

**Aproximações Markovianas e Reamostragem
para Cadeias de Ordem Infinita
com aplicação à Linguística.**

Denise Duarte S. M. Alves

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTER O GRAU
DE
DOUTOR EM ESTATÍSTICA

Área de concentração: Probabilidade
Orientador: Antonio Galves

Esta tese teve o apoio financeiro do programa CAPES-PICDT

**Aproximações Markovianas e Reamostragem
para Cadeias de Ordem Infinita
com aplicação à Linguística.**

Este exemplar corresponde à redação final
da tese devidamente corrigida e defendida
por Denise Duarte Scarpa Magalhães Alves
e aprovada pela comissão julgadora.

São Paulo, 18 de fevereiro de 2003.

Banca examinadora:

- Prof. Dr. J. Antonio Galves (Universidade de São Paulo)
- Profa. Dra Nancy Lopes Garcia (UNICAMP)
- Prof. Dr. Ricardo Maronna (Universidad de la Plata)
- Prof. Dr. Marzio Cassandro (Università di Roma - *La Sapienza*)
- Prof. Dr. Pierre Collet (CNRS, Ecole Polytechnique)

Agradecimentos

Durante a elaboração desta tese tive discussões muito proveitosas em questões relacionadas a Matemática, Estatística e Linguística com Miguel Abadi, Marzio Cassandro, Pierre Collet, Roberto Fernández, Charlotte Galves, Antonio Galves, Jesus Garcia, Nancy L. Garcia, Ricardo Lima, Bernard Prum, Ricardo Maronna e Franck Ramus. Agradeço a todos não apenas pelas discussões, mas principalmente pela generosidade de me ensinar tantas coisas.

Em especial, quero agradecer ao Antonio Galves, meu orientador, pela dedicação ao longo de todo o trabalho.

Sem o apoio da minha família, que tanto me ajudou com meus filhos, teria sido bem mais difícil realizar este trabalho. Fica aqui também o meu muito obrigado à minha turma lá das Minas Gerais, meus pais, irmãs e sogros.

Agradeço também :

- O apoio dos colegas do Instituto de Matemática e Estatística da Universidade Federal de Goiás, onde trabalho.
- Aos professores do IME-USP que contribuíram para a minha formação.
- Ao prof Jorge de Sousa, grande responsável pela minha escolha pela Estatística.
- O apoio das equipes dos seguintes projetos: Projeto Temático FAPESP *n*^o 98/3382-0 (Padrões rítmicos, fixação de parâmetros e mudança linguística), PRONEX 66.2177/1996-6 (Núcleo de Excelência Fenômenos críticos em probabilidade e processos estocásticos) e CNPq 465928/2000-5 (Ferramentas probabilísticas para identificação de padrões aplicadas).

Resumo

Nesta tese apresentamos um novo método de reamostragem para cadeias de ordem infinita assumindo valores em um alfabeto finito. O nosso objetivo é desenvolver uma metodologia que nos permita fazer inferência para a média de uma função real em cadeias de ordem infinita com decaimento exponencial. A base matemática que justifica o procedimento de reamostragem que propomos é um Teorema Central do Limite da Reamostragem para cadeias de ordem infinita. A demonstração deste teorema é baseada em um outro resultado original que é um Teorema Limite Central da Reamostragem para sequências de cadeias de Markov de ordens crescentes.

Como aplicação, utilizamos nosso Teorema Limite Central da Reamostragem para construir intervalos de confiança assintóticos para a média da sonoridade em oito línguas naturais. Fazemos isto para classificar as línguas consideradas em grupos de acordo com a sonoridade.

Abstract

In this dissertation we present a new approach to the bootstrap for chains of infinite order taking values on a finite alphabet. Our goal is to develop an inferential methodology for the mean of a real function of the chain. The mathematical basis of our bootstrap procedure is a Bootstrap Central Limit Theorem. We first prove a Central Limit Theorem for the bootstrap mean in Markov chains with increasing orders. Afterwards we use this Theorem and the results given in Bressaud, Fernández and Galves (1999) to prove a Central Limit Theorem for the bootstrap mean in chains of infinite order with exponential decay. We apply this methodology to derive asymptotical confidence intervals for the mean of the sonority function of a dataset with eight languages.

*Esta tese é fruto de muita dedicação.
Dedicação é coisa que toma muito tempo...
tempo que faltou para a Alice, o Caio e o Oswaldo, meus anjos.
Agora eu volto para vocês...*

Sumário

1	Introdução	9
2	Definições e resultados que serão utilizados	13
2.1	Teorema Limite Central para sequências φ misturadoras	13
2.2	Cadeias de ordem infinita	15
2.3	Acoplamento maximal	17
2.4	Resultados para cadeias de Markov de ordem k	18
2.5	Aproximações markovianas para cadeias de ordem infinita	19
2.6	Tempos de renovação para cadeias de ordem infinita	20
2.7	Relação entre tempos de renovação de uma cadeia de ordem in- finita e das cadeias de Markov aproximantes	22
2.7.1	Aplicação dos tempos de renovação	25
3	Reamostragem para sequências de cadeias de Markov	28
3.1	Reamostragem sequencial	28
3.2	Teorema Limite Central da Reamostragem para $\mu^{[k]*}$	31
3.3	Resultados preliminares	34
3.4	Demonstração do Teorema Limite Central para $\mu^{[k]*}$	45
4	Reamostragem para cadeias de ordem infinita	49
4.1	Procedimento de reamostragem	49
4.2	Teorema Limite Central da Reamostragem para μ_k^*	51
4.3	Demonstração do Teorema Limite Central da Reamostragem para μ_k^*	52
5	Uma aplicação à Linguística	56
5.1	Correlatos do ritmo presentes no sinal acústico	56

5.2	Uma medida de sonoridade como base para a identificação das classes rítmicas	58
5.3	Sonoridade como uma cadeia de ordem infinita	61
5.4	Aplicação do procedimento de reamostragem para estimar a variância da sonoridade	61

Capítulo 1

Introdução

Desde o final dos anos 60 é conhecido o fato de que, para processos com boas propriedades de mistura, a média amostral tem, no limite, uma flutuação normal em torno da média teórica (ver Billingsley, 1967). Mas a variância desta distribuição normal depende das autocorrelações de todas as ordens que são, em geral, desconhecidas. A questão que se coloca é a da estimação desta variância, permitindo assim a construção de uma teoria inferencial para a média teórica.

Nesta tese trabalharemos com uma classe de processos chamado cadeia de ordem infinita que têm longa memória do passado, mas boas propriedades de mistura. Processos desse tipo foram introduzidos na literatura probabilística pelos artigos de Onicescu e Mihoc (1935) e Doebelin e Fortet (1937) e Harris (1955). O tema teve recentemente um grande desenvolvimento (cf. Bressaud, Galves e Fernández 1999a e 1999b, Ferrari e Galves 2000, Ferrari, Maass, Martinez e Ney 2000 e Fernández, Ferrari e Galves 2001).

Reamostragem para sequências de variáveis aleatórias

O caminho para se chegar a uma solução para o problema da estimação da variância foi aberto pelo trabalho pioneiro de Bradley Efron em 1979. Neste trabalho Efron introduziu uma técnica não paramétrica muito importante para o estudo das distribuições amostrais de uma estatística que ele chamou de *bootstrap* (que traduziremos neste texto como reamostragem).

Em 1981, Singh (e independentemente também Bickel e Freedman) demonstrou que a técnica da reamostragem é válida para a estimativa da média da distribuição

no caso i.i.d., ou seja, ele mostrou que a distribuição assintótica da média empírica em torno da média verdadeira, que é uma normal com média zero e variância σ^2 , coincide com a distribuição da média da reamostragem em torno da média empírica. Este resultado diz que podemos utilizar a distribuição empírica da média da reamostragem para construir intervalos para a média verdadeira. Mas Singh também observou que para processos com alguma estrutura de dependência, a variância da média da reamostragem não tem o mesmo comportamento que a variância da média empírica, mostrando que, para esse tipo de processo, algum outro procedimento deveria ser considerado. Este comportamento da média reamostrada se deve principalmente ao fato de que na técnica proposta por Efron a reamostra é construída através de sorteios independentes e com reposição das observações da amostra original, quebrando assim a estrutura natural de dependência entre as variáveis. Sendo assim, uma tentativa natural para contornar o problema seria fazer a reamostragem de blocos inteiros de observações ao invés de uma observação apenas.

Até onde sabemos, o primeiro a propor um esquema de reamostragem em blocos (blockwise bootstrap) foi Künsch em 1989. Ele propõe a construção de blocos de tamanho fixo, a partir da primeira observação. Fixado um inteiro n , o primeiro bloco é formado pelos primeiros n símbolos observados. O segundo bloco é uma translação de um passo para a direita na amostra e assim por diante. Ele propõe uma reamostragem nos moldes de Efron para os blocos assim definidos. Observe que estes blocos têm uma interseção grande e consequentemente uma grande correlação. Este método foi aperfeiçoado por Liu e Singh em 1992. Em 1992 e 1993 Shao e Yu demonstraram vários Teoremas Limite Centrais, condicionados à amostra original, para o estimador da reamostragem em blocos para classes de sequências com diferentes estruturas de dependência. Estes resultados foram melhorados, sempre em relação às taxas de decaimento das correlações, por Naik-Nimbalkar e Rajarshi (1994), Bühlmann (1994) e Peligrad (1998).

Em 1992, Athreya e Fuh propuseram para cadeias de Markov de ordem um, em um espaço de estados enumerável, um procedimento de reamostragem em blocos para a estimação das probabilidades de transição da cadeia. Eles justificam o procedimento com a demonstração de um Teorema Limite Central da Reamostragem para estimativas das probabilidades de transição obtidas através do procedimento de reamostragem por blocos.

Um novo procedimento de reamostragem para cadeias de ordem infinita

Nesta tese consideramos um novo procedimento de reamostragem para cadeias de ordem infinita com espaço de estados finito. O fato que dá suporte ao procedimento de reamostragem que propomos é que cadeias de ordem infinita com decaimento somável são bem aproximadas por sequências de cadeias de Markov de ordens crescentes. Então, primeiro mostramos que o procedimento funciona para sequências de cadeias de Markov com ordens crescentes. Sabemos que o erro da aproximação da cadeia de ordem infinita por uma cadeia de Markov diminui à medida que a ordem da cadeia aumenta. Portanto, para uma ordem k suficientemente grande já quase não distinguimos um processo do outro. Utilizando este fato mostramos que o procedimento de reamostragem também funciona para cadeias de ordem infinita. Por se basear em cadeias de Markov de ordens crescentes, chamaremos o novo procedimento de reamostragem sequencial.

A reamostragem sequencial é baseada nos tempos de retorno dos primeiros símbolos da amostra. Informalmente, o procedimento consiste no seguinte. Fixado um inteiro positivo k , olhamos para as ocorrências sucessivas dos primeiros k símbolos na amostra. Fazemos uma seleção uniforme i.i.d. dos blocos delimitados por essas ocorrências para obtermos a reamostra e calculamos a estatística de interesse, no caso, a média de uma função real. Uma análise preliminar indica que o procedimento de reamostragem sequencial é mais rápido do que aquele proposto por Künsch (1989) para amostras de mesmo tamanho.

A demonstração do Teorema Limite Central da Reamostragem para cadeias de ordem infinita é feita em duas etapas distintas. Primeiro consideramos o procedimento de reamostragem sequencial para estimar a média em cadeias de Markov com ordens crescentes. O tamanho da amostra cresce como função da ordem da cadeia que, por sua vez, tende a infinito. Em cadeias de Markov de ordem k , a propriedade de Markov implica que os blocos de valores delimitados pelas sucessivas ocorrências dos primeiros símbolos são independentes entre si e identicamente distribuídos. Portanto, uma grande vantagem do esquema de reamostragem sequencial é que ele preserva, para cadeias de Markov, toda a estrutura de dependência entre as variáveis. O Teorema Limite Central da Reamostragem para o estimador da média em cadeias de Markov que apresentamos é, em si, um resultado novo importante, uma vez que a ordem da cadeia tende a infinito. A seguir utilizamos os resultados apresentados em Bressaud, Fernández e Galves (1999) sobre velocidade de convergência das aproximações markovianas para estender o

resultado para cadeias de ordem infinita.

Aplicação à Linguística

Apresentamos uma aplicação das ferramentas estatísticas desenvolvidas nesta tese na análise estatística de um corpus de fala de oito línguas naturais gentilmente cedido por Franck Ramus. A análise deste corpus é bastante complexa e abrangente. Entre os artigos que se dedicam a estudá-lo de um ponto de vista estatístico citamos Ramus, Nespore e Mehler (1999), Duarte, Galves, Garcia e Maronna (2001) e Galves, Garcia, Duarte e Galves (2002). Nesta tese o procedimento de amostragem sequencial será utilizado para construir intervalos de confiança para médias de uma função do sinal acústico deste corpus de fala que chamaremos de sonoridade para cada uma das oito línguas. Queremos saber se as médias da sonoridade são todas iguais ou há formação de grupos de línguas.

Esta tese faz parte de um projeto de pesquisa mais ambicioso de modelagem matemática da aquisição e da mudança linguística. Para uma apresentação das idéias centrais deste projeto enviamos o leitor a Cassandro, Collet, Galves, e Galves (1999), Collet, Galves, e Lopes (1995) e Fernández e Galves (2000).

Apresentação da tese

Esta tese está apresentada da seguinte maneira: No Capítulo 2 apresentamos definições e alguns resultados que são importantes na demonstração dos teoremas principais. No Capítulo 3 apresentamos o procedimento de amostragem sequencial para sequências de cadeias de Markov com ordens crescentes e enunciamos e provamos o Teorema Limite Central da Reamostragem para a média amostral. No Capítulo 4 utilizamos os resultados demonstrados nos capítulos 2 e 3 para estender o Teorema Limite Central da Reamostragem para a classe das cadeias de ordem infinita. No Capítulo final mostramos uma aplicação em dados acústicos da metodologia de amostragem aqui apresentada. Primeiramente colocamos o contexto linguístico do problema que tratamos e a seguir apresentamos os resultados obtidos com a nossa metodologia. No final de cada capítulo colocamos alguns apêndices com resultados relacionados que consideramos importantes, mas de leitura não obrigatória. Todos os programas da amostragem foram feitos em linguagem S e rodados com o software S-PLUS. Os programas para o cálculo da sonoridade foram feitos em PERL por Jesus Garcia.

Capítulo 2

Definições e resultados que serão utilizados

Neste capítulo definimos a classe de processos com os quais iremos trabalhar e apresentamos algumas definições e resultados importantes na demonstração do nosso resultado principal. Apresentamos também, na última seção deste capítulo, um resultado novo relacionando os tempos de retorno das cadeias de ordem infinita com os tempos de retorno das cadeias de Markov aproximantes. Começamos lembrando os clássicos resultados dados em Billingsley (1967) para processos com uma certa estrutura de dependência que detalharemos a seguir.

2.1 Teorema Limite Central para sequências φ misturadoras

Seja $(X_n)_{n \in \mathbb{Z}}$ um processo estocástico estacionário. Vamos chamar de $(\Omega, \mathcal{A}, \mathbb{P})$ o espaço de probabilidades onde o processo está definido. Para $a < b$ definimos a σ -álgebra gerada pelas variáveis aleatórias X_a, \dots, X_b como \mathcal{M}_a^b . Chamamos de \mathcal{M}_a^∞ a σ -álgebra gerada pelas variáveis aleatórias $\{X_a, X_{a+1}, \dots\}$ e chamamos de $\mathcal{M}_{-\infty}^a$ a σ -álgebra gerada pelas variáveis aleatórias $\{\dots, X_{a-1}, X_a\}$.

Dizemos que a sequência $(X_n)_{n \in \mathbb{Z}}$ é φ -misturadora se existir uma função $\varphi : \mathbb{N} \rightarrow \mathbb{R}$, com $\varphi(n) \rightarrow 0$, quando $n \rightarrow \infty$ tal que para cada $n \geq 1$, $E_1 \in \mathcal{M}_{-\infty}^0$ e $E_2 \in \mathcal{M}_n^\infty$ tivermos que

$$|\mathbb{P}(E_1, E_2) - \mathbb{P}(E_1)\mathbb{P}(E_2)| \leq \varphi(n)\mathbb{P}(E_1). \quad (1)$$

Dizemos que $(X_n)_{n \in \mathbb{Z}}$ é ψ -misturadora se para uma função $\psi : \mathbb{N} \rightarrow \mathbb{R}$ com $\psi(n) \rightarrow 0$, quando $n \rightarrow \infty$, valer que

$$|\mathbb{P}(E_1, E_2) - \mathbb{P}(E_1)\mathbb{P}(E_2)| \leq \psi(n)\mathbb{P}(E_1)\mathbb{P}(E_2). \quad (2)$$

O Teorema Limite Central para sequências φ misturadoras apresentado em Billingsley (1967) diz o seguinte:

Teorema 2.1. *Seja $(\zeta_n)_{n \in \mathbb{Z}}$ um processo estacionário φ -misturador assumindo valores em \mathbb{R} com $\sum_n \varphi^{1/2}(n) < \infty$ e suponhamos que ζ_0 tenha média zero e variância finita. Então a série*

$$\sigma_\infty^2 = \mathbb{E}(\zeta_0) + 2 \sum_{k=1}^{\infty} \mathbb{E}(\zeta_0 \zeta_k) \quad (3)$$

converge absolutamente. Além disso, chamando $S_n = \zeta_1 + \zeta_2 + \dots + \zeta_n$, se $\sigma_\infty^2 > 0$, então

$$\frac{1}{\sqrt{n}} S_n \rightarrow \mathcal{N}(0, \sigma_\infty^2). \quad (4)$$

Mesmo no caso em que o número de termos usados para estimar a média é aleatório o Teorema Limite Central para a média ainda é válido devido ao seguinte teorema também dado em Billingsley (1967).

Teorema 2.2. *Seja (ν_n) uma variável aleatória com valores inteiros positivos. Suponhamos que*

$$\frac{\nu_n}{a_n} \xrightarrow{P} \theta \quad (5)$$

onde θ é uma constante positiva e a sequência (a_n) tende para o infinito. Se as hipóteses do teorema 2.1 forem satisfeitas, então

$$\frac{1}{\sigma \sqrt{n}} S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

implica

$$\frac{1}{\sigma \sqrt{\nu_n}} S_{\nu_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Este resultado não nos permite fazer inferência sobre μ uma vez que não conhecemos a verdadeira variância σ_∞^2 . No capítulo 4 propomos um procedimento de *reamostragem* baseado em tempos de retorno da primeira sequência de tamanho k da amostra para estimar a variância. Com isso evitamos o problema de estimar as covariâncias de todas as ordens que aparecem no segundo termo da expressão de σ_∞^2 em (3).

2.2 Cadeias de ordem infinita

Definiremos agora a classe de processos com os quais iremos trabalhar ao longo desta tese.

Seja $(X_n)_{n \in \mathbb{Z}}$ um processo estacionário assumindo valores em um alfabeto finito A . Adotaremos aqui o ponto de vista construtivo apresentado em Fernández, Ferrari e Galves (2001). Assim sendo, $(\Omega, \mathcal{A}, \mathbb{P})$ será o espaço de probabilidades no qual está definida a sequência de variáveis aleatórias uniformes e independentes, que daqui para frente chamaremos $(U_n)_{n \in \mathbb{Z}}$, com as quais construiremos todos os processos aqui considerados. Isto aparecerá claramente quando construirmos o acoplamento maximal na próxima seção.

Vamos utilizar a notação

$$p(x_0|x_{-1}, x_{-2}, \dots) = \mathbb{P}(X_0 = x_0 | X_{-1} = x_{-1}, X_{-2} = x_{-2}, \dots)$$

para denotar a versão regular da probabilidade condicional do processo.

Daqui em diante utilizaremos a notação \underline{x} para o passado (\dots, x_2, x_1) e $(x_{-k}, -1)$ para (x_k, \dots, x_{-1}) .

Os processos com os quais trabalharemos diferem entre si de acordo com hipóteses a respeito de sua continuidade e positividade que detalharemos abaixo.

Hipóteses de continuidade

A taxa de continuidade $(\beta_l)_{l \in \mathbb{N}}$ é definida por

$$\beta_l = \sup_{\substack{x_i=y_i \\ i=-l, \dots, 0}} |p(x_0|x_{-1}, x_{-2}, \dots) - p(y_0|y_{-1}, y_{-2}, \dots)|, \quad (6)$$

onde o sup é tomado sobre todos os pares de sequências cujas coordenadas coincidem de $-l$ até 0.

O processo (X_n) é chamado **contínuo** se

$$\lim_{l \rightarrow \infty} \beta_l = 0 . \quad (7)$$

Uma noção mais forte de continuidade, frequentemente utilizada na literatura envolve o que chamamos de taxa de log-continuidade

$$\gamma_s := \sup_{\substack{x_i = y_i \\ i = -s, \dots, 0}} \left| \frac{p(x_0 | x_{-1}, x_{-2}, \dots)}{p(y_0 | y_{-1}, y_{-2}, \dots)} - 1 \right| . \quad (8)$$

Aqui o sup é tomado sobre todos os pares de seqüências cujas coordenadas coincidem de $-s$ até 0.

O processo (X_n) é chamado **log-contínuo** se

$$\lim_{s \rightarrow \infty} \gamma_s \rightarrow 0 .$$

Nesta tese trabalharemos com processos contínuos.

Hipóteses de positividade

Um processo é chamado **fracamente não-nulo** se satisfaz

$$\sum_{a \in A} \inf_{\underline{x}} p(a | \underline{x}) > 0 . \quad (9)$$

E é chamado de **fortemente não-nulo** se satisfaz

$$\inf_{a, \underline{x}} p(a | \underline{x}) > 0 . \quad (10)$$

Nesta tese utilizaremos uma condição mais fraca que fortemente não-nulo que chamaremos de **fortemente não-nulo para seqüências admissíveis**.

Um passado \underline{x} é admissível para um símbolo $a \in A$ se,

$$p(a | \underline{x}) > 0 .$$

Chamamos de \mathcal{A}_a ao conjunto das seqüências admissíveis para o símbolo $a \in A$, isto é $\mathcal{A}_a = \{\underline{x} : p(a | \underline{x}) > 0\}$. Dizemos que um processo é **fortemente não-nulo para seqüências admissíveis** se para cada $a \in A$

$$\inf_{\underline{x} \in \mathcal{A}_a} p(a | \underline{x}) > \delta , \quad (11)$$

com $\delta > 0$.

Em Fernández, Ferrari e Galves (2001), $(X_n)_{n \in \mathbb{Z}}$ é uma cadeia de ordem infinita do

1. **Tipo A** se o processo é contínuo e fracamente não-nulo;
2. **Tipo B** se é log-contínuo e fortemente não-nulo.

Neste texto chamaremos de cadeia de ordem infinita do **tipo C** se ela for contínua e fortemente não-nula para seqüências admissíveis.

2.3 Acoplamento maximal

Todos os resultados que citaremos nas próximas seções e também os resultados apresentados nesta tese utilizam o mesmo acoplamento, chamado *maximal*, que detalharemos a seguir. Lembramos que um *acoplamento* entre dois processos estocásticos $X = (X_n)_{n \in \mathbb{Z}}$ e $Y = (Y_n)_{n \in \mathbb{Z}}$ tomando valores em A é qualquer processo $(\tilde{X}, \tilde{Y}) = (\tilde{X}_n, \tilde{Y}_n)_{n \in \mathbb{Z}}$ tomando valores em $A \times A$ tal que $\tilde{X} \stackrel{\mathcal{D}}{=} X$ e $\tilde{Y} \stackrel{\mathcal{D}}{=} Y$. A seguir definimos um acoplamento maximal entre uma cadeia de ordem infinita e uma cadeia de Markov de ordem k . O acoplamento maximal é definido da seguinte forma: dado um duplo passado $(\underline{x}, \underline{y})$, definimos,

$$\left\{ \begin{array}{l} t_a(\underline{x}, \underline{y}) := p(a | \underline{x}) \wedge p^{[k]}(a | \underline{y}) \\ r_a(\underline{x}, \underline{y}) := (p(a | \underline{x}) - p^{[k]}(a | \underline{y})) \vee 0 \\ s_a(\underline{x}, \underline{y}) := (p^{[k]}(a | \underline{y}) - p(a | \underline{x})) \vee 0 \end{array} \right\}$$

onde $p^{[k]}(a | \underline{x}) = \mathbb{P}(X_0^{[k]} = a | X_{-\infty, -1}^{[k]} = x_{-\infty, -1})$. Observamos que

$$\sum_{a \in A} t_a(\underline{x}, \underline{y}) + \sum_{a \in A} r_a(\underline{x}, \underline{y}) = 1$$

e

$$\sum_{a \in A} t_a(\underline{x}, \underline{y}) + \sum_{a \in A} s_a(\underline{x}, \underline{y}) = 1.$$

Agora suponhamos, sem perda de generalidade que $A = \{1, 2, \dots, |A|\}$. Para cada par de passados $(\underline{x}, \underline{y})$, podemos considerar duas partições de $[0, 1]$, cada uma feita de $2|A|$ intervalos,

$$\{T_1^{\underline{x}, \underline{y}}, \dots, T_{|A|}^{\underline{x}, \underline{y}}, R_1^{\underline{x}, \underline{y}}, \dots, R_{|A|}^{\underline{x}, \underline{y}}\} \text{ e } \{T_1^{\underline{x}, \underline{y}}, \dots, T_{|A|}^{\underline{x}, \underline{y}}, S_1^{\underline{x}, \underline{y}}, \dots, S_{|A|}^{\underline{x}, \underline{y}}\} \quad (12)$$

formado por intervalos de comprimentos

$$|T_a^{\underline{x}, \underline{y}}| = t_a(\underline{x}, \underline{y}), \quad |R_a^{\underline{x}, \underline{y}}| = r_a(\underline{x}, \underline{y}) \quad \text{e} \quad |S_a^{\underline{x}, \underline{y}}| = s_a(\underline{x}, \underline{y}),$$

para todo $a \in A$.

Seja $(U_n)_{n \in \mathbb{Z}}$ a sequência variáveis aleatórias uniformes independentes utilizadas na construção dos processos. Dados dois passados $\underline{x}, \underline{y}$, definimos $Z_n^{\underline{x}, \underline{y}}$ da seguinte maneira. Se $n \leq -1$:

$$Z_0^{\underline{x}, \underline{y}} := \begin{cases} (a, a) & \text{se } U_0 \in T_a^{\underline{x}, \underline{y}}, \\ (a, b) & \text{se } U_0 \in R_a^{\underline{x}, \underline{y}} \cap S_b^{\underline{x}, \underline{y}}. \end{cases}$$

Se $n \geq 0$, definimos por indução:

$$Z_n^{\underline{x}, \underline{y}} := \begin{cases} (a, a), & \text{se } U_n \in T_a^{\underline{x}, \underline{y}}, a \in A \\ (a, b), & \text{se } U_n \in R_a^{\underline{x}, \underline{y}} \cap S_b^{\underline{x}, \underline{y}}, a \neq b. \end{cases}$$

Observamos que a cadeia $Z_n^{\underline{x}, \underline{y}}$ tem como versão regular da probabilidade condicional a matriz

$$\tilde{p}((a, b) \mid (\underline{x}, \underline{y})) := \begin{cases} |T_a^{\underline{x}, \underline{y}}| & \text{se } a = b, \\ |R_a^{\underline{x}, \underline{y}} \cap S_b^{\underline{x}, \underline{y}}| & \text{se } a \neq b. \end{cases}$$

2.4 Resultados para cadeias de Markov de ordem k

Lembramos agora algumas definições e resultados relacionados ao encontro de duas trajetórias de uma cadeia de Markov utilizando o acoplamento maximal. Seja $X^{[k], a_1, k}$ uma cadeia de Markov começando em $X_{-k+1}^{[k]} = a_1, \dots, X_0^{[k]} = a_k$.

Construímos duas trajetórias acopladas da cadeia de Markov,

$$(X_n^{[k], a_1, k})_{n \geq -k+1}$$

e

$$(X_n^{[k], b_1, k})_{n \geq -k+1}.$$

Para cada $k \in \mathbb{N}$, o primeiro instante em que as trajetórias se encontram é definido como

$$\tau^{(a_1, k, b_1, k)} = \inf\{n : X_n^{[k], a_1, k} = X_n^{[k], b_1, k}, n \geq 1\}. \quad (13)$$

Seja

$$\rho_k = \inf_{u \in A^k, v \in A^k} \sum_{a \in A} (p^{[k]}(a|u) \wedge p^{[k]}(a|v)). \quad (14)$$

Temos para cada instante n que (ver Ferrari e Galves, 2000)

$$\mathbb{P}(X_n^{[k],a_{1,k}} = X_n^{[k],b_{1,k}}) \geq \rho_k . \quad (15)$$

Então,

$$\mathbb{P}(\tau^{(a_{1,k},b_{1,k})} > n) \leq (1 - \rho_k)^n . \quad (16)$$

Vamos agora encontrar uma cota inferior para a probabilidade de ocorrência de uma sequência de tamanho k . Esta cota será utilizada em várias demonstrações. Seja $B = \{X_{1,k}^{[k]} = x_{1,k}\}$,

$$\begin{aligned} \mathbb{P}(B) &= \sum_{x_{-k+1,0} \in A^k} \mathbb{P}(X_{-k+1,0}^{[k]} = x_{-k+1,0}) \mathbb{P}(X_k^{[k]} = x_k | X_{k-1,0}^{[k]} = x_{k-1,0}) \\ &\times \mathbb{P}(X_{k-1}^{[k]} = x_{k-1} | X_{k-2,-1}^{[k]} = x_{k-2,-1}) \dots \\ &\times \mathbb{P}(X_1^{[k]} = x_1 | X_{-k+1,0}^{[k]} = x_{-k+1,0}) . \end{aligned} \quad (17)$$

Sequências admissíveis para o símbolo $a \in A$ neste caso são aquelas em que $p^{[k]}(a|x_{k-1,0}) > 0$. Mas a soma para sequências não admissíveis é zero. Podemos então considerar a soma apenas para sequências $x_{-k+1,0}$ admissíveis. Chamaremos o conjunto das sequências admissíveis para cadeias de ordem k de \mathcal{A}_a^k . Então, pela estacionaridade segue que

$$\mathbb{P}(B) \geq \left[\inf_{a \in A, x_{-k+1,0} \in \mathcal{A}_a^k} p^{[k]}(a | X_{-k+1,0}^{[k]} = x_{-k+1,0}) \right]^k . \quad (18)$$

Mas

$$\inf_{a \in A, x_{-k+1,0} \in \mathcal{A}_a^k} p^{[k]}(a | X_{-k+1,0}^{[k]} = x_{-k+1,0}) > \delta .$$

Logo,

$$\mathbb{P}(B) \geq (\delta)^k . \quad (19)$$

2.5 Aproximações markovianas para cadeias de ordem infinita

Nesta seção apresentamos uma estimativa da velocidade de convergência das aproximações markovianas para cadeias de ordem infinita. Se as taxas de continuidade da cadeia de ordem infinita são somáveis é mostrado que a velocidade de convergência é proporcional a estas taxas. O resultado foi primeiramente provado para cadeias do tipo B com taxas log-contínuas somáveis e posteriormente, em Fernández

e Galves (2001), ele é estendido para cadeias do tipo A, com taxas contínuas somáveis.

Definição 2.1. A aproximação markoviana de ordem $k \in \mathbb{N}$ de um processo estacionário $(X_n)_{n \in \mathbb{Z}}$ é a cadeia de Markov estacionária de ordem k tendo como probabilidades de transição,

$$p^{[k]}(b \mid a_{1,k}) := \mathbb{P}(X_{k+1} = b \mid X_j = a_j, 1 \leq j \leq k) \quad (20)$$

para todo inteiro $k \geq 1$ e $a_{1,k}, b \in A$.

Definição 2.2. A distância \bar{d} entre dois processos estacionários X e Y é definida como

$$\bar{d}(X, Y) = \inf \left\{ \mathbb{P}(\tilde{X}_0 \neq \tilde{Y}_0) : (\tilde{X}, \tilde{Y}) \text{ acoplamento estacionário entre } X \text{ e } Y \right\}.$$

Teorema 2.3. Seja $X = (X_n)_{n \in \mathbb{Z}}$ uma cadeia de ordem infinita do tipo A com taxa de continuidade somável $(\beta_l)_{l \geq 1}$. Então existe uma constante $K > 0$ tal que, para todo $k \geq 1$,

$$\bar{d}(X, Y^{[k]}) \leq K \beta_k,$$

onde $Y^{[k]} = (Y_n^{[k]})_{n \in \mathbb{Z}}$ é a aproximação markoviana canônica de ordem k do processo X .

2.6 Tempos de renovação para cadeias de ordem infinita

Introduzidas por Fernández, Ferrari e Galves (2001) as misturas enumeráveis de cadeias de Markov (CMMC) são cadeias de ordem infinita cujas probabilidades de transição são combinações convexas enumeráveis de probabilidades de transição markovianas de ordem crescentes. Isto é, elas são da forma

$$p(a \mid \underline{x}) = \lambda_0 p_0(a) + \sum_{k=0}^{\infty} \lambda_k p_k(a \mid x_{-k, -1}) \quad (21)$$

onde $\lambda_k \geq 0$, $\sum_{k=0}^{\infty} \lambda_k = 1$, e cada para cada $k \geq 1$, $p_0(a \mid x_{-k, -1})$ é uma transição markoviana de ordem k para $k \geq 1$, onde p_0 é uma medida de probabilidade em A . As transições regidas por (21) podem ser pensadas como se resultassem de

dois mecanismos aleatórios independentes. Primeiro um inteiro $k \geq 0$ é escolhido com probabilidade λ_k , e, em seguida, um símbolo é escolhido levando em consideração apenas os k últimos símbolos da cadeia de acordo com a probabilidade de transição de ordem k , p_k . Portanto, cada transição depende na verdade de um número finito, mas aleatório, de estados precedentes. Raftery (1985a, 1985b) estudou uma expressão parecida com 21, mas com a ordem k assumindo apenas valores finitos. Ele utilizou o nome de *mixture transition distribution (MTD) model* (ver também Raftery and Tavaré, 1994).

Uma das vantagens deste tipo de processo é que um algoritmo de simulação pode ser facilmente construído com base em uma sequência dupla de variáveis aleatórias uniformes independentes $(U_i, V_i)_{i \in \mathbb{Z}}$. Usamos a primeira variável aleatória uniforme para escolher a ordem de dependência do passado. Usamos a segunda variável aleatória uniforme para gerar a transição de ordem k .

Em Fernández, Ferrari e Galves (2001) é mostrado que toda cadeia de ordem infinita com taxas contínuas é uma mistura enumerável de cadeias de Markov (CMMC). Vamos usar este fato para mostrar a relação entre tempos de renovação em cadeias de ordem infinita e cadeias de Markov. No texto abaixo transcrevemos algumas definições relacionadas aos tempos de renovação para CMMC's.

Dada uma CMMC, para todo instante $n \in \mathbb{Z}$, definimos as ordens aleatórias L_n como

$$L_n = i, \text{ se } U_n \in]\alpha_{i-1}, \alpha_i], \quad (22)$$

onde $\alpha_j = \sum_{i=0}^j \lambda_i$. A lei de L_n é dada pelos coeficientes de mistura λ_j , $j = 1, 2, \dots$

$$P(L_n = k) = \lambda_n. \quad (23)$$

O **tempo de renovação** para a janela (X_l, \dots, X_m) é dado por

$$\tau[l, m] := \max \left\{ t \leq l : t \leq n - L_n, \text{ para todo } n \in [t, m] \right\} \quad (24)$$

com a convenção de que $\tau[l, m] = -\infty$ se o conjunto do lado direito é vazio. No caso em que $l = m$ escrevemos $\tau[l] := \tau[l, l]$.

Para um l fixo, a sequência de tempos de renovação $\tau[l, m]$ para $m \geq l$ é decrescente. A monotonicidade da sequência $(\tau[l, m])_m$ implica na existência do limite

$$\tau[l, +\infty[:= \lim_{m \rightarrow \infty} \tau[l, m]. \quad (25)$$

Definição 2.3. Se $\tau[l, +\infty[< \infty$ então l é chamado de **tempo de renovação da CMMC**.

É fácil ver que um instante $l \in \mathbb{Z}$ será um tempo de renovação de $(X_n)_{n \in \mathbb{Z}}$ se $L_{l+m} \leq m$ para todo $m \geq 1$.

Sejam $(J_n)_{n \in \mathbb{Z}}$ os tempos ordenados tais que

$$\begin{aligned} J_1 &:= \inf\{j > 0 : \tau[j, +\infty[= j\} \\ J_i &:= \inf\{j > \tau_{i-1} : \tau[j, +\infty[= j\}, \text{ para } i > 1, \\ J_i &:= \sup\{j < \tau_{-i+1} : \tau[j, +\infty[= j\} \text{ para } i \leq 0. \end{aligned} \quad (26)$$

Estes tempos de renovação $(J_n)_{n \in \mathbb{Z}}$ dividem a cadeia $(X_n)_{n \in \mathbb{Z}}$ em blocos independentes (ver, por exemplo, Teorema 2.2.11 em Guiol, 2001). Consequentemente, temos que $(J_n)_{n \in \mathbb{Z}}$ também dividem a sequência $(f(X))_{n \in \mathbb{Z}}$ em blocos independentes, onde $f : A \rightarrow \mathbb{R}$ é uma função qualquer.

2.7 Relação entre tempos de renovação de uma cadeia de ordem infinita e das cadeias de Markov aproximantes

Vamos utilizar o fato de que podemos escrever uma cadeia de ordem infinita como uma CMMC para mostrar que sempre que ocorrer um tempo de renovação na cadeia de ordem infinita este tempo será também de renovação na cadeia de ordem k aproximante, para todo k .

Para uma cadeia de ordem infinita podemos decompor

$$p(a|\underline{x}) = \sum_{j=0}^{\infty} \lambda_j p_j(a|x_{-j,-1}). \quad (27)$$

Em Ferrari, Fernández e Galves(2001), a demonstração de que toda cadeia de ordem finita com taxas de decaimento contínuas é uma CMMC é feita dando uma forma explícita para os λ'_j s e os p'_j s que é dada . Em particular,

$$\lambda_j = \min_{x_{-j,-1}} \sum_{a \in A} r_j(a|x_{-j,-1}), \quad (28)$$

onde

$$\begin{aligned} r_0(a) &= \inf_{\underline{z}} p(a|\underline{z}) \\ r_j(a|x_{-j,-1}) &= \inf_{\underline{z}} p(a|x_{-j,-1}\underline{z}). \end{aligned} \quad (29)$$

Fazemos o mesmo para a aproximação canônica cuja probabilidade de transição é $p^{[k]}$. De forma semelhante podemos decompor $p^{[k]}(a|x_{-i,-1})$ como

$$p^{[k]}(a|\underline{x}) = \sum_{j=0}^k \lambda_j p_j(a|x_{-j,-1}). \quad (30)$$

$$\lambda_j^{[k]} = \min_{x_{-j,-1}} \sum_{a \in A} r_j^{[k]}(a|x_{-j,-1}), \quad (31)$$

onde

$$\begin{aligned} r_0^{[k]}(a) &= \inf_{\underline{z}} p^{[k]}(a|\underline{z}) \\ r_j^{[k]}(a|x_{-j,-1}) &= \inf_{\underline{z}} p^{[k]}(a|x_{-j,-1}\underline{z}), \quad 1 \leq j \leq k. \end{aligned} \quad (32)$$

Lema 2.1.

$$r_j^k(a) \geq r_j(a) \quad j = 1, 2, \dots \quad (33)$$

Demonstração

Por definição

$$r_j^{[k]}(a|x_{-j,-1}) = \inf_{\underline{z}} p^{[k]}(a|x_{-j,-1}\underline{z}) = \inf_{x_{-(j+1),-k}} p^{[k]}(a|x_{-j,-1}x_{-(j+1),-k}) \quad (34)$$

e

$$p^{[k]}(a|\underline{z}) = p(a|z_{-k,-1}). \quad (35)$$

Temos que

$$p(a|x_{-k,-1}) = \sum_{y_{-r,-1}} p(a|x_{-k,-1}, y_{-r,-1}) p(y_{-r,-1}|x_{-k,-1}). \quad (36)$$

Mas $p(a|x_{-k,-1}, y_{-r,-1}) \geq r_j^{[k+r]}(a)$ para todo $j = 1, 2, \dots$ e $\sum_{y_{-r,-1}} p(y_{-r,-1}|x_{-k,-1}) = 1$.

Lembramos que, pela definição (2.1) da cadeia de Markov aproximante,

$$\inf_{z_{-k,-1}} p^{[k]}(a|z_{-k,-1}) = \inf_{z_{-k,-1}} p(a|z_{-k,-1}). \quad (37)$$

Logo

$$p(a|x_{-k,-1}) \geq r_j^{[k+r]}(a). \quad (38)$$

Portanto,

$$r_j^{[k]}(a) \geq r_j^{[k+r]}(a) . \quad (39)$$

Logo a sequência $(r_j^{[n]}(a))_{n \in \mathbb{Z}}$ é monótona decrescente limitada e, portanto converge. Finalmente temos que $r_j^{[k+r]}(a)$ converge para $r_0(a)$, pela continuidade, pois $\beta_l \rightarrow 0$.

Como $\lambda_j = \min_{x_{-j,-1}} \sum_{a \in A} r_j(a|x_{-j,-1})$ e $\lambda_j^{[k]} = \min_{x_{-j,-1}} \sum_{a \in A} r_j^{[k]}(a|x_{-j,-1})$, o Lema 2.1 implica que

$$\sum_i \lambda_i \leq \sum_i \lambda_i^{[k]} \quad (40)$$

Lembrando que $\alpha_j = \sum^j \lambda_i$ e definindo $\alpha_j^{[k]} = \sum^j \lambda_i^{[k]}$, temos que

$$\alpha_j \leq \alpha_j^{[k]}, 1 \leq j \leq k.$$

Seja L_n a ordem aleatória da cadeia de ordem infinita e L_n^k a ordem aleatória da cadeia aproximante de ordem k .

Lema 2.2. *Existe um acoplamento tal que*

$$L_n \geq L_n^k, \text{ para todo } n . \quad (41)$$

Demonstração

Definimos

$$\begin{aligned} L_n &= l, \text{ se } \alpha_{l-1} \leq U_n \leq \alpha_l , \\ L_n^k &= j, \text{ se } \alpha_{j-1}^k \leq U_n \leq \alpha_j^k . \end{aligned} \quad (42)$$

Podemos escrever

$$L_n = \sum_{l=1}^{\infty} l \mathbf{1}_{\alpha_{l-1} \leq U_n \leq \alpha_l} , \quad (43)$$

$$L_n^k = \sum_{j=1}^k j \mathbf{1}_{\alpha_{j-1}^k \leq U_n \leq \alpha_j^k} , \quad (44)$$

Como $\alpha_i \leq \alpha_i^k, 1 \leq i \leq k$, sempre que $L_n = l$, temos que $L_n^k = l$. Portanto a ordem aleatória $L_n \geq L_n^k$.

Teorema 2.4. Se $(L_n)_{n \in \mathbb{Z}}$ e $(L_n^k)_{n \in \mathbb{Z}}$ estão acopladas como em (42), se $n \in \mathbb{Z}$ é ponto de renovação de $(X_n)_{n \in \mathbb{Z}}$, n é também um ponto de renovação de $(X_n^{[k]})_{n \in \mathbb{Z}}$, ou seja, se

$$L_{n+m} \leq m \Rightarrow L_{n+m}^k \leq m, \forall m \in \mathbb{N}^*. \quad (45)$$

Demonstração

Consequência imediata dos lemas (2.1) e (2.2).

2.7.1 Aplicação dos tempos de renovação

Nesta seção apresentamos uma aplicação dos tempos de renovação em cadeias de ordem infinita que utilizaremos mais tarde na demonstração da proposição 3.1 do Capítulo 3.

Sejam $(J_n)_{n \in \mathbb{Z}}$ tempos de renovação de $(X_n)_{n \in \mathbb{Z}}$ dados (26) e $f : A \rightarrow \mathbb{R}$ uma função qualquer. Definimos

$$W_i = f(X_i) - \mathbb{E}\{f(X_i)\}$$

e

$$Y_i = \sum_{j=\tau_i+1}^{\tau_{i+1}} W_j$$

Pela propriedade da renovação, os Y_i 's são variáveis aleatórias i.i.d. de média zero.

Temos que $S_n = \sum_{i=1}^n Y_i$ é um martingal. Por cálculos simples de martingais mostra-se que $S_n^2 - \sum_{i=1}^n Y_i^2$ também é um martingal de média zero. Então

$$\mathbb{E}\left\{\left(\sum_{i=1}^n Y_i\right)^2\right\} = \mathbb{E}\left\{\sum_{i=1}^n Y_i^2\right\}.$$

Segue pelo teorema de parada ótima que, se N for um tempo de parada para o martingal, teremos então que

$$\mathbb{E}\{S_N^2\} = \mathbb{E}\{N\}\mathbb{E}\{Y_1^2\}. \quad (46)$$

Observação Pelos resultados da seção anterior, $(J)_{n \in \mathbb{Z}}$ serão também tempos de renovação da cadeia de ordem k aproximante, para todo k . Então (46) também é válido se considerarmos uma função de uma cadeia de Markov de ordem k , para cada k .

Apêndice A_1 :

Propriedades de mistura das cadeias de Markov

Vamos mostrar agora que, para cada k , a cadeia de Markov ordem k , $(X_n^{[k]})_{n \in \mathbb{Z}}$, é φ -misturadora. Sejam os cilindros

$$A_1 = \{X_{-(l+1)}^{[k]} = a_1, \dots, X_0^{[k]} = a_l\}$$

e

$$A_2 = \{X_{n+1}^{[k]} = b_1, \dots, X_{n+q}^{[k]} = b_q\}.$$

Sem perda de generalidade podemos assumir que $q = kl$. Vamos chamar

$$B_i = \{X_{ik+1, (i+1)k}^{[k]} = b_{ik+1, (i+1)k}, i = 0, 1, \dots, l-1\}.$$

Temos que

$$\begin{aligned} \mathbb{P}(A_2, A_1) &= \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1) = \\ &= \mathbb{P}(A_1)\mathbb{P}(X_{n+1, n+q}^{[k]} = b_{1,q} | X_{-(l+1), 0}^{[k]} = a_{1,l}) = \\ &= \mathbb{P}(A_1)\mathbb{P}(B_1, B_2, \dots, B_l | A_1) \\ &= \mathbb{P}(A_1)\mathbb{P}(B_l | B_{l-1})\mathbb{P}(B_{l-1} | B_{l-2}) \dots \mathbb{P}(B_1 | A_1). \end{aligned}$$

Além disso,

$$\begin{aligned} \mathbb{P}(A_1)\mathbb{P}(A_2) &= \mathbb{P}(A_1)\mathbb{P}(X_{n+1, n+q}^{[k]} = b_{1,q}) = \\ &= \mathbb{P}(A_1)\mathbb{P}(B_1, B_2, \dots, B_l) \\ &= \mathbb{P}(A_1)\mathbb{P}(B_l | B_{l-1})\mathbb{P}(B_{l-1} | B_{l-2}) \dots \mathbb{P}(B_2 | B_1)\mathbb{P}(B_1). \end{aligned}$$

Então,

$$\begin{aligned} &|\mathbb{P}(A_2 | A_1) - \mathbb{P}(A_2)| \leq \\ &\leq [\mathbb{P}(X_{n+1, n+k}^{[k]} = b_{1,k} | X_{-k+1, 0}^{[k]} = a_{1,k}) - \mathbb{P}(X_{n+1, n+k}^{[k]} = b_{1,k})] \\ &\quad \times \mathbb{P}(A_1)\mathbb{P}(B_l | B_{l-1})\mathbb{P}(B_{l-1} | B_{l-2}) \dots \mathbb{P}(B_2 | B_1). \end{aligned} \quad (47)$$

Portanto

$$\begin{aligned} |\mathbb{P}(A_2 | A_1) - \mathbb{P}^k(A_2)| &\leq |\mathbb{P}(X_{n, n+k-1}^{[k]} = b_{1,k} | X_{-(k-1), 0}^{[k]} = a_{1,k}) - \mathbb{P}(X_{n, n+k}^{[k]} = b_{1,k})| \leq \\ &\leq \mathbb{P}(\tau^{(b_{1,k}, a_{1,k})} > n) \leq (1 - \rho_k)^n. \end{aligned} \quad (48)$$

Logo,

$$|\mathbb{P}(A_2 | A_1) - \mathbb{P}^k(A_2)| \leq (1 - \rho_k)^n. \quad (49)$$

Então chamando $(1 - \rho_k)^n = \varphi_k(n)$ fica demonstrada a propriedade φ -misturadora para cada k .

Cadeias de Markov de ordem k são ψ -misturadoras

Para demonstrar que $(X_n^{[k]})_{n \in \mathbb{Z}}$ é ψ -misturadora, retomamos a expressão (47), multiplicamos e dividimos por $\mathbb{P}(B_1)$:

$$\begin{aligned} |\mathbb{P}(A_2 | A_1) - \mathbb{P}^{[k]}(A_2)| &\leq |\mathbb{P}(X_{n,n+k-1}^{[k]} = b_{1,k} | X_{-(k-1),0}^{[k]} = a_{1,k}) - \mathbb{P}(X_{n,n+k}^{[k]} = b_{1,k})| \leq \\ &\leq [\mathbb{P}(X_{n+1,n+k}^{[k]} = b_{1,k} | X_{-k+1,0}^{[k]} = a_{1,k}) - \mathbb{P}(X_{n+1,n+k}^{[k]} = b_{1,k})] \\ &\quad \times \frac{\mathbb{P}(A_1)\mathbb{P}(B_l | B_{l-1})\mathbb{P}(B_{l-1} | B_{l-2}) \dots \mathbb{P}(B_2 | B_1)\mathbb{P}(B_1)}{\mathbb{P}(B_1)}, \end{aligned} \quad (50)$$

$$\leq \frac{(1 - \rho_k)^n \mathbb{P}(A_1)\mathbb{P}(A_2)}{\mathbb{P}(B_1)}, \quad (51)$$

mas, por (19), $\mathbb{P}(B_1) > \delta^k$, portanto, para cada k , fazendo

$$\psi_k(n) = \frac{(1 - \rho_k)^n}{\delta^k},$$

temos que a cadeia de Markov de ordem k é ψ -misturadora.

Capítulo 3

Reamostragem para sequências de cadeias de Markov

Seja uma função $f : A^r \rightarrow \mathbb{R}$. Consideremos uma cadeia de Markov de ordem k , $(X_n^{[k]})_{n \in \mathbb{Z}}$, e seja

$$\mu^{[k]} = \mathbb{E}(f(X_0^{[k]}, \dots, X_r^{[k]})).$$

Sem perda de generalidade, podemos considerar apenas o caso em que $r = 1$. O objetivo deste capítulo será propor um estimador para $\mu^{[k]}$ com boas propriedades assintóticas que nos permita estimar a sua variância.

Observamos que se considerarmos uma amostra $X_1^{[k]}, \dots, X_n^{[k]}$ da cadeia $(X_n^{[k]})$, um estimador natural para $\mu^{[k]}$ é dado por

$$\hat{\mu}_n^{[k]} = \frac{1}{n} \sum_{i=1}^n f(X_i^{[k]}) \quad (1)$$

A consistência de $\hat{\mu}_n^{[k]}$ segue diretamente do teorema ergódico. Como cadeias de Markov são φ -misturadoras, a normalidade assintótica segue do Teorema 2.1 (com algumas hipóteses nos coeficientes de mistura). Mas neste caso não teremos um estimador da variância de $\hat{\mu}_n^{[k]}$.

3.1 Reamostragem sequencial

Em Athreya e Fuh (1992) é apresentado um procedimento de reamostragem para uma única cadeia de Markov de ordem um. Nesta seção este procedimento é generalizado para uma sequência de cadeias de Markov com alcances k crescendo

para infinito. Este resultado é a principal ferramenta para a demonstração do Teorema Limite Central da Reamostragem para cadeias de ordem infinita que discutiremos na Seção 2.2.

Seja $(X_n^{[k]})_{n \in \mathbb{Z}}$ uma sequência de Cadeias de Markov com $k = 1, 2, \dots$, tomando valores no alfabeto finito A . Consideramos que cada cadeia de Markov é homogênea no tempo.

Definimos uma taxa global de mistura para a sequência $(X_n^{[k]})_{n \in \mathbb{Z}}$, $k = 1, 2, \dots$ da seguinte maneira. Seja $\mathcal{C}_{(p,q)}^{[k]}$ o conjunto dos cilindros entre p e $q > p$. Dizemos que uma sequência satisfaz a condição global de mistura se para cada inteiro positivo n , existe $\varphi(n) \rightarrow 0$, com $\sum \varphi(n) < \infty$, tal que

$$\sup_{k,l,m} \sup_{C \in \mathcal{C}_{(-l,0)}^{[k]}} \sup_{D \in \mathcal{C}_{(n,n+m)}^{[k]}} |\mathbb{P}(C | D) - \mathbb{P}(C)| \leq \varphi(n). \quad (2)$$

Daqui em diante consideraremos apenas sequências de cadeias de Markov que satisfazem (2).

O procedimento da reamostragem sequencial consiste no seguinte. Para cada k , consideramos uma amostra $X_1^{[k]}, X_2^{[k]}, \dots$ de $(X_n^{[k]})_{n \in \mathbb{Z}}$ e olhamos para as ocorrências sucessivas dos primeiros k símbolos, $X_1^{[k]}, \dots, X_k^{[k]}$. A propriedade de Markov de ordem k da cadeia $(X_n^{[k]})_{n \in \mathbb{Z}}$ implica que os blocos de valores delimitados pelas sucessivas ocorrências são independentes entre si e identicamente distribuídos. Fazemos uma reamostragem uniforme i.i.d. dos blocos delimitados por essas ocorrências. A seguir apresentamos formalmente estas idéias.

Para cada k definimos a sequência de tempos de retorno $(R_j^{[k]})_{j \in \mathbb{N}}$ como

$$R_{i+1}^{[k]} = \inf \left\{ t > R_i^{[k]} : (X_t^{[k]}, \dots, X_{t+k-1}^{[k]}) = (X_1^{[k]}, \dots, X_k^{[k]}) \right\}, \quad (3)$$

com $R_0^{[k]} = 1$.

O bloco $\xi_i^{[k]}$ dos valores da cadeia $(X_n^{[k]})_{n \in \mathbb{Z}}$ indo de $R_{i-1}^{[k]}$ até $R_i^{[k]} - 1$, $i = 1, \dots, m_k$, é definido como

$$\xi_i^{[k]} = (X_{R_{i-1}^{[k]}}^{[k]}, \dots, X_{R_i^{[k]}-1}^{[k]}). \quad (4)$$

Desta forma obtemos uma sequência de m_k blocos, $\xi_1^{[k]}, \dots, \xi_{m_k}^{[k]}$. O número de ocorrências sucessivas dos primeiros k símbolos, m_k , é uma função de k que explicitaremos oportunamente.

Construímos uma reamostra da cadeia de ordem k fazendo uma seleção independente e uniformemente distribuída destes blocos. Ou seja, para cada k , sejam $I_1(k), \dots, I_{m_k}(k)$ variáveis aleatórias independentes com distribuição uniforme no conjunto $\{1, \dots, m_k\}$. Definimos os blocos reamostrados como

$$\xi_l^{[k]*}(k) = \xi_{I_l(k)}^{[k]},$$

para $l = 1, \dots, m_k$.

Os tempos de retorno na reamostra para $l = 1, \dots, m_k$ são dados por

$$R_l^{[k]*} = R_{l-1}^{[k]*} + R_{I_l(k)+1}^{[k]} - R_{I_l(k)}^{[k]},$$

com

$$R_0^{[k]*} = 1.$$

Então, para $1 \leq n \leq m_k$, cada bloco reamostrado fica

$$\xi_n^{[k]*} = \left(X_{R_{n-1}^{[k]*}}^{[k]*}, \dots, X_{R_n^{[k]*}-1}^{[k]*} \right). \quad (5)$$

A reamostra $X_{n^*}^{[k]*}$ é obtida concatenando os blocos $\xi_1^{[k]*}, \xi_2^{[k]*}, \dots, \xi_{m_k}^{[k]*}$, onde $n^* = \sum_{i=1}^{m_k} |\xi_i^{[k]*}|$.

Propomos um estimador para $\mu^{[k]} = \mathbb{E}(f(X_0^{[k]}))$ da seguinte maneira

$$\hat{\mu}^{[k]} = \frac{1}{R_{m_k}^{[k]}} \sum_{n=1}^{R_{m_k}^{[k]}} f(X_n^{[k]}), \quad (6)$$

e sua contrapartida baseada nos blocos reamostrados como

$$\mu^{[k]*} = \frac{1}{R_{m_k}^{[k]*}} \sum_{n=1}^{R_{m_k}^{[k]*}} f(X_n^{[k]*}). \quad (7)$$

Observação Se os primeiros símbolos da amostra apresentarem uma estrutura periódica (exemplo: abcabc...) pode acontecer de $R_2^{[k]}$ ser menor do que k . Neste caso, $\xi_1^{[k]}$ e $\xi_2^{[k]}$, (e também outros blocos) teriam muitas observações em comum. Para resolver este problema bastaria definir os tempo de retorno da seguinte maneira

$$R_{i+1}^{[k]} = \inf \left\{ t > R_i^{[k]} + k : (X_t^{[k]}, \dots, X_{t+k-1}^{[k]}) = (X_1^{[k]}, \dots, X_k^{[k]}) \right\}. \quad (8)$$

Esta modificação da definição dos $R_i^{[k]}$ não acarreta em nenhuma mudança significativa nas demonstrações. Além disto, em Collet, Galves e Schmitt (1999) é mostrado que a probabilidade de aparecerem seqüências periódicas de tamanho k tende a zero com k . Isto significa que, na prática, tais seqüências não serão grande um problema. Optamos então em deixar a definição dos $R_i^{[k]}$'s como em (1).

3.2 Teorema Limite Central da Reamostragem para $\mu^{[k]*}$

Para os estimadores da média definidos na seção anterior, temos que

Teorema 3.1. *Seja $(X_n^{[k]})_{n \in \mathbb{Z}}$ uma seqüência de cadeias de Markov de ordens $\{k = 1, 2, \dots\}$, respectivamente, satisfazendo a condição global de mistura (2). Então para $m_k = \lceil e^{\alpha k} \rceil$, $\alpha > 3 \ln(1/\delta)$, temos para quase toda seqüência de realizações da cadeia que*

$$\frac{\sqrt{R_{m_k}^{[k]*}}}{\eta_*^{[k]}} \{ \mu^{[k]*} - \hat{\mu}^{[k]} \} \rightarrow \mathcal{N}(0, 1), \quad (9)$$

quando k diverge, onde

$$\eta_*^{[k]} = \left(\frac{\text{var} \left(\sum_j^{R_{m_k}^{[k]*}} f(X_j^{[k]*}) - \hat{\mu}^{[k]} \mid X_1^{[k]}, \dots, X_{R_{m_k}^{[k]}}^{[k]} \right)}{R_{m_k}^{[k]*}} \right)^{1/2}. \quad (10)$$

Observamos que $\eta_*^{[k]}$ é uma versão reamostrada de

$$\eta^{[k]} = \left(\frac{\text{var} \left(\sum_j^{R_{m_k}^{[k]*}} f(X_j^{[k]}) - \mu^{[k]} \right)}{R_{m_k}^{[k]}} \right)^{1/2} \quad (11)$$

e $\eta^{[k]} \rightarrow \sigma$ quando $k \rightarrow \infty$ (ver apêndice A_2).

O Teorema 3.1 é válido para quaisquer seqüências de cadeias de Markov satisfazendo (2). Para utilizar os resultados de aproximação da cadeia de ordem infinita por seqüências de cadeias de Markov apresentados em Bressaud, Fernández

e Galves (1999), precisamos de um tipo especial de sequência de cadeias de Markov, chamada de aproximação canônica, definida no capítulo 2. Mas o Teorema 3.1 também vale para a aproximação canônica, como se vê no teorema abaixo.

Teorema 3.2. *Se $(X_n^{[k]})_{n \in \mathbb{Z}}$ é uma sequência canônica de aproximações markovianas temos que a condição (2) é satisfeita.*

Demonstração

Consideremos a sequência canônica de cadeias de Markov aproximantes $(X_n^{[k]})_{n \in \mathbb{Z}}$ definida em (2.1). Sejam os cilindros

$$A_1 = \{X_{-(l+1)}^{[k]} = a_1, \dots, X_0^{[k]} = a_l\}$$

e

$$A_2 = \{X_{n+1}^{[k]} = b_1, \dots, X_{n+q}^{[k]} = b_q\}.$$

Vamos construir duas trajetórias acopladas da cadeia de Markov,

$$(X_n^{[k], a_{1,k}})_{n \geq -k+1} \text{ e } (X_n^{[k], b_{1,k}})_{n \geq -k+1},$$

usando o acoplamento maximal descrito na seção 2.2.

Lembramos que a probabilidade de que as duas trajetórias se encontrem satisfaz a seguinte desigualdade

$$\mathbb{P}(\tau^{(b_{1,k}, a_{1,k})} > n) \leq (1 - \rho_k)^n. \quad (12)$$

onde $\tau^{(b_{1,k}, a_{1,k})}$, o instante de encontro das trajetórias, e ρ_k são definidos em (13) e (14), respectivamente.

Por (49), temos que

$$\sup_{l, q} \sup_{A_1 \in \mathcal{C}_{-l, 0}^{[k]}} \sup_{A_2 \in \mathcal{C}_{n, n+q}^{[k]}} |\mathbb{P}(A_2 | A_1) - \mathbb{P}^{[k]}(A_2)| \leq (1 - \rho_k)^n. \quad (13)$$

Temos que mostrar então que existe $\rho = \lim_{k \rightarrow \infty} \rho_k$ para definir assim o coeficiente global de mistura para a sequência de cadeias de Markov aproximantes.

Começamos lembrando que

$$|p(a|x_{-\infty, -1}) - p^{[k]}(a|x_{-k, -1})| \leq \beta_k, \quad (14)$$

onde β_k é a taxa de continuidade da cadeia de ordem infinita definida em (6). Este resultado aparece como lema 5 em Bressaud, Fernández e Galves (1999).

portanto

$$|p^{[k]}(a|x_{-k,-1}) - p^{[k+1]}(a|x_{-(k+1),-1})| \leq 2\beta_k . \quad (15)$$

Vamos avaliar $|\rho_k - \rho_{k+1}|$. Começamos avaliando $|\rho_k(x, y) - \rho_{k+1}(x, y)|$, onde

$$\rho_k(x, y) = \sum_{a \in A} (p^{[k]}(a|x_{-k,-1}) \wedge p^{[k]}(a|y_{-k,-1})) .$$

Lembramos que

$$|\alpha \wedge \beta - \alpha' \wedge \beta'| \leq |\alpha - \alpha'| \vee |\beta - \beta'| . \quad (16)$$

Aplicando (16) para as probabilidades de transição em $|\rho_k(x, y) - \rho_{k+1}(x, y)|$, obtemos

$$\begin{aligned} & \left| \sum_{a \in A} (p^{[k]}(a|x_{-k,-1}) \wedge p^{[k]}(a|y_{-k,-1})) - \sum_{a \in A} (p^{[k+1]}(a|x_{-(k+1),-1}) \wedge p^{[k+1]}(a|y_{-(k+1),-1})) \right| \leq \\ & \left| \sum_{a \in A} [(p^{[k]}(a|x_{-k,-1}) \wedge p^{[k]}(a|y_{-k,-1})) - (p^{[k+1]}(a|x_{-(k+1),-1}) \wedge p^{[k+1]}(a|y_{-(k+1),-1}))] \right| \\ & \leq \sum_{a \in A} |(p^{[k]}(a|x_{-k,-1}) \wedge p^{[k]}(a|y_{-k,-1})) - (p^{[k+1]}(a|x_{-(k+1),-1}) \wedge p^{[k+1]}(a|y_{-(k+1),-1}))| \\ & \leq \sum_{a \in A} [|p^{[k]}(a|x_{-k,-1}) - p^{[k+1]}(a|x_{-(k+1),-1})| \vee |p^{[k]}(a|y_{-k,-1}) - p^{[k+1]}(a|y_{-(k+1),-1})|] \\ & = 2|A|\beta_k , \forall \underline{x} \text{ e } \underline{y} . \end{aligned}$$

Logo, para cada par (x, y) , temos que

$$\rho_k(x, y) \rightarrow \rho(x, y). \quad (17)$$

Logo, o limite $\lim_{k \rightarrow \infty} \rho_k = \rho$ existe e para cada ϵ existe $k_0(\epsilon)$ tal que $\forall k \geq k_0(\epsilon)$, $\rho_k \in [\rho - \epsilon, \rho + \epsilon]$ e, conseqüentemente,

$$(1 - (\rho - \epsilon))^n \geq (1 - \rho_k)^n \quad (18)$$

para todo n . Além disto, $\rho > 0$, pois estamos assumindo que a cadeia é do tipo C. Então a condição global de mistura (2) é satisfeita fazendo

$$\varphi(n) = (1 - (\rho - \epsilon))^n . \quad (19)$$

Isto encerra a demonstração do teorema.

3.3 Resultados preliminares

Antes de provar o Teorema 3.1 precisamos de alguns resultados auxiliares.

Seja

$$S_i^{[k]} = \sum_{l=R_{i-1}^{[k]}}^{R_i^{[k]}-1} f(X_l^{[k]}).$$

Informalmente falando, esta é a soma dos valores da função f dentro do bloco $\xi_i^{[k]}$. Sua versão reamostrada é dada por

$$S_i^{[k]*} = S_{I_i}^{[k]} = \sum_{l=R_{i-1}^{[k]*}}^{R_i^{[k]*}-1} f(X_l^{[k]*}).$$

Lembramos que $I_i, i = 1, 2, \dots$ são os índices dos blocos reamostrados obtidos através de sorteios independentes com reposição.

Nos lemas abaixo apresentamos alguns resultados úteis envolvendo estas somas.

Vamos usar a notação $\mathbb{E}^*\{\cdot\}$ para designar $\mathbb{E}\{\cdot | X_1^{[k]}, \dots, X_{R_{m_k}^{[k]}}^{[k]}\}$ e $var^*\{\cdot\}$ para $var\{\cdot | X_1^{[k]}, \dots, X_{R_{m_k}^{[k]}}^{[k]}\}$. Observamos que

$$\mathbb{E}^*\{S_1^{[k]*}\} = \sum_{j=1} S_j^{[k]} \mathbb{P}(I_1 = j) = \frac{1}{m_k} \sum_{j=1} S_j^{[k]}.$$

Lema 3.1. *Temos que*

1.

$$\mathbb{E}^* \left(\sum_{l=1}^{m_k} S_l^{[k]*} - R_{m_k}^{[k]*} \hat{\mu}^{[k]} \right) = 0. \quad (20)$$

2.

$$Var^* \left(\sum_{l=1}^{m_k} S_l^{[k]*} - R_{m_k}^{[k]*} \hat{\mu}^{[k]} \right) = \sum_{l=1}^{m_k} \left[S_l^{[k]} - \hat{\mu}^{[k]} (R_{l+1}^{[k]} - R_l^{[k]}) \right]^2. \quad (21)$$

3.

$$\mathbb{E}^* \left\{ \left[S_1^{[k]*} - (R_2^{[k]*} - R_1^{[k]*}) \hat{\mu}^{[k]} \right]^4 \right\} = \frac{1}{m_k} \sum_{j=1}^{m_k} \left[S_j^{[k]} - (R_{j+1}^{[k]} - R_j^{[k]}) \hat{\mu}^{[k]} \right]^4. \quad (22)$$

Demonstração

Para simplificar a notação, chamamos $(R_{i+1}^{[k]} - R_i^{[k]}) = D_i^k$.

1.

$$\begin{aligned} \mathbb{E}^* \left[\sum_{l=1}^{m_k} S_l^{[k]*} - R_{m_k}^{[k]*} \hat{\mu}^{[k]} \right] &= \mathbb{E}^* \sum_{i=1}^{m_k} (S_{I_i}^{[k]} - D_{I_i}^k \hat{\mu}^{[k]}) = \\ &= m_k \sum_{j=1}^{m_k} \mathbb{P}(I_i = j) (S_j^{[k]} - D_j^k \hat{\mu}^{[k]}) \\ &= \sum_{j=1}^{m_k} (S_j^{[k]} - D_j^k \hat{\mu}^{[k]}) \\ &= \sum_{j=1}^{m_k} S_j^{[k]} - R_{m_k}^{[k]} \hat{\mu}^{[k]} = 0 . \end{aligned}$$

2.

$$\begin{aligned} \text{var}^* \left(\sum_{l=1}^{m_k} S_l^{[k]*} - R_{m_k}^{[k]*} \hat{\mu}^{[k]} \right) &= \mathbb{E}^* \left(\sum_{i=1}^{m_k} (S_{I_i}^{[k]} - D_{I_i}^k \hat{\mu}^{[k]}) \right)^2 \\ &= \sum_{j=1}^{m_k} (S_j^{[k]} - D_j^k \hat{\mu}^{[k]})^2 . \end{aligned}$$

3. Análoga à do item anterior e será omitida.

Lema 3.2. *Seja $(X_n^{[k]})_{n \in \mathbb{Z}}$ uma seqüência de cadeias de Markov de ordem $k \in \mathbb{N}$ crescente. Temos, para $r = 1, 2, \dots$, que*

$$\frac{r!}{\lambda_k^{r-2}} - e^{-C_3 k} \frac{r! \delta^k}{C_4^{r-1}} \leq \mathbb{E}\{(D_1^k)^r\} \leq \frac{r!}{\lambda_k^{r-2}} + e^{-C_3 k} \frac{r! \delta^k}{C_4^{r-1}} . \quad (23)$$

A principal ferramenta na demonstração deste lema é o Teorema 3.3 abaixo que é apresentado em Abadi (2002). Antes de enunciar o teorema precisamos de algumas definições:

Definição 3.1. *Para um subconjunto $B \subset \Omega$ dizemos que $B \in \mathcal{C}_n$ se e somente se*

$$B = \{X_0 = a_0, \dots, X_{n-1} = a_{n-1}\} .$$

Denotamos por $T : \Omega \rightarrow \Omega$ a translação de um passo para a direita. Definimos a periodicidade p_B de B , com respeito à translação T , como sendo

$$p_B = \inf\{n \in \mathbb{N} \mid B \cap T^n(B) \neq \emptyset\}.$$

Denotamos por $\mathcal{R}_p = \mathcal{R}_p(n)$ à família de sequências de tamanho n e periodicidade p_B , isto é

$$\mathcal{R}_p = \{B \in \mathcal{C}_n \mid p_B = p\}.$$

De acordo com esta notação temos o seguinte resultado

Teorema 3.3. *Seja $(Y_n)_{n \in \mathbb{Z}}$ um processo φ -misturador com coeficiente de mistura exponencial. Então, existem constantes estritamente positivas C_1, C_2 e C_3 tais que quaisquer que sejam $n \in \mathbb{N}$ e $y_{1,k} \in \mathcal{R}_p$ a seguinte desigualdade vale para todo $1 \geq t \geq p$*

$$\begin{aligned} & |\mathbb{P}(D_1^k > t \mid Y_{1,k} = y_{1,k}) - \lambda_k e^{-\lambda_k \mathbb{P}(Y_{1,k} = y_{1,k})t}| \\ & \leq C_1 e^{-C_2 t} e^{-C_3 k} \mathbb{P}(Y_{1,k} = y_{1,k}), \end{aligned} \quad (24)$$

onde $\lambda_k = \mathbb{P}(D_1^k > p_{y_{1,k}} \mid Y_{1,k} = y_{1,k})$.

Demonstração

Vamos usar o resultado acima para encontrar cotas inferior e superior para $\mathbb{E}\{D_1^k\}$. Primeiro notemos que

$$\begin{aligned} \mathbb{P}((D_1^k)^r > t) &= \sum_{x_{1,k}} \mathbb{P}(X_{1,k}^{[k]} = x_{1,k}) \mathbb{P}((D_1^k)^r > t \mid X_{1,k}^{[k]} = x_{1,k}) \\ &= \sum_{x_{1,k}} \mathbb{P}(X_{1,k}^{[k]} = x_{1,k}) \mathbb{P}((D_1^k) > t^{1/r} \mid X_{1,k}^{[k]} = x_{1,k}). \end{aligned} \quad (25)$$

Denotando $\mathbb{P}(X^{[k]} = x_{1,k}) = p(x_{1,k})$ e $p(\cdot \mid X_{1,k}^{[k]} = x_{1,k})$ por $p(\cdot \mid x_{1,k})$, temos

$$\begin{aligned} \mathbb{E}\{(D_1^k)^r\} &= \int \mathbb{P}((D_1^k)^r > t) dt \\ &= \int \sum_{x_{1,k}} p(x_{1,k}) \mathbb{P}((D_1^k)^r > t \mid x_{1,k}) dt \\ &= \sum_{x_{1,k}} p(x_{1,k}) \int \mathbb{P}((D_1^k)^r > t \mid x_{1,k}) dt \\ &= \sum_{x_{1,k}} p(x_{1,k}) \int \mathbb{P}((D_1^k) > t^{1/r} \mid x_{1,k}) dt. \end{aligned} \quad (26)$$

Utilizando o Teorema 3.3, temos

$$\mathbb{E}\{(D_1^k)^r\} \leq \sum_{x_{1,k}} p(x_{1,k}) \int [\lambda_k e^{-\lambda_k p(x_{1,k}) t^{1/r}} + C_1 p(x_{1,k}) e^{-C_2 t^{1/r}} e^{-C_3 k}] dt \quad (27)$$

e

$$\mathbb{E}\{(D_1^k)^r\} \geq \sum_{x_{1,k}} p(x_{1,k}) \int [\lambda_k e^{-\lambda_k p(x_{1,k}) t^{1/r}} - C_1 p(x_{1,k}) e^{-C_2 t^{1/r}} e^{-C_3 k}] dt . \quad (28)$$

Apresentamos em detalhes os cálculos para (27). A integral do primeiro termo fica

$$\lambda_k \int e^{-\lambda_k p(x_{1,k}) t^{1/r}} dt = r \int e^{-\lambda_k p(x_{1,k}) s} s^{r-1} ds = \frac{r!}{\lambda_k^{r-2} p(x_{1,k})^{r-1}} . \quad (29)$$

A integral do segundo termo fica

$$C_1 p(x_{1,k}) e^{-C_3 k} \int e^{-C_2 t^{1/r}} dt = C_1 p(x_{1,k}) e^{-C_3 k} r \int e^{-C_2 s} s^{r-1} ds = e^{-C_3 k} \frac{r! C_1 p(x_{1,k})}{C_2^{r-1}} .$$

Logo

$$\mathbb{E}\{(D_1^k)^r\} \leq \sum_{x_{1,k}} p(x_{1,k}) \left[\frac{r!}{\lambda_k^{r-2} p(x_{1,k})^{r-1}} + e^{-C_3 k} \frac{r! C_1 p(x_{1,k})}{C_2^{r-1}} \right] .$$

Por (19) $\delta^k \leq p(x_{1,k}) \leq 1$. Então

$$\mathbb{E}\{(D_1^k)^r\} \leq \sum_{x_{1,k}} p(x_{1,k}) \left[\frac{r!}{\lambda_k^{r-2} \delta^{k(r-1)}} + e^{-C_3 k} \frac{r!}{C_4^{r-1}} \right] ,$$

onde $C_4 = C_1/C_2$.

$$= \frac{r!}{\lambda_k^{r-2} \delta^{k(r-1)}} + e^{-C_3 k} \frac{r!}{C_4^{r-1}} . \quad (30)$$

Analogamente, verificamos que

$$\mathbb{E}\{(D_1^k)^r\} \geq \frac{r!}{\lambda_k^{r-2}} - e^{-C_3 k} \frac{r! \delta^k}{C_4^{r-1}} . \quad (31)$$

Observação Para $r = 1$, de maneira análoga mostra-se que existem constantes C_5 e C_6 tais que

$$\mathbb{E}\{D_1^k\} \leq \frac{1}{\delta^k} + C_5 e^{-C_3 k} . \quad (32)$$

e

$$\mathbb{E}\{D_1^k\} \geq 1 - C_6 e^{-C_3 k} \delta^k . \quad (33)$$

Lema 3.3. Se $m_k = \lceil e^{\alpha k} \rceil$, com $\alpha > (1 - r) \ln(\delta)$, então

$$\frac{\mathbb{E}\{(D_1^k)^r\}}{m_k} \rightarrow 0. \quad (34)$$

Demonstração

Pelo lema (3.2) temos que

$$\frac{\mathbb{E}\{(D_1^k)^r\}}{m_k} \leq \frac{1}{m_k} \left(\frac{r!}{\lambda_k^{r-2} \delta^{k(r-1)}} + e^{-C_3 k} \frac{r!}{C_4^{r-1}} \right) . \quad (35)$$

Como o segundo termo à direita tende a zero com k , temos apenas que mostrar que

$$m_k \lambda_k^{r-2} \delta^{k(r-1)} \rightarrow \infty ,$$

quando $k \rightarrow \infty$. Mas $\lambda_k^{r-2} = \mathbb{P}(D_1^k > p_{y_{1,k}} | Y_{1,k} = y_{1,k})^{r-2}$ converge para uma constante quando $k \rightarrow \infty$. como

$$\begin{aligned} \alpha &> (1 - r) \ln(\delta) , \\ m_k \delta^{k(r-1)} &= e^{k(\alpha + (r-1) \ln \delta)} \rightarrow 0 , \end{aligned}$$

se $k \rightarrow \infty$ concluindo a demonstração.

Observação 1 Para $r = 1$, basta tomarmos $\alpha > \ln(1/\delta)$.

Observação 2 As constantes que aparecem no teorema (3.3) dependem, na verdade, da função φ_k de mistura do processo. Ou seja, são funções $C_k = C_{\varphi_k}$. Mas como vamos considerar apenas processos que satisfazem (2), as constantes serão todas majoradas por uma constante C_φ .

Proposição 3.1. Existe uma constante C tal que

$$\mathbb{E}\left\{ \left[\sum_{n=1}^{D_1^k} (f(X_n^{[k]}) - \mu^{[k]}) \right]^2 \right\} \geq C \mathbb{E}\{D_1^k\} \mathbb{E}\left\{ \left(\sum_{j=J_1+1}^{J_2} W_j \right)^2 \right\} .. \quad (36)$$

Demonstração

Sejam J_1, J_2, \dots os tempos de renovação de $(X_n)_{n \in \mathbb{Z}}$ (também tempos de renovação de $(X_n^{[k]})_{n \in \mathbb{Z}}$). Seja $N = \inf\{J_i > D_1^{[k]}, i = 1, 2, \dots\}$. Observamos que os tempos de renovação não dependem de k . Seja $W_j = (f(X_j^{[k]}) - \mu^{[k]})$, temos que

$$\begin{aligned}
 \mathbb{E}\left\{\left(\sum_{j=1}^{D_1^k} W_j\right)^2\right\} &= \mathbb{E}\left\{\left(\sum_{j=1}^{J_N} W_j + \sum_{j=D_1^k+1}^{J_N} W_j\right)^2\right\} \\
 &= \mathbb{E}\left\{\left(\sum_{j=1}^{J_N} W_j\right)^2\right\} + \mathbb{E}\left\{\left(\sum_{j=D_1^k+1}^{J_N} W_j\right)^2\right\} \\
 &\quad - 2\mathbb{E}\left\{\left(\sum_{j=1}^{J_N} W_j\right)\left(\sum_{j=D_1^k+1}^{J_N} W_j\right)\right\}. \tag{37}
 \end{aligned}$$

Mas

$$\begin{aligned}
 \mathbb{E}\left\{\left(\sum_{j=1}^{J_N} W_j\right)\left(\sum_{j=D_1^k+1}^{J_N} W_j\right)\right\} &= \mathbb{E}\left\{\left(\sum_{j=1}^{D_1^k} W_j + \sum_{j=D_1^k+1}^{J_N} W_j\right)\sum_{j=D_1^k+1}^{J_N} W_j\right\} \\
 &= \mathbb{E}\left\{\left(\sum_{j=1}^{D_1^k} W_j\right)\left(\sum_{j=D_1^k+1}^{J_N} W_j\right)\right\} \\
 &\quad + \mathbb{E}\left\{\left(\sum_{j=D_1^k+1}^{J_N} W_j\right)^2\right\} \\
 &= \mathbb{E}\left\{\left(\sum_{j=D_1^k+1}^{J_N} W_j\right)^2\right\}, \tag{38}
 \end{aligned}$$

pois

$$\mathbb{E}\left\{\left(\sum_{j=1}^{D_1^k} W_j\right)\left(\sum_{j=D_1^k+1}^{J_N} W_j\right)\right\} = \mathbb{E}\left\{\sum_{j=1}^{D_1^k} W_j\right\}\mathbb{E}\left\{\sum_{j=D_1^k+1}^{J_N} W_j\right\} = 0, \tag{39}$$

uma vez que $\mathbb{E}\left\{\sum_{j=1}^{D_1^k} W_j\right\} = 0$.

Logo,

$$\begin{aligned}
\mathbb{E}\left\{\left(\sum_{j=1}^{D_1^k} W_j\right)^2\right\} &= \mathbb{E}\left\{\left(\sum_{j=1}^{J_N} W_j\right)^2\right\} - \mathbb{E}\left\{\left(\sum_{j=D_1^k+1}^{J_N} W_j\right)^2\right\} \\
&\geq \mathbb{E}\left\{\left(\sum_{i=1}^N \sum_{j=J_{i+1}}^{J_{i+1}} W_j\right)^2\right\} - \mathbb{E}\left\{\left(\sum_{j=J_{N-1}}^{J_N} W_j\right)^2\right\} \\
&= \mathbb{E}\left\{\left(\sum_{i=1}^N \sum_{j=J_{i+1}}^{J_{i+1}} W_j\right)^2\right\} - \mathbb{E}\left\{\left(\sum_{j=J_1+1}^{J_2} W_j\right)^2\right\}.
\end{aligned}$$

Como N é tempo de parada, utilizando (46), temos que

$$\mathbb{E}\left\{\left(\sum_{i=1}^N \sum_{j=J_{i+1}}^{J_{i+1}} W_j\right)^2\right\} = \mathbb{E}\{N\} \mathbb{E}\left\{\left(\sum_{j=J_1+1}^{J_2} W_j\right)^2\right\}. \quad (40)$$

Então

$$\mathbb{E}\left\{\left(\sum_{j=1}^{D_1^k} W_j\right)^2\right\} \geq \mathbb{E}\{N-1\} \mathbb{E}\left\{\left(\sum_{j=J_1+1}^{J_2} W_j\right)^2\right\}. \quad (41)$$

Observamos que $\mathbb{E}\{N\} \geq \alpha_k \mathbb{E}\{D_1^k\}$, com $\alpha_k \geq \alpha$ (ver equação (??)). Logo

$$\begin{aligned}
\mathbb{E}\left\{\left(\sum_{j=1}^{D_1^k} W_j\right)^2\right\} &\geq \mathbb{E}\{\alpha_k \mathbb{E}\{D_1^k\} - 1\} \mathbb{E}\left\{\left(\sum_{j=J_1+1}^{J_2} W_j\right)^2\right\} \\
&\geq (\alpha - \epsilon) \mathbb{E}\{D_1^k\} \mathbb{E}\left\{\left(\sum_{j=J_1+1}^{J_2} W_j\right)^2\right\}.
\end{aligned} \quad (42)$$

Proposição 3.2.

$$2\mathbb{E}\{D_1^k\} \sum_{n=1}^{D_1^k} (f(X_n^{[k]}) - \mu^{[k]})(\mu^{[k]} - \hat{\mu}^{[k]}) \rightarrow 0, \quad (43)$$

quando $k \rightarrow \infty$.

Demonstração

Definimos os seguintes estimadores:

$$\hat{\mu}_1^{[k]} = \frac{\sum_{n=1}^{D_1^k} f(X_n^{[k]})}{D_1^k}, \quad (44)$$

$$\hat{\mu}_2^{[k]} = \frac{\sum_{n=D_1^k+1}^{R_{m_k}^{[k]}} f(X_n^{[k]})}{D_2^k + \dots + D_{m_k}^k}. \quad (45)$$

Então

$$\mathbb{E}\left\{D_1^k \sum_{n=1}^{D_1^k} (f(X_n^{[k]}) - \mu^{[k]})(\mu^{[k]} - \hat{\mu}_1^{[k]})\right\} = -\mathbb{E}\{(D_1^k)^2 (\hat{\mu}_1^{[k]} - \mu^{[k]})(\mu^{[k]} - \hat{\mu}_1^{[k]})\}. \quad (46)$$

Somando e subtraindo $\hat{\mu}_2^{[k]}$ no segundo fator, temos

$$-\mathbb{E}\{(D_1^k)^2 (\hat{\mu}_1^{[k]} - \mu^{[k]})(\hat{\mu}_2^{[k]} - \mu^{[k]})\} \quad (47)$$

$$-\mathbb{E}\{(D_1^k)^2 (\hat{\mu}_1^{[k]} - \mu^{[k]})(\hat{\mu}_1^{[k]} - \hat{\mu}_2^{[k]})\}. \quad (48)$$

Ora, $(D_1^k)^2 (\hat{\mu}_1^{[k]} - \mu^{[k]})$ é independente de $(\hat{\mu}_2^{[k]} - \mu^{[k]})$, logo (47) fica

$$-\mathbb{E}\{(D_1^k)^2 (\hat{\mu}_1^{[k]} - \mu^{[k]})\} \mathbb{E}\{(\hat{\mu}_2^{[k]} - \mu^{[k]})\} = 0, \quad (49)$$

pois $\mathbb{E}\{(\hat{\mu}_2^{[k]} - \mu^{[k]})\} = 0$.

Além disto,

$$\hat{\mu}_1^{[k]} - \hat{\mu}_2^{[k]} = (\hat{\mu}_1^{[k]} - \hat{\mu}_2^{[k]}) \frac{D_1^k}{R_{m_k}^{[k]}}. \quad (50)$$

Então o segundo termo (48) fica

$$-\mathbb{E}\left\{\frac{(D_1^k)^3}{R_{m_k}^{[k]}} (\hat{\mu}_1^{[k]} - \mu^{[k]}) (\hat{\mu}_1^{[k]} - \hat{\mu}_2^{[k]})\right\}. \quad (51)$$

Consideremos o valor absoluto de (51):

$$|\mathbb{E}\left\{\frac{(D_1^k)^3}{R_{m_k}^{[k]}} (\hat{\mu}_1^{[k]} - \mu^{[k]}) (\hat{\mu}_1^{[k]} - \hat{\mu}_2^{[k]})\right\}| \leq \quad (52)$$

$$\mathbb{E}\left\{\frac{(D_1^k)^3}{R_{m_k}^{[k]}} |(\hat{\mu}_1^{[k]} - \mu^{[k]}) (\hat{\mu}_1^{[k]} - \hat{\mu}_2^{[k]})|\right\}. \quad (53)$$

Como $\min_{a \in A} f(a) \leq \hat{\mu}_1^{[k]}, \hat{\mu}_2^{[k]} \leq \max_{a \in A} f(a)$, então $|(\hat{\mu}_1^{[k]} - \mu^{[k]})(\hat{\mu}_1^{[k]} - \hat{\mu}_2^{[k]})| \leq C_2$. Logo a expressão acima é menor ou igual a

$$\mathbb{E}\left\{\frac{(D_1^k)^3}{R_{m_k}^{[k]}}\right\} \leq \mathbb{E}\left\{\frac{(D_1^k)^3}{m_k}\right\} \rightarrow 0, \quad (54)$$

peelo lema (3.3), terminando a demonstração da proposição.

Lema 3.4. *Para cada k , seja $(X_n^{[k]})_{n \in \mathbb{Z}}$ uma cadeia de Markov de ordem k . Se a sequência $(X_n^{[k]})_{n \in \mathbb{Z}}$ satisfaz a condição (2), temos então que*

$$\lim_{k \rightarrow \infty} \frac{\mathbb{E}\left(\left(S_1^{[k]} - (R_2^{[k]} - R_1^{[k]})\hat{\mu}^{[k]}\right)^4\right)}{m_k \left(\mathbb{E}\left(\left(S_1^{[k]} - (R_2^{[k]} - R_1^{[k]})\hat{\mu}^{[k]}\right)^2\right)\right)^2} = 0 \quad (55)$$

Demonstração

A idéia da demonstração é encontrar uma majoração para o numerador e uma minoração para o denominador, de tal forma que a nova fração resultante seja maior que a apresentada em (55). A seguir mostramos que esta nova fração converge para zero quando $k \rightarrow \infty$, demonstrando, conseqüentemente, o lema (3.4). Primeiro encontramos uma majoração para o numerador.

$$\mathbb{E}\left(S_1^{[k]} - (D_1^k)\hat{\mu}^{[k]}\right)^4 = \sum_{n=1}^{\infty} \mathbb{E}\left(\mathbf{1}_{\{D_1^k=n\}} \left(\sum_{l=1}^n f(X_l^{[k]}) - n\hat{\mu}^{[k]}\right)^4\right)$$

como o alfabeto A é finito, $\min_{a \in A} f(a) \leq f(X_l^{[k]}) \leq \max_{a \in A} f(a) < \infty$. Também pela definição de $\hat{\mu}^{[k]}$ segue que $\min_{a \in A} f(a) \leq \hat{\mu}^{[k]} \leq \max_{a \in A} f(a)$. Logo

$$(f(X_l^{[k]}) - \hat{\mu}^{[k]}) \leq \max_{a \in A} f(a) - \min_{a \in A} f(a) = C_1$$

é limitado superiormente. Portanto

e

$$\mathbb{E}\left(S_1^{[k]} - (D_1^k)\hat{\mu}^{[k]}\right)^4 \leq C_1^4 \sum_{n=1}^{\infty} n^4 \mathbb{P}(D_1^k = n) = C_3 \mathbb{E}((D_1^k)^4),$$

onde C_1 é uma constante que não depende de k , pois o alfabeto é o mesmo para todo k . Desta maneira encontramos então uma majoração para o numerador. A seguir tratamos de encontrar uma minoração não trivial para o denominador.

Somando e subtraindo $\mu^{[k]}$ no denominador, temos que

$$\mathbb{E}\{(S_1[k] - D_1^k \hat{\mu}^{[k]})^2\} = \mathbb{E}\left\{\left[\sum_{n=1}^{D_1^k} (f(X_n^{[k]}) - \mu^{[k]}) + D_1^k(\mu^{[k]} - \hat{\mu}^{[k]})\right]^2\right\} \quad (56)$$

$$\mathbb{E}\left\{\left[\sum_{n=1}^{D_1^k} (f(X_n^{[k]}) - \mu^{[k]})\right]^2 + [D_1^k(\mu^{[k]} - \hat{\mu}^{[k]})]^2 + 2D_1^k \sum_{n=1}^{D_1^k} (f(X_n^{[k]}) - \mu^{[k]}) (\mu^{[k]} - \hat{\mu}^{[k]})\right\}. \quad (57)$$

Como $[D_1^k(\mu^{[k]} - \hat{\mu}^{[k]})]^2 \geq 0$,

$$\begin{aligned} \mathbb{E}\{(S_1[k] - D_1^k \hat{\mu}^{[k]})^2\} &\geq \mathbb{E}\left\{\left[\sum_{n=1}^{D_1^k} (f(X_n^{[k]}) - \mu^{[k]})\right]^2\right\} \\ &\quad + 2\mathbb{E}\left\{D_1^k \sum_{n=1}^{D_1^k} (f(X_n^{[k]}) - \mu^{[k]}) (\mu^{[k]} - \hat{\mu}^{[k]})\right\}. \end{aligned} \quad (58)$$

Vamos analisar separadamente cada um dos termos da expressão acima.

Neste ponto a demonstração do lema (3.4) se reduz a mostrar que

$$\frac{C_3 \mathbb{E}((D_1^k)^4)}{m_k (C_4 \mathbb{E}((D_1^k) + R_k)^2)} \rightarrow 0, \quad (59)$$

quando $k \rightarrow \infty$, onde o resto R_k tende a zero com k . Mas isto é consequência dos lemas (3.2), (3.3), do teorema (3.3) e das proposições (3.1) e (3.2).

Lema 3.5. *Seja $(X_n^{[k]})_{n \in \mathbb{Z}}$ uma sequência de cadeias de Markov de ordens crescentes k . Se $(X_n^{[k]})_{n \in \mathbb{Z}}$ satisfaz a condição (2), temos então, para quase todas as realizações que*

$$\lim_{k \rightarrow \infty} \frac{\sum_{i=1}^{m_k} \left(S_i^{[k]} - (R_{i+1}^{[k]} - R_i^{[k]}) \hat{\mu}^{[k]} \right)^4}{\left(\sum_{i=1}^{m_k} \left((S_i^{[k]} - (R_{i+1}^{[k]} - R_i^{[k]}) \hat{\mu}^{[k]})^2 \right) \right)^2} = 0. \quad (60)$$

Demonstração

Vamos primeiro mostrar a convergência em probabilidade do limite acima e depois, usando Borel-Cantelli mostramos a convergência quase certa. Fazemos como na demonstração do lema anterior. Primeiro majoramos o numerador e a seguir minoramos o denominador e mostramos que a nova fração, maior que a original, tende a zero quase certamente e o lema segue.

Seja

$$z_i^{[k]} = (S_i^{[k]} - (R_{i+1}^{[k]} - R_i^{[k]})\hat{\mu}^{[k]}).$$

Começamos multiplicando e dividindo (60) por m_k . Podemos então escrever o numerador como

$$f_N^k = \frac{1}{m_k} \sum_{i=1}^{m_k} (z_i^{[k]})^4.$$

Usando a desigualdade de Markov, temos que

$$\mathbb{P}(f_N^k > m_k) \leq \frac{\mathbb{E}\{f_N^k\}}{m_k} = \frac{\mathbb{E}\{(z_i^{[k]})^4\}}{m_k} \leq \frac{C_3 \mathbb{E}\{(D_1^k)^4\}}{m_k} \quad (61)$$

Usando o lema (3.4) temos que (61) converge para zero de uma forma somável. Portanto, por Borel-Cantelli, temos que o numerador é quase certamente limitado superiormente.

O próximo passo é encontrar uma cota inferior para o denominador

$$f_D^k = \frac{1}{m_k} \sum_{j=1}^{m_k} (z_j^{[k]})^2.$$

Para fazer isto usamos a desigualdade de Paley-Zygmund para $0 < \lambda < 1$

$$\mathbb{P}(f_D^k > \lambda \mathbb{E}(f_D^k)) \geq (1 - \lambda)^2 \frac{\mathbb{E}(f_D^k)^2}{\mathbb{E}((f_D^k)^2)}.$$

Uma vez que $z_j^{[k]}$ e $z_1^{[k]}$ têm a mesma distribuição, temos que

$$\mathbb{E}\{(f_D^k)^2\} = \frac{1}{m_k} \mathbb{E}(z_1^{[k]})^4 + \frac{(m_k - 1)}{m_k} \mathbb{E}((z_1^{[k]})^2 (z_2^{[k]})^2).$$

Além disto

$$\mathbb{E}\{f_D^k\} = \mathbb{E}(z_1^{[k]})^2.$$

Isto implica que

$$\begin{aligned} \mathbb{P}(f_D^k > \lambda \mathbb{E}(f_D^k)) &\geq (1 - \lambda)^2 \frac{1}{\frac{\mathbb{E}\left((z_1^{[k]})^2(z_2^{[k]})^2\right)}{\mathbb{E}\left((z_1^{[k]})^2\right)^2} - \frac{\mathbb{E}\left((z_1^{[k]})^2(z_2^{[k]})^2\right)}{m_k \mathbb{E}\left((z_1^{[k]})^2\right)^2} + \frac{\mathbb{E}\left(z_1^{[k]}\right)^4}{m_k \mathbb{E}\left((z_1^{[k]})^2\right)^2}}. \\ &\geq (1 - \lambda)^2 \frac{1}{\frac{\mathbb{E}\left((z_1^{[k]})^4\right)}{\mathbb{E}\left((z_1^{[k]})^2\right)^2} - \frac{\mathbb{E}\left((z_1^{[k]})^2(z_2^{[k]})^2\right)}{m_k \mathbb{E}\left((z_1^{[k]})^2\right)^2} + \frac{\mathbb{E}\left(z_1^{[k]}\right)^4}{m_k \mathbb{E}\left((z_1^{[k]})^2\right)^2}}. \end{aligned}$$

Agora aplicamos os lemas 3.2, 3.4 e novamente Borel Cantelli para concluir que o denominador é quase certamente limitado inferiormente por $\lambda \mathbb{E}\left(z_1^{[k]}\right)^2$. Temos então quase certamente que (60) é menor ou igual a

$$\frac{C_3 \mathbb{E}\{(D_1^k)^4\}}{m_k (C_4 \mathbb{E}\{(D_1^k)\} + R_k)^2}. \quad (62)$$

E reencontramos a expressão (59) do lema (3.4) que tende a zero de maneira somável. Portanto, (60) tende a zero quase certamente terminando assim a demonstração do lema (3.5).

Neste ponto estamos aptos para demonstrar o teorema 3.1.

3.4 Demonstração do Teorema Limite Central para $\mu^{[k]*}$

Podemos reescrever a variável aleatória em (9) como

$$\frac{1}{\eta_*^k} \left\{ \sum_{i=1}^{m_k} S_i^{[k]*} - R_{m_k}^{[k]*} \hat{\mu}^{[k]} \right\} \quad (63)$$

que, por sua vez, pode ser escrito como

$$\frac{1}{\eta_*^k} \left\{ \sum_{i=1}^{m_k} (S_i^{[k]*} - (R_{I_i+1}^{[k]} - R_{I_i}^{[k]}) \hat{\mu}^{[k]}) \right\}. \quad (64)$$

Uma vez que para cada k tomamos uma amostra de m_k blocos, podemos pensar nas variáveis $(S_i^{[k]})_{i,k \in \mathbb{N}}$, como um arranjo duplo. Temos que $(S_i^{[k]})_{i,k \in \mathbb{N}}$ são

independentes, pois o processo $(X_n^{[k]})_{n \in \mathbb{Z}}$ é Markov. Chamando $z_i^{[k]*} = (S_i^{[k]*} - (R_{i+1}^{[k]} - R_i^{[k]})\hat{\mu}^{[k]})$, temos também o arranjo duplo reamostrado $(z_i^{[k]*})_{i,k \in \mathbb{N}}$.

Como $\mathbb{E}^*(z_i^{[k]*}) = 0$ (lema 3.1), pelo Teorema Limite Centralde Lindeberg-Feller para arranjos duplos, o teorema 64 é verdadeiro se a seguinte condição é satisfeita

$$\frac{\mathbb{E}^* \left((z_1^{[k]*})^2 \mathbf{1}_{\{|z_1^{[k]*}| > \epsilon \sqrt{m_k \mathbb{E}^*(z_1^{[k]*})^2}\}} \right)}{\text{var}^*(z_1^{[k]*})^2} \rightarrow 0. \quad (65)$$

Temos que

$$\frac{\mathbb{E}^* \left((z_1^{[k]*})^2 \mathbf{1}_{\{|z_1^{[k]*}| > \epsilon \sqrt{m_k \mathbb{E}^*(z_1^{[k]*})^2}\}} \right)}{\text{var}(z_1^{[k]*})} \leq \quad (66)$$

$$\frac{\sqrt{\mathbb{E}^*(z_1^{[k]*})^4} \sqrt{\mathbb{P}(|z_1^{[k]*}| > \epsilon \sqrt{m_k \mathbb{E}^*(z_1^{[k]*})^2})}}{\text{var}(z_1^{[k]*})} \leq \quad (67)$$

$$\frac{\sqrt{\mathbb{E}^*(z_1^{[k]*})^4} \sqrt{\frac{\text{var}(z_1^{[k]*})}{m_k \epsilon^2 \text{var}(z_1^{[k]*})}}}{\text{var}(z_1^{[k]*})} = \quad (68)$$

$$\frac{\sqrt{\sum_{j=1}^{m_k} \left(S_j^{[k]} - (R_{i+1}^{[k]} - R_i^{[k]})\hat{\mu}^{[k]} \right)^4}}{\left(\epsilon^2 \sum_{j=1}^{m_k} \left((S_j^{[k]} - (R_{i+1}^{[k]} - R_i^{[k]})\hat{\mu}^{[k]})^2 \right) \right)},$$

A última expressão foi obtida usando o lema 3.1. Então, pelo lema (3.5) segue que, quando $k \rightarrow \infty$, a condição (65) vale e fica demonstrado o teorema (64).

Apêndice A₂: Convergência da Variância Sequencial

Vamos mostrar que

$$\eta^{[k]} = \left(\frac{\text{var} \left(\sum_j^{R_{m_k}^{[k]}} f(X_j^{[k]}) - \mu^{[k]} \right)}{R_{m_k}^{[k]}} \right)^{1/2} \quad (69)$$

converge para σ_∞ .

Começamos mostrando a convergência em probabilidade de $\frac{\mathbb{E}\{R_{m_k}^{[k]}\}}{R_{m_k}^{[k]}}$ para 1.

$$\begin{aligned} \mathbb{P}\left(\left|\frac{R_{m_k}^{[k]}}{\mathbb{E}\{R_{m_k}^{[k]}\}} - 1\right| > \epsilon\right) &\leq \frac{\mathbb{E}\{R_{m_k}^{[k]} - \mathbb{E}\{R_{m_k}^{[k]}\}\}^2}{\mathbb{E}\{R_{m_k}^{[k]}\}^2 \epsilon^2} \\ &= \frac{\mathbb{E}\{(R_{m_k}^{[k]})^2\} - \mathbb{E}\{R_{m_k}^{[k]}\}^2}{\mathbb{E}\{R_{m_k}^{[k]}\}^2 \epsilon^2} \\ &= \frac{m_k \mathbb{E}\{(D_1^k)^2\} + m_k(m_k - 1) \mathbb{E}\{D_1^k\}^2 - m_k^2 \mathbb{E}\{D_1^k\}^2}{m_k^2 \mathbb{E}\{D_1^k\}^2 \epsilon^2} \\ &= \frac{\mathbb{E}\{(D_1^k)^2\} - \mathbb{E}\{D_1^k\}^2}{m_k \mathbb{E}\{D_1^k\}^2 \epsilon^2}, \end{aligned} \quad (70)$$

que, pelo lema (3.4), tende a zero de uma maneira somável. Logo $\frac{R_{m_k}^{[k]}}{\mathbb{E}\{R_{m_k}^{[k]}\}}$ converge quase certamente para 1 e, conseqüentemente, $\frac{\mathbb{E}\{R_{m_k}^{[k]}\}}{R_{m_k}^{[k]}}$ também converge quase certamente para 1.

Como $\mathbb{E}\{\sum_j^{R_{m_k}^{[k]}} (f(X_j^{[k]}) - \mu^{[k]})\} = 0$,

$$\begin{aligned}
\text{var} \left(\sum_j^{R_{m_k}^{[k]}} f(X_j^{[k]}) - \mu^{[k]} \right) &= \mathbb{E}\left\{ \sum_j^{R_{m_k}^{[k]}} f(X_j^{[k]}) - \mu^{[k]} \right\}^2 \\
&= \mathbb{E}\left\{ \sum_j^{R_{m_k}^{[k]}} (f(X_j^{[k]}) - \mu^{[k]})^2 \right\} \\
&+ \mathbb{E}\left\{ \sum_{i \neq j}^{R_{m_k}^{[k]}} (f(X_i^{[k]}) - \mu^{[k]}) (f(X_j^{[k]}) - \mu^{[k]}) \right\} \\
&= \mathbb{E}\{R_{m_k}^{[k]}\} \mathbb{E}\{(f(X_j^{[k]}) - \mu^{[k]})^2\} \\
&+ \mathbb{E}\left\{ \sum_{i \neq j}^{R_{m_k}^{[k]}} (f(X_i^{[k]}) - \mu^{[k]}) (f(X_j^{[k]}) - \mu^{[k]}) \right\}.
\end{aligned}$$

Então

$$\begin{aligned}
\frac{\text{var} \left(\sum_j^{R_{m_k}^{[k]}} f(X_j^{[k]}) - \mu^{[k]} \right)}{R_{m_k}^{[k]}} &= \frac{\mathbb{E}\{R_{m_k}^{[k]}\} \mathbb{E}\{(f(X_j^{[k]}) - \mu^{[k]})^2\}}{R_{m_k}^{[k]}} \\
&+ \frac{\mathbb{E}\left\{ \sum_{i \neq j}^{R_{m_k}^{[k]}} (f(X_i^{[k]}) - \mu^{[k]}) (f(X_j^{[k]}) - \mu^{[k]}) \right\}}{R_{m_k}^{[k]}}.
\end{aligned}$$

Chamando

$$\sigma_k^2 = \mathbb{E}\{(f(X_j^{[k]}) - \mu^{[k]})^2\},$$

como $\frac{\mathbb{E}\{R_{m_k}^{[k]}\}}{R_{m_k}^{[k]}}$ tende a 1, temos o resultado esperado uma vez que $\sigma_k^2 \rightarrow \sigma_\infty^2$ e

$$\frac{\mathbb{E}\left\{ \sum_{i \neq j}^{R_{m_k}^{[k]}} (f(X_i^{[k]}) - \mu^{[k]}) (f(X_j^{[k]}) - \mu^{[k]}) \right\}}{R_{m_k}^{[k]}} \rightarrow 0$$

pois as correlações decrescem de maneira somável.

Capítulo 4

Reamostragem para cadeias de ordem infinita

Neste capítulo tratamos da questão de estabelecer um Teorema Limite Central da Reamostragem para a média amostral de funções cilíndricas de uma cadeia de ordem infinita do tipo C.

Seja $f : A \rightarrow \mathbb{R}$ uma função definida em uma cadeia de ordem infinita do tipo C. Sem perda de generalidade, podemos considerar apenas o caso em que $r = 1$. E seja

$$\mu = \mathbb{E}(f(X_0, \dots, X_r)).$$

4.1 Procedimento de reamostragem

Seja X_1, X_2, \dots uma amostra de uma cadeia de ordem infinita do tipo C. O método de reamostragem que utilizamos é o mesmo utilizado no capítulo anterior para seqüências de cadeias de Markov.

Sejam

$$R_{i+1}(k) = \inf \left\{ t > R_i(k) : (X_t, \dots, X_{t+k-1}) = (X_1, \dots, X_k) \right\}, \quad (1)$$

com $R_0(k) = 1$, os tempos de retorno da primeira seqüência de símbolos X_1, \dots, X_k na amostra. Deixamos o processo evoluir até que os k primeiros símbolos apareçam $m_k = [e^{\alpha k}]$ vezes. Chamamos de $\xi_i(k)$ o bloco dos valores da cadeia entre $R_{i-1}(k)$ e $R_i(k) - 1$, isto é

$$\xi_i(k) = (X_{R_{i-1}(k)}, \dots, X_{R_i(k)-1}). \quad (2)$$

Desta forma obtemos uma sequência de m_k blocos $\xi_1(k), \dots, \xi_{m_k}(k)$ formados a partir da amostra inicial. Construímos uma reamostra da cadeia de ordem infinita fazendo uma seleção independente e uniformemente distribuída destes blocos. Formalmente, para cada k , sejam $I_1(k), \dots, I_{m_k}(k)$ variáveis aleatórias independentes com distribuição uniforme no conjunto $\{1, \dots, m_k\}$. Definimos os blocos reamostrados como

$$\xi_l^*(k) = \xi_{I_l(k)}(k),$$

para $l = 1, \dots, m_k$. A k -ésima reamostra é construída concatenando os blocos reamostrados $\xi_l^*(k)$.

Definimos os tempos de retorno reamostrados como

$$R_0^*(k) = 1$$

e para $l = 1, \dots, m_k$

$$R_l^*(k) = R_{l-1}^*(k) + R_{I_l(k)+1}(k) - R_{I_l(k)}(k).$$

O n -ésimo bloco reamostrado, $1 \leq n \leq m_k(k)$, é dado por

$$(X_{R_{n-1}^*(k)}^*(k), \dots, X_{R_n^*(k)-1}^*(k)) = \xi_n^*(k). \quad (3)$$

O estimador natural para μ com uma amostra de tamanho $R_{m_k}(k)$ é dado por:

$$\hat{\mu}_k = \frac{1}{R_{m_k}(k)} \sum_{n=1}^{R_{m_k}(k)} f(X_n) \quad (4)$$

e sua versão reamostrada é dada por:

$$\mu_k^* = \frac{1}{R_{m_k}^*(k)} \sum_{n=1}^{R_{m_k}^*(k)} f(X_n^*) \quad (5)$$

Para o estimador 4 segue que, pelo Teorema 2.1, satisfeitas as condições do teorema, vale o seguinte Teorema Limite Central :

$$\frac{\sqrt{R_{m_k}(k)}}{\sigma_\infty} (\hat{\mu}_k - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad (6)$$

quando k diverge.

A consistência de $\hat{\mu}_k$ segue diretamente do Teorema Ergódico, pois um processo φ – *misturador* é ergódico, portanto

$$\frac{1}{n} \sum_{t=1}^n f(X_t) \rightarrow \mathbb{E}(f(X_0)) \text{ q. c. quando } n \rightarrow \infty. \quad (7)$$

Isto significa que

$$\frac{1}{n} \sum_{t=1}^n X_t = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{X_t}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(\mu, 0) \text{ quando } n \rightarrow \infty. \quad (8)$$

Portanto, usando o teorema (2.2) e o lema (4.2) (ver apêndice), que garante a condição (5), nós temos que

$$\hat{\mu}_k - \mu \xrightarrow{P} 0$$

Observamos que também aqui nem μ , nem σ são conhecidos.

4.2 Teorema Limite Central da Reamostragem para

$$\mu_k^*$$

O principal resultado deste trabalho, que é um Teorema Central do Limite da Reamostragem para o estimador apresentado em 5, é apresentado a seguir.

Teorema 4.1. *Seja (X_n) uma cadeia de ordem infinita do tipo C cujos coeficientes de continuidade satisfazem $-\limsup_{l \rightarrow \infty} l^{-1} \log \beta_l = c > -2 \ln(\delta)$. Então, para $-\ln(\delta) < \alpha$ e $m_k = e^{\alpha k}$, para quase todas as realizações da cadeia (X_n) , temos*

$$\frac{\sqrt{R_{m_k}^*(k)}}{\eta_k^*} (\mu_k^* - \hat{\mu}_k) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad (9)$$

quando k diverge, onde

$$\eta_k^* = \left(\frac{\text{var} \left(\sum_j^{R_{m_k}(k)} f(X_j^*) - \hat{\mu}_k \mid X_1, \dots, X_{R_{m_k}} \right)}{R_{m_k}^*(k)} \right)^{1/2}. \quad (10)$$

Observamos que η_k^* , assim como $R_{m_k}^*(k)$, μ_k^* , e $\hat{\mu}_k$, é uma função da amostra $X_1, \dots, X_{R_{m_k}(k)}$ e são, portanto, todos diretamente calculáveis, dada a amostra. Além disto, a média considerada em (10) é tomada em relação às variáveis aleatórias $I_1(k), \dots, I_{m_k}(k)$, única fonte de aleatoriedade na reamostragem. Uma fórmula explícita para η_k^* é dada no lema (3.1).

4.3 Demonstração do Teorema Limite Central da Reamostragem para μ_k^*

No capítulo anterior mostramos o Teorema Limite Central da Reamostragem para sequências de cadeias de Markov de ordem crescente. Neste capítulo vamos mostrar que o teorema permanece válido no caso de uma cadeia de ordem infinita do tipo C. A demonstração do Teorema Limite Central para μ_k^* , teorema 4.1, principal resultado desta tese, utiliza o resultado do capítulo anterior e também resultados apresentados em Bressaud, Fernández e Galves (1999) que estão descritos no capítulo 2.

Seja $\Delta_{(1, R_{m_k})}$ o conjunto onde $(X_n)_{n \in \mathbb{Z}}$ e $(X_n^{[k]})_{n \in \mathbb{Z}}$ coincidem até o tempo R_{m_k} ,

$$\Delta_{(1, R_{m_k})} := \{X_t^{[k]} = X_t, t = 1 \dots, R_{m_k}\}. \quad (11)$$

O lema abaixo é uma adaptação de um resultado dado em Bressaud, Fernández e Galves (1999a). A diferença é que aqui o tempo em que $(X_n)_{n \in \mathbb{Z}}$ e $(X_n^{[k]})_{n \in \mathbb{Z}}$ coincidem, R_{m_k} , é uma função crescente de k e é uma variável aleatória.

Lema 4.1. *Seja $(X_n)_{n \in \mathbb{Z}}$ uma cadeia de ordem infinita do tipo C com $\beta_k = e^{-ck}$. Se $c > \ln(\delta)$ então temos que*

$$\mathbb{P}(\Delta_{(1, R_{m_k})}) \rightarrow 1 \quad (12)$$

quando $k \rightarrow \infty$.

Demonstração

Em Bressaud, Fernández e Galves (1999) é demonstrado para cadeias do tipo A que

$$\mathbb{P}\{\Delta_{1,k}\} \geq 1 - \frac{1 - (1 - \tilde{\beta}_k)^k}{\prod_{p=0}^{+\infty} (1 - \tilde{\beta}_p)} \geq 1 - C k \tilde{\beta}_k \rightarrow 1, \quad (13)$$

pois $\tilde{\beta}_k = e^{-ck}$, onde $(\tilde{\beta}_n)_{n \in \mathbb{N}}$ é definido como

$$\begin{cases} \tilde{\beta}_0 &= 1 - \inf_{a \in A, \underline{u} \in \underline{A}} p(a|\underline{u}), \\ \tilde{\beta}_n &= \min\left(\tilde{\beta}_0, \frac{\beta_n}{2}\right), \end{cases}$$

onde β_n , definido em (2.1), é a taxa de decaimento da cadeia de ordem infinita. O resultado segue, obviamente, para cadeias do tipo C.

Segue diretamente da demonstração de (13) que $\mathbb{P}(\Delta_{1,f(k)}) \rightarrow 1$, onde $f(k)$ é uma função crescente determinística de k com $f(k)/e^{-ck} \rightarrow 0$.

Para mostrar o resultado para R_{m_k} , observamos que, para $t_k > 1$, função não aleatória de k , vale que

$$\begin{aligned} \mathbb{P}(\Delta_{(1,R_{m_k})}^c) &= \mathbb{P}(\Delta_{(1,R_{m_k})}^c, R_{m_k} > t_k) \\ &+ \mathbb{P}(\Delta_{(1,R_{m_k})}^c, R_{m_k} < 1) \\ &+ \mathbb{P}(\Delta_{(1,R_{m_k})}^c, 1 \leq R_{m_k} \leq t_k) \\ &\leq \mathbb{P}(R_{m_k} > t_k) + \mathbb{P}(R_{m_k} < 1) + \mathbb{P}(\Delta_{(1,t_k)}^c). \end{aligned}$$

Queremos mostrar que $\mathbb{P}(\Delta_{(1,R_{m_k})}^c) \rightarrow 0$. Então vamos analisar cada um dos 3 termos da soma acima.

$$\begin{aligned} \mathbb{P}(R_{m_k} > t_k) &\leq \frac{\mathbb{E}\{R_{m_k}\}}{t_k} = \frac{m_k \mathbb{E}\{R_2 - R_1\}}{t_k} \\ &\leq \frac{Cm_k}{t_k \mathbb{P}(X_{i,k} = x_{i,k})} \leq \frac{Cm_k}{t_k \delta^k} \end{aligned} \quad (14)$$

pois, por (19), $\mathbb{P}(X_{i,k} = x_{i,k}) \geq \delta^k$.

Escolhendo $t_k = e^{\theta k}$, como $m_k = e^{\alpha k}$, temos que

$$\frac{Cm_k}{t_k \delta^k} = C e^{-k(\theta - \ln(1/\delta) - \alpha)}. \quad (15)$$

Portanto, se tomarmos $\theta > \alpha - \ln(1/\delta)$, segue que

$$\mathbb{P}(R_{m_k} > t_k) \rightarrow 0. \quad (16)$$

Observamos que, como neste caso, pela observação 1 do lema (3.3), $\alpha > \ln(1/\delta)$, então $\theta > 2 \ln(1/\delta)$.

Por sua vez, $\mathbb{P}(R_{m_k} < 1) = 0$, uma vez que R_{m_k} assume valores no conjunto $\{m_k, m_k + 1, \dots\}$, pois $(R_{i+1} - R_i) \geq 1$.
 Por último notamos que, se $c > \theta > 2 \ln(1/\delta)$, por (13)

$$\mathbb{P}(\Delta_{(1, t_k)}^c) \leq t_k \tilde{\beta}_k \rightarrow 0. \quad (17)$$

Usaremos o lema (4.1) para mostrar que a diferença entre os estimadores de μ considerando uma amostra da cadeia de Markov de ordem k ou considerando uma cadeia de ordem infinita é desprezível quando a ordem k cresce. O mesmo acontece com a reamostra.

Consideremos os seguintes conjuntos

$$A_\epsilon^k(w) = \{w \in \omega : \frac{\sqrt{R_{m_k}^{[k]*}}}{\eta_k^*} (\hat{\mu}^{[k]} - \hat{\mu}_k) < \epsilon\} \quad (18)$$

$$B_\epsilon^*(w) = \{w \in \omega : \frac{\sqrt{R_{m_k}^{[k]*}}}{\eta_k^*} (\mu_k^* - \mu^{[k]*}) < \epsilon\}. \quad (19)$$

Como $\hat{\mu}^{[k]} - \hat{\mu} = 0$ para todo $w \in \Delta_{1, R_{m_k}}$ temos que $\mathbb{P}(A_\epsilon^k(w)) > \mathbb{P}(\Delta_{1, R_{m_k}})$ e $(\mu_k^* - \mu^{[k]*}) = 0$ para todo $w \in \Delta_{(1, R_{m_k})}$ implica que $\mathbb{P}(B_\epsilon^*(w)) > \mathbb{P}(D_{(1, R_{m_k})})$, então

Corolário 4.1. *Se $c > \alpha > -\ln \delta$ então para $m_k = \lceil e^{\alpha k} \rceil$ temos que*

$$\frac{\sqrt{R_{m_k}^{[k]*}}}{\eta_k^*} (\hat{\mu}_k^{[k]} - \hat{\mu}_k) \rightarrow 0 \text{ em probabilidade quando } k \rightarrow \infty. \quad (20)$$

e

$$\frac{\sqrt{R_{m_k}^{[k]*}}}{\eta_k^*} (\mu_k^* - \mu^{[k]*}) \rightarrow 0 \text{ em probabilidade quando } k \rightarrow \infty. \quad (21)$$

Uma vez que podemos escrever

$$\mu_k^* - \hat{\mu}_k = (\mu^* - \mu^{[k]*}) + (\mu^{[k]*} - \mu^{\hat{[k]}}) + (\mu^{\hat{[k]}} - \hat{\mu}_k),$$

o corolário (4.1) e o teorema (3.1) juntos provam o teorema (4.1).

Apêndice A₃: Condição para utilizar o Teorema 2.2

Seja B_k um cilindro de tamanho k tal que $\mathbb{P}(B_k) > 0$ e seja $\tau_m^{B_k}$ a m -ésima vez que o cilindro B_k aparece. Seja $N_{B_k}(n)$ o número de vezes que o conjunto B_k aparece antes do tempo n , isto é

$$N_{B_k}(n) = \sum_{t=1}^{n-k} \mathbf{1}_{\{B_k \in X_{t,t+k-1}\}}.$$

O lema abaixo mostra que os teorema (2.1) e (2.2) dados em Billingsley (1967) podem ser utilizados.

Lema 4.2. *Suponhamos que exista uma constante $C > 0$ tal que, para qualquer k e qualquer n , temos que*

$$\text{Var}\left(N_{B_k}(n)\right) = \mathbb{E}\left(\left[N_{B_k}(n) - nP(B_k)\right]^2\right) \leq CnP(B_k). \quad (22)$$

Então, se $m_k \rightarrow \infty$, temos que

$$P(B_k)\tau_{m_k}^{B_k}/m_k$$

converge para 1 em probabilidade.

Demonstração

Para cada $1 > \delta > 0$ temos obviamente que

$$\mathbb{P}\left(\tau_m^{B_k} > \frac{m(1+\delta)}{\mathbb{P}(B_k)}\right) = \mathbb{P}\left(N_{B_k}\left(\left[\frac{m(1+\delta)}{\mathbb{P}(B_k)}\right]\right) < m\right)$$

e

$$\mathbb{P}\left(\tau_m^{B_k} < \frac{m(1-\delta)}{\mathbb{P}(B_k)}\right) = \mathbb{P}\left(N_{B_k}\left(\left[\frac{m(1-\delta)}{\mathbb{P}(B_k)}\right]\right) > m\right).$$

Portanto, usando a desigualdade de Bienaymé-Chebyshev temos

$$\mathbb{P}\left(\tau_m^{B_k} > \frac{m(1+\delta)}{\mathbb{P}(B_k)}\right) \leq \mathbb{P}\left(\left|N_{B_k}\left(\left[\frac{m(1+\delta)}{\mathbb{P}(B_k)}\right]\right) - m(1+\delta)\right| > m\delta\right) \leq \frac{Cm}{m^2\delta^2}.$$

Isto implica que

$$\lim_{k \rightarrow \infty} \mathbb{P}\left(\tau_{m_k}^{B_k} > \frac{m_k(1+\delta)}{\mathbb{P}(B_k)}\right) = 0$$

para qualquer $\delta > 0$. Similarmente, para qualquer $0 < \delta < 1$ temos que

$$\lim_{k \rightarrow \infty} \mathbb{P}\left(\tau_{m_k}^{B_k} < \frac{m_k(1-\delta)}{\mathbb{P}(B_k)}\right) = 0.$$

Capítulo 5

Uma aplicação à Linguística

Vamos modelar sequências de valores de uma função do sinal acústico de fala como uma cadeia de ordem infinita. Esta função, que chamamos “sonoridade”, mede a regularidade local do sinal acústico amostrado num conjunto discreto de instantes. O interesse dessa função vem do fato que, conforme mostrado em Galves, Garcia, Duarte e Galves (2002), estatísticas simples dessas sequências de valores permitem identificar as classes rítmicas das línguas consideradas no trabalho pioneiro de Ramus, Nespore e Mehler (1999).

5.1 Correlatos do ritmo presentes no sinal acústico

Uma conjectura existente na literatura linguística diz que as línguas naturais são divididas em classes rítmicas (cf. Abercrombie, 1967). Mas até pouco tempo atrás nenhuma evidência fonética confiável havia sido apresentada para confirmar essa teoria. Em Ramus, Nespore e Mehler (1999) conseguiram pela primeira vez apresentar propriedades estatísticas simples dando evidências de que as classes rítmicas tinham correlatos acústicos retiradas do sinal acústico.

A idéia básica contida no trabalho de Ramus et al (1999) é que a informação relevante sobre o ritmo pode ser extraída de características bem simples do sinal acústico, como tempos gastos em vogais e consoantes. No artigo são analisadas amostras de oito línguas: Polonês, Holandês e Inglês, supostamente pertencentes ao grupo das línguas acentuais; Catalão, Francês, Italiano e Espanhol, supostas representantes das línguas consideradas silábicas; e japonês, representando a classe moraica. São 20 frases para cada língua, sendo 5 frases para cada uma de 4 falantes, sempre mulheres. Para cada uma das línguas analisadas eles

dispunham de dados acústicos que foram segmentados em intervalos vocálicos e intervocálicos. Para cada uma das línguas eles estimaram o desvio padrão do comprimento dos intervalos consonantais (ΔC) e a proporção de tempo passado nas vogais (%V). O resultado encontrado por eles para essas línguas é surpreendente, vindo exatamente de encontro com a conjectura sobre a existência das classes rítmicas. A figura 2 traz o gráfico de %V contra ΔC . Observa-se neste gráfico dois fatos interessantes: a forte correlação linear entre as variáveis (-.93) e a aparente separação das línguas em grupos correspondendo exatamente à hipótese das classes rítmicas.

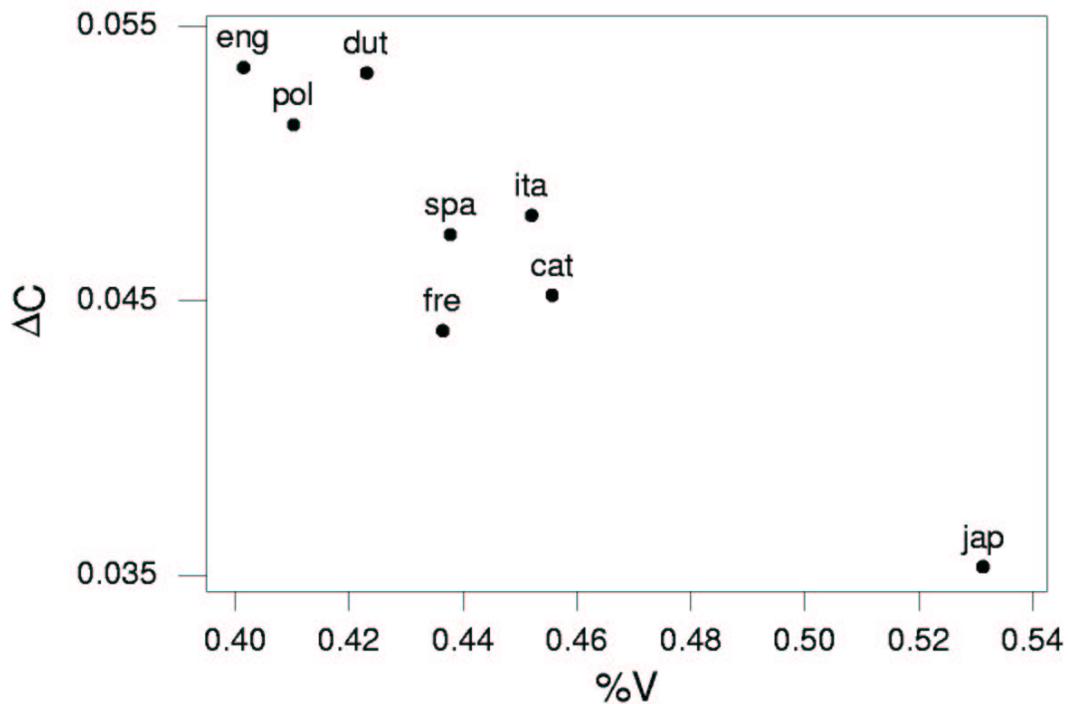


Figura 5.1: Distribuição das línguas no plano (%V, ΔC) (Ramus et al. 1999)

As evidências empíricas apresentadas em Ramus, Nespore e Mehler (1999) são reforçadas pelo trabalho de modelagem feito em Duarte, Galves, Garcia e Maronna (2001). Lá a hipótese de separação das línguas em classes rítmicas é testada através de um modelo paramétrico. A conclusão é que a hipótese de separação

das oito línguas consideradas em três classes rítmicas é compatível com os dados amostrais.

5.2 Uma medida de sonoridade como base para a identificação das classes rítmicas

A abordagem apresentada em Ramus, Nespor e Mehler tem problemas que comprometem a continuidade do estudo para outras línguas. Primeiro, a marcação dos intervalos vocálicos e intervocálicos é feita à mão, o que toma muito tempo, inviabilizando a análise de uma base maior de dados. Depois, a marcação manual depende de decisões difíceis de serem feitas de forma homogênea em larga escala.

Uma nova abordagem para o problema é apresentada em Galves, Garcia, Duarte e Galves (2002). Em vez de estudar durações de intervalos vocálicos e consonantais a proposta é estudar os valores de uma função que mede, em cada instante, a “sonoridade” local do sinal acústico. Esta função assume valores entre 0 e 1. Quanto mais sonora a região, como em vogais e consoantes nasais, mais perto de 1 a sonoridade fica. Quanto menos sonora, ou “obstruída”, como em consoantes fricativas, mais próximo de 0 é o valor da função.

Tecnicamente, definimos $s(t)$, a função sonoridade, onde t é o tempo assumindo valores em $\{ku : k = 1, \dots, T\}$, sendo u o tempo entre duas colunas e T é o número de passos considerados no espectrograma do sinal acústico. Aqui tomamos $u = 2$, onde as unidades são contadas em mili-segundos. Os valores do espectrograma são estimados com uma janela Gaussiana de 25ms. Consideramos somente frequências entre 0 e 800 Hz. Os cálculos foram feitos através do programa Praat (<http://www.praat.org>).

Seja $c_t(i)$ o coeficiente de Fourier da frequência i em torno do instante t do espectrograma. Definimos o espectrograma renormalizado por

$$p_t(i) = \frac{c_t(i)^2}{\sum_f c_t(f)^2}. \quad (1)$$

Isso define uma sequência de medidas de probabilidades $\{p_t : t = 1, \dots, T\}$. Padrões regulares, em geral, correspondem a sequências de medidas de probabilidade próximas entre si. Esta “proximidade” pode ser medida pela entropia relativa, que é definida como:

$$h(p_t | p_{t+1}) = \sum_i p_t(i) \log \left(\frac{p_t(i)}{p_{t+1}(i)} \right). \quad (2)$$

Definimos então $s(t)$ como

$$s(t) = 1 - \min \left(1, \frac{1}{27} \sum_{u=t-4}^{t+4} \sum_{i=1}^3 h(p_u | p_{u-i}) \right). \quad (3)$$

O fato empírico que dá suporte à sonoridade como uma medida do ritmo linguístico é apresentado em Mehler et al (1996). Neste artigo é apresentado um experimento onde bebês recém-nascidos são capazes de distinguir entre línguas consideradas de classes rítmicas distintas, mesmo com o sinal acústico filtrado a 400 Hz. Neste intervalo de frequências, a distinção entre intervalos vocálicos e consonantais não pode ser feita com precisão, semivogais e vogais ensurdecidas se confundem com consoantes, e consoantes sonoras, como nasais, se confundem com vogais. Neste intervalo de frequências permanece simplesmente a distinção entre zonas sonoras e zonas obstruídas do sinal. Isso sugere que os bebês distinguem as classes rítmicas utilizando classificações menos refinadas do que a oposição vogais e consoantes.

As estatísticas que consideraremos são:

1. A sonoridade empírica média, que irá fazer o papel de $\%V$

$$\bar{S} = \frac{1}{T} \sum_{t=1}^T s(t). \quad (4)$$

2. A variação total da sonoridade (que irá fazer o papel de ΔC)

$$\delta S = \frac{1}{T} \sum_{t=1}^T |s(t) - s(t-1)|. \quad (5)$$

A figura 2 apresenta características bem parecidas às da figura 1. Por um lado aparece uma forte correlação linear entre as variáveis. Por outro lado é sugerida a separação das línguas em 3 grupos. Uma explicação para a correlação linear é sugerida em Cassandro, Collet, Duarte, Galves e Garcia (2002), mas está fora do tema desta tese. A separação das línguas em classes que correspondem à hipótese das classes rítmicas é menos evidente que na figura 2 e é exatamente o que queremos testar com a construção de intervalos de confiança para a sonoridade.

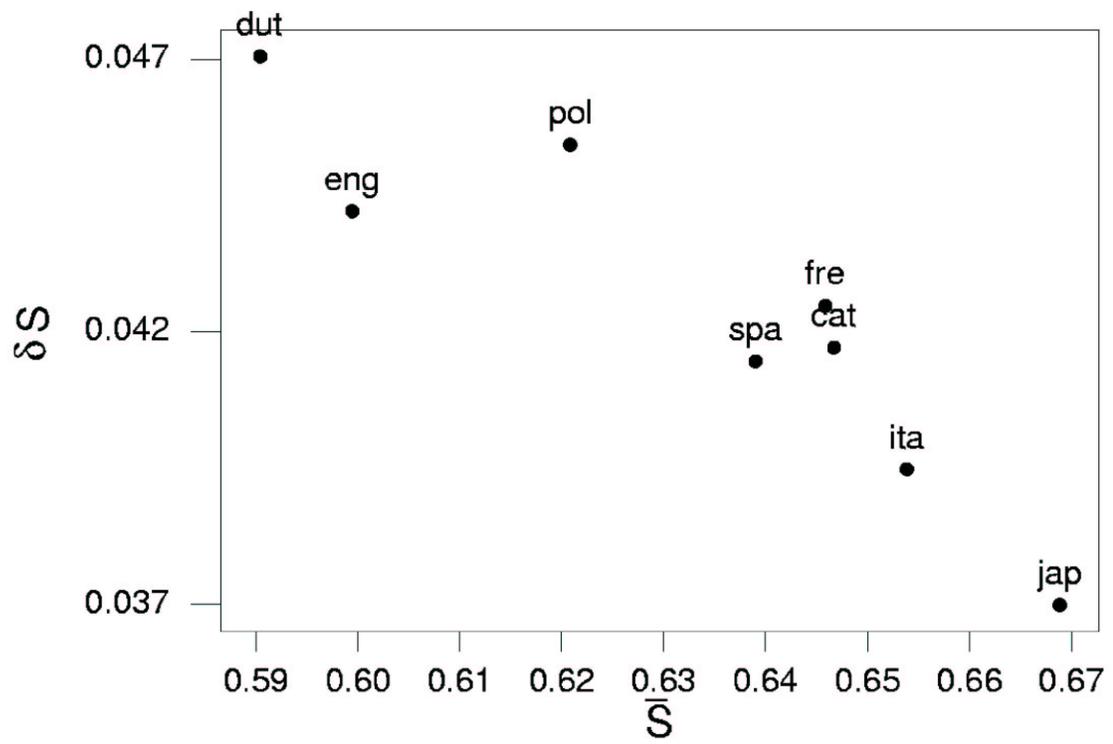


Figura 5.2: Distribuição das línguas no plano (\bar{S} , δO). (Ramus et al. 1999)

5.3 Sonoridade como uma cadeia de ordem infinita

A função sonoridade assume valores no intervalo $[0, 1]$. No entanto, uma análise exploratória mostra que é possível discretizá-la, utilizando 3 regiões correspondentes à sonoridade alta, média e baixa. Um estudo mais aprofundado do assunto é dado em Galves, Garcia, Duarte e Galves (2002).

A sonoridade não se comporta como uma cadeia de Markov de alcance fixo. A probabilidade de o processo continuar em um determinado estado, depende do tempo já passado neste estado e depende da zona de sonoridade (alta/ baixa) em que o processo esteja, que tem tamanho variável. Além disto, uma análise exploratória dada em Duarte, Galves, Garcia e Maronna (2001) mostra que os comprimentos dos intervalos consonantais sucessivos são não correlacionados e não têm distribuição exponencial. A não exponencialidade é mais uma razão para que o processo não seja modelado como uma cadeia de Markov. Já o fato de serem não correlacionados sugere que a cadeia de ordem infinita descrita pela sonoridade tem decaimento rápido das correlações. Lembramos que intervalos consonantais correspondem bem às zona de baixa sonoridade e intervalos vocálicos às de alta sonoridade. Diante do que foi exposto acima acreditamos que é razoável supor a sonoridade como sendo uma cadeia de ordem infinita com perda de memória rápida.

5.4 Aplicação do procedimento de reamostragem para estimar a variância da sonoridade

Nesta seção vamos estimar \bar{S} e utilizar o procedimento da reamostragem sequencial para encontrar uma estimativa da variância de \bar{S} . Para isto precisamos discretizar os valores da sonoridade. Análises feitas previamente em Galves, Garcia, Duarte e Galves (2002) mostram que é suficiente utilizar 3 regiões de sonoridade, a saber: $[0; 0, 16[$, $[0, 16; 0, 64[$ e $[0, 64; 1]$. Em cada região, a primeira idéia é utilizar o ponto médio do intervalo como representativo da região. A saber, o ponto 0, 08 representando a sonoridade baixa, o valor 0, 40 representando a sonoridade média e o valor 0, 72 representando a sonoridade alta. Na prática é mais cômodo representar as regiões com os números 0,1 e 2, que mantém as ordens de grandeza relativas dos valores médios.

Como a distribuição limite da média empírica em torno da média verdadeira é a mesma que da média reamostrada em torno da média empírica, como prova o Teorema Limite Central da Reamostragem, podemos usar os intervalos de confiança

da reamostra para tomada de decisões sobre a média verdadeira, como nos ensinou Efron.

Com as amostras divididas em blocos aplicamos então o procedimento de reamostragem sequencial considerando 1000 reamostras. Desta forma obtivemos uma distribuição empírica das médias da sonoridade reamostradas e uma estimativa do desvio padrão para cada uma. A partir destas estimativas obtivemos um intervalo de confiança para a média verdadeira. O intervalo de confiança individual foi construído de tal forma que o nível de confiança conjunto seja de, no mínimo, 5%.

As tabelas abaixo trazem os valores, para $k = 5, 6, 7, 8$ e 9 , dos seguintes estimadores: média empírica (\bar{x}), reamostrada (x^*), desvio padrão (S) e intervalo de confiança ($IC_{99\%}$) por língua. Trazem também o Número de blocos (m_k) e o tamanho da reamostra ($R_{m_k}^*(k)$).

Tabela 1: Estimativas com $k = 5$.

Lingua	m_k	$R_{m_k}^*(k)$	\bar{x}	x^*	S	$IC_{99\%}$
Japonesonês	1812	27818	0.7135	0.7129	0.02219	(0.6529, 0.7728)
Italiano	705	27756	0.7241	0.7252	0.03101	(0.6415, 0.8084)
Francês	834	27816	0.7424	0.7424	0.02951	(0.6628, 0.8219)
Polonês	916	29742	0.7789	0.7811	0.02598	(0.7109, 0.8512)
Catalão	1357	27918	0.7882	0.7875	0.02124	(0.7302, 0.8449)
Espanhol	797	28643	0.7937	0.7936	0.02941	(0.7141, 0.8731)
Inglês	940	26970	0.8526	0.8512	0.02901	(0.7727, 0.9297)
Holandês	858	28140	0.8621	0.8631	0.02993	(0.7822, 0.9439)

Table 2: Estimativas com $k = 6$.

Lingua	m_k	$R_{m_k}^*(k)$	\bar{x}	x^*	S	$IC_{99\%}$
Japones	1626	27811	0.7135	0.7122	0.02489	(0.6444, 0.7794)
Italiano	527	27755	0.7241	0.7270	0.03030	(0.6453, 0.8087)
Frances	612	27811	0.7424	0.7439	0.02779	(0.6690, 0.8189)
Polones	706	29743	0.7789	0.7793	0.02491	(0.7121, 0.8465)
Catalao	1296	27916	0.7882	0.7863	0.02430	(0.7207, 0.8518)
Espanhol	620	28637	0.7937	0.7964	0.02829	(0.7201, 0.8727)
Ingles	703	26971	0.8526	0.8506	0.02750	(0.7764, 0.9248)
Holandes	645	28140	0.8621	0.8616	0.02891	(0.7836, 0.9395)

Table 3: Estimativas com $k = 7$.

Lang	m_k	$R_{m_k}^*(k)$	\bar{x}	x^*	S	$IC_{99\%}$
Japones	1577	27812	0.7135	0.7126	0.02435	(0.6468, 0.7783)
Italiano	496	27756	0.7241	0.7271	0.02829	(0.6413, 0.8129)
Frances	511	27812	0.7424	0.7434	0.02689	(0.6708, 0.8159)
Polones	623	29744	0.7789	0.7794	0.02652	(0.7078, 0.8510)
Catalao	1234	27910	0.7882	0.7869	0.02359	(0.7234, 0.8505)
Espanhol	576	28640	0.7937	0.7956	0.02892	(0.7175, 0.8737)
Ingles	640	26947	0.8526	0.8568	0.02829	(0.7742, 0.9272)
Holandes	559	28141	0.8621	0.8609	0.030301	(0.7790, 0.9428)

Table 4: Estimativas com $k = 8$.

Lang	m_k	$R_{m_k}^*(k)$	\bar{x}	x^*	S	$IC_{99\%}$
Japones	1318	27807	0.7135	0.7110	0.0242	(0.6454, 0.7763)
Italiano	405	27753	0.7241	0.7262	0.0305	(0.6437, 0.8086)
Frances	419	27809	0.7424	0.7425	0.0254	(0.6737, 0.8112)
Polones	457	29737	0.7789	0.7780	0.0266	(0.7064, 0.8498)
Catalao	1024	27913	0.7882	0.7866	0.0236	(0.7229, 0.8503)
Espanhol	441	28663	0.7937	0.7935	0.0278	(0.7184, 0.8687)
Ingles	489	26969	0.8526	0.8494	0.0296	(0.7695, 0.9292)
Holandes	434	28137	0.8621	0.8631	0.0288	(0.7853, 0.9409)

Table 5: Estimativas com $k = 9$.

Lang	m_k	$R_{m_k}^*(k)$	\bar{x}	x^*	S	$IC_{99\%}$
Japones	995	27811	0.7135	0.7133	0.02352	(0.6498, 0.7769)
Italiano	316	27748	0.7241	0.7256	0.03049	(0.6434, 0.8079)
Frances	310	27811	0.7424	0.7432	0.02779	(0.6676, 0.8161)
Polones	345	29737	0.7789	0.7786	0.02720	(0.7051, 0.8521)
Catalao	851	27910	0.7882	0.7855	0.02420	(0.7229, 0.8503)
Espanhol	383	28630	0.7937	0.7927	0.02929	(0.7137, 0.8718)
Ingles	386	26965	0.8526	0.8514	0.02579	(0.7775, 0.9218)
Holandes	335	28135	0.8621	0.8632	0.02739	(0.7893, 0.9370)

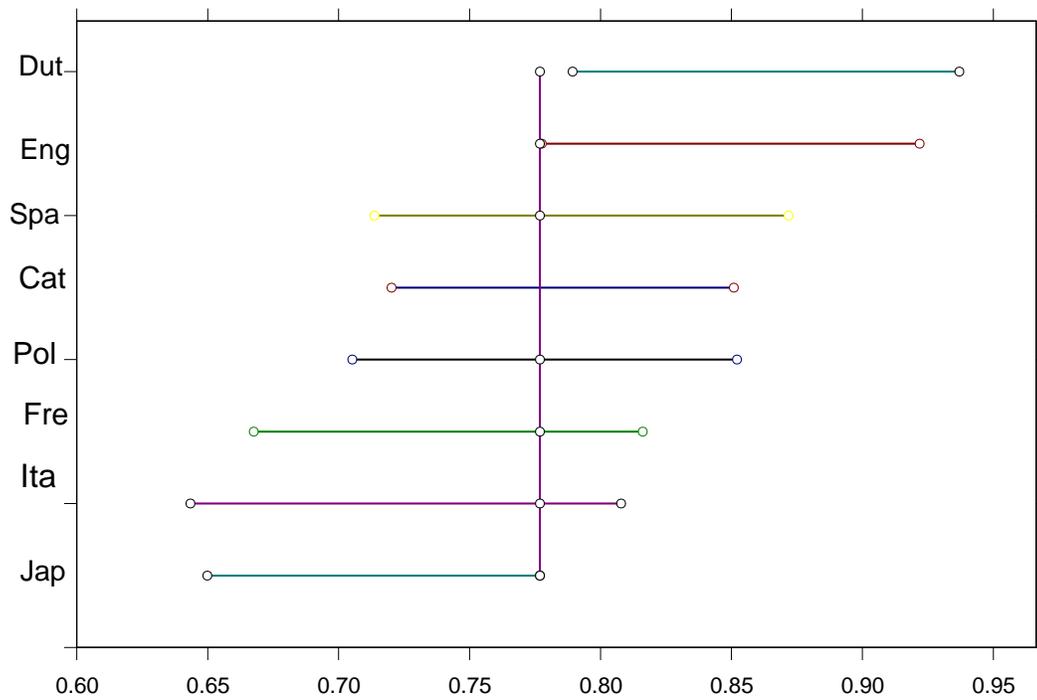


Figura 5.3: *Intervalos de confiança para as 8 línguas considerando o retorno dos 9 primeiros símbolos*

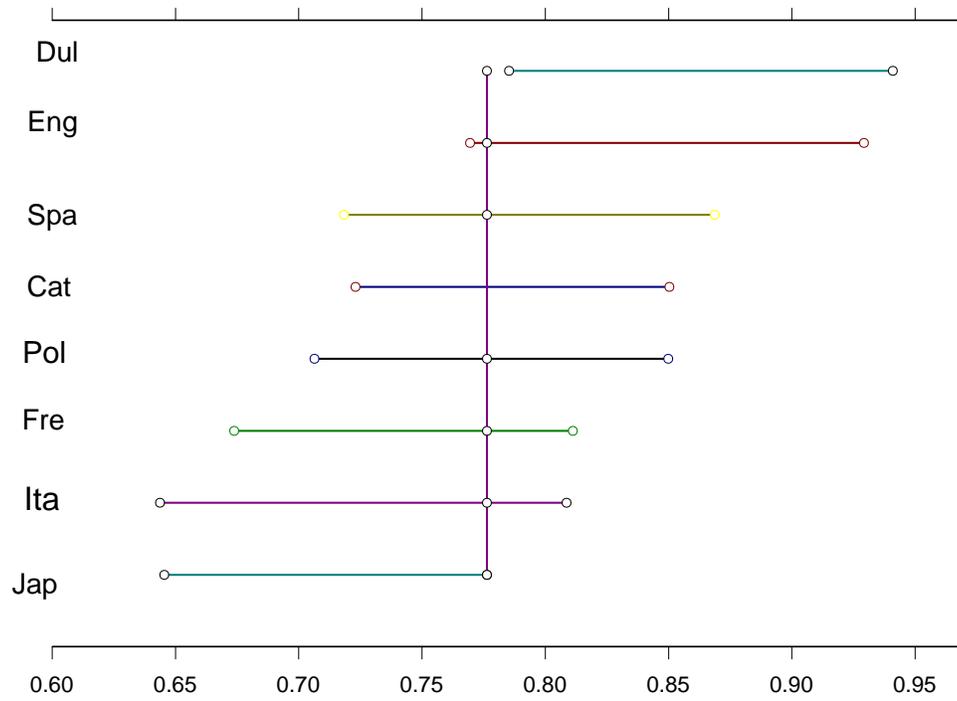


Figura 5.4: *Intervalos de confiança para as 8 línguas considerando o retorno dos 8 primeiros símbolos*

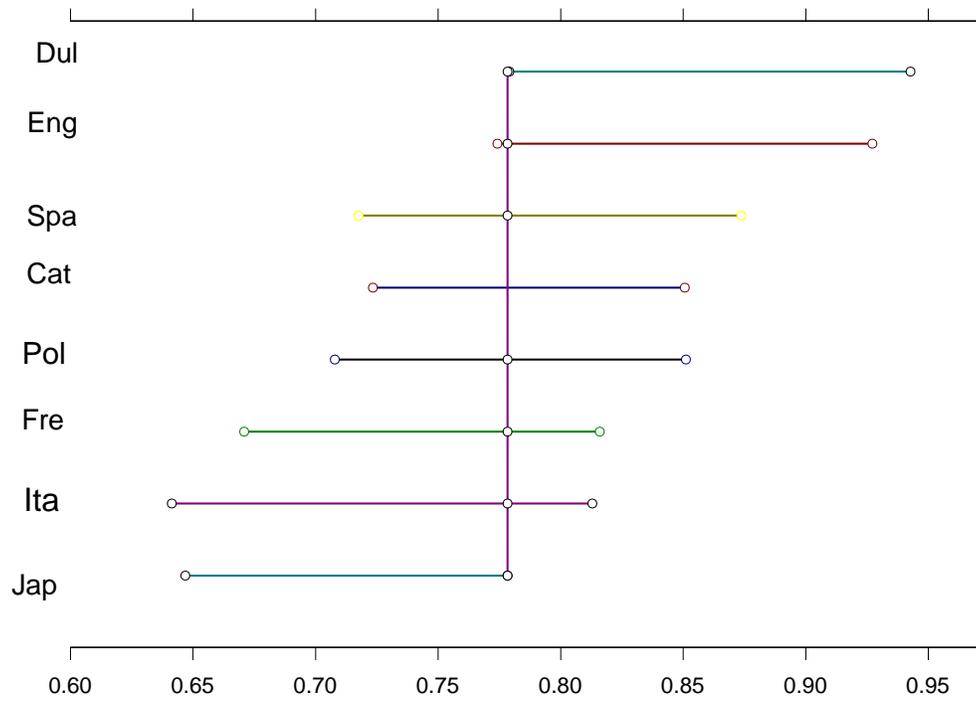


Figura 5.5: *Intervalos de confiança para as 7 línguas considerando o retorno dos 7 primeiros símbolos*

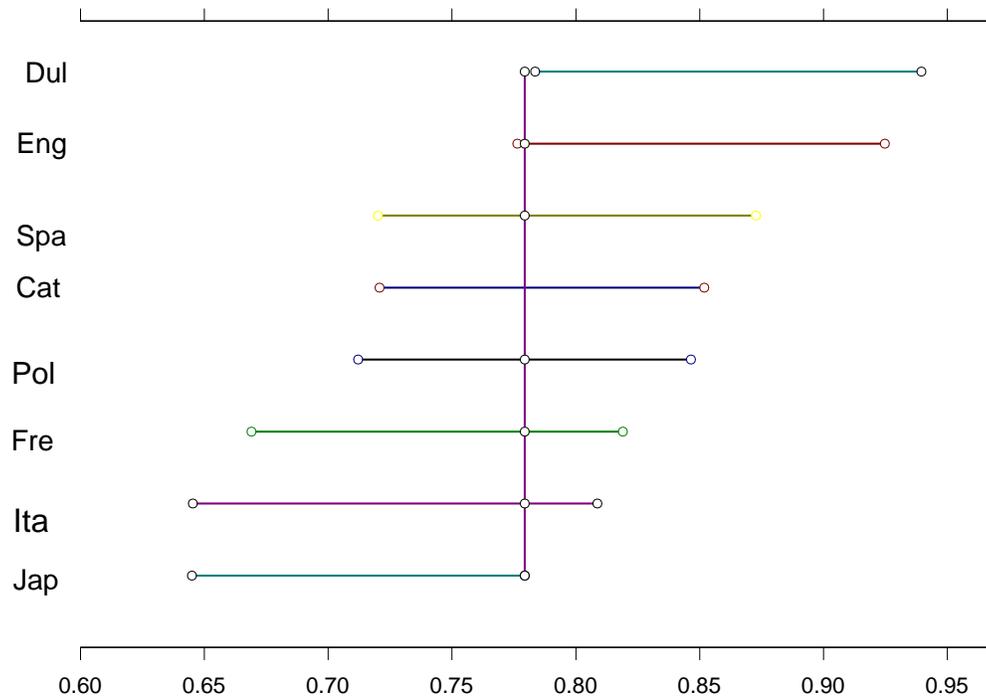


Figura 5.6: *Intervalos de confiança para as 8 línguas considerando o retorno dos 6 primeiros símbolos*

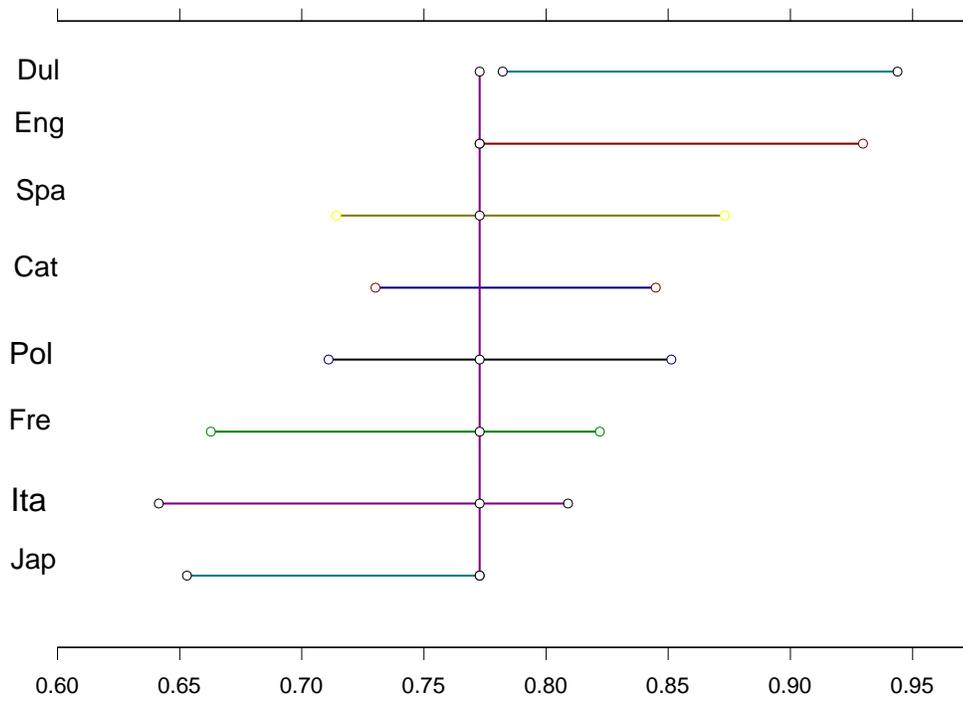


Figura 5.7: *Intervalos de confiança para as 8 línguas considerando o retorno dos 5 primeiros símbolo*

Discussão

De acordo com os valores apresentados nas tabelas acima as únicas línguas que se distinguem em relação à média da sonoridade são Japonês e Holandês. Todas as outras aparecem com valores intermediários que não podem ser considerados distintos de uma ou de outra.

Não encontramos evidências de que a função sonoridade aqui utilizada sirva para agrupar as línguas em concordância com a conjectura das classes rítmicas.

Referências Bibliográficas

- [1] Abadi, M. (2002) Statistics and error terms of occurrence times in mixing process. Manuscrito.
- [2] Abercrombie, D. Elements of general phonetics. *Chicago: Aldine, 1967.*
- [3] Athreya, K. B. e Fuh, C. D.(1992) Bootstrapping Markov chains: countable case. *J. Statist. Plann. Inference*, Vol 33, no 3, 311-331.
- [4] Bickel, P. J. e Freedman, D. A. (1981) Some asymptotic theory for the bootstrap. *Annals of Statistics*, vol 9, 1196-1217.
- [5] Billingsley, P.(1967) Convergence of probability measures
- [6] Bühlmann, P (1994) Blockwise bootstrapped empirical process for stationary sequences. *Annals of Statistics*, vol 22, 995-1012.
- [7] Bressaud, X., R. Fernández e Galves, A.(1999). Speed of d-convergence for Markov approximations of chains with complete connections. A coupling approach. *Stochastic Process. Appl.*, vol. 83 , no.1, 127-138.
- [8] Bressaud, X., R. Fernández e Galves, A.(1999). Decay of correlations for non Hölderian dynamics. A coupling approach. *Electronic Journal of Probability*, vol 4, N3 1-19.
- [9] Cassandro, M., Collet, P., Galves, A. e Ch. Galves (1999). A Statistical-Physics Approach to Language Acquisition and Language Change. *Physica A*, vol. 263, 427-437.
- [10] Collet, P., Galves, A. e A. Lopes (1995). Maximum likelihood and minimum entropy identification of grammars. *Random and Computational Dynamics*, vol. 3, 241-256.
- [11] Collet, P., Galves, A. e Schmitt (1999). Repetition times for gibbsian sources. *Non-linearity*, vol 12, 1225-1237.
- [12] W. Doeblin and R. Fortet. Sur les chaînes à liaisons complètes (1937). *Bull. Soc. Math. France*, vol 65, 132-148.

- [13] Duarte, D., Galves, A., Garcia, N. L. e Maronna, R. (2001) The Statistical analysis of acoustic correlates of speech rhythm. Artigo apresentado no ZiF, Bielefeld.
- [14] Efron, B. (1979) Bootstrap methods: another look at jackknife. *Annals of Statistics*, vol 7, 1-26
- [15] Fernández, R. e A. Galves (2000). Identifying features in the presence of competing evidence. The case of first-language acquisition. Em *Dynamical Systems: from Crystal to Chaos*, World Sci. Ser. Appl. Anal., 52-62.
- [16] Fernández, R. Ferrari, P. e A. Galves (2001). Coupling, renewal and perfect simulations of chain of infinite order, 5^a Escola Brasileira de Probabilidades. Pode ser obtido em <http://www.ime.usp.br/~galves/livro/revised2.pdf>
- [17] P. A. Ferrari, A. Maass, S. Martínez, and P. Ney. Cesàro mean distribution of group automata starting from measures with summable decay, 2000. *Ergodic Th. Dyn. Syst.*, 20, No 6, 1657-1670.
- [18] Galves, A., Garcia, J., Duarte, D. e Galves, Ch. (2002) Sonority as a basis for rhythmic class discrimination. Artigo apresentado no *Speech Prosody 2002*. Aix-en-Provence. (<http://www.lpl.univ-aix.fr/sp2002>).
- [19] Guiol, D. Comportamento assintótico de estimadores da entropia para cadeias de ordem infinita com perda de memória exponencial (2001). Tese de doutorado IME/USP.
- [20] Harris, T. E. (1955) On chains of infinite order. *Pacific J. Math.*, vol 5, 707-24.
- [21] Iosifescu, M. e Grigorescu, S. (1990) *Dependence with Complete Connections and its Applications*. Cambridge University Press, Cambridge, UK.
- [22] Künsch, H. R. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, vol 17, 1217-1241.
- [23] Liu, R. Y. e Singh, H. (1992) Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the limits of the bootstrap* (R. Lepage e L. Billard eds) 225-248, Wiley, N. Y.
- [24] Mehler, J., Dupoux, E., Nazzi, T., Dehane-Lambertz, G. (1996) Coping with linguistic diversity: the infant's viewpoint. *Signal to syntax: bootstrapping from speech to grammar in early acquisition*, J.L. Morgan and K. Demuth eds.
- [25] Naik-Ninbalkar, U. e Rajarshi, M. B. (1994) Validity of blockwise bootstrap for empirical process with stationary observations. *Annals of Statistics*, vol 22, 980-994.

- [26] Onicescu, O. e Mihoc, G. (1935) Sur les chaines de variables statistiques, Bulletin Science. Math, 59, 174-192.
- [27] Peligrad, M.(1998) On the blockwise bootstrap for empirical processes for stationary sequences. Annals of Probability 26, vol 2 877-901.
- [28] Raftery, A.(1985) A model for high-order Markov chains. *J. R. Statist. Soc. B*, 47:528-539.
- [29] Raftery, A. (1985) A new model for discrete-valued time series: autocorrelations and extensions. *Rass. Met. Statist. Appl.*, 3-4:149-162.
- [30] Raftery,A. e Tavaré, A. (1994) Estimation and modelling repeated patterns in high order Markov chains with the mixture transition distribution model. *Appl. Statist.*, 43:179–199, 1994.
- [31] Ramus, F., Nespor, M. e J. Mehler (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.
- [32] Singh, H (1981) On the asymptotic accuracy of Efron’s bootstrap. *Annals of Statistics* vol 9, 1187-1195.
- [33] Shao,Q. e Yu, H. (1993) Bootstrapping the sample mean for stationary mixing sequences. *Stochastic Process and applications* vol 48 175-190.