{-j-2} b | X_{-k}^{-j-1} = u_{-k}^{-j-2} a\right). \tag{3.49}$$

Para comprovar a validade de (3.47), fazemos $a = u_{-j-1}$ e $b = \omega_{-j-1}$ para obter

$$\begin{aligned} p_3 &:= \mathbb{P}\left(Z_{-k}^{-j-1} = \omega_{-k}^{-j-2} b, X_{-k}^{-j-1} = u_{-k}^{-j-2} a\right) \\ &= \mathbb{P}\left(\bigcap X_{-i} = u_{-i}, \bigcap Z_{-i} = u_{-i}, \bigcap X_{-l} = u_{-l}, \bigcap Z_{-l} = \omega_{-l}\right) \end{aligned} \tag{3.50}$$

onde os índices $j+1 \leq i \leq k$ são tais que $u_{-i} = \omega_{-i}$ e os índices $j+1 \leq l \leq k$ são tais que $u_{-l} \neq \omega_{-l}$. Por (3.45), obtemos

$$p_3 = \mathbb{P}\left(\bigcap X_{-i} = u_{-i}, \bigcap Z_{-i} = u_{-i}, \bigcap X_{-l} = u_{-l}, \bigcap Y_{-l} = \omega_{-l}, \bigcap \xi_{-l} = 0\right). \tag{3.51}$$

Pela equivalência (3.42) e por (3.51), tem-se

$$\begin{aligned} p_3 &= \mathbb{P}\left(\bigcap X_{-i} = u_{-i}, \bigcap \xi_{-i} = 1, \bigcap X_{-l} = u_{-l}, \bigcap Y_{-l} = \omega_{-l}, \bigcap \xi_{-l} = 0\right) \\ &+ \mathbb{P}\left(\bigcap X_{-i} = u_{-i}, \bigcap Y_{-i} = u_{-i}, \bigcap \xi_{-i} = 0, \bigcap X_{-l} = u_{-l}, \bigcap Y_{-l} = \omega_{-l}, \bigcap \xi_{-l} = 0\right). \end{aligned}$$

Considerando a última igualdade, (3.50) e a independência entre os processos (X_t), (Y_t) e (ξ_t), obtem-se

$$\begin{aligned}
p_{u_{-k}^{-j-2}} &:= \mathbb{P}\left(Z_{-k}^{-j-1} = \omega_{-k}^{-j-2}b|X_{-k}^{-j-1} = u_{-k}^{-j-2}a\right) \\
&= \mathbb{P}\left(\bigcap \xi_{-i} = 1, \bigcap Y_{-l} = \omega_{-l}, \bigcap \xi_{-l} = 0\right) \\
&+ \mathbb{P}\left(\bigcap Y_{-i} = u_{-i}, \bigcap \xi_{-i} = 0, \bigcap Y_{-l} = \omega_{-l}, \bigcap \xi_{-l} = 0\right).
\end{aligned} \tag{3.52}$$

Assim, por (3.48) e (3.52), temos

$$p_1 = \mathbb{P}\left(X_{-j}^0 = \omega_{-j}^{-1}\omega_0, X_{-j-1} = a, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right).$$

Analogamente, podemos mostrar que

$$p_2 = \mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = a, Z_{-j-1} = b, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)$$

e assim concluir a prova de (3.47). Como na Seção 3.1, temos que

$$p_X\left(\omega_0|\omega_{-\infty}^{-1}\right) - \beta_{j,X} \leq p_X\left(\omega_0|u_{-k}^{-j-2}a\omega_{-j}^{-1}\right) \leq p_X\left(\omega_0|\omega_{-\infty}^{-1}\right) + \beta_{j,X}, \tag{3.53}$$

e a afirmação deste lema segue diretamente. □

Observação 3.11. *Como o processo (Y_t) também satisfaz as condições de não nulidade e taxa de continuidade somável temos, para todo $\omega_{-\infty}^0 \in \mathcal{A}_{-\infty}^0$, todo $j \geq 0$ e qualquer $a \in \mathcal{A}$, que*

$$\inf_{v_{-\infty}^{-j-1}} p_Y\left(\omega_0|v_{-\infty}^{-j-1}\omega_{-j}^{-1}\right) \leq p_Y\left(\omega_0|u_{-k}^{-j-2}a\omega_{-j}^{-1}\right) \leq \sup_{v_{-\infty}^{-j-1}} p_Y\left(\omega_0|v_{-\infty}^{-j-1}\omega_{-j}^{-1}\right). \tag{3.54}$$

Lema 3.12. *Para todo $\varepsilon \in (0,1)$, todo $k \geq 0$ e qualquer ω_{-k}^0, temos*

$$p_Z\left(\omega_0|\omega_{-k}^{-1}\right) \geq \alpha_{min}, \tag{3.55}$$

$$\mathbb{P}\left(X_0 = \omega_0|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \geq \alpha_X \tag{3.56}$$

e

$$\mathbb{P}\left(Y_0 = \omega_0|Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \geq \alpha_Y, \tag{3.57}$$

onde $\alpha_{min} = \min\{\alpha_X, \alpha_Y\}$. Além disso, para todo $0 \leq j \leq k$ temos

$$\mathbb{P}\left(X_{-j-1} = \omega_{-j-1}|X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \geq \alpha_X\beta_X^* \tag{3.58}$$

e

$$\mathbb{P}\left(Y_{-j-1} = \omega_{-j-1}|Y_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \geq \alpha_Y\beta_Y^*. \tag{3.59}$$

Demonstração. Pela definição do processo (Z_t) (conforme (2.4)) e por (3.44), temos

$$p_Z\left(\omega_0|\omega_{-k}^{-1}\right) = (1-\varepsilon)\mathbb{P}\left(X_0=\omega_0|Z_{-k}^{-1}=\omega_{-k}^{-1}\right) + \varepsilon\mathbb{P}\left(Y_0=\omega_0|Z_{-k}^{-1}=\omega_{-k}^{-1}\right). \tag{3.60}$$

Afirmamos que, para provarmos a primeira asserção deste lema é suficiente comprovarmos a segunda e a terceira afirmações. Isto segue de (3.56), (3.57) e (3.60), pois,

$$p_Z\left(\omega_0|\omega_{-k}^{-1}\right) \geq (1-\varepsilon)\alpha_X + \varepsilon\alpha_Y \geq \alpha_{min}.$$

Agora, para atestar (3.56) iremos condicionar nos valores de X_{-l}^{-1}, Y_{-k}^{-1}, considerar a independência entre os processos (ξ_t), (X_t), (Y_t) e as desigualdades em (3.6), para obtermos

$$\mathbb{P}\left(X_0=\omega_0|Z_{-k}^{-1}=\omega_{-k}^{-1}\right) =$$

$$\lim_{l\to+\infty}\frac{\displaystyle\sum_{v_{-k}^{-1}}\sum_{u_{-l}^{-1}} p_X\left(\omega_0|\omega_{-\infty}^{-l-1}u_{-l}^{-1}\right)\mathbb{P}\left(X_{-l}^{-1}=u_{-l}^{-1}|X_{-\infty}^{-l-1}=\omega_{-\infty}^{-l-1}\right)p_Y(v_{-k}^{-1})p_\xi}{\displaystyle\sum_{v_{-k}^{-1}}\sum_{u_{-l}^{-1}}\mathbb{P}\left(X_{-l}^{-1}=u_{-l}^{-1}|X_{-\infty}^{-l-1}=\omega_{-\infty}^{l-1}\right)p_Y(v_{-k}^{-1})p_\xi},$$

onde

$$p_\xi = \mathbb{P}\left(\bigcap_{i,\ -k\leq -i\leq -1:\ u_{-i}\neq v_{-i},\ u_{-i}=\omega_{-i}}\xi_{-i}=1,\ \bigcap_{t,\ -k\leq -t\leq -1:\ u_{-t}\neq v_{-t},\ u_{-t}\neq\omega_{-t}}\xi_{-t}=0\right),$$

com as sequências $u_{-l}^{-1}\in\mathcal{A}_{-l}^{-1}$, $v_{-k}^{-1}\in\mathcal{A}_{-k}^{-1}$, satisfazendo

$$\mathbb{P}\left(X_{-l}^{-1}=u_{-l}^{-1}, Y_{-k}^{-1}=v_{-k}^{-1}, Z_{-k}^{-1}=\omega_{-k}^{-1}\right)\neq\emptyset.$$

Para cada l a expressão no lado direito é limitada inferiormente por α_X. Então, o mesmo vale para o limite quando $l\to+\infty$. De modo análogo, podemos comprovar (3.57). Agora, para provarmos (3.58) e (3.59), note que

$$\mathbb{P}\left(X_{-j-1}=\omega_{-j-1}|X_{-j}^{-1}=\omega_{-j}^{-1}, Z_{-k}^{-j-2}=\omega_{-k}^{-j-2}\right) \tag{3.61}$$

$$=\frac{\displaystyle\sum_{y_{-k}^{-j-2}}\sum_{x_{-k}^{-j-2}}\mathbb{P}\left(Z_{-k}^{-j-2}=\omega_{-k}^{-j-2}|X_{-k}^{-j-2}=x_{-k}^{-j-2}, Y_{-k}^{-j-2}=y_{-k}^{-j-2}\right)p_X\left(x_{-k}^{-j-2}\omega_{-j-1}^{-1}\right)p_Y(y_{-k}^{-j-2})}{\displaystyle\sum_{y_{-k}^{-j-2}}\sum_{x_{-k}^{-j-2}}\mathbb{P}\left(Z_{-k}^{-j-2}=\omega_{-k}^{-j-2}|X_{-k}^{-j-2}=x_{-k}^{-j-2}, Y_{-k}^{-j-2}=y_{-k}^{-j-2}\right)p_X\left(x_{-k}^{-j-2}\omega_{-j}^{-1}\right)p_Y(y_{-k}^{-j-2})},$$

onde as sequências x_{-k}^{-j-2}, $y_{-k}^{-j-2}\in\mathcal{A}_{-k}^{-j-2}$ são tais

$$\left\{X_{-k}^{-j-2}=x_{-k}^{-j-2}, Y_{-k}^{-j-2}=y_{-k}^{-j-2}, Z_{-k}^{-j-2}=\omega_{-k}^{-j-2}\right\}\neq\emptyset.$$

Para atestar (3.61), utilizamos (3.40) e a independência entre os processo (ξ_t), (X_t) e (Y_t) para obter

$$\begin{aligned} p_3 &= \mathbb{P}\left(X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \\ &= \sum_{\bar{y}_{-k}^{-j-2}} \sum_{\bar{x}_{-k}^{-j-2}} \mathbb{P}\left(Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}, X_{-k}^{-j-2} = \bar{x}_{-k}^{-j-2}, X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Y_{-k}^{-j-2} = \bar{y}_{-k}^{-j-2}\right) \\ &= \sum_{y_{-k}^{-j-2}} \sum_{x_{-k}^{-j-2}} \mathbb{P}\left(\bigcap \xi_{-l} = 1, \bigcap \xi_{-i} = 0\right) p_X\left(x_{-k}^{-j-2}\omega_{-j-1}^{-1}\right) p_Y(y_{-k}^{-j-2}) \end{aligned}$$

$$= \sum_{y_{-k}^{-j-2}} \sum_{x_{-k}^{-j-2}} \mathbb{P}\left(Z_{-k}^{-j-2} = \omega_{-k}^{-j-2} | X_{-k}^{-j-2} = x_{-k}^{-j-2}, Y_{-k}^{-j-2} = y_{-k}^{-j-2}\right) p_X\left(x_{-k}^{-j-2}\omega_{-j-1}^{-1}\right) p_Y(y_{-k}^{-j-2}),$$

onde $\bar{x}_{-k}^{-j-2}$, $\bar{y}_{-k}^{-j-2} \in \mathcal{A}_{-k}^{-j-2}$, os índices $j+2 \leq l \leq k$ são tais que $x_{-l} \neq y_{-l}$ e $\omega_{-l} = x_{-l}$, enquanto os índices $j+2 \leq i \leq k$ são tais que $x_{-i} \neq y_{-i}$. Analogamente, podemos comprovar que

$$\mathbb{P}\left(X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)$$

$$= \sum_{y_{-k}^{-j-2}} \sum_{x_{-k}^{-j-2}} \mathbb{P}\left(Z_{-k}^{-j-2} = \omega_{-k}^{-j-2} | X_{-k}^{-j-2} = x_{-k}^{-j-2}, Y_{-k}^{-j-2} = y_{-k}^{-j-2}\right) p_X\left(x_{-k}^{-j-2}\omega_{-j}^{-1}\right) p_Y(y_{-k}^{-j-2}),$$

e, com isso, atestar a validade de (3.61). Além disso, como na demonstração do Lema 3.2, temos

$$\frac{p_X\left(x_{-k}^{-j-2}\omega_{-j-1}^{-1}\right)}{p_X\left(x_{-k}^{-j-2}\omega_{-j}^{-1}\right)} \geq \alpha_X \beta_X^*, \tag{3.62}$$

pois o processo (X_t) satisfaz as hipóteses de não nulidade e taxa de continuidade somável. Logo, considerando a identidade (3.61) e a desigualdade (3.62), obtemos

$$\mathbb{P}\left(X_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \geq \alpha_X \beta_X^*.$$

De modo análogo, podemos mostrar que

$$\mathbb{P}\left(Y_{-j-1} = \omega_{-j-1} | Y_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \geq \alpha_Y \beta_Y^*.$$

Isto finaliza a prova do Lema 3.12. □

Lema 3.13. *Para todo $\varepsilon \in (0,1)$, quaisquer $k > j \geq 0$ e todo ω_{-k}^0, temos*

$$\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}\right) \leq \frac{\varepsilon}{\alpha\beta^*_{min}},$$

onde $\omega'_{-j-1} \neq \omega_{-j-1}$ e $\alpha\beta^*_{min} = \min\{\alpha_X \beta_X^*, \alpha_Y\}$.

Demonstração. Temos, como na demonstração do Lema 3.3, que

$$\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}\right)$$

$$\begin{aligned} &= \frac{\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1}, Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}{\mathbb{P}\left(Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)} \\ &\leq \frac{\varepsilon}{\mathbb{P}\left(Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right)}. \end{aligned} \tag{3.63}$$

Agora, pela independência entre os processos (X_t), (Y_t) e da desiguadade (3.57) do Lema 3.12, segue

$$\begin{aligned} \mathbb{P}\left(Y_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) &= \mathbb{P}\left(Y_{-j-1} = \omega_{-j-1} | Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \\ &\geq \alpha_Y, \end{aligned} \tag{3.64}$$

Assim, considerando (3.64), a inequação (3.58) do Lema 3.12 e a equivalência entre eventos (3.44), tem-se

$$\begin{aligned} \alpha\beta^*_{min} &\leq \varepsilon\mathbb{P}\left(Y_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \\ &+ (1-\varepsilon)\mathbb{P}\left(X_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right) \\ &= \mathbb{P}\left(Z_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}\right), \end{aligned} \tag{3.65}$$

onde $\alpha\beta^*_{min} = \min\{\alpha_X\beta^*_X, \alpha_Y\}$. Portanto, por (3.63) e (3.65), obtemos

$$\mathbb{P}\left(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}\right) \leq \frac{\varepsilon}{\alpha\beta^*_{min}}.$$

Isto conclui a prova do Lema 3.13. $\square$

Demonstração do Teorema 2.10. *Primeiramente, para todo $a \in \mathcal{A}$, para todo $\omega_{-k}^{-1} \in \mathcal{A}_{-k}^{-1}$ e pela equivalência entre eventos (3.44), temos*

$$p_Z(a|\omega_{-k}^{-1}) = (1-\varepsilon)\mathbb{P}\left(X_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}\right) + \varepsilon\mathbb{P}\left(Y_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}\right). \tag{3.66}$$

Portanto, das propriedades da função módulo e de (3.66), obtem-se

$$\left| p_Z\left(a|\omega_{-k}^{-1}\right) - \mathbb{P}\left(X_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}\right)\right| \leq 2\varepsilon. \tag{3.67}$$

Se $k = 0$ o Teorema 2.10 está provado. Para $k \geq 1$, seja

$$\begin{aligned} s_1 \quad := \quad & \sum_{j=0}^{k-1} \Big[\mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ - \quad & \mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}) \Big] . \end{aligned} \tag{3.68}$$

Então, como em Collet, Galves e Leonardi (2008), tem-se

$$s_1 = \mathbb{P}(X_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}) - \mathbb{P}(X_0 = a | X_{-k}^{-1} = \omega_{-k}^{-1}). \tag{3.69}$$

Denotaremos por $s_{2,j}$ cada paracela da soma (3.68), ou seja,

$$s_{2,j} \quad = \quad \mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) - \mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}).$$

Limitaremos superiormente cada termo $s_{2,j}$ da soma (3.68) separadamente. Para tanto, note, como em Collet, Galves e Leonardi (2008), que

$$\begin{aligned} s_{2,j} \quad = \quad & \sum_{b \in \mathcal{A}} \Big[\mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = b, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ - \quad & \mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}) \Big] \\ \times \quad & \mathbb{P}(X_{-j-1} = b | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}). \end{aligned} \tag{3.70}$$

Como $|\mathcal{A}| = N$, esta a soma (3.70) possui N parcelas. Utilizamos as propriedades da função módulo para limitarmos os termos em que $b = \omega'_{-j-1}$, com $\omega'_{-j-1} \neq \omega_{-j-1}$, por

$$\begin{aligned} c_{1,j} \quad := \quad & \Big| \mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = \omega'_{-j-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ - \quad & \mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}) \Big| \\ \times \quad & \mathbb{P}(X_{-j-1} = \omega'_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}). \end{aligned} \tag{3.71}$$

Por (3.71), pelos Lemas 3.10 e 3.13, obtemos

$$c_{1,j} \leq \frac{2\varepsilon \beta_{j,X}}{\alpha \beta^*_{min}}, \tag{3.72}$$

onde, em (3.72), consideramos as desigualdades em (3.6) e (3.54). Agora, a parcela correspondendo a $b = \omega_{-j-1}$ na soma (3.70), pode ser limitada superiormente por

$$\begin{aligned} c_{2,j} \quad := \quad & \Big| \mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = \omega_{-j-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ - \quad & \mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}) \Big| \\ \times \quad & \mathbb{P}(X_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}). \end{aligned} \tag{3.73}$$

Como em Collet, Galves e Leonardi (2008), temos que $c_{2,j} \leq s_{3,j}$, *onde*

$$\begin{aligned} s_{3,j} &:= \sum_{c \in \mathcal{A}} \Big| \mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = \omega_{-j-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ &- \mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}, Z_{-j-1} = c) \Big| \\ &\times \mathbb{P}(X_{-j-1} = \omega_{-j-1} | X_{-j}^{-1} = \omega_{-j}^{-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ &\times \mathbb{P}(Z_{-j-1} = c | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}). \end{aligned} \tag{3.74}$$

Note que a parcela de (3.74) correspondendo a $c = \omega_{-j-1}$ *é nula. Podemos limitar superiormente os termos do somatório (3.74) em que* $c = \omega'_{-j-1}$, *com* $\omega'_{-j-1} \neq \omega_{-j-1}$, *por*

$$\begin{aligned} c_{3,j} &:= \Big| \mathbb{P}(X_0 = a | X_{-j}^{-1} = \omega_{-j}^{-1}, X_{-j-1} = \omega_{-j-1}, Z_{-k}^{-j-1} = \omega_{-k}^{-j-1}) \\ &- \mathbb{P}(X_0 = a | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}, Z_{-j-1} = \omega'_{-j-1}) \Big| \\ &\times \mathbb{P}(Z_{-j-1} = \omega'_{-j-1} | X_{-j-1}^{-1} = \omega_{-j-1}^{-1}, Z_{-k}^{-j-2} = \omega_{-k}^{-j-2}). \end{aligned} \tag{3.75}$$

Por (3.75), pelo Lema 3.10, pelas desigualdades em (3.6) e (3.54), obtemos

$$s_{3,j} \leq 2\varepsilon\beta_{j,X}. \tag{3.76}$$

Assim, de (3.68), (3.70), (3.72) e (3.76), tem-se

$$s_1 \leq \sum_{j=0}^{k-1} \left[(N-1) \frac{2\varepsilon\beta_{j,X}}{\alpha\beta^*_{min}} + (N-1) 2\varepsilon\beta_{j,X} \right]. \tag{3.77}$$

Da desigualdade triangular, obtemos

$$\begin{aligned} \left| p_Z\left(a|\omega_{-k}^{-1}\right) - p_X\left(a|\omega_{-k}^{-1}\right) \right| &\leq \left| p_Z\left(a|\omega_{-k}^{-1}\right) - \mathbb{P}\left(X_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \right| + \\ &+ \left| \mathbb{P}\left(X_0 = a | X_{-k}^{-1} = \omega_{-k}^{-1}\right) - \mathbb{P}\left(X_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}\right) \right|. \end{aligned}$$

Portanto, considerando a última desigualdade, (3.67), (3.69) e (3.77), obtemos

$$\left| \mathbb{P}(X_0 = a | Z_{-k}^{-1} = \omega_{-k}^{-1}) - \mathbb{P}(X_0 = a | X_{-k}^{-1} = \omega_{-k}^{-1}) \right| \leq 2\varepsilon + (N-1) \left[\frac{2\varepsilon\beta_X}{\alpha\beta^*_{min}} + 2\varepsilon\beta_X \right]$$

e o teorema está provado.

3.4 Demonstração do Teorema 2.11

Lembramos, primeiramente, que tanto o processo (X_t) quanto o processo (Y_t) satisfazem as hipóteses de não nulidade e taxa de continuidade somável. Por isso, utilizando o Lema 3.4 existem sequências somáveis, digamos $(\rho_{l,X})_{l\in\mathbb{N}}$ e $(\rho_{j,Y})_{j\in\mathbb{N}}$, satisfazendo

$$\sum_{l\geq 1}\rho_{l,X} \leq 1+\frac{2\beta_X}{\alpha_X},$$

$$\sum_{s\geq 1}\rho_{s,Y} \leq 1+\frac{2\beta_Y}{\alpha_Y},$$

tais que para todo $i \geq 1$, todo $k > i$, todo $j \geq 1$ e qualquer sequência finita ω_1^j, valem as seguintes desigualdades

$$\sup_{x_1^i\in\mathcal{A}^i}\left|\mathbb{P}\left({X_k}^{k+j-1}=\omega_1^j \mid X_1^i = x_1^i\right) - p_X\left(\omega_1^j\right)\right| \leq \sum_{l=0}^{j-1}\rho_{k-i+l,X},$$

$$\sup_{y_1^i\in\mathcal{A}^i}\left|\mathbb{P}\left({Y_k}^{k+j-1}=\omega_1^j \mid Y_1^i = y_1^i\right) - p_Y\left(\omega_1^j\right)\right| \leq \sum_{s=0}^{j-1}\rho_{k-i+s,Y},$$

sendo que as constantes α_X, α_Y, β_X e β_Y são as mesmas que definimos na Seção 2.1. O restante da demonstração do Teorema 2.11 baseia-se na prova de quatro lemas.

As desigualdades acima serão utilizadas para provar o seguinte resultado envolvendo as mesmas quantidades α_X, α_Y, β_X, β_Y, $(\rho_{l,X})_{l\in\mathbb{N}}$ e $(\rho_{j,Y})_{j\in\mathbb{N}}$.

Lema 3.14. *Existem duas sequências somáveis $(\rho_{l,X})_{l\in\mathbb{N}}$ e $(\rho_{s,Y})_{s\in\mathbb{N}}$, satisfazendo*

$$\sum_{l\geq 1}\rho_{l,X} \leq 1+\frac{2\beta_X}{\alpha_X},$$

$$\sum_{s\geq 1}\rho_{s,Y} \leq 1+\frac{2\beta_Y}{\alpha_Y},$$

tais que para todo $i \geq 1$, todo $k \geq i$, todo $j \geq 1$ e qualquer sequência finita ω_1^j, vale a seguinte desigualdade

$$\sup_{x_1^i,\ y_1^i\in\mathcal{A}^i;\ \theta_1^i\in\{0,1\}^i}\left|\mathbb{P}\left(Z_k^{k+j-1}=\omega_1^j \mid X_1^i = x_1^i, Y_1^i = y_1^i, \xi_1^i=\theta_1^i\right) - p_Z\left(\omega_1^j\right)\right| \leq \frac{2\sum_{l=0}^{j-1}\rho_{k-i+l,\ max}}{(1-\varepsilon)^j},$$

onde

$$\sum_{l=0}^{j-1}\rho_{k-i+l,\ max} = \max\left\{\sum_{l=0}^{j-1}\rho_{k-i+l,X}, \sum_{s=0}^{j-1}\rho_{k-i+s,Y}\right\}.$$

Demonstração. Para quaisquer $x_1^i,\ y_1^i \in \mathcal{A}^i$ e qualquer $\theta_1^i \in \{0,1\}^i$ temos, pela independência entre os processos (X_t), (Y_t) e (ξ_t), que

$$\left|\mathbb{P}\left({Z_k}^{k+j-1}=\omega_1^j \mid X_1^i=x_1^i, Y_1^i=y_1^i, \xi_1^i=\theta_1^i\right)-p_Z\left(\omega_1^j\right)\right|=$$

$$|\sum \mathbb{P}\left(X_k^{k+j-1}=x_k^{k+j-1}, Y_k^{k+j-1}=y_k^{k+j-1}, Z_k^{k+j-1}=\omega_1^j | X_1^i=x_1^i, Y_1^i=y_1^i, \xi_1^i=\theta_1^i\right)-p_Z\left(\omega_1^j\right)|$$

$$=\sum \mathbb{P}\left(Z_k^{k+j-1}=\omega_1^j | X_k^{k+j-1}=x_k^{k+j-1}, Y_k^{k+j-1}=y_k^{k+j-1}\right)$$

$$\times|\mathbb{P}\left(X_k^{k+j-1}=x_k^{k+j-1}, Y_k^{k+j-1}=y_k^{k+j-1} | X_1^i=x_1^i, Y_1^i=y_1^i\right)-p_X(x_k^{k+j-1})p_Y(y_k^{k+j-1})|,$$

onde os somatórios acima são sobre x_k^{k+j-1}, $y_k^{k+j-1} \in \mathcal{A}_1^j$. Somando e subtraindo

$$p_Y(y_k^{k+j-1}|y_1^i)p_X(x_k^{k+j-1})$$

nos seus correspondentes termos em módulo e considerando o Lema 3.4, a primeira expressão pode ser limitada superiormente por

$$\frac{2\sum_{l=0}^{j-1}\rho_{k-i+l,\ max}}{(1-\xi)^j}.$$

Isto conclui a prova deste lema. □

À partir da demontração do último resultado e das definições da Seção 2.1, aplicaremos a mesma técnica utilizada na demonstração do Teorema 2.7.

A prova do próximo lema baseia-se na Proposição 4 de Dedecker e Doukhan (2003) que foi enunciada na Seção 3.2 (ver Proposição 3.6).

Lema 3.15. *Para toda sequência finita ω e todo $t>0$, temos*

$$\mathbb{P}\left(|N_n(\omega)-(n-l(\omega)+1)p_Z(\omega)|>t\right)\leq e^{\frac{1}{e}}\exp\left[-\frac{-t^2(1-\varepsilon)^{l(\omega)}}{4e[n-l(\omega)+1]l(\omega)\left(1+\frac{\beta}{\alpha}\right)_{max}}\right].$$

Além disso, para todo $a\in\mathcal{A}$ e para todo $n>\frac{N+1}{tq(\omega)}+l(\omega)$, temos

$$\mathbb{P}\left(|\hat{p_{Z}}_n(a|\omega)-p_Z(a|\omega)|>t\right)$$

$$\leq (N+1)e^{\frac{1}{e}}\exp\left\{-\left[t-\frac{N+1}{(n-l(\omega))p_Z(\omega)}\right]^2\frac{[p_Z(\omega)]^2(n-l(\omega))(1-\varepsilon)^{l(\omega a)}}{32N^2el(\omega a)\left(1+\frac{\beta}{\alpha}\right)_{max}}\right\}.$$

Demonstração. Observamos, para toda sequência finita $\omega_1^j\in\mathcal{A}^j$, que

$$N_n(\omega_1^j)=\sum_{t=0}^{n-j}\prod_{1\leq i\leq j}\left[\mathbf{1}_{\{X_{t+i}=\omega_i\}}\mathbf{1}_{\{\xi_{t+i}=1\}}+\mathbf{1}_{\{Y_{t+i}=\omega_i\}}\mathbf{1}_{\{\xi_{t+i}=0\}}\right].$$

Definimos o processo $\{U_t : \ t \in \mathbb{Z}\}$ por

$$U_t = \prod_{1 \leq i \leq j} \left[\mathbf{1}_{\{X_{t+i}=\omega_i\}} \mathbf{1}_{\{\xi_{t+i}=1\}} + \mathbf{1}_{\{Y_{t+i}=\omega_i\}} \mathbf{1}_{\{\xi_{t+i}=0\}} \right] - p_Z(\omega_1^j).$$

e denotamos por $\mathcal{M}_i$ a σ-algebra gerada por $U_0, ..., U_i$. Aplicando a Proposição 3.6, obtemos

$$\begin{aligned}
\left\| N_n(\omega_1^j) - (n-j+1)p_Z(\omega_1^j) \right\|_r &\leq \left(2r \sum_{t=0}^{n-j} \max_{t \leq l \leq n-j} \left\| U_t \sum_{k=t}^{l} \mathbb{E}(U_k | \mathcal{M}_t) \right\|_{r/2} \right)^{1/2} \\
&\leq \left(2r \sum_{t=0}^{n-j} \|U_t\|_{r/2} \sum_{k=t}^{l} \|\mathbb{E}(U_k | \mathcal{M}_t)\|_\infty \right)^{1/2},
\end{aligned}$$

para todo $r \geq 2$. Note que $\|U_t\|_{r/2} \leq 1$. Por outro lado, para quaisquer $x_1^{t+j}, y_1^{t+j} \in \mathcal{A}_1^{t+j}$ e qualquer $\theta_1^{t+j} \in \{0,1\}^{t+j}$, temos

$$\begin{aligned}
\|\mathbb{E}(U_k | \mathcal{M}_t)\|_\infty &= \sup_{\sigma_0^t} \left| \mathbb{E}(U_k | U_0^t = \sigma_0^t) \right| \\
&= \sup_{x_1^{t+j},\ y_1^{t+j},\ \theta_1^{t+j}} \left| \mathbb{E}(U_k | X_1^{t+j} = x_1^{t+j}, Y_1^{t+j} = y_1^{t+j}, \xi_1^{t+j} = \theta_1^{t+j}) \right|
\end{aligned}$$

$$= \sup_{x_1^{t+j},\ y_1^{t+j},\ \theta_1^{t+j}} \left| \mathbb{P}(Z_{k+j}^{k+1} = \omega_j^1 | X_1^{t+j} = x_1^{t+j}, X_1^{t+j} = y_1^{t+j}, \xi_1^{t+j} = \theta_1^{t+j}) - p_Z(\omega_1^j) \right|.$$

Assim, pelo Lema 3.14, obtemos a seguinte desigualdade

$$\|N_n(\omega) - (n-j+1)p_Z(\omega)\|_r \leq \left[\frac{8r}{(1-\varepsilon)^{l(\omega)}} (n - l(\omega) + 1) l(\omega) \left(1 + \frac{\beta}{\alpha}\right)_{max} \right]^{1/2},$$

sendo $\left(1 + \frac{\beta}{\alpha}\right)_{max} = \max\left\{ \left(1 + \frac{\beta_X}{\alpha_X}\right), \left(1 + \frac{\beta_Y}{\alpha_Y}\right) \right\}$. Seja

$$B = \frac{8}{(1-\varepsilon)^{l(\omega)}} (n - l(\omega) + 1) l(\omega) \left(1 + \frac{\beta}{\alpha}\right)_{max}.$$

Logo, como em Dedecker and Prieur (2005), temos

$$\begin{aligned}
\mathbb{P}\left(\left| N_n(\omega_1^j) - (n-j+1)p_Z(\omega_1^j) \right| > t \right) &\leq \min\left(1, \frac{\mathbb{E}\left(\left| N_n(\omega_1^j) - (n-j+1)p_Z(\omega_1^j) \right|^r \right)}{t^r} \right) \\
&\leq \min\left(1, \left[\frac{rB}{t^2} \right]^{\frac{r}{2}} \right).
\end{aligned}$$

para todo $t > 0$. A função $r \to (cr)^{\frac{r}{2}}$ possui um ponto de mínimo em $r_0 = \frac{1}{ec}$. Portanto, como em

Galves e Leonard (2008), comparando o valor desta função com 1 e r_0 com 2, podemos inferir que

$$\min\left(1, \left[\frac{rB}{t^2}\right]^{\frac{r}{2}}\right) \leq \exp\left\{-\frac{t^2}{eB} + e^{-1}\right\}.$$

Portanto,

$$\mathbb{P}\left(|N_n(\omega) - (n - l(\omega) + 1)p_Z(\omega)| > t\right) \leq e^{\frac{1}{e}} \exp\left\{-\frac{t^2(1-\varepsilon)^{l(\omega)}}{8e(n - l(\omega) + 1)l(\omega)\left(1 + \frac{\beta}{\alpha}\right)_{max}}\right\}.$$

Para provar a segunda afirmação deste lema, note que

$$\left|p_Z(a|\omega) - \frac{(n - l(\omega))p_Z(\omega a) + 1}{(n - l(\omega))p_Z(\omega) + |\mathcal{A}|}\right| \leq \frac{|\mathcal{A}| + 1}{(n - l(\omega))p_Z(\omega)}.$$

Assim, para todo $n \geq \frac{|\mathcal{A}|+1}{tq(\omega)} + l(\omega)$, temos

$$\mathbb{P}\left(|p_Z(a|\omega) - \hat{p_Z}(a|\omega)| > t\right)$$

$$\leq \mathbb{P}\left(\left|\frac{N_n(\omega a) + 1}{N_n(\omega .) + |\mathcal{A}|} - \frac{(n - l(\omega))p_Z(\omega a) + 1}{(n - l(\omega))p_Z(\omega) + |\mathcal{A}|}\right| > t - \frac{|\mathcal{A}| + 1}{(n - l(\omega))p_Z(\omega)}\right).$$

Seja $t' = t - \frac{|\mathcal{A}|+1}{(n-l(\omega))p_Z(\omega)}$. Logo,

$$\mathbb{P}\left(\left|\frac{N_n(\omega a)+1}{N_n(\omega .)+|\mathcal{A}|}-\frac{(n-l(\omega))p_Z(\omega a)+1}{(n-l(\omega))p_Z(\omega)+|\mathcal{A}|}\right|>t'\right)$$

$$\begin{aligned} \leq\ & \mathbb{P}\left(|N_n(\omega a)-(n-l(\omega))p_Z(\omega a)|>\frac{t'}{2}\left[(n-l(\omega))p_Z(\omega)+|\mathcal{A}|\right]\right) \\ +\ & \sum_{b\in\mathcal{A}}\mathbb{P}\left(|N_n(\omega b)-(n-l(\omega))p_Z(\omega b)|>\frac{t'}{2|\mathcal{A}|}\left[(n-l(\omega))p_Z(\omega)+|\mathcal{A}|\right]\right). \end{aligned}$$

Agora, podemos aplicar a cota referente à primeira asserção deste lema para limitar superiormente a última soma por

$$(N+1)e^{\frac{1}{e}}\exp\left\{-\left[t-\frac{N+1}{(n-l(\omega))p_Z(\omega)}\right]^2\frac{[p_Z(\omega)]^2(n-l(\omega))(1-\varepsilon)^{l(\omega a)}}{32N^2e(l(\omega a))\left(1+\frac{\beta}{\alpha}\right)_{max}}\right\}.$$

□

Lema 3.16. *Para todo* $\delta>4\left[1+\frac{2(N-1)\beta_X}{\min(1,\alpha\beta^*_{min})}\right]\varepsilon$*, todo* $\omega\in\mathcal{T}_X$*, todo* $u\omega\in\hat{\mathcal{T}}_n^{\delta,d}$ *e*

$$n>\frac{2(N+1)}{\left\{\delta-4\varepsilon\left[1+\frac{2(N-1)\beta_X}{\min(1,\alpha\beta^*_{min})}\right]\right\}\alpha^d_{min}}+d$$

temos

$$\mathbb{P}(\Delta_n(u\omega)>\delta)\leq 2N(N+1)e^{\frac{1}{e}}\exp\left\{-(n-d)\frac{\left[\frac{\delta}{2}-\bar{k}\right]^2\alpha^{2d}_{min}(1-\varepsilon)^{d+1}}{32N^2e(d+1)\left(1+\frac{\beta}{\alpha}\right)_{max}}\right\},$$

onde

$$\bar{k}=2\varepsilon\left[1+\frac{2(N-1)\beta_X}{\min(1,\alpha\beta^*_{min})}\right]+\frac{N+1}{(n-d)\alpha^d_{min}}.$$

Demonstração. Recordamos que (ver Seção 2.1)

$$\Delta_n(u\omega)=\max_{a\in\mathcal{A}}\left|\hat{p_{Z_n}}(a|u\omega)-\hat{p_{Z_n}}(a|suf(u\omega))\right|.$$

Note que o fato $\omega\in\mathcal{T}_X$ implica que para toda sequência finita u e todo símbolo $a\in\mathcal{A}$ temos $p_X(a|u\omega)=p_X(a|suf(u\omega))$. Daí,

$$\begin{aligned} \left|\hat{p_{Z_n}}(a|u\omega)-\hat{p_{Z_n}}(a|suf(u\omega))\right| \leq\ & \left|\hat{p_{Z_n}}(a|u\omega)-p_Z(a|u\omega)\right|+\left|p_X(a|u\omega)-p_Z(a|u\omega)\right| \\ +\ & \left|p_X(a|suf(u\omega))-p_Z(a|suf(u\omega))\right| \\ +\ & \left|p_Z(a|suf(u\omega))-\hat{p_Z}(a|suf(u\omega))\right|. \end{aligned}$$

Considerando a última desigualdade e o Teorema 2.10, temos

$$\mathbb{P}\left(\Delta_n(u\omega) > \delta\right)$$

$$\leq \sum_{a\in\mathcal{A}} \left[\mathbb{P}\left(|\hat{p_{Z_n}}(a|u\omega) - p_Z(a|suf(u\omega))| > \frac{\delta}{2} - 2\varepsilon\left[1 + \frac{2(N-1)\beta_X}{\min\left(1, \alpha\beta^*_{min}\right)}\right]\right)\right.$$
$$\left. + \, \mathbb{P}\left(|p_Z(a|suf(u\omega)) - \hat{p_Z}(a|suf(u\omega))| > \frac{\delta}{2} - 2\varepsilon\left[1 + \frac{2(N-1)\beta_X}{\min\left(1, \alpha\beta^*_{min}\right)}\right]\right)\right].$$

Agora, para

$$n > \frac{2(N+1)}{\left\{\delta - 4\varepsilon\left[1 + \frac{2(N-1)\beta_X}{\min\left(1, \alpha\beta^*_{min}\right)}\right]\right\}\alpha^d_{min}} + d$$

podemos, utilizando o Lema 3.15, limitar superiormente $\mathbb{P}\left(\Delta_n(u\omega) > \delta\right)$ por

$$\mathbb{P}(\Delta_n(u\omega) > \delta) \leq 2N(N+1)e^{\frac{1}{e}} \exp\left\{-(n-d)\frac{\left[\frac{\delta}{2} - \bar{k}\right]^2 \alpha^{2d}_{min}(1-\varepsilon)^{d+1}}{32N^2 e(d+1)\left(1+\frac{\beta}{\alpha}\right)_{max}}\right\}.$$

□

No próximo resultado, D_k e C_k são como definimos na Seção 2.1.

Lema 3.17. *Existe d tal que para todo $\delta < D_d - 4\varepsilon\left[1 + \frac{2(N-1)\beta_X}{\min\left(1,\ \alpha\beta^*_{min}\right)}\right]$, todo $\omega \in \hat{\mathcal{T}}^{\delta,d}_n$, com $l(\omega) < K$, $\omega \notin \mathcal{T}_X$, e todo*

$$n > \frac{2(N+1)}{\left\{D_d - \delta - 4\varepsilon\left[1 + \frac{2(N-1)\beta_X}{\min\left(1, \alpha\beta^*_{min}\right)}\right]\right\}\alpha^d} + d$$

temos

$$\mathbb{P}\left(\bigcap_{u\omega\in T|_d} \{\Delta_n(u\omega) \leq \delta\}\right) \leq 2(N+1)e^{\frac{1}{e}} \exp\left[-(n-d)\frac{\left[D_d - \delta - \bar{k}\right]^2 \alpha^{2d}_{min}(1-\varepsilon)^d}{128N^2 e(d+1)\left(1+\frac{\beta}{\alpha}\right)_{max}}\right]$$

onde $\bar{k}$ é como no Lema 3.16.

Demonstração. Seja

$$d = \max_{u\notin\mathcal{T}_X,\ l(u)<K} \min\{k:\ \text{existe } \omega \in C_k \text{ com } \omega \succ u\}.$$

Logo, existe $u\bar{\omega} \in \mathcal{T}_X|_d$ tal que $p_X(a|u\bar{\omega}) \neq p_X(a|suf(u\bar{\omega}))$ para algum $a \in \mathcal{A}$. Temos que

$$\mathbb{P}(\bigcap_{u\omega\in T|_d} \{\Delta_n(u\omega) \leq \delta\}) \leq \mathbb{P}\left(\Delta_n(u\bar{\omega}) \leq \delta\right).$$

Observamos, para todo $a \in \mathcal{A}$, que

$$\begin{aligned}
|\hat{p_{Z_n}}(a|u\bar{\omega}) - \hat{p_{Z_n}}(a|suf(u\bar{\omega}))| &\geq |p_X(a|u\bar{\omega}) - p_X(a|suf(u\bar{\omega}))| \\
&- |\hat{p_{Z_n}}(a|suf(u\bar{\omega})) - p_Z(a|suf(u\bar{\omega}))| \\
&- |p_Z(a|u\bar{\omega}) - p_X(a|u\bar{\omega})| - |\hat{p_{Z_n}}(a|u\bar{\omega}) - p_Z(a|u\bar{\omega})| \\
&- |p_Z(a|suf(u\bar{\omega})) - p_X(a|suf(u\bar{\omega}))| .
\end{aligned}$$

Assim, podemos escrever

$$\begin{aligned}
\Delta_n(u\bar{\omega}) &\geq D_d - 4\varepsilon\left[1 + \frac{2(N-1)\beta_X}{\min(1, \alpha\beta^*_{min})}\right] \\
&- |\hat{p_{Z_n}}(a|u\bar{\omega}) - p_Z(a|u\bar{\omega})| - |\hat{p_{Z_n}}(a|suf(u\bar{\omega})) - p_Z(a|suf(u\bar{\omega}))| .
\end{aligned}$$

Considerando a última inequação, tem-se

$$\begin{aligned}
\mathbb{P}\left(\Delta_n(u\bar{\omega}) \leq \delta\right) &\leq \mathbb{P}(\bigcap_{a\in\mathcal{A}}\{|\hat{p_{Z_n}}(a|u\bar{\omega}) - p_Z(a|u\bar{\omega})| \geq \frac{D_d - \delta - 4\varepsilon\left[1 + \frac{2(N-1)\beta_X}{\min(1,\alpha\beta^*_{min})}\right]}{2}\}) \\
&+ \mathbb{P}(\bigcap_{a\in\mathcal{A}}\{|\hat{p_{Z_n}}(a|u\bar{\omega}) - p_Z(a|u\bar{\omega})| \geq \frac{D_d - \delta - 4\varepsilon\left[1 + \frac{2(N-1)\beta_X}{\min(1,\alpha\beta^*_{min})}\right]}{2}\}).
\end{aligned}$$

Se $\delta < D_d - 4\varepsilon\left[1 + \frac{2(N-1)\beta_X}{\min(1,\alpha\beta^*_{min})}\right]$ e

$$n > \frac{2(N+1)}{\left\{D_d - \delta - 4\varepsilon\left[1 + \frac{2(N-1)\beta_X}{\min(1,\alpha\beta^*_{min})}\right]\right\}\alpha^d} + d$$

podemos utilizar o Lema 3.15 para obtermos

$$\mathbb{P}\left(\bigcap_{u\omega\in\mathcal{T}_X|_d}\{\Delta_n(u\omega) \leq \delta\}\right) \leq 2(N+1)e^{\frac{1}{e}}\exp\left[-(n-d)\frac{\left[D_d - \delta - \bar{k}\right]^2\alpha^{2d}_{min}(1-\varepsilon)^d}{128N^2e(d+1)\left(1+\frac{\beta}{\alpha}\right)_{max}}\right].$$

Isto conclui a prova do Lema 3.17. □

Agora, provaremos o resultado principal desta seção.

Demonstração do Teorema 2.11. *Definimos*

$$O^{K,d}_{n,\delta} = \bigcup_{\substack{\omega\in\mathcal{T}_X \\ l(\omega)<K}} \bigcup_{u\omega\in\hat{\mathcal{T}}^{\delta,d}_n} \{\Delta_n(u\omega) > \delta\},$$

e

$$U_{n,\delta}^{K,d} = \bigcup_{\substack{\omega \in \hat{\mathcal{T}}_n^{\delta,d} \\ l(\omega)<K}} \bigcap_{u\omega \in \mathcal{T}_X|_d} \{\Delta_n(u\omega) \leq \delta\}.$$

Logo, se $d < n$ temos

$$\left\{\hat{\mathcal{T}}_n^{\delta,d}|_K \neq \mathcal{T}_X|_K\right\} \subseteq O_{n,\delta}^{K,d} \cup U_{n,\delta}^{K,d}.$$

Portanto,

$$\mathbb{P}\left(\hat{\mathcal{T}}_n^{\delta,d}|_K \neq \mathcal{T}_X|_K\right) \leq \sum_{\substack{\omega \in \mathcal{T}_X \\ l(\omega)<K}} \sum_{u\omega \in \hat{\mathcal{T}}_{3n}^{\delta,d}} \mathbb{P}\left(\Delta_n(u\omega) > \delta\right) + \sum_{\substack{\omega \in \hat{\mathcal{T}}_n^{\delta,d} \\ l(\omega)<K}} \mathbb{P}\left(\bigcap_{u\omega \in \mathcal{T}_X|_d} \{\Delta_n(u\omega) \leq \delta\}\right).$$

Por meio dos Lemas 3.16 e 3.17 obtemos a desigualdade

$$\mathbb{P}\left(\hat{\mathcal{T}}_n^{\delta,d}|_K \neq \mathcal{T}_X|_K\right) \leq c_2 exp\left\{-c_3(n-d)\right\},$$

onde $c_2 = 48N^d(N+1)e^{\frac{1}{e}}$, $c_3 = \frac{\left[\min(D_d-\delta,\delta)-2\bar{k}\right]^2 \alpha^{2d}}{128N^2 e(d+1)\left(1+\frac{\beta}{\alpha}\right)_{min}}$ *e*

$$n > \frac{2(N+1)}{\left\{\min\left\{D_d-\delta,\delta\right\} - 4\varepsilon\left[1+\frac{2(N-1)\beta_X}{\min\left(1,\alpha\beta^*_{min}\right)}\right]\right\}\alpha^d} + d.$$

Assim, concluímos a prova do Teorema 2.11.

Demonstração do Corolário 2.12. *Segue do Teorema 2.11, do Primeiro Lema de Borel-Cantelli e do fato de que as cotas para o erro de estimação da árvore de contextos truncada são somáveis em n para escolhas apropriadas de d e δ.*

Capítulo 4

Comparações

Neste capítulo vamos comparar os resultados obtidos nesse trabalho com os correspondentes apresentados em Collet, Galves e Leonardi (2008). Para tanto, consideraremos que $(X_t)_{t\in\mathbb{Z}}$ e $(Y_t)_{t\in\mathbb{Z}}$ são processos estocásticos, estacionários e ergódicos sobre o alfabeto finito $\mathcal{A} = \{0, 1\}$. Como na Seção 2.1, as cadeias (X_t) e (Y_t) são supostas independentes. Além disso, vamos supor que estes os processos satisfazem as condições de não nulidade e taxa de continuidade somável com as mesmas constante α_X e β_X.

Seja $(\xi_t)_{t\in\mathbb{Z}}$ uma sequência de variáveis aleatórias i.i.d. tomando valores em $\{0, 1\}$, independente dos processos (X_t) e (Y_t), com

$$\mathbb{P}(\xi_t = 1) = 1 - \varepsilon,$$

onde ε é um parâmetro de ruído fixado em $(0, 1)$. Para a e b em $\{0, 1\}$, definimos

$$a \oplus b = a + b \ (\mathrm{mod}\ 2).$$

Agora, definimos as cadeias estocasticamente perturbadas $(Z_{1,t})_{t\in\mathbb{Z}}$, $(Z_{2,t})_{t\in\mathbb{Z}}$ e $(Z_{3,t})_{t\in\mathbb{Z}}$ por

$$Z_{1,t} = X_t \oplus (1 - \xi_t), \tag{4.1}$$

$$Z_{2,t} = X_t.\xi_t, \tag{4.2}$$

$$Z_{3,t} = \begin{cases} X_t, & \text{se } \xi_t = 1, \\ Y_t, & \text{se } \xi_t = 0, \end{cases} \tag{4.3}$$

onde o modelo (4.1) foi proposto em Collet, Galves e Leonardi (2008).

No modelo (4.1), o processo $(Z_{1,t})$ será diferente do processo (X_t) sempre que $\xi_t = 0$, o que ocorre com probabilidade ε. Neste modelo, é possível que qualquer dos símbolos do processo original (X_t) seja modificado pelo efeito Bernoulli (ξ_t). O processo $(Z_{2,t})$, definido em (4.2), será igual ao processo (X_t) com probabilidade alta $1-\varepsilon$. Além disso, neste modelo, apenas o símbolo 1 do processo (X_t) possui probabilidade positiva de ser modificado pelo processo (ξ_t). O terceiro modelo de contaminação, definido em (4.3), é tal que em cada instante do tempo o processo $(Z_{3,t})$ ou será igual ao processo (X_t), com probabilidade alta $1-\varepsilon$, ou será igual ao processo (Y_t), com probabilidade pequena ε.

O Teorema 1 de Collet, Galves e Leonardi (2008), afirma, para (X_t) e $(Z_{1,t})$ como acima, que para todo $\varepsilon \in (0,1)$ e para todo $k \geq 0$, temos

$$\sup_{a\in\mathcal{A},\ \omega_{-k}^{-1}\in\mathcal{A}_{-k}^{-1}} \left|q_1\left(a|\omega_{-k}^{-1}\right) - p_X\left(a|\omega_{-k}^{-1}\right)\right| \leq \varepsilon\left[1+\frac{4\beta_X}{\min\left(1,\alpha_X\beta_X^*\right)}\right], \tag{4.4}$$

onde $\beta_X^* = \prod_{k=0}^{+\infty}(1-\beta_{k,X}) < +\infty$ e $q_1(.)$ é a lei do processo $(Z_{1,t})$. Nosso Teorema 2.5 diz que se (X_t) e $(Z_{2,t})$ são como acima então, para todo $\varepsilon \in (0,1)$ e para todo $k \geq 0$, temos

$$\sup_{a\in\mathcal{A},\ \omega_{-k}^{-1}\in\mathcal{A}_{-k}^{-1}} \left|q_2\left(a|\omega_{-k}^{-1}\right) - p_X\left(a|\omega_{-k}^{-1}\right)\right| \leq \varepsilon\left[1+\frac{4\beta_X}{\min\left(1,(1+\varepsilon)\alpha_X\beta_X^*\right)}\right], \tag{4.5}$$

onde $q_2(.)$ é a lei do processo $(Z_{2,t})$. O Teorema 2.10 deste trabalho afirma que se (X_t), (Y_t) e $(Z_{3,t})$ são como acima então, para todo $\varepsilon \in (0,1)$ e para todo $k \geq 0$, temos

$$\sup_{a\in\mathcal{A},\ \omega_{-k}^{-1}\in\mathcal{A}_{-k}^{-1}} \left|q_3\left(a|\omega_{-k}^{-1}\right) - p_X\left(a|\omega_{-k}^{-1}\right)\right| \leq \varepsilon\left[2+\frac{4\beta_X}{\min\left(1,\alpha\beta_{min}^*\right)}\right], \tag{4.6}$$

onde $\alpha\beta_{min}^* = \min\{\alpha_X\beta_X^*, \alpha_X\}$ e $q_3(.)$ é a lei do processo $(Z_{3,t})$. Agora, sejam

$$k_1 := \varepsilon\left[1+\frac{4\beta_X}{\min\left(1,\alpha_X\beta_X^*\right)}\right],$$

$$k_2 := \varepsilon\left[1+\frac{4\beta_X}{\min\left(1,(1+\varepsilon)\alpha_X\beta_X^*\right)}\right],$$

$$k_3 := \varepsilon\left[2+\frac{4\beta_X}{\min\left(1,\alpha\beta_{min}^*\right)}\right].$$

Com o intuito de comparar k_1, k_2 e k_3 consideramos, inicialmente, que

$$\alpha_X\beta_X^* < 1. \tag{4.7}$$

Neste caso, podemos escrever

$$\min\left(1,\alpha\beta_{min}^*\right) \leq \min\left(1,\alpha_X\beta_X^*\right) \leq \min\left(1,(1+\varepsilon)\alpha_X\beta_X^*\right). \tag{4.8}$$

Para atestar a primeira desiguadade em (4.8), note que $\alpha\beta^*_{min} = \alpha_X\beta^*_X$ se $\beta^*_X < 1$. No caso em que $\beta^*_X \geq 1$, então $\alpha\beta^*_{min} = \alpha_X \leq \alpha_X\beta^*_X$. A segunda desiguadade em (4.8) é válida pois $\alpha_X\beta^*_X < (1+\varepsilon)\alpha_X\beta^*_X$. Logo, por (4.8), temos

$$k_2 \leq k_1 \leq k_3, \tag{4.9}$$

quando vale (4.7). No caso em que

$$\alpha_X\beta^*_X \geq 1, \tag{4.10}$$

as desigualdades em (4.8) são mantidas e, consequentemente, também vale (4.9).

Observamos que a desigualdade $k_2 \leq k_1$ era esperada pelas definições das cadeias $(Z_{1,t})$ e $(Z_{2,t})$. Para ver isso, note que a equação (4.1) diz que tanto o símbolo 0 como o símbolo 1 do processo (X_t) possuem probabilidade positiva ε de serem alterados pelo ruído Bernoulli. Enquanto (4.2) nos diz que apenas o símbolo 1 do processo (X_t) possui probabilidade ε de ser alterado por (ξ_t). Portanto, o processo $(Z_{2,t})$ está mais próximo da cadeia (X_t) que o processo $(Z_{1,t})$.

Referências Bibliográficas

[1] Collet, P., Galves, A., Leonardi, F., *Random Perturbations of Stochastic Processes with Unbounded Variable Length Memory,* Eletronic Journal of Probability, Vol. 13, Paper no. 48, 1345–1361, (2008).

[2] Galves, A., Maume-Deschamps, V., Schmitt, B., *Exponential inequalities for VLMC empirical trees,* ESAIM Prob. Stat. (accepted), (2006).

[3] Csiszár, I., Talata, Z., *Context tree estimation for not necessarily finite memory processes, via BIC and MDL,* IEEE Trans. Inform. Theory 52(3), 1007–1016, (2006).

[4] Dedecker, J., Doukhan, P., *A new covariance inequality and applications,* Stochastic Process. Appl. 106(1), 63–80, (2003).

[5] Dedecker, J., Prieur, C., *New dependence coefficients. examples and applications to statistics,* Probab. Theory Relatated Fields 132: 203–236, (2005).

[6] Duarte, D., Galves, A., Garcia, N., *Markov approximation and consistent estimation of unbounded probabilistic suffix trees,* Bull. Braz. Math. Soc. 37(4): 581–592, (2006).

[8] Fernández, R., Galves, A., *Markov approximations of chains of infinite order,* Bull. Braz. Math. Soc., vol. 33, n° 3: 1–12, (2002).

[8] Fernández, R., Ferrari, P., Galves, A., *Coupling, renewal and perfect simulation of chains of infinite order.* Notes for a minicourse at the V^{th} Brazilian School of Probability, (2001).

[9] Ferrari, F. Wyner, A., *Estimation of general stationary processes by variable length Markov chains,* Scand. J. Statist. 30(3): 459–480, (2003).

[10] Galves, A., Leonardi, F., *Exponential inequalities for empirical unbounded context trees,* Vol. 60 of Progress in Probability, Birkhauser, 257–270, (2008).

[11] Galves, A., Löcherbach, E., *Stochastic chains with memory of variable length,* TICSP Series 38: 117–133, (2008).

[12] Rissanen, J., *A universal data compression system,* IEEE Trans. Inform. Theory 29(5): 656–664, (1983).

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

[14] Willems, F. M., Shtarkov, Y. M., Tjalkens, T. J., *The context-tree weighting method: basic properties,* IEEE Trans. Inform. Theory IT-44: 653–664, (1995).