

TESE DE DOUTORADO Nº 189
UMA MEDIDA DE CONTRIBUIÇÃO LOCAL PARA PROCESSOS FORA DE
REGIME

Silvio Alves de Souza
DATA DA DEFESA: 30/04/2015

Universidade Federal de Minas Gerais
Escola de Engenharia
Programa de Pós-Graduação em Engenharia Elétrica

**UMA MEDIDA DE CONTRIBUIÇÃO LOCAL PARA PROCESSOS
FORA DE REGIME**

Silvio Alves de Souza

Tese de Doutorado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do Título de Doutor em Engenharia Elétrica.

Orientador: Prof. Eduardo Mazoni Andrade Marçal Mendes

Belo Horizonte - MG

Abril de 2015


"Uma Medida de Contribuição Local para Processos Fora de Regime"

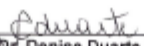
Silvio Alves de Souza

Tese de Doutorado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Doutor em Engenharia Elétrica.

Aprovada em 30 de abril de 2015.


Por:

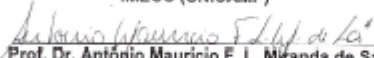

Prof. Dr. Eduardo Mazoni Andrade Marçal Mendes
DELT (UFMG) - Orientador


Prof. Dr. Denise Duarte Scarpa Magalhaes Alves
EST (UFMG)


Prof. Dr. Hani Camille Yehia
DELT (UFMG)


Prof. Dr. Carlos Andrey Maia
DEE (UFMG)


Prof. Dr. Jesus Enrique Garcia
IMECC (UNICAMP)


Prof. Dr. Antonio Mauricio F. L. Miranda de Sa
COPPE (UFRJ)

Departamento de Engenharia Eletrônica
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, CEP 31270-901, Belo Horizonte, MG, Brasil
Fone: +55 3499-3470



Uma Medida de Contribuição Local Para Sequências de Processos Fora de Regime Permanente

Silvio Alves de Souza

Orientador: Prof. Dr. Eduardo Mazoni Andrade Marçal Mendes
Co-Orientadora: Profa. Dra. Denise Duarte

Belo Horizonte, 30 de abril de 2015

Departamento de Engenharia Eletrônica
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, CEP 31270-901, Belo Horizonte, MG, Brasil
Fone: +55 3499-3470



Uma Medida de Contribuição Local Para Sequências de Processos Fora de Regime Permanente

Silvio Alves de Souza

Tese de doutorado submetida à banca examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Eletrônica da Universidade Federal de Minas Gerais, como parte dos requisitos necessários à obtenção do grau de Doutor em Engenharia Elétrica.

Orientador: Prof. Dr. Eduardo Mazoni Andrade Marçal Mendes
Co-Orientadora: Profa. Dra. Denise Duarte

Belo Horizonte, 30 de abril de 2015

Dedicatória

A Deus, por esta oportunidade e pela força de seguir sempre em frente.

A Antônio Victor e Maria da Conceição, meus pais.

A Camila Alves, minha filha.

A Rosânia, minha esposa.

A todos os meus irmãos, irmãs, cunhados e cunhadas.

A todos os meus sobrinhos e sobrinhas.

A todos os meus familiares.

Aos meus orientadores Denise e Eduardo, pela oportunidade deste trabalho e por tudo que contribuíram para minha aprendizagem.

A todos aqueles que me ajudaram.

À vida.

Agradecimentos

No livro "Parábolas Eternas" encontrei duas frases que marcam este momento:

Feliz é aquele que tem a capacidade de sonhar e sentir, em seus sonhos, a beleza da vida.

Todos têm a oportunidade de aprender o caminho, mas só aqueles que fazem o percurso chegam.

A primeira frase nos traz a beleza de sonhar. Sonhar em algo que acreditamos ser possível. Sonhar com este momento, que por anos dedicamos. Foram muitos os momentos de sacrifícios pessoais e tantos outros de cansaço. Momentos de dedicação em que deixei de realizar muitas coisas. Mas algumas coisas eu nunca deixei, pois são momentos que não voltam. Um deles é viver a felicidade de estar ao lado de minha família e de aproveitar o crescimento de minha filha.

A segunda frase nos mostra a dádiva da perseverança. Para mim o curso de doutorado foi um sonho que se tornará realidade. Não esperava chegar a tão longe, mas a força e determinação me guiaram.

Neste momento agradeço a todos:

Inicialmente agradeço a Deus. Por ter nascido, vivido e ter chegado até aqui. Sei que sua misericórdia está sendo grande. E peço que ajude a todos.

Agradeço a Rosânia e Camila, esposa e filha, por estar em meu lado e por compreender, por muitas vezes, minha ausência.

Aos meus pais, irmãos e familiares, pela força.

A todos os professores, desde o jardim ao doutorado, que de uma forma ou de outra contribuíram para este momento.

Aos professores Eduardo Mazoni e Denise Duarte por acreditarem em mim. Por me darem esta oportunidade e pela paciência dedicada a mim.

Epígrafe

*“A pedra segura.
O espinho previne.
O fel remedeia.
O fogo refunde.
O lixo fertiliza.
O temporal purifica a atmosfera.
O sofrimento redime.
A enfermidade adverte.
O sacrificio enriquece a vida.
A morte renova sempre.
Aprendamos , assim, a louvar o Senhor pelas bênçãos que nos confere.
Bom é o calor que modifica, bom é o frio que conserva.
A alegria que estimula é irmã da dor que aperfeiçoa.
Roguemos à Providência Celeste suficiente luz para que nossos olhos identifiquem o celeiro da graça em que nos encontramos.
É a cegueira íntima que nos faz tropeçar em obstáculos, onde só existe o favor divino.
...
”*

Mensagem: *Rendamos graças*, Psicografada por Francisco Cândido Xavier. Extraída do livro *Pão Nosso*.

Sumário

Resumo	xiii
Abstract	xv
Lista de Figuras	xviii
Lista de Tabelas	xix
Lista de Símbolos	xxi
Lista de Abreviações	1
1 Introdução	3
1.1 Objetivos	7
1.2 Estrutura do Trabalho	7
2 Fundamentação Teórica	9
2.1 Cadeia de Markov Discreta	9
2.2 Árvore Probabilística de Contexto	11
2.3 Automata Finito Probabilístico	14
2.4 Automata Probabilístico de Sufixo	15
2.5 Estimação de PCT via BIC	17
2.6 Cadeia de Markov Contínua	21
2.6.1 Núcleo Estimador	22
3 Revisão Bibliográfica	25

4	Contribuição Local em Sequências de PCT para Espaço de Estados Discreto	41
4.1	Emulação de uma Árvore Probabilística de Contexto por um Automato Probabilístico de Sufixo	41
4.2	Estimação da Matriz de Transição Cruzada	42
4.3	Sequências de Automatos Probabilísticos de Sufixos	47
4.3.1	Cadeias de Markov de Ordem 1	47
4.3.2	Cadeia de Markov de Ordem L	48
4.3.3	Modelagem Utilizando Árvore Probabilística de Contexto	49
4.4	Algoritmo Proposto	51
4.5	Medida de Proximidade entre Distribuições de Probabilidade	52
4.6	Aplicações do Modelo Proposto	57
4.7	Estimação dos Coeficientes de Contribuição Local dos Ritmos em Textos Escritos	60
4.7.1	Coeficiente de Contribuição Local de Textos Escritos por Mesmo Autor	61
4.7.2	Autores Nascidos em Períodos de Tempo Próximos	62
4.7.3	Autores nascidos em Diferentes Períodos de Tempo	63
4.8	Estimação da Variância dos Coeficientes de Contribuição	63
4.8.1	Estimativa da Variância dos Textos	64
5	Contribuição Local em Sequências de Cadeia de Markov com Espaço de Estados Contínuo	67
5.1	Modelo: Combinação Linear de Densidades de Transição Marginais	67
5.2	Função a ser Otimizada	70
5.3	Estimação das Funções de Densidade Marginais	72
5.4	Janela Ótima	74
5.5	Implementação	76
5.6	Simulações	77
5.6.1	Testando o Tipo de Intervalo	78
5.6.2	Testando o Tamanho de Amostra Efetivo N^*	83
5.6.3	Testando o Número de Intervalos w	84

	xi
5.6.4	Testando a Precisão ϵ_h 86
5.6.5	Comportamento dos λ 's em Cadeias de Markov com Intervalo Absorvente 86
5.7	Estimação da Variância dos Coeficientes de Contribuição 90
5.7.1	Estimativa da Variância da Cadeia 90
5.8	Conclusões Sobre as Simulações 91
5.9	Aplicação 92
5.9.1	Controle de Manutenção 93
5.9.2	A Técnica de Envelope de Aceleração 94
5.9.3	A Máquina de Trefilar 16032 95
5.9.4	Análise dos dados da máquina de Trefilar 16032 99
6	Perspectivas de Continuidade do Trabalho de Pesquisa 105
7	Conclusões 107
8	Anexo 1 109
8.1	Matrizes de Frequências de Transições Estimadas da Aplicação do Capítulo 4 109
8.1.1	Matriz de Frequência de Transição dos Textos G003 e G004 109
8.1.2	Matriz de Frequência de Transição dos Textos V002 e V004 111
8.1.3	Matriz de Frequência de Transição dos Textos A003 e A004 112
8.1.4	Matriz de Frequência de Transição dos textos G008 e O001 114
9	Anexo 2 117
9.1	Dados Rolamentos Máquina de Trefilar 117
10	Anexo 3 121
10.1	Reflexões Sobre Processo de Otimização Minimax 121
10.1.1	Valores dos λ 's Iguais a 0,5 122
10.1.2	Valores dos λ 's Diferentes de 0,5 123

11 Anexo 4	127
11.1 Cartas, Almeida Garrett (Garrett, A.) - Texto G003	127
Bibliografia	132

Resumo

Neste trabalho, consideramos um conjunto com um número finito de sequências, tanto discretas quanto contínuas. Nosso objetivo é medir instantaneamente a contribuição da distribuição de probabilidade marginal de cada sequência deste conjunto para o cálculo da distribuição de probabilidade marginal de outra sequência, no caso discreto. No caso contínuo, o objetivo é medir a contribuição de cada função de densidade de transição de sequência deste conjunto para a função de transição marginal de outra, instantaneamente.

No caso discreto, consideramos sequências categóricas discretas geradas por fontes com o mesmo alfabeto. A contribuição encontrada é medida pelo peso de um parâmetro combinação linear especial envolvendo informações de todas as sequências. Os coeficientes desta combinação são denominados *coeficientes de contribuição local* e representam o quanto cada distribuição de probabilidade marginal contribui para o cálculo de determinada distribuição de probabilidade marginal. Por isso, é possível saber em tempo real quão provável é um resultado em uma sequência baseando-se nos resultados de todas as outras sequências. A metodologia aqui proposta estende os modelos dados em Ching et al. [2002] de cadeias de Markov de ordem fixa para modelos de árvores probabilísticas de contextos. Um algoritmo para estimar a contribuição local de cada sequência de distribuições marginais evoluindo com tempo é apresentado. Os resultados das simulações mostram que a distribuição marginal estimada utilizando o modelo está perto da verdadeira. A metodologia proposta é aplicada a um corpus eletrônico Português histórico de textos escritos em Português por autores nascidos entre 1380 e 1845 codificada de acordo com características rítmicas.

No caso contínuo, aplicamos a metodologia a um conjunto de sequências com espaço de estado contínuo e em tempo discreto com estrutura de dependência Marko-

viana sobre o passado. Tal medida não é uma medida de (dis) similaridade entre séries temporais, mas uma medida da contribuição de uma seqüência para outra seqüência. Para isto, a função de densidade de transição marginal cruzada ao longo das seqüências contínuas de Markov é definida e um modelo que exprime a função de densidade de transição marginal de uma determinada cadeia como uma combinação linear de todas as outras funções de densidade de transição marginal é proposto. Esta contribuição é medida pelo peso de um parâmetro combinação linear especial envolvendo informações de todas as outras funções de densidade. Desta forma, a densidade de transição marginal leva em conta as informações de todas as outras seqüências. Os coeficientes desta combinação também são denominados *coeficientes de contribuição local* e representam o quanto cada função de densidade de transição contribui para o cálculo de determinada função de densidade de transição. Como resultado, um algoritmo para estimar a contribuição local de cada seqüência de funções de densidade é apresentado. Para estimar as funções de densidade marginais são utilizados núcleos estimadores. As simulações, considerando 2 cadeias, são apresentadas de modo a verificar a sensibilidade do modelo proposto e um exemplo usando níveis de vibração dos rolamentos dos motores da máquina de trefilar também é apresentado.

Abstract

In this work we consider a set with a finite number of sequences, discrete and continuous. The main goal in this work is to measure instantaneously the contribution of each sequence of this set to the marginal probability distribution of another, instantaneously, in the discrete case. In the continuous case, the aim is to measure the contribution of each transition density function of these sequences to the marginal transition function of another one instantaneously.

In the discrete case we consider categorical sequences generated by sources with the same alphabet. The contribution found is measured by the weight parameter of a special linear combination involving information of all sequences. The coefficients of this combination are denominated *local contribution coefficients* and represent how much each marginal probability distribution contributes to calculating of another. Therefore it is possible to know in real time how probable an outcome is based on the outcomes of all other sequences. The methodology proposed here extends the models given in Ching et al. [2002] for Markov chains to Probabilistic Context Tree models. An algorithm to estimate the local contribution of each sequence to marginal distributions as time evolves is presented. Simulation results show that the estimated marginal distribution using the model is close to the true one. The proposed methodology is applied to a Historical Portuguese electronic corpus of texts written in Portuguese by authors born between 1380 and 1845 codified according to rhythmic features.

In the continuous case a methodology to measure the contribution of one sequence (time series) to another sequence using the marginal probability of the process is proposed. A set of sequences with continuous space state and discrete time with Markovian dependence structure on the past is considered. Such a measure is not a measure of

(dis)similarity between time series, but rather a measure of contribution of one sequence towards another sequence. To this end, the cross marginal transition density function across continuous Markov chains is defined and a model that expresses the marginal transition density function of a determined chain as a linear combination of all the other marginal transition density functions is proposed. This contribution is measured by the weight parameter of a special linear combination involving information of all the other density functions. In this way the marginal transition density takes into account information of all other sequences. The coefficients of this combination are also denominated *local contribution coefficients* and represent how much each transition density function contributes to the calculation of the determined transition density function. As a result an algorithm to estimate the local contribution of each sequence of density functions is presented. In order to estimate the marginal density functions a kernel-based estimator is used. Simulations, considering 2 chains, are presented so as to check the sensitivity of the proposed model and also we presented an example using vibration levels of the bearings of the wire-drawing machine motor.

Lista de Figuras

2.1	Árvore Probabilística de Contexto	12
2.2	Exemplo de uma PSA. A linha sólida representa 0 e a linha tracejada representa 1 como o próximo símbolo.	16
2.3	Densidade conjunta marginal, $q_n(x,y)$, da cadeia X para ela mesma considerando os pontos x e y	24
4.1	PCT (a) e seu equivalente PSA (b)	43
4.2	Transição cruzada: (a) X para Y e (b) Y para X	44
4.3	Árvore probabilística de contexto, τ_1 , utilizada para gerar as cadeias	54
4.4	Árvore probabilística de contexto, τ_2 , utilizada para gerar as cadeias	54
4.5	Divergência de Kullback-Leibler entre a verdadeira distribuição marginal de uma PCT e a distribuição marginal estimada para τ_1 em (a) e (b), e τ_2 em (c) e (d).	56
4.6	PCT estimada dos textos do Corpus of Historical Portuguese Tycho	59
5.1	Densidade conjunta, $q_n^{(jk)}(x,y)$, da cadeia k para a cadeia j considerando os pontos x e y . k representa a cadeia X e j representa a cadeia Y	73
5.2	Média ergódica das cadeias X e Y	81
5.3	Função de densidade de probabilidade: (a) cadeia X e (b) cadeia Y	82
5.4	(a) Comportamento da cadeia X e (b) Comportamento da cadeia Y	89
5.5	Modelo de uma Máquina de trefilar	96
5.6	Foto de uma Máquina de trefilar	96
5.7	Foto de um rebobinador de uma Máquina de trefilar	96
5.8	Caixa de lubrificante e de fieira de uma Máquina de trefilar	97
5.9	(a) Vista em corte longitudinal e (b) foto real de uma fieira uma Máquina de trefilar	98
5.10	Pontos de medição de vibração do motor de uma Máquina de trefilar	98
5.11	Espectro do rolamento dianteiro do bloco 1 da máquina de trefilar 16032	99
5.12	Espectro do rolamento dianteiro do bloco 2 da máquina de trefilar 16032	100
5.13	Espectro do rolamento dianteiro do bloco 3 da máquina de trefilar 16032	100
5.14	Espectro do rolamento dianteiro do bloco 4 da máquina de trefilar 16032	101
5.15	Espectro do rolamento dianteiro do bloco 5 da máquina de trefilar 16032	101

5.16	Espectro do rolamento dianteiro do bloco 6 da máquina de trefilar 16032	102
5.17	Espectro do rolamento dianteiro do bloco 7 da máquina de trefilar 16032	102
10.1	Representação Gráfica do modelo proposto.	122
10.2	Gráfico das restrições.	123
10.3	Gráfico das restrições.	124
10.4	Gráfico das restrições.	124
10.5	Gráfico das restrições.	125
10.6	Gráfico das restrições.	125

Lista de Tabelas

2.1	Probabilidade de transição da árvore (τ, p)	13
2.2	Função de transição, α , para os estados de E e probabilidade de transição, γ , entre parênteses, de acordo com o próximo símbolo ser 0 ou 1.	16
4.1	Matriz de probabilidade de transição quadrada da PSA da Figura 4.1.	43
4.2	Textos escritos apresentados por Código, Ano, Autor e Título.	57
5.1	Valores dos λ 's por tamanho de amostra efetiva N^*	84
5.2	Valores dos λ 's para cada número de intervalo w	85
9.1	Valores dos níveis de aceleração, em G, do rolamento dianteiro dos blocos da máquina de trefilar 16032,	117

Lista de Símbolos

\mathbb{Z}	Espaço dos números inteiros;
\mathbb{N}	Espaço dos números naturais;
\mathbb{R}	Espaço dos números reais;
\square	Fim de demonstração;
E	Espaço de estados de uma PCT;
\mathbb{A}	Espaço de estados da cadeia;
h_{op}	Janela ótima;
D_{KL}	Divergência de Kullbach-Leibler;
P_{ij}	Matriz de probabilidade de transição do estado j para o estado i ;
$F^{(ij)}$	Matriz de frequência de transição do estado j para o estado i ;
μ_n^k	Distribuição de probabilidade marginal de estados da k -ésima sequência no passo n ;
f_n	Função de densidade de probabilidade estacionária marginal no tempo n ;
t_n	Função de densidade de probabilidade de transição marginal no tempo n ;
q	Função de densidade de probabilidade conjunta marginal no tempo n ;

Lista de Abreviatura

PCT	Árvores Probabilísticas de Contexto
PST	Árvore Probabilística de Sufixo
PSA	Automato de Sufixo Probabilístico
VLMC	Cadeia de Markov de comprimento variável
BIC	Critério de Informação Bayesiana
SMC	Critério de Smallest Maximizer
MISE	Erro Quadrático Integrado Médio
ISE	Erro Quadrático Integrado

Introdução

Modelagem por Cadeias de Markov tem sido aplicada a muitos campos da ciência. Suas aplicações podem ser encontradas na Física, Química, Genética, Economia, Teoria da Informação, Engenharia e muitos outros (ver, por exemplo, (Avery and Henderson [1999], Hamilton [1989], Gill [2007], Robert and Casella [2004] e Akimov et al. [2007])). Sistemas Markovianos aparecem em Termodinâmica e Mecânica Estatística, sempre que as probabilidades são usadas para representar detalhes desconhecidos do sistema. O famoso artigo, *A mathematical theory of communication*, de Claude Shannon em 1948 criou o campo da teoria da informação, iniciou com a introdução do conceito de entropia por meio de um modelo de Markov do idioma inglês. O PageRank de uma página web usado pelo Google é definida por uma cadeia de Markov (Page et al. [1999]). Em Estatística, cadeias de Markov também se tornaram muito importantes para gerar sequências de números aleatórios para refletir com precisão complicadas distribuições de probabilidades desejadas, por meio de cadeia de Markov Monte Carlo (MCMC).

Em muitas áreas diferentes como mercado de ações, línguas naturais, médicas e biológicas, redes neurais e muitas outras utilizam dados de séries temporais (ver, por exemplo, (Vrugt et al. [2013], Lladser et al. [2013], Andersen et al. [2014], Bhattacharya [2014] e Pranevicius et al. [2014])). Como consequência, muito esforço tem sido dedicado à introdução de novas metodologias para a classificação, agrupamento e aproximação das séries de tempo (Faloutsos [1994]). A fim de realizar estas tarefas algumas medidas para o cálculo das distâncias entre duas ou mais séries de tempo foram propostas na literatura. Exemplos dessas medidas são várias. Por exemplo, a distância euclidiana (ED) entre duas sequências é dada por (Faloutsos [1994]) onde os autores apresentam um método de indexação eficiente para localizar subsequências unidimensionais dentro de uma coleção de sequências, de modo que estas subsequen-

cias apresentam um padrão dentro de uma tolerância especificada. O objetivo é mapear cada sequência de dados em um pequeno conjunto de retângulos multidimensionais no espaço recurso. Esses retângulos podem ser facilmente indexados usando métodos de acessos espaciais tradicionais com o uso de uma janela deslizante ao longo da sequência de dados afim de extrair suas características. Um outro exemplo é a distância definida como Dynamic Time Warping (DTW) (Bemdt [1994]) no qual os autores, por meio do conhecimento de bases de dados, apresentam ferramentas computacionais para explorar grandes arquivos de dados. Muitos desses dados são temporais, como os preços das ações ou dados de telemetria da NASA. Os mesmos autores descrevem algumas experiências preliminares com abordagem de programação dinâmica e apresentam um algoritmo de detecção de padrões baseado na técnica de deformação dinâmica aplicado ao reconhecimento de fala. Keogh and Ratanamahatana [2004] também utilizam a distância DTW, pois é mais robusta do que a distância euclidiana ao trabalhar com séries temporais e apresentam uma nova técnica para a indexação exata da distância DTW.

Vlachos et al. [2002] investigam técnicas para análise e recuperação de trajetórias de objetos em espaços euclidianos de duas ou três dimensões com presença de grande quantidade de ruídos. Estes dados surgem ao trabalhar com experiências de mobilidade dos animais, reconhecimento de sinais de linguagem, o uso do telefone móvel, na computação móvel, sensores e tecnologia de GPS. Considerando que, neste caso, muitas métricas falham, apresentam uma medida de similaridade não métricas denominada Longest Common Subsequence (LCSS), que é muito mais robusta aos ruídos e, além disso, fornece uma noção intuitiva de semelhança entre trajetórias, dando mais peso para as porções semelhantes das sequências. Os autores comparam estes novos métodos para as funções euclidianas e Dynamic Time Warping (DTW) e mostram a superioridade do método, especialmente sob a forte presença de ruído.

Morse and Patel [2007] propõem um novo algoritmo denominado Fast Time Series Evaluation (FTSE) que utiliza uma medida de similaridade denominada Sequence Weighted ALignmEnt (Swale) que pode ser utilizada para avaliar técnicas de valor limiar, incluindo LCSS e DTW. Os autores comparam as métricas Swale, LCSS e DTW

(entre outras) e mostram que, geralmente, a Swale é mais precisa. Chen et al. [2007] propõem uma medida de distância de deformação denominada Spatial Assembling Distance (SpADe), eficiente em detecção de padrões contínuos, adequados no tratamento de mudança de escala e na dimensão da amplitude, o qual os métodos LCSS e DTW são ineficazes. Abfalg et al. [2006] apresentam uma novo conceito de pesquisa por similaridade com base em consultas ao limiar em séries temporais (threshold queries) e uma medida de distância que mapeia cada intervalo de uma seqüência para o mais próximo (mais semelhante) intervalo da outra seqüência e vice-versa. Segundo os autores o método é eficiente e utiliza o fato de que apenas informações parciais da série é necessário no momento da consulta. Wang et al. [2012] fazem uma comparação destas medidas de distância para os dados de séries temporais.

O presente trabalho se concentra, entretanto, em uma medida da contribuição de uma seqüência (série temporal) para outra seqüência usando a probabilidade marginal do processo, em vez de semelhança das observações nas seqüências. Tal medida não é uma medida de similaridade entre as seqüências categóricas, mas uma medida da contribuição de uma seqüência para outra seqüência. Portanto, o objetivo não é comparar o desempenho desta medida de contribuição com outras medidas de distância. Há muitas medidas propostas na literatura para calcular distâncias entre duas ou mais seqüências. O leitor interessado é referido a (Wang et al. [2012]).

A modelagem com cadeias de Markov pode levar a um número grande de parâmetros quando a ordem de dependência aos valores do passado não é pequena. Por outro lado, a modelagem utilizando Árvore Probabilística de contexto (PCT), modelo proposto por Rissanen [1983] na teoria da informação e introduzido por Bühlmann and Wyner [1999] na literatura estatística, pode resolver este problema pois torna possível reduzir o número de parâmetros livres no modelo. Considerando uma cadeia de Markov de ordem r e um alfabeto finito A de tamanho $|A|$ o número de parâmetros livres é $|A|^r(|A| - 1)$, que cresce exponencialmente com r . Na modelagem utilizando PCT o número de parâmetros livres é no máximo da ordem de $|A|^r$ se todos os passados forem considerados, o que equivale a um modelo de cadeias de Markov de ordem r , mas, em geral, esse número é muito menor, uma vez que, nesse modelo, pode-se considerar

várias probabilidades de transição sendo iguais.

Neste trabalho, uma de nossas contribuições, é utilizar PCT (árvore probabilística de contexto) e seu PSA (automato probabilístico de sufixo) equivalente como uma forma de contornar o problema do aumento do número de parâmetros da cadeia quando a ordem de dependência no passado não é pequena. Apresentamos uma metodologia para estimar a contribuição de cada PSA em relação a outros. Esta contribuição é medida pelo parâmetro peso da combinação linear que envolve as informações de todas os outros PSA's. Assim podemos verificar o quanto uma distribuições de probabilidade marginal é influenciada pelas demais. Embora o número de trabalhos publicados com aplicações de PCT seja muito menor que o número utilizando cadeias de Markov, alguns deles mostram que a aplicação do primeiro, em alguns campos, resultou em ganho quando comparado com a aplicação do último (Ver, por exemplo, (Bejerano [2004], Browning [2006], Leonardi [2006], Bühlmann and Wyner [1999])).

Para estimar uma Árvore Probabilística de contexto e as probabilidades de transição, Rissanen [1983] introduz o *Algoritmo de Contexto*. Csiszar and Talata [2006], utilizando o *Critério de Informação Bayesiana* (BIC), propõem uma metodologia para podar as árvores candidatas e estimar a PCT em tempo real. Para o caso de amostra finita, Galves et al. [2012] apresentam uma metodologia para estimar os contextos da árvore. Optamos em utilizar a árvore estimada por Csiszar and Talata [2006], pelo fato de ser uma ótima metodologia e nosso interesse não é em estimar árvores mas estimar a contribuição entre os PSA's.

No caso contínuo, apresentamos uma metodologia para estimar a contribuição das densidades de transição marginais de uma sequência em relação a outras sequências, como outra contribuição deste trabalho. Neste caso, estimamos as densidades via núcleo estimador. Esta contribuição é medida pelo parâmetro peso da combinação linear que envolve as informações de todas as outras densidades de transição marginais. Assim podemos verificar o quanto uma densidade de transição marginal é influenciada pelas demais.

O conceito de *função de densidade de transição cruzada marginal* entre séries temporais é introduzido e é o principal instrumento utilizado na metodologia proposta neste

trabalho. Esta densidade de transição permite calcular a probabilidade de uma determinada sequência de pertencer a um intervalo fixo, a cada tempo, dada a informação sobre a saída da outra sequência no tempo anterior. Para estimar estas densidades de transições cruzadas marginais uma modificação nos estimadores do kernel habituais é proposto.

Após este breve relato sobre cadeias de Markov e PCT os objetivos, a motivação e a estrutura desta tese são apresentados.

1.1 Objetivos

Os objetivos deste trabalho são:

1. Encontrar o coeficiente de contribuição local das medidas marginais de duas ou mais sequências com espaço de estados discretos utilizando PCT;
2. Encontrar o coeficiente de contribuição local de duas ou mais sequências de cadeias de Markov com espaço de estados contínuo e tempo discreto de passo 1;
3. Aplicar as metodologias propostas a dados reais tanto para o caso discreto quanto para o caso contínuo.

1.2 Estrutura do Trabalho

O trabalho está apresentado da seguinte maneira: no capítulo 2 apresentamos algumas definições importantes sobre o assunto e principais resultados; no capítulo 3 apresentamos uma revisão bibliográfica sobre os assuntos tratados neste trabalho; trabalhamos com cadeias discretas e aplicamos a metodologia a um conjunto de dados linguísticos, apresentados no capítulo 4; no capítulo 5 apresentamos alguns resultados preliminares sobre o caso contínuo; no capítulo 6 descrevemos as perspectivas de continuidade do Trabalho de pesquisa; no capítulo 7 trazemos algumas conclusões; no capítulo 8 apresentamos as matrizes de frequência de transição utilizadas nas aplicações do capítulo 4; no capítulo 9 temos os anexos 2 e no capítulo 10 temos uma breve

discussão sobre o processo de otimização. No capítulo 11 apresentamos uma parte de um texto de Almeida Garrett, como exemplo. Por fim as referências bibliográficas.

Fundamentação Teórica

No presente capítulo apresentamos as principais definições e resultados necessários para a compreensão do tema.

2.1 Cadeia de Markov Discreta

Definição 2.1.1 *Um processo estocástico é uma família X_t de variáveis aleatórias tomadas ao longo do tempo $t \in T$. Se $T = \{0, 1, \dots\}$ então X_t é um processo estocástico com tempo discreto e se $T = [0, \infty)$ então X_t é um processo estocástico com tempo contínuo.*

Um processo estocástico X_t assume determinados valores para cada $t \in T$, digamos $X_t = a_t$. Estes valores pertencem a um conjunto \mathbb{A} denominado *espaço de estados* ou *alfabeto*. Este espaço de estados pode ser discreto ou contínuo. O processo estocástico é discreto quando o conjunto \mathbb{A} é discreto e o processo é contínuo quando o conjunto \mathbb{A} é contínuo.

Definição 2.1.2 *Um processo estocástico $(X_t)_{t \in T}$ com espaço de estados \mathbb{A} e $T = \{0, 1, \dots\}$ é uma Cadeia de Markov de ordem k se*

$$P(X_t = x_t | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots, X_1 = x_1) = P(X_t = x_t | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots, X_{t-k} = x_{t-k}) \quad (2.1)$$

onde $P(X_t = x_t | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots, X_1 = x_1)$ é a probabilidade condicional de X_t dado $\{X_{t-1}, X_{t-2}, \dots, X_1\}$ para todo $k \geq 1$ e para todo $x_i \in \mathbb{A}$ com $i = \{1, 2, \dots, t\}$.

Observe que a probabilidade da cadeia assumir determinado valor no tempo t depende de k termos anteriores a t . Em particular, uma cadeia tem ordem 1 se:

$$P(X_t = x_t | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots, X_1 = x_1) = P(X_t = x_t | X_{t-1} = x_{t-1}).$$

Definição 2.1.3 Uma cadeia de Markov $(X_t)_{t \in T}$ é dita estacionária se para todo r :

$$P(X_0 = x_0, X_1 = x_1, \dots, X_t = x_t) = P(X_r = x_0, X_{r+1} = x_1, \dots, X_{t+r} = x_t) \quad (2.2)$$

Em uma cadeia de Markov de tamanho n as probabilidades condicionais de ir para determinado estado j partindo de um estado i são dadas por uma matriz denominada *Matriz de Probabilidade de Transição*, P . Assim cada elemento desta matriz, p_{ij} , é definido como

$$p_{ji} = P(X_{n+1} = j | X_n = i) \quad (2.3)$$

onde $\sum_{j=1}^n p_{ji} = 1$.

Uma cadeia de Markov tem medida invariante π se π for solução da equação $\pi P = \pi$, onde P é a matriz de probabilidade de transição de estados.

Um fato importante para cadeias de Markov com espaço discreto $\mathbb{A} = \{0, 1, 2, \dots, N\}$ e matriz de probabilidade P é que $\sum_{j=0}^N p_{ij}(t) = 1$ para todo t . A soma dos elementos de cada coluna na matriz de transição de probabilidade é igual a 1.

A seguir apresentamos o teorema de Perron Frobenius. Sua demonstração é dada por Berman and Plemmons [1979]. Ele será utilizado como ferramenta para justificar o modelo apresentado no capítulo 4.

Teorema 2.1.1 (Perron Frobenius) *Seja Q uma matriz quadrada não-negativa e irredutível de ordem m . Então*

1. Q tem autovalor real positivo, θ , igual a seu raio espectral, i.e. $\theta = \max_k |\theta_k(Q)|$ onde $\theta_k(Q)$ denota o k -th autovalor de Q .
2. Seu correspondente autovetor z , suas entradas são reais e positivas, de modo que $Qz = \theta z$.

3. θ é um autovalor simples de Q .

Pelo teorema de Perron Frobenius, a matriz de transição P da cadeia de Markov tem um autovalor igual a 1 e todos os outros autovalores de P tem módulo menor ou igual a 1. Como corolário deste mesmo teorema, vemos que existe um vetor positivo z de modo que $Pz = z$. Este vetor z é denominado *vetor de probabilidade estacionária de P* , isto é $\sum_{i=1}^m z_i = 1$. Além disso, z_i é a probabilidade de estado estacionária que o sistema está no estado i .

2.2 Árvore Probabilística de Contexto

Ao trabalhar com cadeias de Markov de ordem r tomando valores em um alfabeto \mathbb{A} , finito, o número de parâmetros do modelo é de $|\mathbb{A}|^r(|\mathbb{A}| - 1)$, onde $|\mathbb{A}|$ é a cardinalidade do conjunto \mathbb{A} e $|\mathbb{A}|^r$ é o conjunto de todas as sequências de tamanho r . Assim, se r for grande, teremos um grande número de parâmetros a estimar. Uma alternativa é utilizar *Árvore Probabilística de Contexto* (PCT), também conhecida como cadeia de memória de comprimento variável (VLMC).

Introduzida por Rissanen [1983], uma PCT é uma classe de cadeias estocásticas com memória de comprimento variável. Ao contrário dos modelos de cadeias de Markov de ordem fixa k , onde cada variável no tempo t depende de um número fixo de variáveis no passado, não precisamos considerar todos os passados de tamanho k para estimar o próximo símbolo da cadeia. Precisamos apenas de uma subsequência finita do passado, denominada *Contexto*, para estimar o próximo símbolo. Estes contextos podem ser representados através de uma árvore enumerável denominada *árvore de contexto*, representada por τ . Assim a árvore τ é formada pela união de todos os contextos, $\tau = \cup_{i=1}^k A^i$, onde A^i é cada contexto de τ .

Definição 2.2.1 *Uma árvore probabilística de contexto, ou PCT, é um par (τ, p) onde $p = \{p(\cdot|w) : w \in \tau\}$ é uma família de medidas de probabilidade em \mathbb{A} associada a cada contexto w tal que $\{w \in \tau\}$.*

Seja \mathbb{A} um alfabeto finito e \mathcal{A} o conjunto de todas as possíveis sequências sobre \mathbb{A} . Considere uma sequência finita $w_1^k = w_1 \dots w_{k-1} w_k$ com $w_i \in \mathbb{A}$ e defina uv a

concatenação de duas subsequências finitas u e v com $u, v \in \mathcal{A}$. Quando existir uma subsequência u de modo que $w_1^k = uv$ dizemos que a subsequência v é um sufixo de w_1^k , denotada por $v = \text{suf}(w_1^k)$. A sequência vazia é denotada por \emptyset .

Se tivermos uma sequência semi-infinita $w_{-\infty}^{-1} = \dots w_{-2}w_{-1}$, um sufixo de $w_{-\infty}^{-1}$ é definido da mesma maneira.

Exemplo 2.2.1 Considere a sequência $w = 112112$. O ramo $v = 2112$ é um sufixo de w pois existe $u = 11$ de modo que $w = uv$.

O comprimento máximo da árvore τ é dado por $l(\tau) = \max\{l(\omega) : \omega \in \tau\}$ onde $l(\omega)$ é o comprimento do contexto w . Assim o comprimento é dado como o tamanho do maior contexto.

Um exemplo de árvore probabilística de contexto é mostrada na Figura 2.1. O alfabeto é formado pelos elementos $\mathbb{A} = \{0,1\}$ e as probabilidades condicionais, dado cada contexto, são apresentadas entre parênteses. O primeiro valor é a probabilidade do próximo símbolo ser 0 e o segundo valor é a probabilidade do próximo símbolo ser 1.

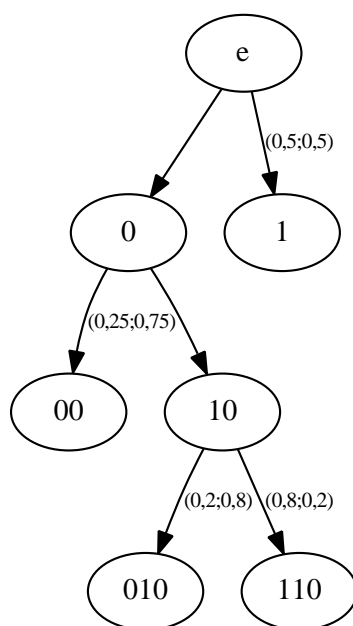


Figura 2.1: Árvore Probabilística de Contexto

Os contextos são definidos por

$$w_{-l}^{-1} = \begin{cases} 0,0 & \text{se } w_{-1} = 0, w_{-2} = 0 \\ 0,1,0 & \text{se } w_{-1} = 0, w_{-2} = 1; w_{-3} = 0 \\ 1,1,0 & \text{se } w_{-1} = 0, w_{-2} = 1; w_{-3} = 1 \\ 1 & \text{se } w_{-1} = 1 \end{cases}$$

Isto significa, por exemplo, que os caminhos "0,0,0" e "1,0,0" tem a mesma informação sobre o presente que o passado "00". Em termos da probabilidade condicional

$$\begin{aligned} P(X_0 = . | X_{-3} = 0, X_{-2} = 0, X_{-1} = 0) &= P(X_0 = . | X_{-3} = 1, X_{-2} = 0, X_{-1} = 0) \\ &= P(X_0 = . | X_{-2} = 0, X_{-1} = 0). \end{aligned}$$

Para cada contexto, ou ramo, a probabilidade condicional é associada a cada estado $a \in \mathbb{A}$. A Tabela 2.1 apresenta as probabilidades condicionais referentes à árvore da Figura 2.1.

Tabela 2.1: Probabilidade de transição da árvore (τ, p)

w	$p(0 w)$	$p(1 w)$
00	0,25	0,75
010	0,2	0,8
110	0,8	0,2
1	0,5	0,5

O modelo PCT com alfabeto \mathbb{A} tem $|\tau|(|\mathbb{A}| - 1)$ parâmetros a serem estimados, onde $|\tau|$ é a quantidade de contextos. O modelo cadeia de Markov de ordem r e alfabeto \mathbb{A} tem $|\mathbb{A}|^r(|\mathbb{A}| - 1)$ parâmetros a serem estimados. Como $|\tau| \ll |\mathbb{A}|^r$, em geral, o modelo PCT é mais parcimonioso, isto é, ele tem menos parâmetros a serem estimados.

Considere uma PCT com alfabeto finito \mathbb{A} em que $|\mathbb{A}|$ é a cardinalidade de \mathbb{A} . Um ramo $s = \omega_m \omega_{m+1} \dots \omega_n$ onde $(\omega_i \in \mathbb{A}, m \leq i \leq n)$ é denotado por ω_m^n cujo tamanho é $l(s) = n - m + 1$. Um ramo vazio é definido como $l(\emptyset) = 0$.

Uma PCT (τ, p) pode ser vista como uma árvore onde a raiz é o valor do processo no presente e os ramos são os contextos que são a parte relevante do passado para

prever o presente. Cada nó pode ter $|\mathbb{A}|$ ramos. Quando isso acontece a árvore é dita ser completa. Assim, uma árvore τ com número finito de galhos, será dita *completa* se τ define uma partição de $\mathbb{A}^{\{1,2,\dots\}}$. Neste caso cada elemento da partição coincide com o conjunto das sequências em $\mathbb{A}^{\{1,2,\dots\}}$ tendo ω_1^k como sufixo, para algum $\omega_1^k \in \tau$.

Definição 2.2.2 *Uma árvore de contexto τ é irredutível se ela satisfaz as seguintes condições:*

1. *Propriedade sufixo: Se $w_1^k \in \tau$ é um sufixo com $k \in \mathbb{N}$ então não existe uma sequência $v_1^i \in \tau$ com $i < k$ de modo que $w_j = v_j$ para $j = \{1, 2, \dots, i\}$.*
2. *Irredutibilidade. Nenhuma sequência pertencente a τ pode ser substituída por um sufixo adequado sem violar a propriedade sufixo.*

A propriedade de sufixo nos diz que dado um sufixo $s_1 \dots s_{k-1}s_k$ pertencente a τ não podemos ter outra sequência $v_1 \dots v_{j-1}v_j$ pertencente a τ , com $j < k$, cujos termos são iguais aos termos do sufixo dado, ou seja $s_i = v_i$ para $i = \{1, 2, \dots, j\}$.

Considerando os contextos de τ , a propriedade de sufixo nos diz que se $s_1 \dots s_{k-1}s_k$ é um contexto então $s_1 \dots s_{k-2}$ não é um contexto de τ .

Neste trabalho apenas árvores irredutíveis serão consideradas.

2.3 Automata Finito Probabilístico

Este modelo foi introduzido por Ron et al. [1996] e é estreitamente relacionado com a PCT.

Definição 2.3.1 *Um Automata Finito Probabilístico (PFA) é uma 5-quíntupla $(E, \mathbb{A}, \alpha, \gamma, \pi)$, onde E é um conjunto finito de estados, \mathbb{A} é um alfabeto, $\alpha : E \times \mathbb{A} \mapsto E$ é a função de transição, $\gamma : E \times \mathbb{A} \mapsto [0,1]$ é a função de probabilidade do próximo símbolo e $\pi : E \mapsto [0,1]$ é a distribuição inicial sobre os estados de E .*

As funções γ e π satisfazem às seguintes condições:

1. para cada $e \in E$ temos $\sum_{a \in \mathbb{A}} \gamma(e, a) = 1$;
2. $\sum_{e \in E} \pi(e) = 1$.

Note que α pode ser estendido e ser definido em $E \times \mathbb{A}$ como:

$$\alpha(\alpha(e, s_1 s_2 \dots s_{l-1}), s_l) = \alpha(\alpha(e, \text{suf}(s)), s_l).$$

2.4 Automata Probabilístico de Sufixo

Definição 2.4.1 *Seja \mathcal{A} o conjunto de todas as possíveis sequências em \mathbb{A} . Um automata probabilístico de sufixo (PSA) é uma subclasse de um PFA de modo que:*

1. *cada estado de E em uma PSA é definido por uma sequência de comprimento finito em \mathcal{A}*
 e
2. *para cada dois estados $e_1 \in E$ e $e_2 \in E$ e para cada símbolo $a \in \mathbb{A}$, se $\alpha(e_1, a) = e_2$ então*
 $e_2 = e_1 a$.

Ron et al. [1996] propõem um método para estimar uma PCT usando um *Automata Probabilístico de Sufixo* (PSA). Os autores mostram que existe uma equivalência entre estes dois modelos.

Considere o grafo subjacente definido por E e $\alpha(\cdot, \cdot)$ onde cada estado pode ser alcançado de quaisquer outros estados com probabilidade não nula. Deste modo um PSA pode ser pensado como uma cadeia de Markov com alfabeto E . Se o PSA é assumido aperiódico, então existe uma *distribuição estacionária* Π_{PSA} da PSA. Para cada par de estados q e q' então

$$\Pi_{PSA}(q) = \sum_{q': \delta(q', a)=q} \Pi_{PSA}(q') \gamma(q', a). \quad (2.4)$$

Para qualquer $L \geq 0$ a subclasse de PSA's em que cada estado é nomeado por uma sequência de comprimento máximo L é denotado por $L-PSA$. Quando E inclui todas as sequências em \mathbb{A}^L temos uma cadeia de Markov de ordem L nomeada $L_{full}-PSA$. Uma vez que cada $L-PSA$ pode ser estendida a uma equivalente $L_{full}-PSA$ ele pode sempre ser descrito por uma cadeia de Markov de ordem L . Alternativamente, desde que os estados de um $L-PSA$ podem ser identificados como apenas um pequeno subconjunto

de $\mathbb{A}^{\leq L}$ e muitos dos sufixos nomeados podem ser menores do que L , ele pode ser visto como uma cadeia de Markov com comprimento variável.

Figura 2.2 mostra um exemplo de um PSA onde $\mathbb{A} = \{0,1\}$, $E = \{11,01,00,110,010\}$. A Tabela 2.2 apresenta a função de transição α de um estado de E para outro e a probabilidade de transição γ de um estado de E para um símbolo de \mathbb{A} .

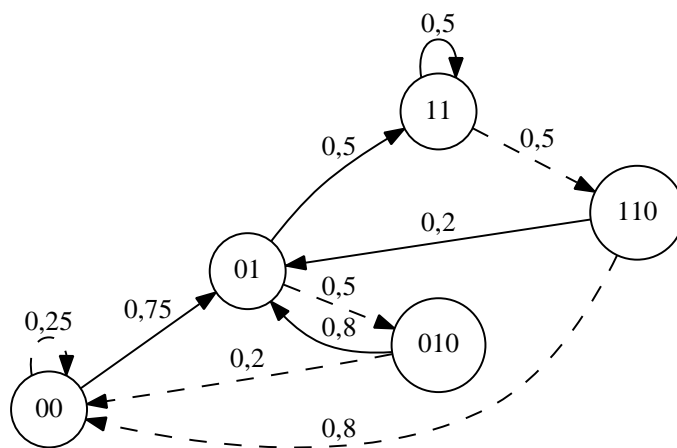


Figura 2.2: Exemplo de uma PSA. A linha sólida representa 0 e a linha tracejada representa 1 como o próximo símbolo.

Tabela 2.2: Função de transição, α , para os estados de E e probabilidade de transição, γ , entre parênteses, de acordo com o próximo símbolo ser 0 ou 1.

E	Próximo símbolo	
	0	1
00	00 [0,25]	01 [0,75]
01	010 [0,50]	11 [0,50]
11	110 [0,50]	11 [0,50]
110	00 [0,80]	01 [0,20]
010	00 [0,20]	01 [0,80]

Estimar uma L_{full} - PSA (cadeia de Markov de ordem L) é simples, uma vez que o conjunto dos estado é conhecido e a função de transição α é unicamente definida. A função de probabilidade do próximo símbolo γ é a única quantidade a ser estimada. No caso de cadeias de Markov de comprimento variável o problema de estimação é uma tarefa muito mais difícil, já que os estados do PSA são desconhecidos a princípio.

Na próxima seção um algoritmo eficiente para estimar uma PCT é apresentado.

A vantagem de se trabalhar com uma PSA é que ela é uma matriz quadrada de probabilidades.

2.5 Estimação de PCT via BIC

Alguns algoritmos são encontrados na literatura com objetivo de estimar uma PCT e também suas probabilidades de transição. O algoritmo de contexto foi apresentado inicialmente por Rissanen [1983]. Este algoritmo foi implementado por Bühlmann and Wyner [1999] que propõem um critério para podar a árvore máxima considerando todas as sequências que aparecem na amostra até um número fixo r vezes. A estimação da PCT é feita através de uma sequência de testes da razão de verossimilhança e depende da definição de um ponto de corte a ser escolhido pelo usuário. A consistência do algoritmo de contexto foi provado por Rissanen [1983] e Duarte et al. [2006] apresentam uma versão simplificada do teorema de consistência do algoritmo de contexto. Csiszar and Talata [2006] propõem um estimador baseado no Critério de Informação Bayesiana (BIC) introduzido por Schwarz [1978] para podar as árvores candidatas e estimar a PCT. Eles mostraram que deste modo a PCT pode ser consistentemente estimada em tempo linear.

Seguindo a abordagem apresentada em Csiszar and Talata [2006], Galves et al. [2012] introduz o Smallest Maximizer Criterion (SMC) para estimar os contextos de uma PCT para uma amostra finita. Este critério também estima a PCT de forma consistente. Garivier and Leonardi [2011] se referem a esta classe de algoritmos baseado no BIC como "penalização de máxima verossimilhança". O algoritmo SMC estima a melhor árvore dentro de uma classe de árvores selecionadas sem a necessidade de escolher uma constante. No entanto, para selecionar as árvores candidatas, é necessário utilizar o algoritmo BIC. Dado uma árvore τ , é preciso fazer podas nesta árvore de modo a encontrar a melhor árvore ou melhor modelo.

A seguir mostraremos como é estimada uma PCT utilizando a metodologia apresentada em Csiszar and Talata [2006]. Os autores utilizam o *Critério de Informação*

Bayesiano (BIC), utilizado para selecionar modelos, que penaliza a função de Máxima Verossimilhança pelo número de parâmetros a serem estimados. Em um modelo PCT, cada probabilidade de transição é um parâmetro a ser estimado. Assim, modelos com pequena amostra e muitos parâmetros podem não ser bem estimados.

Definição 2.5.1 Considere uma amostra x_1^n . O número de ocorrências do ramo $s \in \mathbb{A}^{l(s)}$ seguido pela letra $a \in \mathbb{A}$ é definido como:

$$N_n(s,a) = |\{i : D(n) < i \leq n, x_{i-l(s)}^{i-1} = s, x_i = a\}| \quad (2.5)$$

onde $D(n)$ é o comprimento máximo de s .

Definição 2.5.2 Considere uma amostra x_1^n . O número de ocorrências do ramo $s \in \mathbb{A}^{l(s)}$ é definido como:

$$N_n(s) = |\{i : D(n) < i \leq n, x_{i-l(s)}^{i-1} = s\}| \quad (2.6)$$

onde $D(n) = (\log(n))$ é o comprimento máximo de s .

Considerando as definições anteriores podemos concluir que

$$\sum_{a \in \mathbb{A}} N_n(s,a) = N_n(s) \quad (2.7)$$

e

$$\sum_{s \in \tau} N_n(s) = n - D(n). \quad (2.8)$$

A definição do estimador BIC e o teorema de convergência deste estimador, apresentados por Csiszar and Talata [2006], são dados a seguir.

Definição 2.5.3 O estimador BIC para uma árvore τ considerando uma amostra x_1^n é dado por:

$$BIC_\tau(x_1^n) = -\log ML_\tau(x_1^n) + \frac{(|\mathbb{A}| - 1)|\tau|}{2} \log n \quad (2.9)$$

onde $ML_\tau(x_1^n)$ é a máxima verossimilhança estimada da árvore τ .

Teorema 2.5.1 *No caso de $d(\tau_0) < \infty$, o estimador BIC para a árvore τ_0 :*

$$\widehat{\tau}_{BIC}(x_1^n) = \operatorname{argmin}_{\tau \in F_1(x_1^n, D(n)) \cap \tau^*} \{BIC_\tau(x_1^n)\} \quad (2.10)$$

onde $D(n) = o(\log(n))$ é o comprimento máximo de uma string, satisfaz

$$\widehat{\tau}_{BIC}(x_1^n) = \tau_0 \quad (2.11)$$

quase certamente quando $n \rightarrow \infty$.

No teorema acima $d(\tau_0)$ é a dimensão da árvore que se deseja estimar τ_0 , τ^* é o conjunto de todas as árvores irredutíveis e $F_1(x_1^n, D(n))$ é a família de todas as possíveis árvores irredutíveis com frequência 1, ou seja, $N_n(s) \geq 1$.

O estimador BIC, no teorema anterior, pode ser reescrito como:

$$\widehat{\tau}_{BIC}(x_1^n) = \operatorname{argmax}_{\tau \in F_1(x_1^n, D(n)) \cap \tau^*} \prod_{s \in \tau} \tilde{\mathbb{P}}_s(x_1^n) \quad (2.12)$$

$$\text{onde } \tilde{\mathbb{P}}_s(x_1^n) = n^{-\frac{|\mathbb{A}| - 1}{2}} \tilde{\mathbb{P}}_{ML,s}(x_1^n).$$

O critério BIC atribui uma pontuação a cada modelo candidato. O melhor modelo é aquele de maior pontuação.

Csiszar and Talata [2006] utilizam $\frac{(|\mathbb{A}| - 1)|\tau|}{2}$ como constante de penalização. No entanto eles mostram que o BIC é consistente para qualquer valor desta constante.

Neste trabalho, utilizamos o algoritmo BIC implementado em R e nomeado algoritmo PCT (Prates, Duarte and Santos, 2012).

Csiszar and Talata [2006] enunciam e demonstram o resultado a seguir. Este resultado nos dá a convergência do vetor das probabilidades condicionais empíricas.

Proposição 2.5.1 *O vetor das probabilidades condicionais empíricas são dadas por:*

$$\widehat{\mathbb{P}}_{\widehat{\tau}}(a|s) = \frac{N_n(s,a)}{N_n(s)}, \quad (2.13)$$

com $a \in \mathbb{A}$ e $s \in \widehat{\tau}$, que converge para as verdadeiras probabilidades condicionais $P(a|s)$, $a \in \mathbb{A}$

e $s \in \tau_0$, quase certamente quando $n \rightarrow \infty$, onde $\widehat{\tau}$ é o estimador BIC.

A máxima verossimilhança $ML_\tau(x_1^n)$ da árvore τ é dada por

$$ML_\tau(x_1^n) = \prod_{s \in \tau} \tilde{\mathbb{P}}_{ML,s}(x_1^n). \quad (2.14)$$

em que

$$\tilde{\mathbb{P}}_{ML,s}(x_1^n) = \begin{cases} \prod_{a \in \mathbb{A}} \left(\frac{N_n(s,a)}{N_n(s)} \right)^{N_n(s,a)} & , \text{se } N_n(s) \geq 1 \\ 1 & , \text{se } N_n(s) = 0. \end{cases} \quad (2.15)$$

Considere uma amostra x_1^n de τ . Seja S_D o conjunto de todos os ramos com tamanho máximo igual a $D = D(n)$. Seja $s \in S_D$ um ramo tal que $N_n(s) \geq 1$. Definindo de forma recursiva começando pelas folhas da PCT completa \mathbb{A}^D , temos:

$$V_s^D(x_1^n) = \begin{cases} \max \left\{ \tilde{\mathbb{P}}_s(x_1^n), \prod_{a \in \mathbb{A}: N_n(as) \geq 1} V_{as}^D(x_1^n) \right\} & , \text{se } 0 \leq l(s) < D \\ \tilde{\mathbb{P}}_s(x_1^n) & , \text{se } l(s) = D. \end{cases} \quad (2.16)$$

e

$$\chi_s^D(x_1^n) = \begin{cases} 1 & , \text{se } 0 \leq l(s) < D, \text{ e } \prod_{a \in \mathbb{A}: N_n(as) \geq 1} V_{as}^D(x_1^n) > \tilde{\mathbb{P}}_s(x_1^n) \\ 0 & , \text{se } 0 \leq l(s) < D, \text{ e } \prod_{a \in \mathbb{A}: N_n(as) \geq 1} V_{as}^D(x_1^n) \leq \tilde{\mathbb{P}}_s(x_1^n) \\ 0 & , \text{se } l(s) = D. \end{cases} \quad (2.17)$$

Com os indicadores $V_s^D(x_1^n)$ e $\chi_s^D(x_1^n)$, para cada $s \in S_D$, $D = D(n)$ a estimação da árvore $\tau_s^D(x_1^n)$ consiste dos ramos $u \geq s$, de modo que:

$$\tau_s^D(x_1^n) = \begin{cases} u \in \{S_D : \chi_u^D(x_1^n) = 0, \chi_v^D(x_1^n) = 1\}, \forall s \leq v \leq u & , \text{se } \chi_s^D(x_1^n) = 1 \\ s & , \text{se } \chi_s^D(x_1^n) = 0. \end{cases} \quad (2.18)$$

Proposição 2.5.2 *A PCT estimada utilizando o estimador $\tilde{\tau}_{\text{BIC}}(x_1^n)$, dado pela equação 2.12, é igual a estimação da PCT começando da raiz ($s = \phi$), ou seja, $\tilde{\tau}_{\text{BIC}}(x_1^n) = \tau_{\phi}^D(x_1^n)$.*

As demonstrações se encontram em Csiszar and Talata [2006].

2.6 Cadeia de Markov Contínua

Nesta seção descrevemos cadeias de Markov com espaço de estados contínuo e tempo discreto.

Definição 2.6.1 *Seja S um conjunto não vazio, Σ uma σ -álgebra de um subconjunto de S . Uma função $P(x,A) : S \times \Sigma \rightarrow [0,1]$ é chamada função de densidade de transição de probabilidade em $(S \times \Sigma)$ se*

1. $\forall x \in S, P(x, \cdot)$ é uma medida de probabilidade em Σ ;
2. $\forall A \in \Sigma, P(\cdot, A)$ é uma função Σ -mensurável.

Definição 2.6.2 *Seja S um conjunto não-vazio. Uma sequência X de S de variáveis aleatórias é uma cadeia de Markov homogênea com função de transição P se, e somente se, para cada $n \geq 0$ e $A \in \Sigma$, onde $\Sigma = \sigma(X_0, X_1, \dots, X_n)$*

$$\begin{aligned} P(X_{n+1} \in A | \sigma(X_0, X_1, \dots, X_n)) &= P(X_{n+1} \in A | \sigma(X_n)) \\ &= P(X_n, A) \end{aligned}$$

com probabilidade 1 onde para qualquer coleção F de variáveis aleatórias $\sigma(F)$ representa uma σ -álgebra gerada por variáveis aleatórias em F e $P(\cdot | \sigma(F))$ é a probabilidade condicional dada a σ -álgebra.

A forma iterativa para P^{n+1} de P com $n > 0$ é dada por

$$P^{(n+1)}(x,A) = \int_A P^{(n)}(y,A)P(x,dy)$$

onde

$$\begin{cases} P^{(0)}(x,A) = 1 & ,\text{se } x \in A \\ P^{(0)}(x,A) = 0 & ,\text{se } x \in A^c \end{cases}$$

e $P^{(1)} = P$.

Também para $n > 0$, $P^{(n)}$ satisfaz as equações de Chapman-Kolmogorov,

$$P^{(m+n)}(x,A) = \int_A P^{(n)}(y,A)P^{(m)}(x,dy).$$

Definição 2.6.3 *Seja X uma cadeia de Markov em (S,Σ) com função de transição $P(x,A)$. Uma medida de probabilidade π em S,Σ é uma distribuição estacionária para a cadeia de Markov se para cada $A \in \Sigma$,*

$$\pi(A) = \int_S P(x,A)\pi(dx) \quad (2.19)$$

2.6.1 Núcleo Estimador

Um estimador kernel é um estimador não-paramétrico para a função de densidade de probabilidade. O estimador kernel e suas propriedades são estudados por vários autores entre eles Rosenblatt [1956], Roussas [1969], Athreya and Atuncar [1998], Dorea [2002] and Campos and Dorea [2005]. Roussas [1969] assume que um processo Markoviano estacionário $\{X_i : i = 1,2,\dots\}$ tem função de densidade inicial $f(x)$, função de densidade conjunta $q(x,y)$ e função de densidade de transição $t(y|x)$. Considerando uma amostra de $n + 1$ elementos, $\{X_i : i = 1,2,\dots,n + 1\}$, os estimadores para as funções de densidade são dados por

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \quad (2.20)$$

$$q_n(x,y) = \frac{1}{nh_n^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)K\left(\frac{y - X_{i+1}}{h_n}\right) \quad (2.21)$$

$$t_n(y|x) = \frac{q_n(x,y)}{f_n(x)} \quad (2.22)$$

se $f_n(x) \neq 0$. Caso $f_n(x) = 0$ então $t_n(y|x) = h_n$,
onde

1. $h_n \rightarrow 0$ quando $n \rightarrow \infty$,
2. $K(x)$ é a função kernel que satisfaz as seguintes condições:

(a) $K(\cdot)$ é limitada, ou seja,

$$|K(x)| \leq M < \infty, \quad (2.23)$$

$$\int |K(x)| dx < \infty \quad (2.24)$$

(b)

$$|x|K(x) \rightarrow 0 \quad (2.25)$$

quando $|x| \rightarrow \infty$

(c)

$$\int K(x) dx = 1 \quad (2.26)$$

Existem várias escolhas para o núcleo K . Neste trabalho optamos pelo núcleo normal. Esta escolha foi pelo fato da normal poder ser padronizada e assim ser mais fácil encontrar as probabilidades.

Para estimar h_n é utilizado o método plug-in apresentado por Chiu [1991].

O teorema a seguir é apresentado por Roussas [1969] e traz a convergência das funções definidas em 2.20 e 2.21.

Teorema 2.6.1 *Considere que K satisfaça as condições 2.23 até 2.26 e $h_n \rightarrow 0$ quando $n \rightarrow \infty$. As variáveis aleatórias $f_n(x)$ e $q_n(x,y)$ são estimativas assintoticamente não viciadas de $f(x)$ e $q(x,y)$, respectivamente, ou seja:*

$$\begin{aligned} E(f_n(x)) &\rightarrow f(x) \text{ quando } n \rightarrow \infty, x \in C(f) \text{ e} \\ E(q_n(x,y)) &\rightarrow q(x,y) \text{ quando } n \rightarrow \infty, (x,y) \in C(q) \end{aligned}$$

onde $C(f)$ e $C(q)$ representam o conjunto de pontos de continuidade de f e q , respectivamente.

Figura 2.3 mostra um exemplo de como a densidade conjunta, definida na Equação 2.21, é calculada considerando uma amostra $X = \{X_1, X_2, \dots, X_{n-1}, X_n\}$. Nesta figura a reta vertical representa o domínio da sequência X , limitada por $\max(X) - h_n$ e $\min(X) + h_n$ para evitar problema de fronteira na estimação. Para quaisquer dois pontos do domínio, x e y , é calculado o valor da densidade conjunta neste dois pontos, $q(x, y)$. Considerando o ponto x efetuamos a subtração de cada ponto da amostra, ou seja, $\{x - X_1, x - X_2, \dots, x - X_n\}$. Analogamente, considerando o ponto y , temos $\{y - X_1, y - X_2, \dots, y - X_n\}$. Assim temos que

$$q_n(x, y) = \frac{1}{nh_n^2} \left\{ K\left(\frac{x - X_1}{h_n}\right) K\left(\frac{y - X_2}{h_n}\right) + K\left(\frac{x - X_2}{h_n}\right) K\left(\frac{y - X_3}{h_n}\right) + \dots + K\left(\frac{x - X_{n-1}}{h_n}\right) K\left(\frac{y - X_n}{h_n}\right) \right\}$$

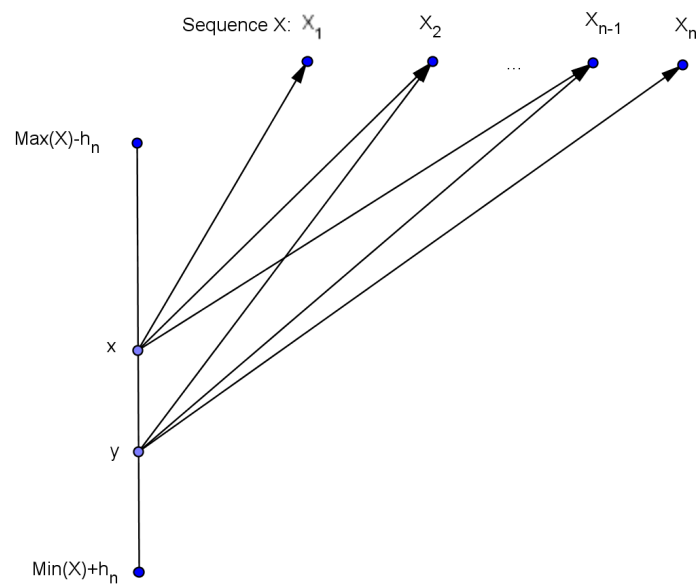


Figura 2.3: Densidade conjunta marginal, $q_n(x, y)$, da cadeia X para ela mesma considerando os pontos x e y .

Revisão Bibliográfica

Para um modelo convencional de Cadeia de Markov de ordem l e m estados, o número de parâmetros é da ordem de $(m - 1)m^l$. Esta quantidade de parâmetros é um ponto desanimador de se trabalhar com estes modelos. Com o objetivo de reduzir o número de parâmetros em modelos de Markov de ordem superior, Raftery [1985] apresenta um modelo de Cadeia de Markov de ordem superior l . Nesse modelo, o número de parâmetros é da ordem de $l + m^2$. O autor considera apenas uma sequência de dados.

Considerando uma sequência $\{X_t : t \in N\}$ onde N é um conjunto de inteiros positivos assumindo valores em $\{1, 2, \dots, m\}$, o modelo proposto por Raftery [1985] é dado por

$$P[X_t = j_0 | X_{t-1} = j_1, \dots, X_{t-l} = j_l] = \sum_{i=1}^l \lambda_i Q_{j_0 j_i} \quad (3.1)$$

onde

1. $\sum_{i=1}^l \lambda_i = 1$;
2. $0 \leq \sum_{i=1}^l \lambda_i q_{j k_i} \leq 1$

para $j, k_1, \dots, k_l = \{1, \dots, m\}$.

A matriz Q na equação 3.1 é quadrada e não negativa e a soma de suas colunas é igual a 1.

Ching et al. [2002] apresenta um modelo com s sequências o qual descreve a distribuição de probabilidade marginal de uma dada sequência, no tempo $n + 1$, como combinação linear de todas as outras distribuições de probabilidade das outras sequências, no tempo n . A ordem do modelo é fixa, cada sequência é composta por m estados possíveis e as cadeias têm espaço de estados discreto e tempo discreto.

A seguir descreveremos o modelo proposto pelos autores.

Considere um conjunto de s seqüências $\{X_1, X_2, \dots, X_s\}$. Sejam $Q^{(jk)}$ as matrizes de probabilidade de transição dos estados da k -ésima seqüência para os estados da j -ésima seqüência onde $j = \{1, 2, \dots, s\}$ e $k = \{1, 2, \dots, s\}$. Seja μ_n^k a distribuição de probabilidade marginal de estados da k -ésima seqüência no passo n . O modelo proposto é

$$\mu_{n+1}^{(j)} = \sum_{k=1}^s \lambda_{jk} Q^{(jk)} \mu_n^{(k)} \quad (3.2)$$

onde

1. $\lambda_{jk} \geq 0$;
2. $1 \leq j, k \leq s$ e
3. $\sum_{k=1}^s \lambda_{jk} = 1$

para $j = \{1, 2, \dots, s\}$.

Neste caso $\mu_0^{(j)}$ é a distribuição de probabilidade inicial da j -ésima seqüência. $Q^{(jk)}(a_r|a_s)$ com $a_r, a_s \in \mathbb{A}$ é definida por

$$Q^{(jk)}(a_r|a_s) = P(X_j(t+1) = a_r | X_k(t) = a_s).$$

Detalhes da estimação das matrizes $Q^{(jk)}$ poderão ser vistos na seção 4.2.

Através de programação linear, os autores encontram os λ 's que satisfazem à seguinte equação

$$\min_{\lambda} \max_i \left\| \left[\sum_{k=1}^s \lambda_{jk} Q^{(jk)} \mu_n^{(k)} - \mu_n^{(j)} \right]_i \right\|, \quad (3.3)$$

sujeito a $\sum_{k=1}^s \lambda_{jk} = 1$ e $\lambda_{jk} \geq 0$.

Ching et al. [2002] apresentam uma proposição em que as matrizes $Q^* = \lambda_{jk} Q^{(jk)}$ têm autovalor máximo igual a 1. Isto é garantido pelo teorema de Perron-Frobenius (Ver capítulo 2) uma vez que esta matriz são quadradas com entradas positivas, e é fundamental no procedimento de estimação dos parâmetros.

Proposição 3.0.1 Se $\lambda_{jk} > 0$ para $1 \leq j, k \leq s$, então a matriz $Q^* = \lambda_{jk}Q^{(jk)}$ tem autovalor igual a 1 e os autovalores de Q^* tem módulo menor ou igual a 1.

Demonstração: Ver Ching et al. [2002].

Ching et al. [2004] apresentam uma generalização do modelo apresentado por Raftery [1985], pois considera que as matrizes Q_i podem ser diferentes, reescrevendo o modelo 3.1 da seguinte maneira

$$\mu_{t+n+1} = \sum_{i=1}^n \lambda_i Q_i \mu_{t+n+1-i} \quad (3.4)$$

onde

1. $\lambda_i \geq 0$ de modo que $\sum_{i=1}^n \lambda_i = 1$;
2. Q_i é a matriz de probabilidade de transição em i passos, ou seja,

$$Q_i(a_r|a_s) = P(X(t+i) = a_r | X(t) = a_s);$$

3. $\mu_{t+n+1-i}$ é a distribuição de probabilidade de estado no tempo $t+n+1-i$.

Quando $Q_1 = Q_2 = \dots = Q_n$ então o modelo 3.4 será exatamente o modelo apresentado por Raftery [1985].

A suposição de que os λ 's sejam não-negativos é fator importante para garantir que $\mu_{t+n+1-i}$ seja distribuição de probabilidade, e isso não foi suposto em Raftery [1985].

Os autores também utilizam programação linear para a estimação dos λ 's, resolvendo o problema

$$\min_{\lambda} \max_k \left\| \left[\sum_{i=1}^n \lambda_i Q_i \mu_{t+n+1-i} - \mu_{t+n+1} \right] \right\|_k, \quad (3.5)$$

sujeito a $\sum_{i=1}^n \lambda_i = 1$ e $\lambda_i \geq 0$ para todo $i = \{1, 2, \dots, n\}$.

Tomando como referência o trabalho de Ching et al. [2002], Ching et al. [2008] apresentam um modelo para estimar os pesos da combinação linear das medidas de distribuições marginais de duas ou mais sequências. Neste modelo os autores trabalham com passo $h \geq 1$. Cada sequência tem m estados possíveis. Os autores utilizam cadeias com espaço de estados discreto e tempo discreto.

Dado um modelo convencional de Cadeia de Markov Multivariado com s sequências $\{X_1, X_2, \dots, X_s\}$, m estados e ordem n o número de parâmetros a serem estimados é da ordem de $O(m^{ns})$ e no modelo proposto por Ching et al. [2008] o número de parâmetros a serem estimados é da ordem de $O(ns^2m^2)$. O modelo definido em Ching et al. [2008] é dado por

$$\mu_{r+1}^{(j)} = \sum_{k=1}^s \sum_{h=1}^n \lambda_{jk}^{(h)} P_h^{(jk)} \mu_{r-h+1}^{(k)} \quad (3.6)$$

para $j = \{1, 2, \dots, s\}$, $r = \{n-1, n, \dots\}$ com distribuições iniciais $\mu_0^{(k)}, \mu_1^{(k)}, \dots, \mu_{n-1}^{(k)}$ com $k = \{1, 2, \dots, s\}$ onde

1. $\lambda_{jk}^{(h)} \geq 0$;
2. $1 \leq j, k \leq s$;
3. $1 \leq h \leq n$ e
4. $\sum_{k=1}^s \sum_{h=1}^n \lambda_{jk}^{(h)} = 1$

para $j = \{1, 2, \dots, s\}$.

Neste modelo assumem que a distribuição de probabilidade marginal da j -ésima sequência no tempo $t = r+1$ depende das distribuições marginais de todas as sequências (incluindo ela mesma) no tempo $t = \{r, r-1, r-2, \dots, r-n+1\}$. Definem $P_h^{(ij)}$ como

$$P_h^{(ij)}(y|x) = P(X_i(t+h) = y | X_j(t) = x);$$

onde $i = \{1, 2, \dots, s\}$, $j = \{1, 2, \dots, s\}$ e $h = \{1, 2, \dots, n\}$ e $\mu_{r+1}^{(j)}$ é a distribuição marginal da sequência j no tempo $r + 1$.

Para estimar $P_h^{(ij)}$ inicialmente encontram-se as matrizes de frequências de transição $F_h^{(ij)}$ onde $i = \{1, 2, \dots, s\}$, $j = \{1, 2, \dots, s\}$ e $h = \{1, 2, \dots, n\}$. Neste caso $F_h^{(ii)}$ representa a matriz de transição da i -ésima sequência para ela mesma em h passos e $F_h^{(ij)}$ representa a matriz de transição da j -ésima sequência para a i -ésima sequência em h passos. De posse das matrizes de transição os autores encontram $\widehat{P}_h^{(ij)}$.

Os autores utilizam programação linear para estimar os λ 's através da expressão

$$\min_{\lambda_{ij}} \max_i \left\| \left[\sum_{k=1}^s \sum_{h=1}^n \lambda_{jk}^{(h)} P_h^{(jk)} \mu_n^{(k)} - \mu_n^{(j)} \right]_i \right\|, \quad (3.7)$$

sujeito a $\sum_{k=1}^s \sum_{h=1}^n \lambda_{jk}^{(h)} = 1$ e $\lambda_{jk}^{(h)} \geq 0$ com $h, k = \{1, 2, \dots, s\}$.

Rissanen [1983], introduz o conceito de PCT e apresenta um algoritmo para estimar a PCT dada uma sequência de dados. O autor o denomina de algoritmo de compressão de dados universal, capaz de codificar dados conhecendo as probabilidades. O algoritmo estima os contextos gerados pelos símbolos da sequência e que são relevantes para avaliar o próximo símbolo da sequência. Estes contextos são únicos para cada símbolo e têm tamanhos variáveis (ordem). Em seu algoritmo, o autor considera cadeia binária.

Bühlmann and Wyner [1999] propõem uma nova versão do algoritmo de árvore de contexto proposto por Rissanen [1983]. Este algoritmo apresenta novas propriedades assintóticas quando estimamos uma PCT. Eles consideram uma árvore de contexto máxima e fazem podas de baixo para cima verificando se cada elemento (nó terminal) é irrelevante. Apresentam uma nova forma de fazer bootstrap utilizando PCT para séries temporais categóricas estacionárias. Aplicam o algoritmo proposto para estimar sequências de DNA do genoma *Drosophila*.

Ron et al. [1996] propõem um algoritmo baseado no PSA (automata probabilístico de sufixo) que identifica estruturas de uma sequência de dados e utiliza estas estruturas para construir modelos. O algoritmo considera cadeia de Markov de ordem variável l . O algoritmo proposto estima a árvore de cima para baixo, ou seja, iniciam com o nó raiz e seus filhos. Gradativamente estes filhos geram outros filhos e assim sucessivamente

até uma profundidade máxima L e no máximo $L \times |E|$ filhos, onde L é a ordem da cadeia e $|E|$ é o tamanho do conjunto de estados.

O algoritmo utiliza a divergência de Kullbach-Leibler D_{KL} para medir a distância entre duas distribuições de probabilidade. Sejam M uma PSA e τ uma PCT. Considere P_M^N a distribuição de probabilidade da PSA de comprimento N e P_τ^N a distribuição de probabilidade da PCT de comprimento N . Então

$$D_{KL}[P_M^N][P_\tau^N] = \sum_{r \in \Sigma^N} P_M^N(r) \log \frac{P_M^N(r)}{P_\tau^N(r)} \quad (3.8)$$

onde $N > 0$ e $P_M^N(r)$ é a distribuição de probabilidade da PSA de comprimento N da string r .

O algoritmo baseia-se na idéia de que dado um L - PSA, $M = (E, \Sigma, T, \gamma, \pi)$, pode-se encontrar uma PCT τ_M equivalente de profundidade máxima L e no máximo $L \cdot |E|$ folhas. Onde

1. $|E|$ é o tamanho do conjunto de estados;
2. Σ é o alfabeto;
3. T é a função de transição;
4. γ é a função de probabilidade do próximo simbolo e
5. π é a distribuição de probabilidade inicial.

Os autores utilizam uma amostra para construir a função de probabilidade empirica $\tilde{P}(\cdot)$ e a partir desta amostra encontrar a árvore $\bar{\tau}$, que com grande probabilidade será uma sub-árvore de τ . Então constroem $\hat{\tau}$ baseado em $\bar{\tau}$ e $\tilde{P}(\cdot)$.

Duas aplicações são apresentadas: a primeira para identificar a estrutura de um texto em inglês e depois utilizam esta estrutura para corrigir um texto. Neste caso utilizaram a bíblia como referência. O alfabeto era formado pelas letras em inglês e os caracteres em branco. Para a fase de aprendizagem do algoritmo os autores retiraram o livro da Genesis e rodaram o algoritmo em todo o restante da bíblia. Nesta fase foi gerado uma PCT com menos do que 3000 nós. Esta PCT foi transformada em uma PSA. Depois,

com a *PSA* definida, utilizaram um trecho do livro da Genesis, que foi corrompido, para corrigí-lo. Na outra aplicação propõem um modelo estocástico simples para base de DNA. O alfabeto de DNA é composto por quatro tipos de nucleótidos: A, C, T, G. Construem dois *PSA*'s diferentes, um para as regiões codificantes, e um para as regiões intergênicas. Os modelos foram construídos com base em 250 fitas de DNA diferentes de cada tipo cujos comprimentos variam de 20 a vários milhares de bases. O *PSA* que modela as regiões codificantes tem 65 estados e o *PSA* que modela as regiões intergênicas tem 81 estados.

Os autores fizeram uma comparação entre a performance utilizando o *PSA* e modelando usando cadeia de Markov de ordem 0,1,2 e 3. A performance foi medida utilizando a log-verossimilhança negativa obtida pela variação dos modelos. Observaram que utilizando *PSA* a log-verossimilhança negativa foi a menor em todos os casos e para cadeias de Markov de ordem 2 ou mais o *PSA* tem menos estados.

Bejerano [2001] apresenta um método para modelar famílias de proteínas usando PCT. O método baseia-se na identificação de padrões de significado biológico em um conjunto de sequência de proteínas. Segundo os autores, o método é muito mais rápido do que utilizar cadeia de Markov oculta. O conhecimento de padrões novos auxiliará no conhecimento de novas famílias de proteínas. "O modelo baseia-se na identificação de aparências significativas de segmentos curtos entre muitas das sequências de proteína de entrada, independentemente das posições relativas destes segmentos dentro das proteínas diferentes". Estes segmentos têm a propriedade de memória curta, ou seja, para estimar o próximo símbolo x_i após o segmento basta observar os L símbolos anteriores. Para os autores, a modelagem utilizando cadeia de Markov seria inconveniente pelo fato do crescimento exponencial do número de parâmetros de acordo com a ordem do modelo.

Utilizando a idéia de PCT, em uma forma modificada, Leonardi [2006] apresenta uma variação do algoritmo da PCT denominado árvores esparsas probabilística de sufixos (SPST) na modelagem de cadeia de Markov esparsa (SMC). Este algoritmo identifica as equivalências entre os contextos de uma PCT. A identificação destas equivalências pode melhorar a taxa de classificação da PCT. A autora utiliza este algoritmo

para a modelagem de sequências esparsas de proteínas.

Csiszar and Talata [2006] utilizando os estimadores via Bayesian Information Criterion (BIC) e KT (critério de Krichevsky-Trofimov) encontram estimadores fortemente consistentes para árvores de contexto. O estimador BIC é descrito na seção 2.5 e o estimador KT dado uma amostra $x_1^n = \{x_1, x_2, \dots, x_n\}$ e $d(\tau_0) < \infty$ é definido por

$$\widehat{\tau}_{KT}(x_1^n) = \operatorname{argmin}_{\tau \in F_1(x_1^n, D(n)) \cap \tau} KT_\tau(x_1^n) \quad (3.9)$$

onde $KT_\tau(x_1^n)$ é o critério KT para uma árvore factível dado por

$$KT_\tau(x_1^n) = -\log P_{KT,\tau}(x_1^n) \quad (3.10)$$

onde $P_{KT,\tau}(x_1^n)$ é a probabilidade KT da amostra x_1^n correspondente a τ .

Estas árvores podem não ser completas e podem ter comprimento infinito. Utilizam otimização de árvores de contexto hipoteticamente finita.

Em Duarte et al. [2006], os autores apresentam uma versão simplificada do teorema de consistência do algoritmo de contexto, utilizado para criar os contextos gerados pelos símbolos da sequência, apresentado por Rissanen [1983] para calcular o comprimento do contexto necessário para estimar o próximo símbolo da cadeia de uma amostra finita. Os autores consideram cadeias de ordem infinita e o tamanho do sufixo do passado utilizado para estimar o próximo símbolo é finito. Assumem que a probabilidade de transição de cada sufixo tem decaimento exponencial.

O algoritmo apresentado permite aproximar uma cadeia de ordem infinita por uma cadeia de Markov gerada por uma PCT com ordem crescente. Existe uma consistência fraca para o algoritmo que calcula a quantidade de contexto necessário para estimar o próximo símbolo.

Duarte et al. [2006] também utilizam a estatística de razão de verossimilhança para verificar se uma amostra é consistente com uma PCT. Também utilizam a definição de contexto candidato com aproximação de Markov canônica de ordem k para estender o resultado apresentado em Rissanen [1983] para PCT ilimitada. Mostram que existe um acoplamento entre X_i e X_i^k onde X_i é uma cadeia consistente com uma PCT com taxa

de continuidade somável e X_t^k é uma aproximação de Markov canônica de ordem k .

Em Galves et al. [2012], os autores utilizam árvore de contexto para recuperar ritmos linguísticos de textos escritos. Um fato curioso foi que o primeiro estudo sobre classe rítmica foi em 1996, quando pesquisadores concluíram que bebês recém nascidos eram capazes de identificar classes rítmicas.

Neste estudo foram utilizados textos digitais do Português Brasileiro e Português Europeu. A amostra é composta de 80 artigos selecionados aleatoriamente das edições de 1994 e 1995 da Folha de São Paulo, diário do Brasil e Portugal. O objetivo é saber se é possível encontrar características rítmicas diferentes entre Português Brasileiro e Português Europeu. Utilizam cadeias de Markov de comprimento variável. O Critério de Smallest Maximizer (SMC) é introduzido e Galves et al. [2012] provam que este critério é consistente. O critério SMC através de seu algoritmo seleciona, entre as árvores campeãs, que são aquelas que maximizam a verossimilhança da amostra para cada valor de determinada constante, a menor árvore. Para cada valor de uma dada constante c obtém-se uma árvore de contexto estimada pelo método BIC. Assim, o algoritmo SMC seleciona as árvores estimadas pelo BIC. O critério SMC não depende de nenhuma constante, enquanto que o BIC depende de determinada constante variando no intervalo $[0, \infty)$.

As cadeia de Markov podem ter espaço de estados discreto e contínuo. Até aqui falamos apenas em cadeias de espaço discreto. No entanto, como proposta deste trabalho, pretendemos também trabalhar com cadeias de espaço contínuo. Quando temos cadeias de Markov com espaço de estados discreto podemos encontrar a matriz de probabilidade de transição de estados. Esta matriz, em geral, é quadrada e positiva. Já quando temos cadeia de Markov com espaço contínuo encontramos função de densidade. Vamos então, inicialmente, fazer um pequeno apanhado do que encontramos na literatura sobre cadeias de Markov com espaço de estados contínuo.

Em Rosenblatt [1956] o autor apresenta um estimador não-paramétrico para a função de densidade de probabilidade univariada. Considera amostras aleatórias iid (independentes e identicamente distribuídos) com função de densidade contínua f . Apresenta uma estimativa S , não viciada, para f . Assume-se que a função definida por

S é Borel mensurável, que $S \geq 0$ desde que $f \geq 0$ e que seja simétrica de x_1, \dots, x_n .

Segundo o autor, S é definida como

$$S(y; X_1, \dots, X_n) = f_n(y) = \frac{F_n(y+h) - F_n(y-h)}{2h} \quad (3.11)$$

onde $F_n(y)$ é a função de distribuição amostral e $h = h_n$ é uma função da amostra de modo que $h \rightarrow 0$ quando $n \rightarrow \infty$. O autor analisa o comportamento assintótico de S utilizando o erro quadrático médio.

Uma classe de estimadores para f_n é dada por:

$$\widehat{f}_n(y) = \int_{-\infty}^{\infty} w_n(y-u) dF_n(u) = \frac{1}{n} \sum_{j=1}^n w_n(y - X_j) \quad (3.12)$$

onde w_n é uma função não-negativa satisfazendo

$$\int_{-\infty}^{\infty} w_n(u) d(u) = 1. \quad (3.13)$$

Estes estimadores são também funções de densidade.

Roussas [1969] apresenta estimadores não-paramétricos para processos Markovianos estacionários. Neste trabalho o autor supõe que o processo satisfaz a condição de Doeblin, é ergódico e cíclico.

A densidade de transição de um processo de Markov estacionário é dado por $t = \frac{q}{p}$, onde p é a densidade inicial e q é a densidade conjunta com respeito a medida de Lebesgue.

Considerando $(\varepsilon_m, \mathfrak{J}^{(m)})$ o espaço Euclidiano m -dimensional com o correspondente σ -álgebra de Borel e $(\mathfrak{K}, \mathfrak{J})$ a reta real, define-se a função $K : (\varepsilon_m, \mathfrak{J}^{(m)}) \rightarrow (\mathfrak{K}, \mathfrak{J})$ mensuráveis tais que

1. $|K_{(z)}| \leq M_1 (< \infty)$, $z \in \varepsilon$;
2. $\|z\|^m |K(z)| \rightarrow 0$; $z \rightarrow \infty$;
3. $\int |K(z)| dz < \infty$.

Define-se que uma função K é mensurável a Borel \mathfrak{J} se

$$K^{-1}(B) = \{(x_1, \dots, x_n) \in \mathfrak{X}^n : K(x_1, \dots, x_n) \in B\} \in \mathfrak{Y}^n, \forall B \in \mathfrak{Y}. \quad (3.14)$$

Baseando-se na função K o autor apresenta estimadores assintóticos não viciados de p e q dados por

$$\widehat{p}_n(x) = \frac{1}{nh_1(n)} \sum_{j=1}^n K_1\left(\frac{x - X_j}{h_1(n)}\right) \quad (3.15)$$

com $x \in \varepsilon_1$ e

$$\widehat{q}_n(y) = \frac{1}{nh_2^2(n)} \sum_{j=1}^n K_2\left(\frac{y - Y_j}{h_2(n)}\right), \quad (3.16)$$

com $y \in \varepsilon_2$, onde

1. $Y_j = (X_j, X_{j+1})$;
2. $j = \{1, \dots, n\}$;
3. $h_1(n) \rightarrow 0$ e
4. $h_2(n) \rightarrow 0$ quando $n \rightarrow \infty$.

Neste caso assume-se que

1. h_1 e h_2 são constantes positivas,
2. $K_1 : (\varepsilon_1, \mathfrak{Y}^{(1)}) \rightarrow (\mathfrak{X}, \mathfrak{Y})$ e
3. $K_2 : (\varepsilon_2, \mathfrak{Y}^{(2)}) \rightarrow (\mathfrak{X}, \mathfrak{Y})$.

Nos casos acima é imposta a seguinte condição

$$\int (K(z)) dz = 1.$$

O autor mostra que os estimadores p_n e q_n são consistentes em média quadrática. Mostra que t_n converge em probabilidade para t quando $n \rightarrow \infty$.

Athreya and Atuncar [1998] fazem um estudo sobre as propriedades dos estimadores, tipo Kernel, para as funções de densidade de transição e densidade estacionária para uma cadeia de Markov homogênea recorrente com espaço contínuo. Os autores utilizam técnicas de processos regenerativos que possibilitam reduzir hipóteses sobre a cadeia de Markov e a hipótese de recorrência de Harris.

Um processo é chamado regenerativo se, dado um tempo aleatório N , temos que a evolução do processo antes de N e depois de N são independentes e o processo depois de N tem distribuição independente dos termos anteriores a N .

A técnica é utilizada pois não utiliza as hipóteses de recorrência de Doeblin, estacionariedade e decaimento de mistura. Os resultados podem ser aplicados para qualquer distribuição inicial. Um processo autoregressivo não satisfaz a recorrência de Doeblin. Já a recorrência de Doeblin implica que a $Cov(X_k, X_{k+r}) \rightarrow 0$ quando $r \rightarrow \infty$.

Os autores trabalham com núcleos estimadores, considerando um tempo n , para as funções de densidade de transição, $t(y|x)$, função de densidade conjunta, $q(x,y)$, e a função de densidade da distribuição de probabilidade estacionária com respeito à medida de Lebesgue, $f(x)$. Estes estimadores são os mesmos apresentados por Roussas [1969], onde $h_1(n) = h_2(n) = \delta_n$. Athreya and Atuncar [1998] assumem que $t_n(y|x) = \delta_{xy}$ caso $f_n(x) = 0$.

Apresentam propriedades de consistência para as funções f_n , q_n e t_n . Mostram que $f_n \rightarrow f$, $q_n \rightarrow q$ e $t_n \rightarrow t$ em probabilidade.

As condições de regularidade de K são as mesmas apresentadas no capítulo 2.

Os autores apresentam ainda propriedades de normalidade assintóticas das funções acima descritas.

Dorea [2002] propõe um método de encontrar classe de estimadores fortemente consistentes para a função de densidade de transição $t(x,y)$ de uma cadeia de Markov que possui densidade de transição com respeito a uma medida σ -finita ν . O estimador de $t_n(x,y)$ é dado por

$$\widehat{t}_n(x, y) = \frac{\sum_{k=1}^n W(h, x, X_k) W(h, y, X_{k+1})}{\sum_{j=1}^n W(h, x, X_j)} \quad (3.17)$$

onde $W(h, x, \cdot)$ são famílias de funções pesos.

O estimador para a função de densidade, $f_n(x)$, é

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{k=1}^n W(h, x, X_k), h = h_n \quad (3.18)$$

Quando ν refere-se a medida de Lebesgue, então

$$W(h, x, y) = \frac{1}{h} K\left(\frac{x - y}{h}\right) \quad (3.19)$$

onde K satisfaz as condições de regularidade apresentadas no capítulo 2.

A consistência do estimador $\widehat{t}_n(x, y)$ é demonstrada.

Campos and Dorea [2005] trabalham a estimação da função de densidade estacionária para cadeias de Markov com espaço de estados gerais. O que difere de Dorea [2002] é que neste artigo as autoras trabalham com o caso iid.

Inicialmente supõem cadeias estritamente estacionárias e φ -mixing.

Uma cadeia $\{X_n\}_{n \geq 0}$ satisfaz a condição de φ -mixing se $\forall A \in \mathbb{F}_0^k, \forall B \in \mathbb{F}_{k+n}^\infty, k \geq 0$ e $n \geq 1$

$$|P(A \cap B) - P(A)P(B)| < \varphi(n)P(A) \quad (3.20)$$

quando $\varphi(n) \rightarrow 0$ e $n \rightarrow \infty$, onde $\mathbb{F}_l^{l+m} = \sigma(X_l, \dots, X_{l+m})$.

As autoras provam a consistência, consistência média quadrática e a forte consistência da f_n . Na demonstração da consistência forte não é necessário supor que a cadeia seja estritamente estacionária. Demonstram também a normalidade assintótica do estimador de f_n .

Substituem as suposições de estacionariedade e φ -mixing por ergodicidade geométrica. Mostram que o estimador para f_n é consistente e provam a normalidade assintótica da f_n .

Chiu [1991] propõe uma metodologia para encontrar a janela ótima ao estimar funções de densidade. Como medida de performance para $\widehat{f}_h(x)$ utiliza o *erro quadrático integrado* (ISE) e *erro quadrático integrado médio* (MISE). O autor define o ISE como

$$ISE(h) = \int [\widehat{f}_h(x) - f(x)]^2 dx \quad (3.21)$$

e o MISE é dado por

$$MISE(h) = E[\int [\widehat{f}_h(x) - f(x)]]. \quad (3.22)$$

Para o cálculo do ISE quanto do MISE é necessário o conhecimento de f o qual não conhecemos. O autor apresenta um estimador não viciado para o MISE, conhecido como *least squares cross-validation*, $CV_n(\beta)$, onde β é a janela e n é o tamanho da amostra

$$CV_n(\beta) = \int \widehat{f}_\beta^2(x) dx - 2n^{-1} \sum_{j=1}^n \widehat{f}_{\beta,j}(X_j). \quad (3.23)$$

onde $\widehat{f}_{\beta,j}$ é a estimativa da função de densidade com a j -ésima observação excluída da amostra.

Para estimar a janela ótima o autor utiliza um estimador da função característica da amostra, $\tilde{\phi}(\lambda)$. Baseando no valor de Λ que é o primeiro valor de λ tal que $|\tilde{\phi}(\lambda)|^2 \leq \frac{c}{n}$ para $c > 1$ a janela ótima é estimada utilizando o método plug in. O método plug-in utiliza uma estimativa \widehat{G} de G , que é desconhecida, para encontrar a janela ótima. Neste caso a janela ótima é dada por

$$\widehat{h}_{op} = k_2^{-2/5} [\int K(t)^2 dt]^{1/5} [\widehat{G}]^{-1/5} n^{-1/5} \quad (3.24)$$

onde

$$\widehat{G} = \frac{1}{\pi} \int_0^\Lambda \lambda^4 [|\tilde{\phi}(\lambda)|^2 - \frac{1}{n}] d\lambda \quad (3.25)$$

Na equação 3.24, k_2 é a variância do núcleo escolhido, ou seja, $[\int t^2 K(t) dt]$ onde $K(t)$ é o núcleo escolhido.

Os casos de estimadores para cadeias não estacionárias são tratados em Phillips and Park [1998], Bertuccioli and How [2008] e Cho et al. [2008]. No caso de processo autorregressivo de primeira ordem não estacionário Phillips and Park [1998] desenvolvem uma teoria assintótica para o estimador tipo kernel definido na função 3.26. Trabalham com séries temporais bivariada (y_t, u_t) definida por

$$y_t = m(y_{t-1}) + u_t$$

onde $m(y_{t-1}) = y_{t-1}$ para $t = \{1, 2, \dots, n\}$ e $u_t = N(0, \sigma^2)$ iid. A função média condicional é dada por

$$m(x) = E(y_t | y_{t-1} = x).$$

O objetivo é estimar a densidade

$$\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{t=1}^n K\left(\frac{x - y_{t-1}}{h_n}\right) \quad (3.26)$$

Eles mostram que $\sqrt{n}\widehat{f}_n(x)$ tem um significado de estimativa da densidade mesmo no caso de não-estacionário. utilizam Martingale e movimento Browniano.

Este levantamento bibliográfico será útil para entendermos o desenvolvimento do conhecimento sobre os temas que serão abordados neste trabalho.

Contribuição Local em Sequências de PCT para Espaço de Estados Discreto

Ching et al. [2002] apresenta um modelo onde r sequências discretas são modeladas utilizando cadeias de Markov. Para estas sequências, a distribuição de probabilidade marginal de determinada sequência, no tempo $n + 1$, é uma combinação linear das demais, no tempo n . Nossa contribuição, apresentada aqui, é utilizar PCT em vez de cadeias de Markov. A vantagem de utilizar a PCT é que esse modelo permite trabalhar com ordem variada, ao contrário da cadeia de Markov que tem ordem fixa. Assim cada contexto da PCT pode ter comprimento variado o que permite uma melhor estimativa do próximo símbolo da cadeia. Para cada PCT estimada encontramos o PSA equivalente. Apresentamos então um modelo em função destes PSA's.

4.1 Emulação de uma Árvore Probabilística de Contexto por um Automato Probabilístico de Sufixo

Ron et al. [1996] mostraram que para cada PSA existe uma PCT equivalente que não é muito maior. Seu objetivo foi o de estimar uma PCT por meio de uma PSA uma vez que este problema era de interesse naquele momento. O objetivo aqui é o oposto. Utilizando uma forma consistente para estimar a PCT, por exemplo baseado no BIC, o objetivo é encontrar o espaço de estados E de uma PSA e como consequência uma matriz de probabilidades de transição quadrada.

A equivalência entre uma PSA e uma PCT apresentada em Ron et al. [1996], cujos

resultados são demonstrados, é dado pelos seguintes teoremas.

Teorema 4.1.1 *Para cada L – PSA = $(E, \mathbb{A}, \alpha, \gamma, \pi)$ existe uma PCT equivalente τ de profundidade máxima L e no máximo $L \cdot |E|$ nós.*

Teorema 4.1.2 *Para cada τ de profundidade L sobre \mathbb{A} , existe uma PFA equivalente M_τ , com no máximo $L \cdot |\alpha|$ estados. Além disso, se para cada sequência s nomeando um nó na árvore τ tivermos*

$$P_\tau(s) = \sum_{a \in \mathbb{A}} P_\tau(as), \quad (4.1)$$

então existe uma PSA equivalente para cada sequência s nomeando um nó na árvore τ .

Estes resultados são para as árvores completas, ou seja, cada nó interno tem $|\mathbb{A}|$ ramos, mas se a árvore estimada não é completa então ela poderá ser completada com ramos cujas probabilidades condicionais são iguais a zero. Assim, na figura 4.1 parte (a) temos a PCT o qual denominamos incompleta pois não tem todos os ramos em alguns nós. Neste caso podemos completar, por exemplo, os contextos 000 e 100 com probabilidade zero. O mesmo podem ser feitos com os contextos 001, 101, 011 e 111.

Uma vez que um modelo PCT é ajustado aos dados, podemos encontrar o PSA equivalente que é uma cadeia de Markov com alfabeto definido pelo conjunto dos contextos da PCT. Uma rotina chamada PCT-PSA foi implementada em R para realizar essa transformação. Como resultado, uma matriz de transição quadrada com as probabilidades de transição de uma PCT pode ser obtida. A Figura 4.1 mostra a PSA equivalente de uma PCT. A matriz de probabilidades de transição da PSA é dada na Tabela 4.1. Os novos contextos e as probabilidades de transição, gerados a partir dos estados "0" e "1", são dados na tabela 2.2 entre colchetes.

4.2 Estimação da Matriz de Transição Cruzada

Na metodologia proposta neste trabalho é necessário estimar as *Matrizes de Transição Cruzadas*.

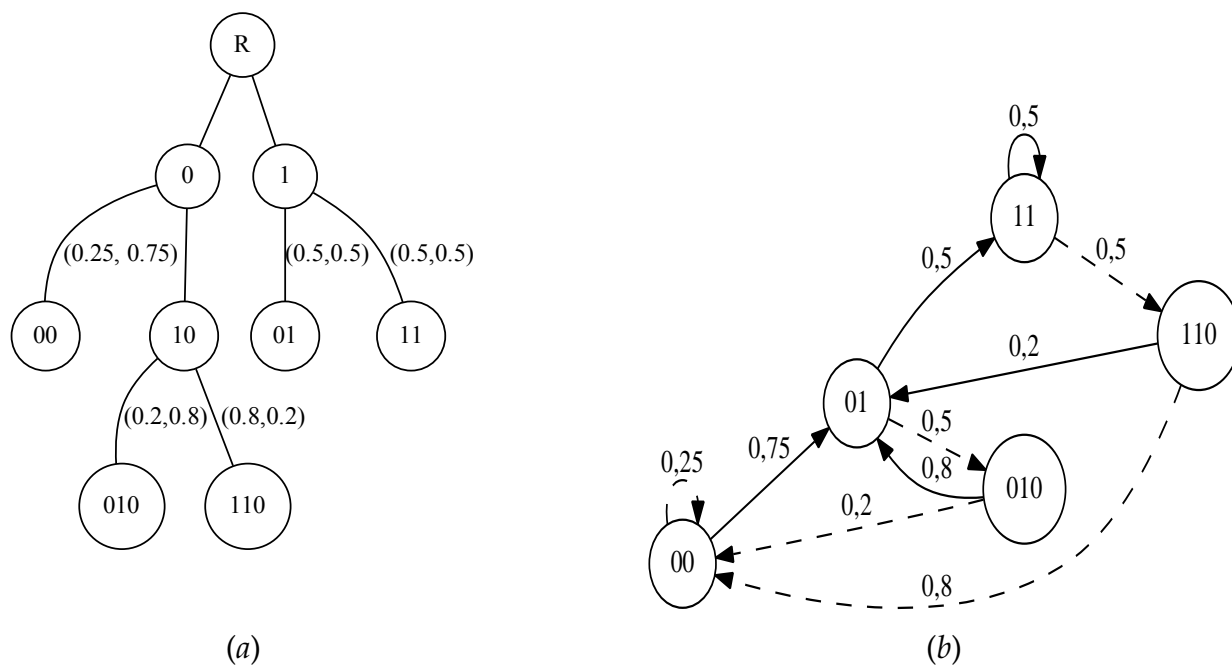


Figura 4.1: PCT (a) e seu equivalente PSA (b)

Tabela 4.1: Matriz de probabilidade de transição quadrada da PSA da Figura 4.1.

E	00	01	11	110	010
00	0,25	0	0	0,8	0,2
01	0,75	0,5	0	0,2	0,8
11	0	0	0,5	0	0
110	0	0	0,5	0	0
010	0	0,5	0	0	0

Para o caso de 2 cadeias de Markov de ordem 1, X e Y , sobre \mathbb{A} , seja $Q^{(xy)}(a_i|a_j)$ com $a_i, a_j \in \mathbb{A}$, a matriz de probabilidade de transição do estado a_j na sequência Y no tempo t para o estado a_i na sequência X no tempo $t + 1$. Analogamente $Q^{(yx)}(a_j|a_i)$ com $a_i, a_j \in \mathbb{A}$ é a matriz de probabilidade de transição do estado a_i na sequência X no tempo t para o estado a_j na sequência Y no tempo $t + 1$. Então

$$Q^{(xy)}(a_i|a_j) = P\{X_{t+1} = a_i | Y_t = a_j\} \quad (4.2)$$

$$Q^{(yx)}(a_j|a_i) = P\{Y_{t+1} = a_j | X_t = a_i\}. \quad (4.3)$$

Também denotamos

$$Q^{(xx)}(a_j|a_i) = P\{X_{t+1} = a_j | X_t = a_i\} \quad (4.4)$$

$$Q^{(yy)}(a_j|a_i) = P\{Y_{t+1} = a_j | Y_t = a_i\}. \quad (4.5)$$

Exemplo 4.2.1 Um exemplo de procedimento de estimação destas matrizes pode ser visto na Figura 4.2

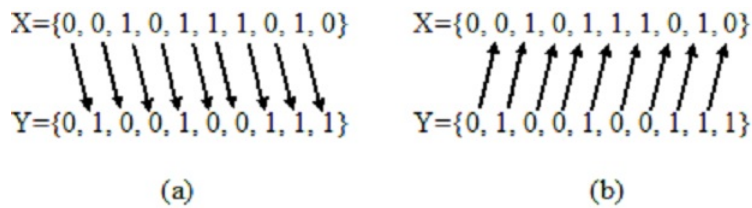


Figura 4.2: Transição cruzada: (a) X para Y e (b) Y para X .

A matriz de transição de frequência de X para Y , $\widehat{F}^{(yx)}$ é

$$\widehat{F}^{(yx)} = \begin{bmatrix} n_{11}^{(yx)} & n_{12}^{(yx)} \\ n_{21}^{(yx)} & n_{22}^{(yx)} \end{bmatrix}$$

onde $n_{ij}^{(yx)}$ é a contagem de frequência do estado j no tempo t na cadeia X para o estado i no tempo $t + 1$ na cadeia Y .

Analogamente a matriz de transição de frequência de Y para X , $\widehat{F}^{(xy)}$ é

$$\widehat{F}^{(xy)} = \begin{bmatrix} n_{11}^{(xy)} & n_{12}^{(xy)} \\ n_{21}^{(xy)} & n_{22}^{(xy)} \end{bmatrix}$$

O resultado a seguir apresenta a convergência dos estimadores para a probabilidade de transição.

Teorema 4.2.1 *Sejam X e Y duas cadeias de Markov de ordem 1 e tamanho n . Considere o estimador de máxima verossimilhança da probabilidade de transição*

$$\widehat{P}(i|j) = \frac{n_{ij}}{n_j}$$

com $i, j \in \mathbb{A}$, onde $n_{ij} = \sum_{t=1}^{n-1} 1_{\{X_t=j, Y_{t+1}=i\}}$ e $n_j = \sum_{i=1}^{n-1} n_{ij}$.

Então $\widehat{P}(i|j) \rightarrow P(i|j)$ onde $P(i|j) = \frac{P(X=i, Y=j)}{P(Y=j)}$

Prova: Temos que

$$\begin{aligned} \widehat{P}(i|j) &= \widehat{P}(X=i|Y=j) \\ &= \frac{\widehat{P}(X=i, Y=j)}{\widehat{P}(Y=j)} \\ &= \frac{n_{ij}}{n_j} \\ &= \frac{n_{ij}}{n_j} \cdot \frac{n}{n} \\ &= \frac{n_{ij}}{n} \cdot \frac{n}{n_j}. \end{aligned}$$

Como $\frac{n_{ij}}{n} \rightarrow P(i, j)$ e $\frac{n}{n} \rightarrow P(j)$ então $\widehat{P}(i|j) \rightarrow P(i|j)$. \square

Dada duas cadeias de Markov X e Y de tamanho n , a probabilidade conjunta entre estas duas cadeias é dada por

$$P(Y_1, Y_2, \dots, Y_n, X_1, X_2, \dots, X_n) = P(Y_n | Y_1, \dots, Y_{n-1}, X_1, X_2, \dots, X_n) P(Y_1, Y_2, \dots, Y_{n-1}, X_1, X_2, \dots, X_n) \quad (4.6)$$

Neste caso a probabilidade conjunta vai depender da definição utilizada para a probabilidade condicional $P(Y_n | Y_1, \dots, Y_{n-1}, X_1, X_2, \dots, X_n)$. O resultado a seguir apresenta

uma definição da probabilidade condicional, utilizada neste trabalho, e consequentemente o cálculo da função de verossimilhança para a probabilidade conjunta.

Teorema 4.2.2 *Sejam X e Y duas amostras de cadeias de Markov de tamanho n . Suponha que*

$$\begin{aligned} P(Y_1|X_1, \dots, X_n) &= P(Y_1) \\ P(Y_n|Y_1, Y_2, \dots, Y_{n-1}, X_1, X_2, \dots, X_n) &= P(Y_n|X_{n-1}) \end{aligned}$$

Então

$$P(Y_1, Y_2, \dots, Y_n, X_1, X_2, \dots, X_n) = \left[\prod_{i=2}^n P(Y_i|X_{i-1}) \right] \left[\prod_{j=2}^n P(X_j|X_{j-1}) \right] P(X_1) P(Y_1). \quad (4.7)$$

Prova: Considere duas cadeias $X = \{X_1, X_2, \dots, X_n\}$ e $Y = \{Y_1, Y_2, \dots, Y_n\}$. Seja

$$P(Y_1, Y_2, \dots, Y_n, X_1, X_2, \dots, X_n) = P(Y_1^n, X_1^n).$$

Então

$$\begin{aligned} P(Y_1^n, X_1^n) &= P(Y_n|Y_1, Y_2, \dots, Y_{n-1}, X_1, X_2, \dots, X_n) \cdot P(Y_1, Y_2, \dots, Y_{n-1}, X_1, X_2, \dots, X_n) \\ &= P(Y_n|X_{n-1}) \cdot P(Y_1, Y_2, \dots, Y_{n-1}, X_1, X_2, \dots, X_n) \\ &= P(Y_n|X_{n-1}) \cdot P(Y_{n-1}|Y_1, Y_2, \dots, Y_{n-2}, X_1, X_2, \dots, X_n) \cdot P(Y_1, Y_2, \dots, Y_{n-2}, X_1, X_2, \dots, X_n) \\ &= P(Y_n|X_{n-1}) \cdot P(Y_{n-1}|X_{n-2}) \cdot P(Y_1, Y_2, \dots, Y_{n-2}, X_1, X_2, \dots, X_n) \\ &= P(Y_n|X_{n-1}) \cdot P(Y_{n-1}|X_{n-2}) \dots P(Y_2|X_1) \cdot P(Y_1|X_1, X_2, \dots, X_n) \cdot P(X_1, X_2, \dots, X_n) \\ &= \left[\prod_{i=2}^n P(Y_i|X_{i-1}) \right] \cdot P(Y_1) \cdot P(X_1, X_2, \dots, X_n) \end{aligned}$$

Como X é uma cadeia de Markov, então $P(X_1, X_2, \dots, X_n) = \left[\prod_{j=2}^n P(X_j|X_{j-1}) \right] P(X_1)$. Assim

$$P(Y_1^n, X_1^n) = \left[\prod_{i=2}^n P(Y_i|X_{i-1}) \right] \left[\prod_{j=2}^n P(X_j|X_{j-1}) \right] P(X_1) P(Y_1).$$

cqd

O teorema 4.2.2 nos diz que a probabilidade conjunta das cadeias X e Y pode ser determinada pelo produto das probabilidades cruzadas $\prod_{i=2}^n P(Y_i|X_{i-1})$ pela distribui-

ção inicial da cadeia Y , que é conhecida, e pela conjunta da cadeia X , $P(X_1, X_2, \dots, X_n)$, que também é conhecida.

4.3 Sequências de Automatos Probabilísticos de Sufixos

4.3.1 Cadeias de Markov de Ordem 1

Considere r sequências de cadeia de Markov sobre $\mathbb{A} = \{1, 2, \dots, m\}$. Cada sequência tem uma matriz de probabilidade de transição $Q^{(kk)}$ com $k = \{1, 2, \dots, r\}$, e considerando duas sequências distintas temos $Q^{(jk)}$ como sendo a matriz de probabilidade de transição cruzada dos estados da k -ésima sequência para os estados da j -ésima sequência onde $k = \{1, 2, \dots, r\}$ e $j = \{1, 2, \dots, r\}$. Para um modelo de Markov de 1ª ordem a equação 3.2 pode ser escrita, na forma matricial, como

$$\mu_{n+1} = \begin{pmatrix} \mu_{n+1}^{(1)} \\ \mu_{n+1}^{(2)} \\ \vdots \\ \mu_{n+1}^{(r)} \end{pmatrix} = \begin{pmatrix} \lambda_{11}Q^{(11)} & \lambda_{12}Q^{(12)} & \dots & \lambda_{1s}Q^{(1r)} \\ \lambda_{21}Q^{(21)} & \lambda_{22}Q^{(22)} & \dots & \lambda_{2s}Q^{(2r)} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{r1}Q^{(r1)} & \lambda_{r2}Q^{(r2)} & \dots & \lambda_{rr}Q^{(rr)} \end{pmatrix} \begin{pmatrix} \mu_n^{(1)} \\ \mu_n^{(2)} \\ \vdots \\ \mu_n^{(r)} \end{pmatrix} = B\mu_n$$

O parâmetro de peso λ tem uma interpretação interessante. Cada λ_{jk} pode ser visto como uma medida local de contribuição de cada sequência k no cálculo da distribuição marginal da j -ésima sequência, $\mu_{n+1}^{(j)}$. Chamamos cada parâmetro de *coeficiente de contribuição*. Observe que este coeficiente, em geral, não é simétrico, isto é, $\lambda_{jk} \neq \lambda_{kj}$.

A distribuição de probabilidade da j -ésima sequência no passo $(n+1)$, $\mu_{n+1}^{(j)}$, depende do peso médio de $B^{(jk)}\mu_n^{(k)}$.

As distribuições de probabilidade de estados μ_n e as matrizes de probabilidade de transição $Q^{(hk)}$ são estimadas utilizando os estimadores canônicos: $\widehat{\mu} = \frac{N_i}{n}$ e $\widehat{q}_{ij} = \frac{N_{ij}}{N_i}$, onde N_i é o número de ocorrências na amostra do estado a_i e N_{ij} é o número de ocorrências na amostra do estado a_j seguido pelo estado a_i .

A matriz de probabilidade de transição P de uma cadeia de Markov tem um autovalor igual a 1 e todos os outros autovalores de P tem módulo menor que ou igual a 1. Como corolário, pode ser visto que existe um vetor positivo z de modo que $Pz = z$. Este

vetor z é chamado vetor de probabilidade estacionária de P , isto é, $\sum_{i=1}^m z_i = 1$. Além disso, z_i é a probabilidade de estado estacionária de que o sistema esteja no estado i .

Este resultado sugere que λ_{ij} pode ser estimado de modo que $\|B\widehat{\mu} - \widehat{\mu}\|$ é mínimo em alguma norma $\|\cdot\|$. Assim o procedimento minimax funciona corretamente.

Estimamos também os λ_{ij} utilizando a norma 2, ou seja,

$$\min_{\lambda} \frac{1}{2} \|\lambda Q \widehat{\mu} - \widehat{\mu}\|_2^2. \quad (4.8)$$

Neste caso, utilizando norma 2, os valores dos coeficientes encontrados foram os mesmos utilizando programação linear, minimax.

Observação: É importante resaltar que o ponto fundamental no processo de estimação é que a matriz quadrada B tem autovalor menor que ou igual a 1.

As estimativas são obtidas usando um algoritmo denominado *contribuição local* implementado em R e Matlab.

4.3.2 Cadeia de Markov de Ordem L

Um modelo de ordem 1 pode ser muito restritivo pois ignora a informação de todo o passado da cadeia para prever o futuro. Então em vez de r sequências de cadeias de Markov de ordem 1 podemos considerar r sequências de cadeias de Markov de ordem L . Note que é possível escrever uma $L_{full}PSA$ como uma cadeia de Markov de ordem L , que pode ser vista como uma cadeia de Markov de ordem 1 sobre \mathbb{A}^L . Deste modo, o procedimento minimax pode ser aplicado diretamente para sequências cadeias de Markov de ordem L ou $L_{full}PSA$. Seja $Q_L^{(hk)}$ a matriz de probabilidade de transição cruzada dos estados da k -ésima sequência para os estados da h -ésima sequência. A mesma estrutura pode ser proposta para a distribuição de probabilidade de estado do seguinte modo

$$\mu_{n+1}^{(j)} = \sum_{k=1}^r \lambda_{jk} Q_L^{(jk)} \mu_n^{(k)} \quad (4.9)$$

onde $j = \{1, 2, \dots, r\}$, $n = \{0, 1, \dots\}$, $\lambda_{jk} \geq 0$ e $\sum_{k=1}^r \lambda_{jk} = 1$. Nesta equação $\mu_0^{(j)}$ é a distribuição de probabilidade inicial da j -ésima sequência.

4.3.3 Modelagem Utilizando Árvore Probabilística de Contexto

Também é possível pensar em PCT de ordem L como uma cadeia de Markov de ordem L . Para este fim, é necessário completar a árvore com os ramos ausentes até a ordem L associando a mesma probabilidade de transição aos nós filhos que o seu pai. Desta forma a cadeia de Markov de ordem 1 equivalente terá novamente $|\mathbb{A}|^L$ estados compostos por todos os símbolos de tamanho L com matriz de transição $Q_L(x,y)$. A diferença entre este caso e os $L_{full} - PSA$ casos é que aqui muitas probabilidades de transição são iguais e portanto há menos parâmetros para estimar. O método de estimação minimax pode também ser aplicado neste caso a fim de estimar os parâmetros de contribuição num caso multivariado. No entanto isto não é necessário. Não há necessidade de completar todos os ramos de uma árvore de ordem L para vê-la como uma cadeia de Markov de ordem 1. Basta considerar uma PCT como um PSA baseado no fato de que os teoremas 4.1.1 e 4.1.2 garantem que há uma equivalência entre eles. Desta forma, em vez de ter $|\mathbb{A}|^L(|\mathbb{A}| - 1)$ parâmetros para estimar apenas $|\tau|(|\mathbb{A}| - 1)$ são necessários. Em geral esse é um número bem menor de parâmetros, pois em uma PSA com espaço de estados E , formado pelos contextos da PCT, espera-se que $E \leq |\mathbb{A}|^L$ e a matriz de transição Q_L vai ser uma matriz $E \times E$ em vez de $|\mathbb{A}|^L \times |\mathbb{A}|^L$. Portanto, dado uma PCT, encontramos o seu PSA equivalente e, em seguida, as transições de probabilidade de estado dos estados em $|\mathbb{A}|$ com base na transição de probabilidade de estado dos estados em E .

Para sequências de PCT sobre o mesmo alfabeto E os coeficientes de contribuição podem ser estimados usando a matriz quadrada de probabilidade de transição cruzada da PSA, $Q_M^{(jk)}$, em vez de $Q_\tau^{(jk)}$ que não é uma matriz quadrada. Então para estado $e_i \in E$ na sequência j podemos escrever a probabilidade de estados $\mu(e_i)^{(j)}$ como

$$\mu_{n+1}^{(j)} = \sum_{k=1}^r \lambda_{jk} Q_M^{(jk)} \mu_n^{(k)} \quad (4.10)$$

onde $j = \{1, 2, \dots, r\}$, $n = \{0, 1, \dots\}$, $\lambda_{jk} \geq 0$ e $\sum_{k=1}^r \lambda_{jk} = 1$. Nesta equação $\mu_0^{(j)}$ é a distribuição de probabilidade inicial da j -ésima sequência.

A equação 4.10 pode ser reescrita, na forma matricial, como

$$\begin{pmatrix} \mu_{n+1}^{(1)} \\ \mu_{n+1}^{(2)} \\ \vdots \\ \mu_{n+1}^{(r)} \end{pmatrix} = \begin{pmatrix} \lambda_{11}Q_M^{(11)} & \lambda_{12}Q_M^{(12)} & \dots & \lambda_{1r}Q_M^{(1r)} \\ \lambda_{21}Q_M^{(21)} & \lambda_{22}Q_M^{(22)} & \dots & \lambda_{2r}Q_M^{(2r)} \\ \vdots & \vdots & \dots & \vdots \\ \lambda_{r1}Q_M^{(r1)} & \lambda_{r2}Q_M^{(r2)} & \dots & \lambda_{rr}Q_M^{(rr)} \end{pmatrix} \begin{pmatrix} \mu_n^{(1)} \\ \mu_n^{(2)} \\ \vdots \\ \mu_n^{(r)} \end{pmatrix} = B_M \mu_n$$

ou seja,

$$\mu_{n+1} = B_M \mu_n.$$

Teorema 4.1 Se $\lambda_{ij} \geq 0$ para $1 \leq i, j \leq r$ então a matriz B_M tem autovalor igual a 1 e os outros autovalores têm módulo menor que ou igual a 1.

Prova: Considere

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1r} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2r} \\ \vdots & \vdots & \dots & \vdots \\ \lambda_{r1} & \lambda_{r2} & \dots & \lambda_{rr} \end{pmatrix}$$

onde $\sum_{k=1}^r \lambda_{jk} = 1$. Como $\lambda_{ij} \geq 0$ então Λ é não-negativa e irredutível. Pelo teorema de Perron-Frobenius existe um vetor

$$Y = [Y_1, Y_2, \dots, Y_r]$$

tal que

$$Y\Lambda = \lambda Y.$$

Considerando autovalor igual a 1 então $Y\Lambda = Y$.

Seja $1_r = [1, 1, \dots, 1]$ um vetor unitário de tamanho r . Então $1_r Q_M^{(ij)} = 1_r$ com $1 \leq i, j \leq r$ pois, dado

$$Q_M^{(ij)} = \begin{pmatrix} Q_M^{(11)} & Q_M^{(12)} & \dots & Q_M^{(1r)} \\ Q_M^{(21)} & Q_M^{(22)} & \dots & Q_M^{(2r)} \\ \vdots & \vdots & \dots & \vdots \\ Q_M^{(r1)} & Q_M^{(r2)} & \dots & Q_M^{(rr)} \end{pmatrix}$$

então

$$1_r Q_M^{(ij)} = \left(\sum_{i=1}^r Q_M^{(i1)} \quad \sum_{i=1}^r Q_M^{(i2)} \quad \dots \quad \sum_{i=1}^r Q_M^{(ir)} \right) = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} = 1_r.$$

Temos que

$$[Y_1 1_r, Y_2 1_r, \dots, Y_r 1_r] B_M = [Y_1 1_r, Y_2 1_r, \dots, Y_r 1_r].$$

Portanto

$$Y \Lambda = \lambda Y.$$

Logo 1 é autovalor de B_M . Falta, assim, mostrar que os outros autovalores são menores ou iguais a 1. Seja $Z = [Z_1, Z_2, \dots, Z_r] \in R^r$ onde $Z_i = [z_1, z_2, \dots, z_r]$ com $1 \leq i \leq r$. Defina

$$\begin{aligned} \|Z\|_v &= \max_{1 \leq i \leq r} \{\|Z_i\|_1 : Z = [Z_1, Z_2, \dots, Z_r]\} \\ \|Z\|_v &= \max_{1 \leq i \leq r} \{\|Z_1\|, \|Z_2\|, \dots, \|Z_r\|\}. \end{aligned}$$

Considere

$$\|B_M\|_N = \sup \{\|B_M Z\|_v : \|Z\|_v = 1\}.$$

É fácil verificar que

$$\|B_M Z\|_i \leq 1.$$

Assim $\|B_M\|_N \leq 1$, ou seja, todos os demais autovalores de B_M tem modulo menor que ou igual a 1.

c.q.d.

Esse resultado dá suporte ao uso do método minimax como forma de estimar os λ 's.

4.4 Algoritmo Proposto

Nesta seção descrevemos um algoritmo para estimar os coeficientes de contribuição de uma sequência de PCT. Para cada amostra a melhor árvore é estimada. Os contextos de cada árvore são tomados como o espaço de estados E de sua PSA equivalente. As amostras são reescritas em termos dos estados $e_i \in E$ a fim de obter as probabilidades de transição da seqüência k para a seqüência j , $\widehat{Q}_M^{(jk)}$. Se existirem r seqüências PSA então r^2 matrizes de transição cruzadas devem ser determinadas, incluindo as transições de uma seqüência para ela mesma. Finalmente, o *coeficiente de contribuição* $\lambda_{\tau_{ij}}$ é determinado utilizando programação linear. Todo o procedimento é dado no seguinte algoritmo:

1. Dados r amostras estimar as PCT's τ_1, \dots, τ_r

$$X_1^1, \dots, X_{n_1}^1 \rightarrow \widehat{\tau}_1$$

,

...

$$X_1^r, \dots, X_{n_r}^r \rightarrow \widehat{\tau}_r$$

.

2. Para cada amostra i escrever $X_1^i, \dots, X_{n_i}^i$ como $e_1^i, e_2^i, \dots, e_{n_i}^i, t = \min(n_i, n_j)$.
3. Estimar os PSA M_i equivalentes para cada $\tau_i, 1 \leq i \leq r$.
4. Estimar as matrizes de transição cruzada $Q_M^{(jk)}$.
5. Encontrar os *coeficientes de contribuição* λ_{jk} usando

$$\min_{\lambda} \max_i \left\| \left[\sum_{k=1}^r \lambda_{jk} \widehat{Q}_M^{(jk)} \widehat{\mu}_n^k - \widehat{\mu}_n^j \right]_i \right\|, \quad (4.11)$$

sujeito a $\sum_{k=1}^r \lambda_{jk} = 1$ e $\lambda_{jk} \geq 0$.

6. Escrever o modelo estimado.

Observação: O PCT-PSA e o algoritmo Contribuição foram desenvolvidos pelos autores em R e Matlab e estão disponíveis mediante solicitação. O algoritmo PCT, em R, foi desenvolvido por outros autores que gentilmente nos permitem usá-lo e também estão disponíveis mediante solicitação.

4.5 Medida de Proximidade entre Distribuições de Probabilidade

Nesta seção nosso objetivo é verificar, para cada tempo t , se a distribuição de probabilidade marginal estimada de uma PCT, $\widehat{\mu}^r(t)$, obtida através do modelo proposto

está próxima da verdadeira distribuição de probabilidade marginal desta PCT, $\mu^\tau(t)$. Se estas duas distribuições de probabilidade estão próximas em algum sentido razoável então este fato indica que o modelo é consistente. A divergência de Kulback-Leibler é utilizada como medida de proximidade entre estas duas distribuições de probabilidade considerando um conjunto de dados simulados.

Definição 4.5.1 A divergência de Kullback-Leibler D_{KL} entre duas distribuições $\widehat{\mu}^\tau$ e μ^τ é definida por

$$D_{KL}(\widehat{\mu}^\tau || \mu^\tau) = \sum_{i=1}^E \widehat{\mu}^\tau(i) \log \frac{\widehat{\mu}^\tau(i)}{\mu^\tau(i)}$$

Para isso foram realizadas $t = 1.000$ iterações. Para cada iteração são consideradas X_1^i, \dots, X_N^i amostras de tamanho N da i -ésima PCT, $1 \leq i \leq r$. Uma parte desta amostra, de tamanho n , é utilizada como aprendizagem e determinação dos coeficientes de contribuição, $1 < n \ll T$, onde T é o momento em que a cadeia entra em regime permanente. Após determinar os coeficientes de contribuição estima-se o restante da cadeia, $N - n$ elementos, utilizando o modelo proposto em 4.10 dado por

$$\mu_{n+1}^{(j)} = \sum_{k=1}^r \lambda_{jk} Q_M^{(jk)} \mu_n^{(k)}. \quad (4.12)$$

Para cada elemento adicional estimado na cadeia, os $N - n$ elementos restantes, estima-se a divergência KL considerando as distribuições marginais estimadas e a distribuição verdadeira.

Várias simulações foram realizadas considerando as mudanças nos parâmetros como o tamanho do alfabeto, ordem da PCT e as probabilidades de transição. Como os resultados das simulações foram muito similares para todas as alterações nos parâmetros, apenas um exemplo é apresentado. Neste exemplo geramos duas sequências de dados a partir das árvores apresentadas nas figuras 4.3 e 4.4.

A primeira sequência foi gerada a partir da árvore τ_1 e a segunda sequência foi gerada a partir da árvore τ_2 . O PSA equivalente da árvore τ_1 , M_1 , e da árvore τ_2 , M_2 , são apresentados a seguir.

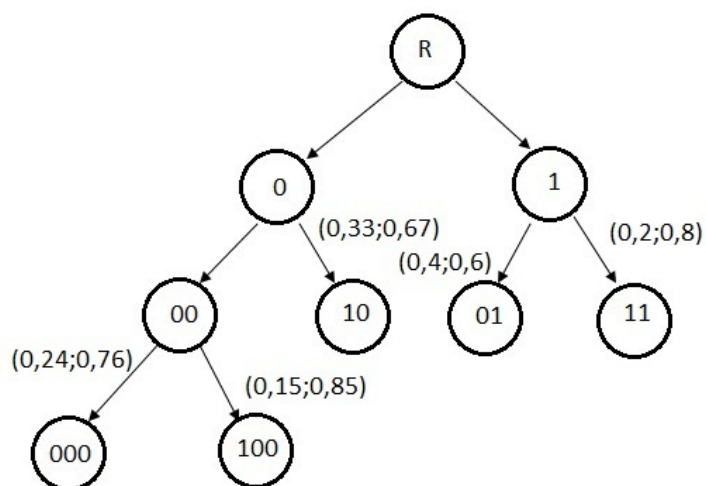


Figura 4.3: Árvore probabilística de contexto, τ_1 , utilizada para gerar as cadeias

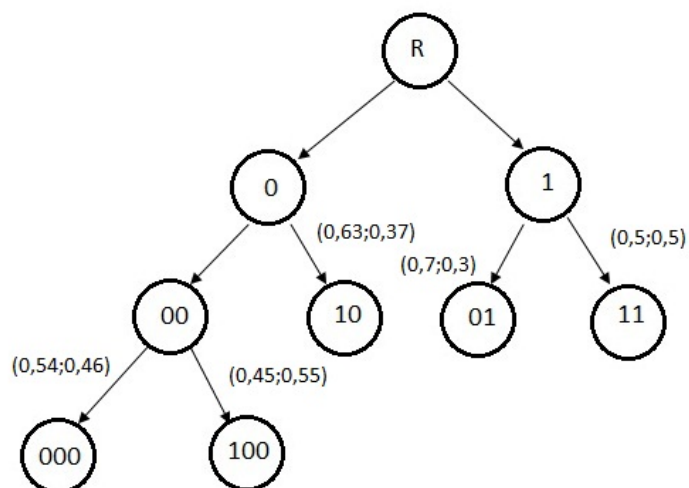


Figura 4.4: Árvore probabilística de contexto, τ_2 , utilizada para gerar as cadeias

$$M_1 = \begin{bmatrix} 0,24 & 0,15 & 0 & 0 & 0 \\ 0 & 0 & 0,33 & 0 & 0 \\ 0 & 0 & 0 & 0,4 & 0,2 \\ 0,76 & 0,85 & 0,67 & 0 & 0 \\ 0 & 0 & 0 & 0,6 & 0,8 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 0,54 & 0,45 & 0 & 0 & 0 \\ 0 & 0 & 0,63 & 0 & 0 \\ 0 & 0 & 0 & 0,7 & 0,5 \\ 0,46 & 0,55 & 0,37 & 0 & 0 \\ 0 & 0 & 0 & 0,3 & 0,5 \end{bmatrix}$$

As distribuições iniciais verdadeiras de M_1 e M_2 são dadas por $\mu_0^{\tau_1} = (0,02; 0,06; 0,18; 0,18; 0,56)$ e $\mu_0^{\tau_2} = (0,05; 0,10; 0,20; 0,20; 0,45)$, respectivamente.

A figura 4.5 apresenta os resultados considerando $N = 10.000$ e $n = 1.000$. Podemos verificar em (a) e em (c) que a divergência KL tende a zero muito rápido para ambas as árvores. As figuras 4.5(b) e 4.5(d) mostram um zoom considerando as 10 primeiras divergências estimadas, para melhor visualização.

Pode-se observar que a divergência KL é muito próxima de zero. Isto significa que o modelo proporciona uma boa estimativa para probabilidades marginais de PCT's.

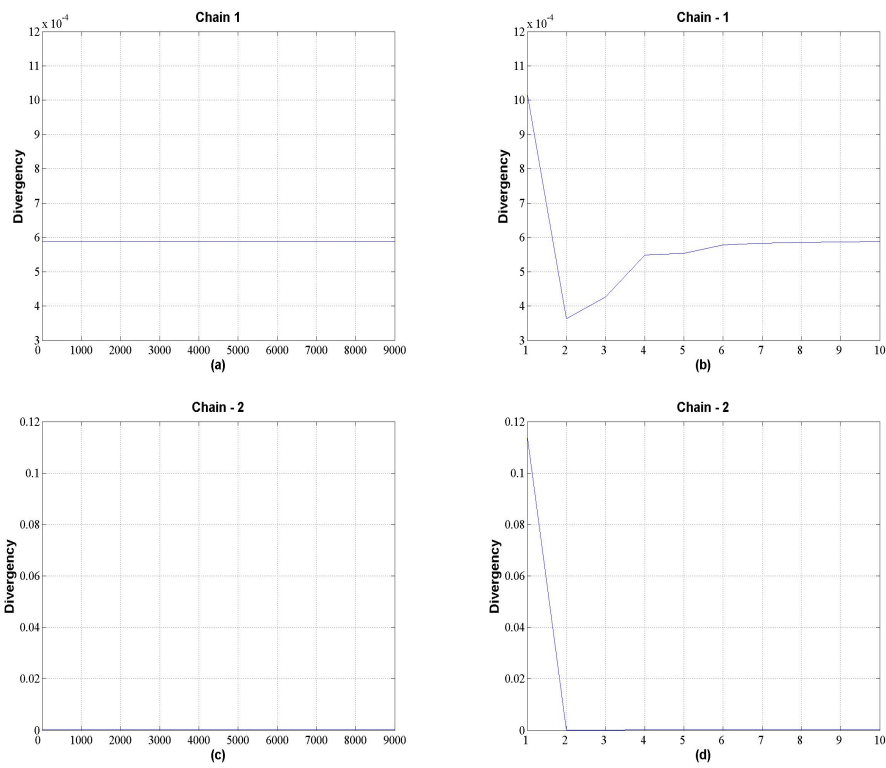


Figura 4.5: Divergência de Kullback-Leibler entre a verdadeira distribuição marginal de uma PCT e a distribuição marginal estimada para τ_1 em (a) e (b), e τ_2 em (c) e (d).

4.6 Aplicações do Modelo Proposto

O corpus eletrônico Tycho Brahe Parsed Corpus of Historical Portuguese é formado por textos escritos em Português por autores nascidos entre 1380 e 1845 e pode ser encontrado em (<http://www.tycho.iel.unicamp.br/tycho>). Nosso objetivo é comparar alguns desses textos, através do coeficiente de contribuição de um texto em outro, de acordo com o ritmo de cada texto utilizando a metodologia proposta. Textos escritos pelo mesmo autor, autores nascidos por volta do mesmo período de tempo e de textos escritos por autores nascidos em períodos de tempo diferentes são comparados. No 1º caso esperamos grande semelhança entre os textos, no 2º caso esperamos semelhança moderada e no 3º caso esperamos ausência de semelhança. A Tabela 4.2 apresenta os textos para que a comparação seja feita. A primeira coluna representa o código do texto, a segunda coluna mostra o ano em que cada autor nasceu, a terceira tem o nome do autor e a última mostra o título do texto.

Tabela 4.2: Textos escritos apresentados por Código, Ano, Autor e Título.

Code	Year	Author	Title
G008	1502	Pero Magalhães de Gandavo	História da Província de Santa Cruz
V002	1608	Antônio Vieira	Cartas
V004	1608	Antônio Vieira	História do Futuro
C003	1631	Antônio das Chagas	Cartas Espirituais
A004	1750	Marquesa de Alorna	Cartas
G003	1799	J. B. da Silva L. de Almeida Garret	Cartas, Almeida Garret
G004	1799	J. B. da Silva L. de Almeida Garret	Teatro, Almeida Garret
A003	1802	Marquês da Fronteira e d'Alorna	Memórias do Marquês da Fronteira e d'Alorna
O001	1836	Ramalho Ortigão	Cartas a Emília

Os textos deste corpus foram modernizados, ou seja, foram modificados de modo a estabelecer um critério ortográfico e gramatical que permitisse codificar as palavras de acordo com a tonalidade das sílabas. Cada sílaba dos textos foi classificada como tônica ou átona e se é o início de palavra prosódica ou não.

A seguinte lista mostra como cada sílaba foi codificada.

- 0: sílaba átona no meio ou final de palavras
- 1: sílaba tônica no meio ou final de palavras

- 2: sílaba átona no início de palavras
- 3: sílaba tônica no início de palavras
- 4: fim de sentença

Este código foi proposto pela equipe linguística do projeto *padrões rítmicos, fixação de parâmetros e mudança lingüística* do Instituto de Estudos Linguísticos da Unicamp. A codificação foi feita pelo programa *Silaba* desenvolvido em PERL por Miguel Galves (2008).

Como exemplo da codificação as duas primeiras sentenças do texto *Memórias do Marquês D'Alorna (1802)* são dadas a seguir:

"Darei principio a estas minhas memorias pela minha genealogia.
Descendo em linha recta de D. Fernão Martins Mascarenhas, Senhor
de Lavre, Commendador de Mertola e Almodovar, Alcaide-Mor de
Montemor-ó-Novu e Alcacer do Sal, e Capitão de ginetes dos Snrs."

Após a codificação temos:

```
"212102103021020102000104210210302121212010212102001201020010200120001  
002010201200120102001033212103201213030212012010200102013212102001200001302001020120104"
```

Uma PCT é um bom modelo para estes dados porque existe dependência na posição das sílabas na frase em relação às sílabas anteriores, mas o comprimento desta dependência pode variar de uma sequência de sílabas para outra. Por exemplo, é natural supor que cada vez que um símbolo "4" aparece, marca o fim da sentença, o processo perde memória e começa novamente. Portanto, este símbolo marca um momento de renovação do processo. Por outro lado, se um símbolo "0" ocorrer no tempo t , não há informações suficientes para prever o próximo símbolo porque uma sílaba átona pôde aparecer seguida por qualquer outro símbolo, mas com probabilidades diferentes. No entanto, se uma sequência de 3 sílabas átonas ocorre, é pouco provável que ocorra outra sílaba átona como próximo símbolo.

Usando este código, a PCT de cada texto pode ser estimada. Para todos os textos a árvore estimada foi a mesma e sua estrutura é apresentada na figura 4.6. Observa-se que esta árvore é formada por 11 ramos ou contextos.

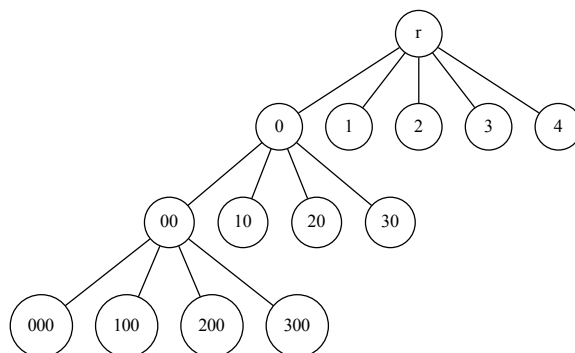


Figura 4.6: PCT estimada dos textos do Corpus of Historical Portuguese Tycho

Uma vez que a PCT é estimada ela pode ser transformada em uma PSA. O espaço de estados, ou conjunto de contextos, da PSA para estes dados é dado por $E = \{000, 100, 200, 300, 10, 20, 30, 1, 2, 3, 4\}$.

Para determinar o tamanho de amostra necessário para estimar as matrizes de probabilidade utilizamos

$$n = \frac{Z_{\frac{\alpha}{2}}^2 p(1-p)}{\varepsilon^2} \quad (4.13)$$

onde

1. α é o nível de significância;
2. m é a ordem da matriz de probabilidade de transição;
3. p é a probabilidade de sucesso e
4. ε é o erro cometido.

Neste caso adotamos $p = 0,5$ pois não conhecemos a variância verdadeira, $\varepsilon = 0,01$ e $\alpha = 0,05$. Como os PCT têm 11 contextos, então as matrizes de probabilidade têm

ordem 11, o que implica $m = 11$. Assim estimamos $Z_{\frac{\alpha}{2m^2}}^2 = 3,50$. Substituindo os valores em 4.13 encontramos $n = 32.000$ elementos.

Este tamanho de amostra é mais do que suficiente para proporcionar boas estimativas para as matrizes de probabilidade de transição, mas está longe de atingir estacionariedade para qualquer texto. O texto que atingiu estacionariedade em menos tempo foi o texto G004 com 57.001 elementos cujo tamanho real era de 57.646 elementos. Somente o texto G008, com 40.314 elementos, não atingiu a estacionariedade. Todos os demais textos atingiram a estacionariedade acima de 57.001 elementos. Para determinar o ponto de corte, ponto de estacionariedade estimada, foi utilizada a diferença entre as distribuições marginais no tempo $t + 1$ e t . O menor t tal que a diferença seja menor que 0,0001 foi considerado como ponto de corte.

4.7 Estimação dos Coeficientes de Contribuição Local dos Ritmos em Textos Escritos

Uma vez que a PCT τ_i para cada texto na amostra é estimada, o PSA equivalente pode ser determinado e, em seguida, o algoritmo descrito na seção 4.4 pode ser aplicado para alguns pares de textos escolhidos. O principal objetivo é estimar a contribuição entre os textos escritos pelo mesmo autor ou por autores diferentes em diferentes períodos de tempo ou não. Esta semelhança é expressa pelo coeficiente de contribuição local. Para esse fim, escolhemos como amostras 2 textos de Almeida Garret (1799), 2 textos de Antonio Vieira (1608), textos escritos por diferentes autores mas por pequena diferença de tempo: Marquês D'Alorna (1802) e Marquesa D'Alorna (1750) e, finalmente, textos escritos por diferentes autores mas por grande diferença de tempo: Gandavo (1502) e Ramalho Urtigão (1836).

Para cada par de textos considere:

$\mu_n^{(1)}$ = distribuição de probabilidade marginal do texto 1 no tempo n ;

$\mu_n^{(2)}$ = distribuição de probabilidade marginal do texto 2 no tempo n ;

$Q^{(11)}$ = matriz de probabilidade de transição do texto 1 para ele mesmo;

$Q^{(12)}$ = matriz de probabilidade de transição cruzada do texto 2 para o texto 1;

4.7 Estimação dos Coeficientes de Contribuição Local dos Ritmos em Textos Escritos 61

$Q^{(21)}$ = matriz de probabilidade de transição cruzada do texto 1 para o texto 2;

$Q^{(22)}$ = matriz de probabilidade de transição do texto 2 para ele mesmo;

As matrizes de frequência de transição estimadas para todos os textos considerados nesta aplicação são apresentadas no capítulo 8 .

4.7.1 Coeficiente de Contribuição Local de Textos Escritos por Mesmo Autor

Textos Escritos por Almeida Garret

Nesta seção apresentamos os resultados encontrados ao analisarmos os dois textos, G003 e G004, de Almeida Garret, nascido no ano de 1799.

As contribuições locais, λ 's, estimadas para estes textos são:

$$\lambda_{(G003,G003)} = 0,6278;$$

$$\lambda_{(G003,G004)} = 0,3722;$$

$$\lambda_{(G004,G003)} = 0,5945;$$

$$\lambda_{(G004,G004)} = 0,4055.$$

Aplicando a metodologia proposta temos

$$\begin{cases} \mu_{n+1}^{(G003)} = 0,6278\widehat{Q}^{(G003,G003)}\mu_n^{(G003)} + 0,3722\widehat{Q}^{(G003,G004)}\mu_n^{(G004)} \\ \mu_{n+1}^{(G004)} = 0,5945\widehat{Q}^{(G004,G003)}\mu_n^{(G003)} + 0,4055\widehat{Q}^{(G004,G004)}\mu_n^{(G004)} \end{cases}$$

As relações acima mostram que $\widehat{Q}^{(G004,G003)}\mu_n^{(G003)}$ tem um peso de 0,5945 no cálculo de $\mu_{n+1}^{(G004)}$. Isto sugere que a forma rítmica do texto G003 tem contribuição moderada à forma rítmica do texto G004.

Já a forma rítmica do texto G004 tem pouca contribuição na forma rítmica do texto G003.

Textos escritos por Antonio Vieira, 1608 (V002 e V004)

Nesta seção apresentamos os resultados encontrados ao analisarmos os dois textos, V002 e V004, de Antonio Vieira, nascido no ano de 1608.

As contribuições locais, λ 's, estimadas para estes textos são:

$$\lambda_{(V002,V002)} = 0;$$

$$\lambda_{(V002,V004)} = 1;$$

$$\lambda_{(V004,V002)} = 1;$$

$$\lambda_{(V004,V004)} = 0;$$

Aplicando a metodologia temos

$$\begin{cases} \mu_{n+1}^{(V002)} = \widehat{Q}^{(V002,V004)} \mu_n^{(V004)} \\ \mu_{n+1}^{(V004)} = \widehat{Q}^{(V004,V002)} \mu_n^{(V002)} \end{cases}$$

Observe que a forma rítmica do texto V002 tem forte contribuição à forma rítmica do texto V004, o mesmo acontecendo entre os textos V004 e V002.

4.7.2 Autores Nascidos em Períodos de Tempo Próximos

Nesta seção analisamos os textos escritos por Marquês D'Alorna (texto A003), nascido em 1802, e Marquesa D'Alorna (texto A004), nascida em 1750.

As contribuições locais, λ 's, estimadas para estes textos são:

$$\lambda_{(A003,A003)} = 0;$$

$$\lambda_{(A003,A004)} = 1;$$

$$\lambda_{(A004,A003)} = 0,9750;$$

$$\lambda_{(A004,A004)} = 0,0250.$$

Aplicando a metodologia proposta encontramos

$$\begin{cases} \mu_{n+1}^{(A003)} = \widehat{Q}^{(A003,A004)} \mu_n^{(A004)} \\ \mu_{n+1}^{(A004)} = 0,9750 \widehat{Q}^{(A004,A003)} \mu_n^{(A003)} + 0,0250 \widehat{Q}^{(A004,A004)} \mu_n^{(A004)} \end{cases}$$

Pode-se observar que há uma forte contribuição entre as formas rítmicas dos dois textos.

4.7.3 Autores nascidos em Diferentes Períodos de Tempo

Nesta seção analisamos os textos escritos por Gandavo (texto G008), nascido em 1502, e Ramalho Urtigão (texto O001), nascido em 1836.

As contribuições locais, λ 's, estimadas para estes textos são:

$$\begin{aligned}\lambda_{(G008,G008)} &= 1; \\ \lambda_{(G008,O001)} &= 0; \\ \lambda_{(O001,G008)} &= 0,2578; \\ \lambda_{(O001,O001)} &= 0,7422;\end{aligned}$$

Aplicando a metodologia encontramos

$$\begin{cases} \mu_{n+1}^{(G008)} = \widehat{Q}^{(G008,G008)} \mu_n^{(G008)} \\ \mu_{n+1}^{(O001)} = 0,2578 \widehat{Q}^{(O001,G008)} \mu_n^{(G008)} + 0,7422 \widehat{Q}^{(O001,O001)} \mu_n^{(O001)} \end{cases}$$

Observamos que a forma rítmica do texto G008 não contribui com a forma rítmica do texto O001. Já a forma rítmica do texto O001 tem pouca contribuição com a forma rítmica do texto G008.

Neste capítulo estendemos a metodologia aplicada a cadeias de Markov para PCT. Para isto utilizamos o PSA equivalente a cada PCT. De acordo com a metodologia proposta conseguimos estabelecer a contribuição local entre as formas rítmicas dos textos apresentados. Percebemos que alguns textos não há contribuição, outros há pouca e outros há forte contribuição.

4.8 Estimação da Variância dos Coeficientes de Contribuição

O objetivo desta seção é estimar a variância dos coeficientes de contribuição.

A variância do estimador de mínimos quadrados dos coeficientes λ 's é estimada através de

$$\sigma^2(\lambda) = \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1} \quad (4.14)$$

onde R é o vetor de restrições de igualdade, $R\lambda = r$, e σ^2 é o estimador da variância de cada texto que, neste caso, foi considerada a cadeia formada pelos contextos relevantes. As restrições consideradas são: $\lambda_{11} + \lambda_{12} = 1$ e $\lambda_{21} + \lambda_{22} = 1$. O estimador 4.14 pode ser encontrado em (<http://www2.econ.iastate.edu/>) "Restricted least squares, hypothesis testing, and prediction in the classical linear regression model".

4.8.1 Estimativa da Variância dos Textos

Para estimar σ^2 de cada texto T utilizamos block bootstrap, apresentado por Evaristo [2010], dado por: considere uma amostra de cadeia de Markov de ordem 1 de tamanho n , $Y = \{Y_1, Y_2, \dots, Y_n\}$. A cadeia Y será sub-dividida em N blocos de tamanho fixo L de modo que $n = N \cdot L$.

Sejam $\epsilon_1, \epsilon_2, \dots, \epsilon_N$ o conjunto de blocos onde $\epsilon_i = \{Y_i, \dots, Y_{i+L}\}$ com $i = 1, 2, \dots, N$. Assim a nova reamostra será definida como $Y_j^* = \{\epsilon_1^*, \epsilon_2^*, \dots, \epsilon_N^*\}$ onde ϵ_i^* é o i -ésimo bloco sorteado com $j = \{1, 2, \dots, N\}$ e $i = \{1, 2, \dots, N\}$.

Para cada reamostra estimamos a média \bar{Y}_j^* . Portanto temos um conjunto com N médias, $\bar{Y}^* = \{\bar{Y}_1^*, \bar{Y}_2^*, \dots, \bar{Y}_N^*\}$. A variância estimada da cadeia é definida por $\hat{\sigma}^2(Y) = \sigma^2(\bar{Y}^*)$, ou seja, a variância do conjunto das médias \bar{Y}^* .

Estimativa da Variância dos Coeficientes de Contribuição entre os Textos V002 e V004

Neste caso o vetor X foi definido como $[P_{(V002, V002)}\mu_{V002}; P_{(V002, V004)}\mu_{V004}]$ para a estimativa dos $\lambda_{(V002, V002)}$ e $\lambda_{(V002, V004)}$. Para as estimativas dos $\lambda_{(V004, V002)}$ e $\lambda_{(V004, V004)}$ utilizamos o vetor X como $[P_{(V004, V002)}\mu_{V002}; P_{(V004, V004)}\mu_{V004}]$.

Para as estimativas de $\sigma^2(V002)$ e $\sigma^2(V004)$ utilizamos 4.14 e consideramos amostras de tamanho $N = 32.000$ e 500 reamostras de cada texto. Neste caso as estimativas foram de $\sigma^2(V002) = 0.00024$ e $\sigma^2(V004) = 0.0003$.

Os resultados das variâncias estimadas dos λ 's foram:

$$\widehat{\sigma}^2(\lambda_{(V002,V002)}) = 0.00035;$$

$$\widehat{\sigma}^2(\lambda_{(V002,V004)}) = 0.00035;$$

$$\widehat{\sigma}^2(\lambda_{(V004,V002)}) = 0.00045;$$

$$\widehat{\sigma}^2(\lambda_{(V004,V004)}) = 0.00045;$$

Estimativa da Variância dos Coeficientes de Contribuição entre os Textos G003 e G004

Neste caso o vetor X foi definido como $[P_{(G003,G003)}\mu_{G003}; P_{(G003,G004)}\mu_{G004}]$ para a estimativa dos $\lambda_{(G003,G003)}$ e $\lambda_{(G003,G004)}$. Para as estimativas dos $\lambda_{(G004,G003)}$ e $\lambda_{(G004,G004)}$ utilizamos o vetor X como $[P_{(G004,G003)}\mu_{G003}; P_{(G004,G004)}\mu_{G004}]$. Para as estimativas de $\sigma^2(G003)$ e $\sigma^2(G004)$ utilizamos 4.14 e consideramos amostras de tamanho $N = 32.000$ e 500 reamostras de cada texto. Neste caso as estimativas foram de $\sigma^2(G003) = 0.00029$ e $\sigma^2(G004) = 0.0003$.

Os resultados das variâncias estimadas dos λ 's foram:

$$\widehat{\sigma}^2(\lambda_{(G003,G003)}) = 0.00045;$$

$$\widehat{\sigma}^2(\lambda_{(G003,G004)}) = 0.00045;$$

$$\widehat{\sigma}^2(\lambda_{(G004,G003)}) = 0.00048;$$

$$\widehat{\sigma}^2(\lambda_{(G004,G004)}) = 0.00048;$$

Contribuição Local em Sequências de Cadeia de Markov com Espaço de Estados Contínuo

Neste capítulo o objetivo é estender a metodologia desenvolvida para sequências de cadeias de Markov com espaço de estados discretos, apresentada no capítulo 4, para sequências de cadeias de Markov com espaço de estado contínuo. Apresentamos uma metodologia que estima a contribuição entre densidades de transição marginais de cadeias de Markov contínuas. Esta contribuição é medida pelos coeficientes da combinação linear que envolve as informações de todas as outras densidades de transição marginais. Assim podemos verificar o quanto uma densidade de transição marginal é influenciada pelas demais.

5.1 Modelo: Combinação Linear de Densidades de Transição Marginais

Seja $\{X_i : i = 1, 2, \dots, r\}$ um conjunto com r sequências com espaço de estados contínuo e tempo discreto. Definimos: $t_n^{(jj)}(y|x)$ como a densidade de transição marginal da j -ésima sequência para ela mesma no tempo n e $t_n^{(jk)}(y|x)$ como a densidade de transição marginal da k -ésima sequência para a j -ésima sequência no tempo n . Para encontrar a densidade no tempo $n + 1$ propomos o seguinte modelo:

$$t_{n+1}^{(jj)}(y|x) = \sum_{k=1}^r \lambda_{(jk)} t_n^{(jk)}(y|x), \quad (5.1)$$

onde $\lambda_{(jk)} \geq 0$, $\sum_{k=1}^r \lambda_{(jk)} = 1$ para $j = \{1, 2, \dots, r\}$.

As funções de densidade de transição $t_n^{(jk)}(y|x)$ com $j = \{1, 2, \dots, r\}$ e $k = \{1, 2, \dots, r\}$ denotamos *densidades de transições cruzadas marginais*.

Na equação 5.1 os coeficientes $\lambda_{(jk)}$ com $j = \{1, 2, \dots, r\}$ e $k = \{1, 2, \dots, r\}$ são denominados *coeficientes de contribuições locais*. Eles representam o quanto cada densidade de transição $t_n^{(jk)}$ contribui para o cálculo da densidade de transição $t_{n+1}^{(jj)}$.

No caso discreto, temos a combinação entre as matrizes de probabilidade de transição e a distribuição de probabilidade, $P^{jk} \mu^j$. No caso contínuo temos a combinação entre as densidades de transição, t^{jk} .

A equação 5.1 pode ser reescrita, na forma matricial, como

$$\begin{pmatrix} t_{n+1}^{(11)} \\ t_{n+1}^{(12)} \\ \vdots \\ t_{n+1}^{(1r)} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1r} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2r} \\ \vdots & & & \\ \lambda_{r1} & \lambda_{r2} & \dots & \lambda_{rr} \end{pmatrix} \begin{pmatrix} t_n^{(11)} \\ t_n^{(12)} \\ \vdots \\ t_n^{(1r)} \end{pmatrix}$$

Proposição 5.1.1 Considere a equação 5.1 e seja $P_{n+1}^{(j)}(x, A)$ definido por

$$P_{n+1}^{(j)}(x, A) = \int_A t_{n+1}^{(jj)}(y|x) dy. \quad (5.2)$$

Então

1. Dado x , $P_{n+1}^{(j)}(x, \cdot)$ é uma medida de probabilidade.
2. Dado A , $P_{n+1}^{(j)}(\cdot, A)$ é uma função mensurável.

Demonstração: Para $j = 1, 2, \dots, s$

$$\begin{aligned} P_{n+1}^{(j)}(x, A) &= \int_A \left(\sum_{k=1}^s \lambda_{(jk)} t_n^{(jk)}(y|x) \right) dy \\ &= \int_A (\lambda_{(j1)} t_n^{(j1)}(y|x) + \dots + \lambda_{(js)} t_n^{(js)}(y|x)) dy \\ &= \lambda_{(j1)} \int_A t_n^{(j1)}(y|x) dy + \dots + \lambda_{(js)} \int_A t_n^{(js)}(y|x) dy. \end{aligned}$$

Demonstração do item (1): Note que

$$0 \leq \int_A t^{(jk)}(y|x) dy \leq 1$$

portanto

$$0 \leq \lambda_{(jk)} \int_A t^{(jk)}(y|x) dy \leq \lambda_{(jk)}$$

para $k = \{1, 2, \dots, s\}$ e

$$0 \leq \lambda_{(j1)} \int_A t^{(j1)}(y|x) dy + \dots + \lambda_{(js)} \int_A t^{(js)}(y|x) dy \leq \lambda_{(j1)} + \dots + \lambda_{(js)}$$

$$0 \leq \lambda_{(j1)} \int_A t^{(j1)}(y|x) dy + \dots + \lambda_{(js)} \int_A t^{(js)}(y|x) dy \leq 1$$

$$0 \leq P_{n+1}^{(j)}(x, A) \leq 1$$

Se $A = \Omega$, todo o espaço amostral,

$$\int_A t^{(jk)}(y|x)dy = 1$$

para $k = \{1,2 \dots, s\}$, então

$$\lambda_{(jk)} \int_A t^{(jk)}(y|x)dy = \lambda_{(jk)}.$$

Portanto

$$\begin{aligned} \lambda_{(j1)} \int_A t^{(j1)}(y|x)dy + \dots + \lambda_{(js)} \int_A t^{(js)}(y|x)dy &= \lambda_{(j1)} + \dots + \lambda_{(js)} \\ \lambda_{(j1)} \int_A t^{(j1)}(y|x)dy + \dots + \lambda_{(js)} \int_A t^{(js)}(y|x)dy &= 1 \\ P_{n+1}^{(j)}(x,A) &= 1. \end{aligned}$$

Demonstração do item (2): Visto que $t^{(jk)}(y|x)$ são densidades de transição marginais elas também são funções mensuráveis e $\int_A t^{(jk)}(y|x)dy$ é mensurável em A para $k = \{1,2 \dots, s\}$. Note que $\lambda_{(jk)} \int_A t^{(jk)}(y|x)dy$ é também mensurável em A e como a soma de funções mensuráveis é mensurável então $P_{n+1}^{(j)}(\cdot, A)$ é mensurável em A .

5.2 Função a ser Otimizada

Da mesma forma que no caso discreto, o cálculo dos coeficientes de contribuição, dados na equação 5.1, é feito utilizando otimização linear denominado *Minimax* com apoio do MatLab.

Dado $P_{n+1}^{(j)}(x,A)$ e $f_n^{(j)}(x)$ podemos escrever

$$\int_A f_n^{(j)}(z)dz = \int_{-\infty}^{+\infty} P_{n+1}^{(j)}(x,A) f_n^{(j)}(x)dx. \quad (5.3)$$

Substituindo as equações 5.1 e 5.2 na equação 5.3 encontramos

$$\int_A f_n^{(j)}(z)dz = \int_{-\infty}^{+\infty} \left[\int_A \left(\sum_{k=1}^r \lambda_{(jk)} t_n^{(jk)}(y|x) \right) dy \right] f_n^{(j)}(x)dx$$

onde $j = \{1, 2, \dots, r\}$. Então

$$\left| \int_A f_n^{(j)}(z)dz - \int_{-\infty}^{+\infty} \left[\int_A \left(\sum_{k=1}^r \lambda_{(jk)} t_n^{(jk)}(y|x) \right) dy \right] f_n^{(j)}(x)dx \right| = 0$$

Para encontrar os coeficientes de contribuição local do modelo proposto minimiza-se em λ e maximiza-se em A a seguinte função:

$$\min_{\lambda} \max_A \left| \int_A f_n^{(j)}(x)dx - \int_{-\infty}^{+\infty} \left[\int_A \left(\sum_{k=1}^r \lambda_{(jk)} t_n^{(jk)}(y|x) \right) dy \right] f_n^{(j)}(x)dx \right| \quad (5.4)$$

onde $j = \{1, 2, \dots, r\}$.

Na função 5.4 o intervalo A deve ser cuidadosamente determinado pois ele deve estar contido no domínio de todas as cadeias, ou seja, ele deve estar na intercessão de todos os domínios. Na seção 5.6 daremos mais detalhes sobre o conjunto A .

Como exemplo, desenvolvendo a equação 5.4 considerando $j = 2$ e obtemos as seguintes expressões a serem resolvidas em relação a λ 's:

$$\begin{cases} \int_A f^{(1)}(x)dx = \lambda_{11} \int_{-\infty}^{+\infty} \left(\int_A t^{(11)}(x,y)dy \right) f^{(1)}(x)dx + \lambda_{12} \int_{-\infty}^{+\infty} \left(\int_A t^{(12)}(x,y)dy \right) f^{(2)}(x)dx \\ \int_A f^{(2)}(x)dx = \lambda_{21} \int_{-\infty}^{+\infty} \left(\int_A t^{(21)}(x,y)dy \right) f^{(1)}(x)dx + \lambda_{22} \int_{-\infty}^{+\infty} \left(\int_A t^{(22)}(x,y)dy \right) f^{(2)}(x)dx \end{cases}$$

A estimação das funções de densidade de probabilidade que são utilizadas na função 5.4 serão feitas utilizando núcleo estimador.

Estimamos também os λ_{ij} utilizando a norma 2, ou seja,

$$\min_{\lambda} \frac{1}{2} \left\| \int_A f_n^{(j)}(x)dx - \int_{-\infty}^{+\infty} \left[\int_A \left(\sum_{k=1}^r \lambda_{(jk)} t_n^{(jk)}(y|x) \right) dy \right] f_n^{(j)}(x)dx \right\|_2^2. \quad (5.5)$$

Neste caso os resultados obtidos foram os mesmos utilizando o minimax.

5.3 Estimação das Funções de Densidade Marginais

Para estimar os coeficientes do modelo proposto é necessário definir as funções de densidade marginais. A estimação da função de densidade conjunta marginal $q_n(x, y)$, dada na equação 2.21, é adaptada para possibilitar a definição da função de densidade de transição cruzada marginal $t_n^{(jk)}(y|x)$. A seguir apresentamos os núcleos estimadores das densidades marginais.

Definição 5.3.1 *Considere r cadeias de Markov contínuas. As funções de densidade marginais no tempo n são definidos como:*

$$t_n^{(jk)}(y|x) = \frac{q_n^{(jk)}(x, y)}{f_n^{(k)}(x)} \quad (5.6)$$

onde

$$f_n^{(k)}(x) = \frac{1}{nh_n} \sum_{i=0}^n K\left(\frac{x - X_i}{h_n}\right) \quad (5.7)$$

e

$$q_n^{(jk)}(x, y) = \frac{1}{nh_n^2} \sum_{i=0}^{n-1} K\left(\frac{x - X_i}{h_n}\right) K\left(\frac{y - Y_{i+1}}{h_n}\right). \quad (5.8)$$

Se $f_n^{(k)}(x) = 0$ então $t_n^{(jk)}(y|x) = h_n$, onde $j \neq k$, $X_i \in k$, $Y_i \in j$, $j = \{1, 2, \dots, r\}$ e $k = \{1, 2, \dots, r\}$ e h_n é a janela ótima.

Roussas [1969] mostra que $f_n^{(k)}(x) \rightarrow f^{(k)}(x)$ e $q_n^{(jj)}(x, y) \rightarrow q^{(jj)}(x, y)$.

Quando $j = k$ utilizamos os estimadores apresentados nas equações 2.20 a 2.22 .

Para estimar a janela ótima h_n utilizamos o método plug-in apresentado por Chiu [1991]. Na seção 5.4 daremos maiores detalhes sobre a estimação da janela ótima.

A figura 5.1 mostra um exemplo, considerando apenas duas seqüências, de como a densidade conjunta, dada pela equação 5.8, é calculada. Sejam $X = \{X_1, X_2, \dots, X_{n-1}, X_n\}$ e $Y = \{Y_1, Y_2, \dots, Y_{n-1}, Y_n\}$ duas seqüências. Nesta figura a reta vertical da esquerda representa o domínio da seqüência X e a reta vertical da direita representa o domínio da seqüência Y . Considerando um ponto qualquer em cada domínio, x e y , é calculado o valor da densidade conjunta considerando $x - X_i$ e $y - Y_{i+1}$ para cada $i = \{1, 2, \dots, n\}$. O valor da densidade conjunta marginal $q_n^{(jk)}(x, y)$ para estes pontos x e y é dada por

$$q_n^{(jk)}(x, y) = \frac{1}{nh_n^2} \left[K\left(\frac{x - X_1}{h_n}\right)K\left(\frac{y - Y_2}{h_n}\right) + K\left(\frac{x - X_2}{h_n}\right)K\left(\frac{y - Y_3}{h_n}\right) + \dots + K\left(\frac{x - X_{n-1}}{h_n}\right)K\left(\frac{y - Y_n}{h_n}\right) \right].$$

onde k representa a cadeia X e j representa a cadeia Y .

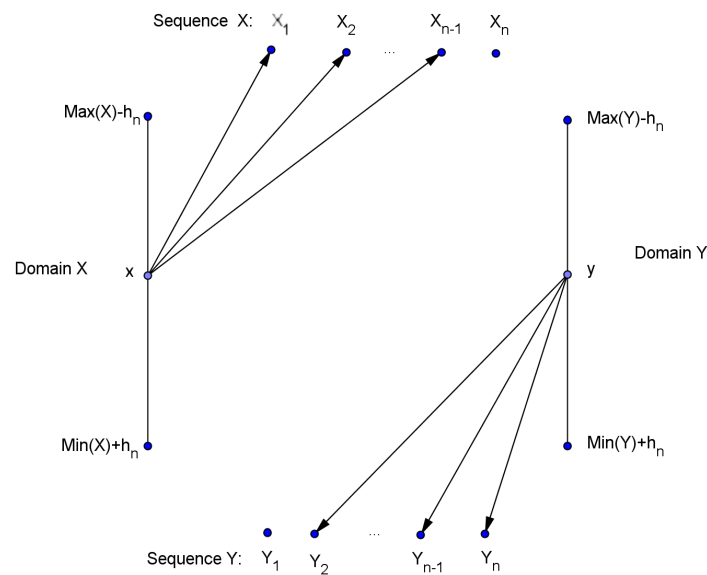


Figura 5.1: Densidade conjunta, $q_n^{(jk)}(x, y)$, da cadeia k para a cadeia j considerando os pontos x e y . k representa a cadeia X e j representa a cadeia Y .

Obs.: Para resolver o problema de estimação na fronteira a seqüência X é limitada entre $max(X) - h_n$ e $min(X) + h_n$ e a seqüência Y é limitada entre $max(Y) - h_n$ e $min(Y) + h_n$.

Neste trabalho consideramos o comprimento n da cadeia é menor que o ponto de convergência de cada cadeia utilizada para estimar o coeficiente de contribuição local, ou seja, utilizamos a parte da seqüência antes que ela atinja a estacionariedade. Assim, considerando cada uma das r seqüências $X_j = \{X_{j1}, X_{j2}, \dots, X_{j(n-1)}, X_{jn}\}$ de tamanho n e

ponto de convergência N com $N < n$, a parte da cadeia considerada é $\{X_{j1}, X_{j2}, \dots, X_{j(N-1)}\}$ para $j = \{1, 2, \dots, r\}$.

Com objetivo de ter uma boa estimativa da função de densidade determinamos o tamanho de amostra mínimo necessário para estimar f . Para cada tamanho de amostra n , uma boa estimativa da função f é considerada por f_β de modo que $f_\beta = f_N$, onde N é o menor n que satisfaz

$$\max |f_{n+1} - f_n| \leq \beta. \quad (5.9)$$

onde $\beta > 0$.

Para cada tamanho de amostra n , consideramos o tamanho de amostra mínimo necessário para estimar a função de densidade f como o menor n que satisfaz

$$\max |f_\beta - f_n| \leq \epsilon \quad (5.10)$$

onde $\epsilon > 0$.

Para resolver o problema de vício dos estimadores de f e q na fronteira optamos por restringir o domínio de estimação destas funções. Assim, o domínio de cada função é definido por

$$[Min + h; Max - h]$$

onde h é a janela ótima estimada, Min e Max são o menor e o maior valor de cada cadeia, respectivamente.

5.4 Janela Ótima

Para a aplicação da metodologia do núcleo estimador é de fundamental importância a escolha de um parâmetro de suavização ou janela h . A função desta janela é controlar o grau de suavização dos dados amostrais. Segundo Chiu [1991] esta janela não pode ser muito pequena nem muito grande. Se for muito pequena poderá apresentar um aumento na variância assintótica do estimador, se muito grande diminui a variância assintótica, mas aumenta o vício assintótico. Sendo assim preocupamos em encontrar

a janela ideal, ou *janela ótima*. A largura desta janela ótima é da ordem de $n^{-\frac{1}{5}}$ quando estimamos as funções de densidade.

Para estimar a janela ótima Chiu [1991] utiliza um estimador da função característica da amostra. Esta função característica será utilizada para estimar a derivada segunda da função de densidade f , \widehat{G} . Esta estimativa é feita pois não conhecemos a função f e consequentemente f'' .

Considerando uma amostra $X = \{X_1, X_2, \dots, X_n\}$ a função característica é dada por

$$\tilde{\phi}(\lambda) = n^{-1} \sum_{j=1}^n \exp(i\lambda X_j) \quad (5.11)$$

e portanto a estimativa de f'' é dada por

$$\widehat{G} = \frac{1}{\pi} \int_0^\Lambda \lambda^4 [|\tilde{\phi}(\lambda)|^2 - \frac{1}{n}] d\lambda \quad (5.12)$$

onde Λ é o primeiro valor de λ tal que $|\tilde{\phi}(\lambda)|^2 \leq \frac{c}{n}$ para $c > 1$.

De acordo com Chiu [1991] a janela ótima, utilizando o método plug in, é dada por

$$h_{op} = k_2^{-2/5} \left[\int K(t)^2 dt \right]^{1/5} \left[\int f''(x)^2 dx \right]^{-1/5} n^{-1/5}. \quad (5.13)$$

Podemos reescrever a equação 5.13 como

$$\hat{h}_{op} = k_2^{-2/5} \left[\int K(t)^2 dt \right]^{1/5} [\widehat{G}]^{-1/5} n^{-1/5} \quad (5.14)$$

onde k_2 é a variância do núcleo escolhido, ou seja, $[\int t^2 k(t) dt]$ e K é o núcleo escolhido.

Damasceno, E. C. em sua tese de doutorado, "Escolha do parâmetro de suavidade em estimação funcional", UFMG - 2000, faz uma análise do método Plug in para os valores $c = 1, 2$ e 3 e conclui que para $c = 3$ a variância do estimador é menor. Segundo ele, o valor de c influencia o resultado de f mesmo quando a função é suficientemente suave. Baseando nesta informação adotamos $c = 3$ neste trabalho.

5.5 Implementação

Nesta seção apresentamos um algoritmo para estimar os coeficientes de contribuição local apresentados na seção 5.1. Este algoritmo foi implementado no software MatLab e baseia-se em uma rotina, proposta por Atuncar et al. [2006], para estimar as funções de densidade f , q e t usando o método plug-in. Algumas adaptações foram feitas na estimação da função de densidade q com objetivo de estimar a função de densidade de transição cruzada t . Este algoritmo foi elaborado considerando duas cadeias, mas pode facilmente ser estendido para s cadeias. Os passos do algoritmo são:

1. Sejam $X = \{X_1, X_2, \dots, X_n\}$ e $Y = \{Y_1, Y_2, \dots, Y_n\}$ duas cadeias contínuas de tamanho n com distribuições iniciais $\pi_X(0)$ e $\pi_Y(0)$;
2. para cada cadeia, estimar o ponto de convergência da cadeia X , N_X , e da cadeia Y , N_Y . Para estimar os pontos N_X e N_Y é utilizado um critério de parada de modo que $N_X = \min\{N\}$ e $N_Y = \min\{N'\}$ satisfazendo

$$\begin{aligned} |\bar{X}_{N+1} - \bar{X}_N| &\leq \epsilon_c, \\ |\bar{Y}_{N'+1} - \bar{Y}_{N'}| &\leq \epsilon_c \end{aligned}$$

onde

- (a) ϵ_c é o erro de convergência,
- (b) \bar{X}_{N+1} e \bar{X}_N são as médias ergódicas da cadeia X considerando amostras de tamanho $N + 1$ e N elementos, respectivamente, isto é, $\bar{X}_{N+1} = \frac{1}{N+1} \sum_{i=1}^{N+1} X_i$ e $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$,

- (c) $\bar{Y}_{N'+1}$ e $\bar{Y}_{N'}$ são as médias ergódicas da cadeia Y considerando amostras de tamanho $N + 1$ e N elementos, respectivamente, isto é, $\bar{Y}_{N'+1} = \frac{1}{N'+1} \sum_{i=1}^{N'+1} Y_i$ e $\bar{Y}_{N'} = \frac{1}{N'} \sum_{i=1}^{N'} Y_i$.

Tome $N_c = \min\{N_X, N_Y\}$ como o tamanho das cadeias a serem utilizadas;

3. estimar as funções de densidade marginais f_n , q_n e t_n considerando tamanho de amostra mínimo e ponto de convergência de f_n ;
4. estimar os coeficientes de contribuição local.

O núcleo estimador utilizado nas estimativas das funções de densidade marginais foi $K \sim N(0,1)$.

Na próxima seção apresentamos várias simulações, considerando $r = 2$ cadeias.

5.6 Simulações

Nesta seção vamos realizar algumas simulações com o intuito de verificar a sensibilidade do modelo proposto em relação a algumas variáveis. Nosso objetivo principal é verificar se os valores dos coeficientes de contribuição, λ 's, são alterados quando variamos alguns parâmetros.

Em cada simulação apresentamos as configurações utilizadas com o intuito de orientar o leitor.

Foram geradas duas cadeias de Markov de tamanho N . O primeiro elemento de cada cadeia foi gerado aleatoriamente.

Alguns parâmetros utilizados nas simulações são fixos e outros variam de acordo com o objetivo de cada simulação. No início de cada simulação apresentamos os parâmetros que sofreram modificações. Os parâmetros fixos em todas as simulações são descritos a seguir:

1. margem de erro na convergência das cadeias: $\epsilon_c = 0,00001$;
2. margem de erro na convergência de f_n : $\beta = 0,001$;

3. margem de erro no cálculo do tamanho de amostra mínimo para estimar f_n :
 $\epsilon_f = 0,003$;
4. step utilizado na estimativa da convergência de f_n : $r = 10$.

Os parâmetros que variam de acordo com o objetivo de cada simulação são:

1. tamanho de amostra gerada: N ;
2. burn-in das cadeias: N_c ;
3. tamanho de amostra necessário para convergência de f_n : N_{corte} ;
4. tamanho de amostra mínimo para estimar f_n : N_{min} ;
5. tamanho da amostra utilizado para estimar os λ 's: N^* ;
6. precisão na estimação de h_n : ϵ_h ;
7. número de pontos em que a densidade foi avaliada: w ;
8. tamanho da partição do intervalo A : ϵ .

Nas simulações verificaremos o comportamento dos coeficientes de contribuição local λ 's quando alteramos os valores de: N^* , ϵ_h , w e o tipo do intervalo A .

Na próxima seção iremos verificar o comportamento dos λ 's quando o tipo do intervalo A é alterado.

5.6.1 Testando o Tipo de Intervalo

O objetivo desta simulação é verificar se os valores dos λ 's são modificados pelo intervalo de integração na equação 5.4.

Sejam $X = \{x_1, x_2, \dots, x_N\}$ e $Y = \{y_1, y_2, \dots, y_N\}$ duas cadeias de Markov geradas da seguinte maneira:

Primeiro cenário: Dependência gerada pela adição do mesmo erro aleatório em ambas as cadeias.

1. Para $j = 1$:

$$x_1 \sim N(0,1);$$

$$y_1 \sim N(2,1);$$

2. Para $j > 1$:

$$\phi \sim N(0,1)$$

$$x_j = ax_{j-1} + b + \phi;$$

$$y_j = cy_{j-1} + d + \phi;$$

onde $a, c \in (0,1)$ e $b, d \in R$.

Segundo cenário: Dependência gerada da segunda cadeia definida como a normal cuja média é o valor da primeira cadeia no tempo anterior.

$$\phi \sim N(0,1)$$

$$x_j = ax_{j-1} + b + \phi;$$

$$\beta = N(x_j, 1)$$

$$y_j = cy_{j-1} + d + \beta;$$

onde $a, c \in (0,1)$ e $b, d \in R$.

As simulações foram feitas para vários valores de a, b, c e d , mas como os resultados são todos equivalentes apenas uma configuração é apresentada. Assim consideramos

1. Para $j = 1$:

$$x_1 \sim N(0,1);$$

$$y_1 \sim N(2,1);$$

2. Para $j > 1$:

$$\phi \sim N(0,1)$$

$$x_j = 0,9x_{j-1} + 2 + \phi;$$

$$y_j = 0,5y_{j-1} + 1 + \phi;$$

Os extremos do intervalo A , $\min(A)$ e $\max(A)$, são dados por:

$$\min(A) = \max\{\min(X), \min(Y)\}$$

e

$$\max(A) = \min\{\max(X), \max(Y)\}$$

O intervalo A é dividido em w partes iguais. Cada parte tem comprimento

$$\epsilon = [\max(A) - \min(A)]/w.$$

Neste trabalho são considerados 4 tipos diferentes de intervalo A definidos a seguir:

1. Intervalo centralizado:

$$[\min(A) + n \cdot \epsilon; \max(A) - n \cdot \epsilon]$$

onde $1 \leq n \leq w/2$.

2. Variando o limite superior:

$$[\min(A); \min(A) + n \cdot \epsilon]$$

onde $w/2 \leq n \leq w - 1$.

3. Intervalo fixo - grande amplitude:

$$[\min(A) + n \cdot \epsilon; \min(A) + (n + 3) \cdot \epsilon]$$

onde $1 \leq n \leq w/2$.

4. Intervalo fixo - pequena amplitude:

$$[\min(A) + n \cdot \epsilon; \min(A) + (n + 1) \cdot \epsilon]$$

onde $1 \leq n \leq w/2$.

Nesta simulação adotamos: $N = 10.000$, $N_c = 3.690$, $N_{corte} = 3.600$, $N_{min} = 3.300$, $\epsilon_h = 0,01$, $w = 10$ e $\epsilon = 0,8192$. Nesta simulação o valor de $N^* = (N_{corte} + N_{min})/2$.

Os valores dos λ 's para todos os 4 tipos de intervalo A citados anteriormente são:

$$\lambda_{11} = 0,5090;$$

$$\lambda_{12} = 0,4910;$$

$$\lambda_{21} = 0,6760;$$

$$\lambda_{22} = 0,3240;$$

Os valores dos λ 's não são alterados quando modificamos o tipo do intervalo A . Em todas as outras simulações, indicadas no primeiro e segundo cenário, os resultados dos λ 's também não foram alterados pela mudança do intervalo A .

Figura 5.2 mostra o comportamento da média ergódica considerando as cadeias X e Y . A média ergódica foi utilizada para determinar o burn-in nas cadeias.

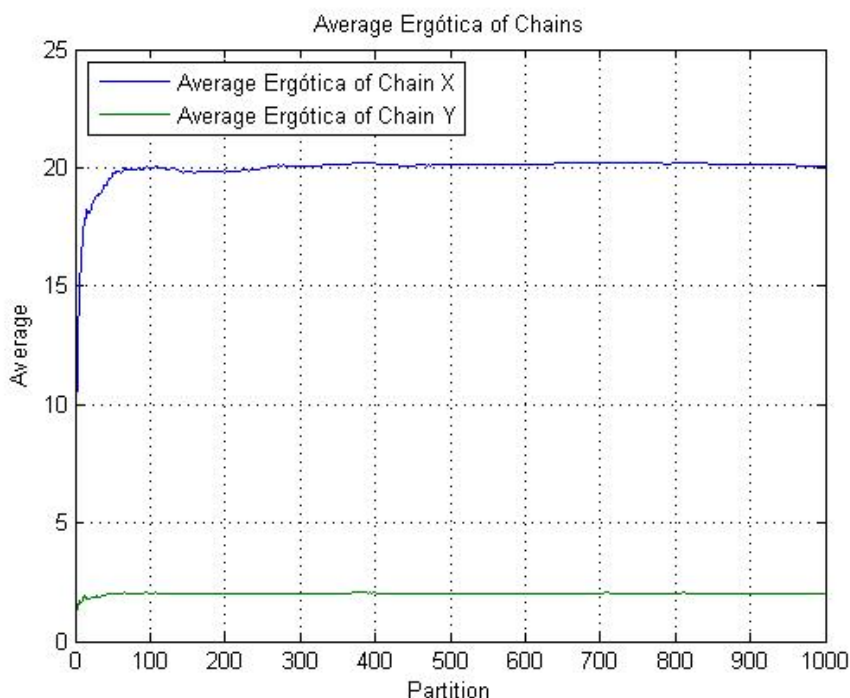


Figura 5.2: Média ergódica das cadeias X e Y

Figura 5.3 apresenta o comportamento da função de densidade de probabilidade: (a) da cadeia X e (b) da cadeia Y . Em cada figura temos 3 funções de densidade:

$f_n(Utilized)$ é a função de densidade utilizada na estimação dos λ 's, $f_n(cut)$ é a função de densidade estacionária estimada ou f_β e $f_n(Minimalsample)$ é a função de densidade utilizando amostra mínima. Podemos observar que, em cada gráfico, as funções de densidade são bem estimadas.

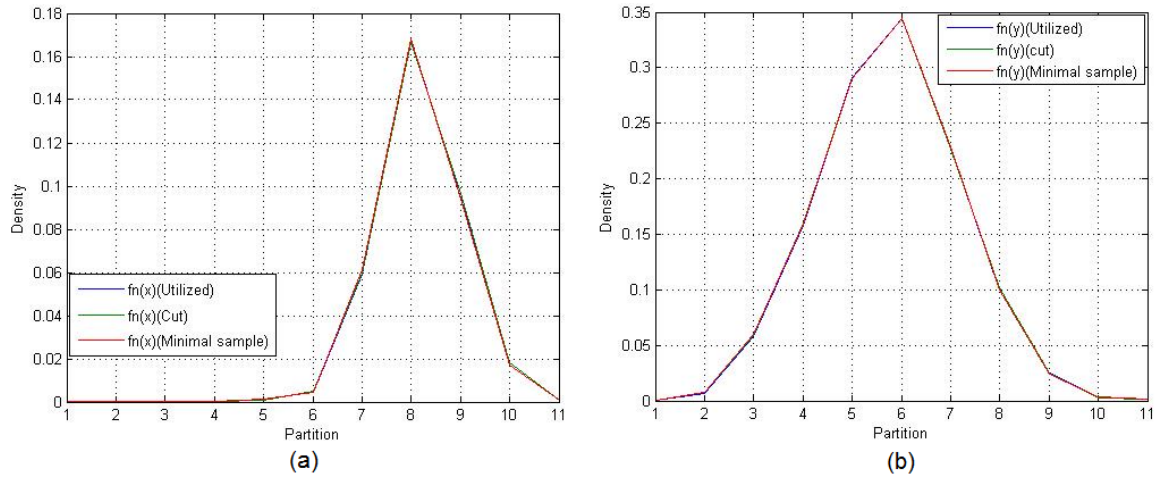


Figura 5.3: Função de densidade de probabilidade: (a) cadeia X e (b) cadeia Y

Para todos os tipos de intervalo A , e considerando as mesmas cadeias, os resultados são

$$\begin{cases} t_{n+1}^{(11)}(y|x) = 0,5090t_n^{(11)}(y|x) + 0,4910t_n^{(12)}(y|x) \\ t_{n+1}^{(22)}(y|x) = 0,6760t_n^{(21)}(y|x) + 0,3240t_n^{(22)}(y|x) \end{cases}$$

Observamos que a densidade de transição cruzada marginal $t_n^{(12)}$ contribui no cálculo da densidade de transição marginal $t_{n+1}^{(11)}$. Neste caso o coeficiente de contribuição local é de $\lambda_{12} = 0,4910$.

A densidade de transição cruzada marginal $t_n^{(21)}$ também contribui no cálculo da densidade de transição marginal $t_{n+1}^{(22)}$.

Na próxima seção verificaremos o comportamento dos λ 's quando o tamanho de amostra N^* é variado.

5.6.2 Testando o Tamanho de Amostra Efetivo N^*

O objetivo desta simulação é verificar o comportamento dos λ 's quando o tamanho de amostra N^* é alterado. O intervalo $(N_{corte} - N_{min})$ foi dividido em 4 partes. Assim

$$N_i^* = N_{min} + \frac{(N_{corte} - N_{min}) * i}{4}$$

onde $i = \{0,1,2,3\}$.

Sejam $X = \{x_1, x_2, \dots, x_N\}$ e $Y = \{y_1, y_2, \dots, y_N\}$ duas amostras de cadeias de Markov geradas da seguinte maneira :

1. Para $j = 1$:

$$x_1 \sim N(0,1);$$

$$y_1 \sim N(2,1);$$

2. Para $j > 1$:

$$\phi \sim N(0,1)$$

$$x_j = 0,9x_{j-1} + 2 + \phi;$$

$$y_j = 0,4y_{j-1} + 1 + \phi;$$

Os tipos de intervalo utilizados nesta simulação são os mesmos utilizados na simulação da seção 5.6.1, tipo 4 considerando $n = 1$.

Nesta simulação consideramos: $N = 20.000$, $N_c = 8.420$, $N_{corte} = 5.860$, $N_{min} = 5.210$, $\epsilon_h = 0,01$, $w = 20$ e $\epsilon = 0,1061$

Tabela 5.1 mostra os resultados dos λ 's para cada tamanho de amostra efetiva N^* . Observamos que os valores dos λ 's são alterados de acordo com o tamanho de amostra N^* .

O resultado para $N = 5.535$ é

$$\begin{cases} t_{n+1}^{(11)}(y|x) = 0,6449t_n^{(11)}(y|x) + 0,3551t_n^{(12)}(y|x) \\ t_{n+1}^{(22)}(y|x) = 0,7482t_n^{(21)}(y|x) + 0,2518t_n^{(22)}(y|x) \end{cases}$$

Tabela 5.1: Valores dos λ 's por tamanho de amostra efetiva N^*

Coeficientes	N^*			
	5.373	5.535	5.698	5.860
λ_{11}	0,5967	0,6449	0,6493	0,5404
λ_{12}	0,4033	0,3551	0,3507	0,4596
λ_{21}	0,7391	0,7482	0,7513	0,8943
λ_{22}	0,2609	0,2518	0,2487	0,1057

Observamos que a densidade de transição cruzada marginal $t_n^{(12)}$ contribui no cálculo da densidade de transição marginal $t_{n+1}^{(11)}$ e o peso desta contribuição é de $\lambda_{12} = 0,3551$. O mesmo acontece com a densidade $t_n^{(21)}$ com peso de $\lambda_{21} = 0,3551$ no cálculo de $t_{n+1}^{(22)}$.

Na próxima seção apresentamos uma simulação mostrando o comportamento dos λ 's quando o número de intervalos w é alterado.

5.6.3 Testando o Número de Intervalos w

Neste estudo de simulação, o número de intervalos w foi variado para verificar como os coeficientes de contribuição se comportam. O valor de w é utilizado como número de partições do domínio da função de densidade. São estimadas as funções de densidade f_n , q_n e t_n para cada valor de w . Também estimamos os valores dos coeficientes de contribuição, λ 's, para cada valor de w .

Considere h a janela ótima estimada. A amplitude $(max - min - 2h)$ foi dividida em w intervalos iguais onde cada intervalo tem comprimento Δ definido por

$$\Delta = (max - min - 2h)/w,$$

Sejam $X = \{x_1, x_2, \dots, x_N\}$ e $Y = \{y_1, y_2, \dots, y_N\}$ duas cadeias de Markov geradas da seguinte maneira :

1. Para $j = 1$:

$$x_1 \sim N(0,1);$$

$$y_1 \sim N(2,1);$$

2. Para $j > 1$:

$$\phi \sim N(0,1)$$

$$x_j = 0,8x_{j-1} + 3 + \phi;$$

$$y_j = 0,3y_{j-1} + 2 + \phi;$$

Os tipos de intervalo utilizados nesta simulação são os mesmos utilizados na simulação da seção 5.6.2.

Para determinar $N^* = (N_{min} + N_{corte})/2$ o número de intervalos foi fixado em $w = 30$. Após determinar N^* o número de intervalos foi modificado de acordo com os valores apresentados na tabela 5.2.

Os valores dos parâmetros variáveis são: $N = 10.000$, $N_c = 3.900$, $N_{corte} = 3.870$, $N_{min} = 3.790$, $\epsilon_h = 0,01$ e $\epsilon = 0,1393$.

Tabela 5.2 apresenta os resultados dos λ 's para cada número de intervalos w . Observamos que o valor de w modifica os valores dos λ 's.

Tabela 5.2: Valores dos λ 's para cada número de intervalo w

Coeficientes	w			
	10	20	30	40
λ_{11}	1	1	1	1
λ_{12}	0	0	0	0
λ_{21}	0,8407	0,8656	0,8658	0,8767
λ_{22}	0,1593	0,1344	0,1342	0,1233

De acordo com a tabela 5.2, o resultado para $w = 40$ é dado por:

$$\begin{cases} t_{n+1}^{(11)}(y|x) = t_n^{(11)}(y|x) \\ t_{n+1}^{(22)}(y|x) = 0,8767t_n^{(21)}(y|x) + 0,1233t_n^{(22)}(y|x) \end{cases}$$

Observamos que a densidade de transição cruzada marginal $t_n^{(12)}$ não contribui no cálculo da densidade de transição marginal $t_{n+1}^{(11)}$, neste caso a densidade $t_{n+1}^{(11)}$ só depende dela mesma no tempo n . Já a densidade $t_n^{(21)}$ contribui fortemente com peso de $\lambda_{21} = 0,8767$ no cálculo de $t_{n+1}^{(22)}$.

Na próxima seção apresentamos uma simulação que mostra o comportamento dos λ 's quando a precisão h , ϵ_h , é variada.

5.6.4 Testando a Precisão ϵ_h

Nesta simulação é testada a precisão ϵ_h que é utilizada na estimação da janela ótima h . Testamos os seguintes valores para a precisão: $\epsilon_h = \{0,006; 0,008; 0,01; 0,012; 0,014; 0,016; 0,018; 0,02\}$. Para cada valor de ϵ_h foi calculado os valores dos λ 's.

Para determinar $N^* = (N_{min} + N_{corte})/2$ a precisão inicial foi fixada em $\epsilon_h = 0,01$. Após determinar N^* variamos os valores da precisão. O tipo de intervalo A e a lei de formação das cadeias são os mesmos da simulação da seção 5.6.2.

Os parâmetros variáveis são: $N = 10.000$, $N_c = 3.160$, $N_{corte} = 3.120$, $N_{min} = 3.010$, $w = 30$ e $\epsilon = 0,1932$.

Para todos os valores de ϵ_h os resultados dos λ 's são:

$$\begin{aligned}\lambda_{11} &= 1; \\ \lambda_{12} &= 0; \\ \lambda_{21} &= 1; \\ \lambda_{22} &= 0;\end{aligned}$$

Note os os resultados dos λ 's não são modificados pela variação na precisão. A densidade de transição cruzada marginal $t_n^{(12)}$ não contribui no cálculo da densidade de transição marginal $t_{n+1}^{(11)}$. Contudo, a densidade de transição cruzada marginal $t_n^{(21)}$ contribui fortemente para o cálculo da densidade de transição $t_{n+1}^{(22)}$. Neste caso a densidade de transição $t_n^{(22)}$ é totalmente explicada pela densidade de transição $t_{n+1}^{(21)}$.

Em todas as outras simulações os λ 's não foram modificados pela variação da precisão.

5.6.5 Comportamento dos λ 's em Cadeias de Markov com Intervalo Absorvente

Nesta seção apresentamos uma simulação quando as cadeias têm um estado absorvente. Considerando uma cadeia de Markov define-se:

Definição: O estado j é acessível a partir de um estado i se, e somente se, existe um $s > 0$ de modo que $P_{ij}(s) > 0$.

Definição: O estado i comunica com o estado j se, e somente se, i é acessível a partir de j e j é acessível a partir de i .

Pela definição anterior, um estado j comunica com o estado i se podemos transitar do estado i para o estado j e vice versa.

Definição: O estado i é absorvente se ele não comunica com nenhum outro estado j .

Pela definição anterior, quando uma cadeia de Markov entra em um estado absorvente ela não mais sai deste estado. Assim a cadeia permanece infinitamente neste estado. Neste caso, para estimar os coeficientes de contribuição λ 's é necessário utilizar a parte da cadeia anterior ao ponto de absorção.

Inúmeras simulações foram feitas. Por simplicidade apresentaremos apenas uma. Nesta simulação é utilizado o intervalo centralizado e os núcleos são dados pelas equações 5.6, 5.7 e 5.8.

Sejam $X = \{x_1, x_2, \dots, x_N\}$ e $Y = \{y_1, y_2, \dots, y_N\}$ duas cadeias de Markov geradas da seguinte maneira :

1. Para $j = 1$:

$$x_1 \sim U(0,1);$$

$$y_1 \sim U(0,1);$$

2. Para $j > 1$:

$$\phi \sim U(0,1)$$

$$x_j = 0,9x_{j-1} + \phi;$$

$$y_j = 0,5y_{j-1} + \phi;$$

Adotamos os intervalos de absorção como sendo $]6, +\infty[$. Assim, se $x_j > 6$ então consideramos $x_i = 0$ e se $y_j > 6$ consideramos $y_i = 0$ para todo $i > j$.

Os valores dos parâmetros variáveis são: $N = 1.000$, $N_c = 86$, $N_{corte} = 80$, $N_{min} = 80$, $w = 10$ e $\epsilon = 0,1$. Adotamos $N^* = (N_{corte} + N_{min})/2$. O ponto considerado como o ponto em que a cadeia alcança o intervalo absorvente é $N_c - 1 = 85$.

Os resultados dos λ 's foram:

$$\lambda_{11} = 0,8357;$$

$$\lambda_{12} = 0,1643;$$

$$\lambda_{21} = 1;$$

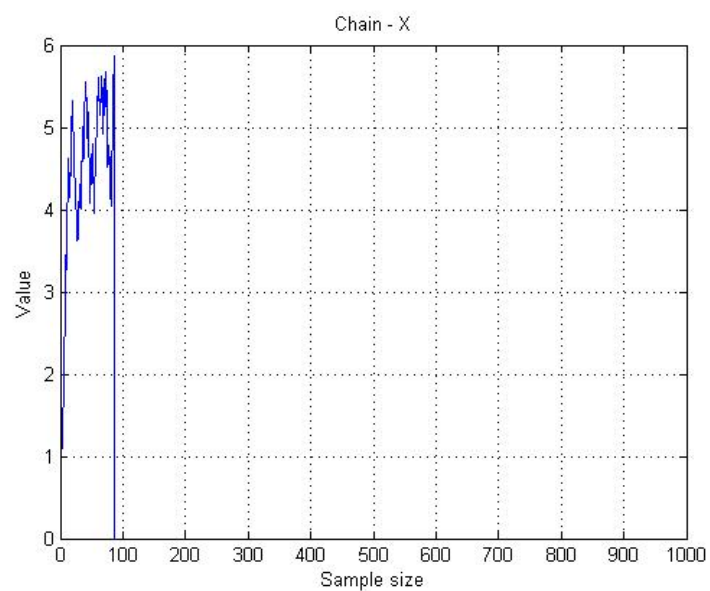
$$\lambda_{22} = 0;$$

Figura 5.4 mostra o comportamento das cadeias X e Y .

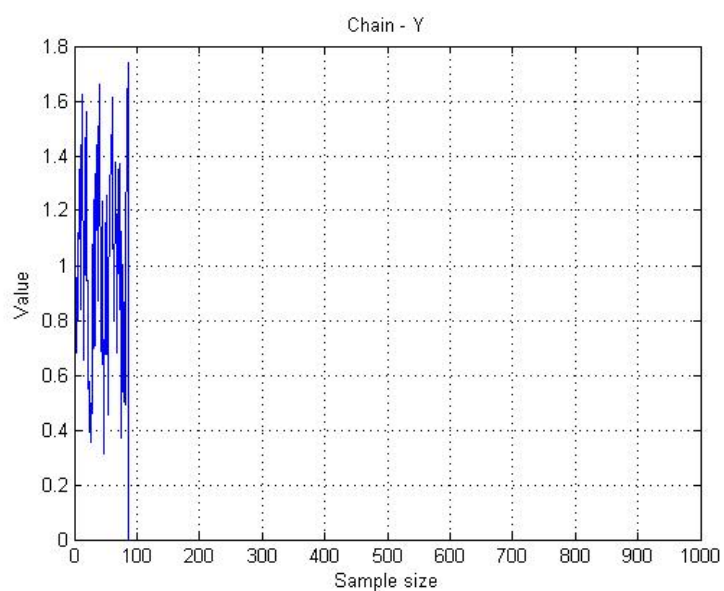
Neste caso encontramos

$$\begin{cases} t_{n+1}^{(11)}(y|x) = 0,8357t_n^{(11)}(y|x) + 0,1643t_n^{(12)}(y|x) \\ t_{n+1}^{(22)}(y|x) = t_n^{(21)}(y|x) \end{cases}$$

Observamos que a densidade de transição $t_n^{(11)}(y|x)$ contribui fortemente para o cálculo de $t_{n+1}^{(11)}(y|x)$ cujo coeficiente de contribuição igual a $\lambda_{11} = 0,8357$. Já a densidade $t_n^{(21)}(y|x)$ contribui totalmente para o cálculo da densidade $t_{n+1}^{(22)}(y|x)$.



(a)



(b)

Figura 5.4: (a) Comportamento da cadeia X e (b) Comportamento da cadeia Y

5.7 Estimação da Variância dos Coeficientes de Contribuição

O objetivo desta seção é estimar a variância dos coeficientes de contribuição.

A variância do estimador de mínimos quadrados dos coeficientes λ 's é estimada através de

$$\sigma^2(\lambda) = \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1} \quad (5.15)$$

onde R é o vetor de restrições de igualdade, $R\lambda = r$, e σ^2 é o estimador da variância da cadeia. As restrições consideradas são: $\lambda_{11} + \lambda_{12} = 1$ e $\lambda_{21} + \lambda_{22} = 1$. O estimador 5.15 pode ser encontrado em (<http://www2.econ.iastate.edu/>) "Restricted least squares, hypothesis testing, and prediction in the classical linear regression model".

5.7.1 Estimativa da Variância da Cadeia

Para estimar σ^2 de uma cadeia X utilizamos block bootstrap, apresentado na seção 4.8.1.

Nesta simulação consideramos $X = \{x_1, x_2, \dots, x_N\}$ e $Y = \{y_1, y_2, \dots, y_N\}$ duas amostras de cadeias de Markov de ordem 1 geradas da seguinte maneira:

1. Para $j = 1$:

$$x_1 \sim N(0,1);$$

$$y_1 \sim N(0,1);$$

2. Para $j > 1$:

$$\phi \sim N(0,1)$$

$$x_j = 0,6x_{j-1} + 1 + \phi;$$

$$y_j = 0,4y_{j-1} + 2 + \phi;$$

Para as estimativas de $\sigma^2(X)$ e $\sigma^2(Y)$ utilizamos 5.15 e foram geradas amostras de tamanho $N = 10.000$ e 500 reamostras. Neste caso o vetor X é definido como

$$\left[\int_{-\infty}^{+\infty} \left[\int_A (\lambda_{(XX)} t_n^{(XX)}(y|x)) dy \right] f_n^{(X)}(x) dx; \int_{-\infty}^{+\infty} \left[\int_A (\lambda_{(XY)} t_n^{(XY)}(y|x)) dy \right] f_n^{(Y)}(x) dx \right]$$

para a estimativa dos $\lambda_{(XX)}$ e $\lambda_{(XY)}$. Para as estimativas dos $\lambda_{(YX)}$ e $\lambda_{(YY)}$ utilizamos o vetor X como

$$\left[\int_{-\infty}^{+\infty} \left[\int_A (\lambda_{(YX)} t_n^{(YX)}(y|x)) dy \right] f_n^{(X)}(x) dx; \int_{-\infty}^{+\infty} \left[\int_A (\lambda_{(YY)} t_n^{(YY)}(y|x)) dy \right] f_n^{(Y)}(x) dx \right]$$

Os tipos de intervalo utilizados nesta simulação são os mesmos utilizados na simulação da seção 5.6.1, tipo 4 considerando $n = 1$.

Nesta simulação consideramos: $N_c = 1.850$, $N_{corte} = 1.840$, $N_{min} = 1.840$ e $w = 10$. O tamanho de cada bloco foi de 153 elementos e assim considerou-se 12 blocos.

Os resultados dos λ 's foram:

$$\lambda_{XX} = 0;$$

$$\lambda_{XY} = 1;$$

$$\lambda_{YX} = 0;$$

$$\lambda_{YY} = 1;$$

Os resultados das variâncias estimadas dos λ 's foram:

$$\widehat{\sigma}^2(\lambda_{XX}) = 0.0008;$$

$$\widehat{\sigma}^2(\lambda_{XY}) = 0.0008;$$

$$\widehat{\sigma}^2(\lambda_{YX}) = 0.0002;$$

$$\widehat{\sigma}^2(\lambda_{YY}) = 0.0002;$$

Outras simulações foram feitas cujos resultados da estimação da variância foram semelhantes. Percebemos que o estimador tem variância muito pequena.

5.8 Conclusões Sobre as Simulações

De acordo com as simulações concluímos que os parâmetros λ 's sofreram alteração quando variamos:

1. o tamanho da amostra efetiva N^* ;
2. o número de pontos em que a densidade foi avaliada: w ;

Os parâmetros λ 's não sofreram alteração quando variamos:

1. o tipo de intervalo A ;
2. a precisão no cálculo da janela ótima ϵ_h .

De acordo com as simulações percebe-se que o modelo independe do tipo de intervalo A e da precisão ϵ_h . Isto é importante pois estas variáveis não afetam na estimação dos coeficientes de contribuição podendo torna-las fixas durante o processo de estimação.

Já as variáveis N^* e w influenciam na estimação dos coeficientes de contribuição. É importante que, neste caso, de acordo com a aplicação do modelo a ser utilizada, se faça uma escolha de qual valor destes parâmetros devem serem adotados.

Propomos uma metodologia de encontrar os coeficientes de contribuição entre cadeias de Markov não estacionárias ou absorventes. Isto será muito útil, como veremos na seção 5.9 pois muitos processos ainda não entraram em regime estacionário. Assim é importante estudar estes processos antes de atingir o regime estacionário.

5.9 Aplicação

Ilustramos a metodologia descrita com uma aplicação a dados reais. Essa aplicação só foi possível devido à colaboração da VE Engenharia que nos forneceu dados da máquina de trefilar 16032 localizada na Arcelor Mital na cidade de Sabará. Um dos trabalhos realizados pela VE Engenharia é o monitoramento e diagnóstico de defeitos em componentes desta máquina de trefilar através da manutenção preditiva. Analisamos os dados de medições dos valores globais de aceleração do rolamento dianteiro de cada motor da máquina de trefilar 16032.

5.9.1 Controle de Manutenção

A manutenção é uma prática que visa, entre outras coisas, à conservação e ao correto funcionamento de equipamentos e componentes.

Quando tratamos de política de manutenção podemos ter:

1. Manutenção corretiva: aquela em que espera ocorrer a quebra do componente ou defeito do equipamento para poder efetuar a troca ou conserto.
2. Manutenção preventiva: aquela manutenção programada.
3. Manutenção preditiva: aquela realizada de acordo com as condições do componente ou equipamento.
4. Manutenção próativa: aquela em que busca-se a eliminação das causas de um possível defeito.

Ao executar uma manutenção espera-se:

1. Redução dos custos de manutenção devido à melhor previsão de falhas e programação de intervenção evitando a parada não programada.
2. Aumento da eficiência das intervenções proporcionando uma manutenção de qualidade.
3. Aumento da disponibilidade do equipamento melhorando a produção.
4. Aumento da confiabilidade do componente ou equipamento.

A vibração, definida como "movimento oscilatório de um corpo, excitado por uma força, em relação a um ponto de referência", irá contribuir para o diagnóstico de possíveis defeitos em componentes e equipamentos. Na abordagem da manutenção preditiva uma das possibilidades é a utilização da "Análise da Vibração", que é uma técnica que investiga a causa das vibrações em sistemas mecânicos através da análise das amplitudes de uma determinada frequência.

Como unidade de amplitude de vibração temos a "Aceleração" expressa em "G" no qual $1G = 9,81m/s^2$.

Dentre as fontes mais comuns de vibração em equipamentos ou componentes temos:

1. Desbalanceamento de massas rotativas.
2. Rotores excêntricos ou empenados.
3. Eixo empenado.
4. Desalinhamentos em geral.
5. Ressonâncias.
6. Folgas.
7. Correias desgastadas ou frouxas.
8. Problemas Elétricos em Sistemas e Motores.
9. Rolamentos danificados ou inadequados.

5.9.2 A Técnica de Envelope de Aceleração

A Demodulação de sinais de alta frequência ou Técnica de Envelope, ferramenta de análise de vibrações para detecção de falhas e diagnóstico de falhas em rolamentos, desenvolvida pela SKF, é considerada atualmente uma das mais poderosas ferramentas na avaliação das condições mecânicas destes elementos. Sua grande vantagem consiste na detecção de defeitos, mesmo incipientes, permitindo o acompanhamento da condição do rolamento mesmo a partir do início da falha.

Considere um rolamento que possui um pequeno defeito em sua pista externa. Cada vez que uma esfera passa por este defeito temos um choque entre a superfície defeituosa e a superfície da esfera. Se um rolamento possui oito esferas, a cada rotação do trem de esferas temos oito choques sobre o defeito. Admitindo que o espaçamento entre as esferas no trem é o mesmo, teremos esses choques se repetindo periodicamente. Estes choques são de curtíssima duração, repetindo-se em baixa frequência e irão excitar as frequências de ressonância (alta frequência) do mancal ou da estrutura do equipamento onde o rolamento está montado. É possível comparar o fenômeno ao das badaladas de um sino: quando o abalo atinge o sino temos um choque e o som de alta

intensidade que se prolonga mesmo após o choque com um rápido decaimento. Tal fenômeno não se restringe apenas à pista externa do rolamento, mas também à pista interna, à gaiola e aos elementos rolantes deste. Desta forma, podemos concluir que defeitos em rolamentos geram séries de choques que se repetem de forma periódica e que cada rolamento defeituoso produzirá tais choques em frequências características. Estas frequências podem ser calculadas em função das características geométricas do rolamento.

As carcaças de mancais se caracterizam por serem estruturas bastante rígidas, cuja resposta dinâmica em altas frequências independe de sua fundação. Isso indica que a influência de outras fontes de vibração será muito pequena nesta faixa de frequência, isto é, não teremos o mascaramento dos sinais de rolamento como acontece em frequências mais baixas. Sendo assim, teremos uma maior qualidade na análise dos sinais em alta frequência, embora sem maior facilidade de diagnóstico.

A técnica de envelope nada mais é que um processo eletrônico na qual os sinais de defeitos de rolamentos que aparecem com destaque em alta frequência são transferidos para uma faixa de baixa frequência na qual pode ser detectado com maior nitidez.

5.9.3 A Máquina de Trefilar 16032

A título de informação adicional a função da máquina de trefilar é diminuir o diâmetro transversal do arame. O arame é reduzido ao passar por fieiras. Ele é arrastado por bobinas, num total de 7, que são acionadas por motores. Em cada bobina temos os sensores que controlam a velocidade do motor com o objetivo de regular o volume de arame na bobina. Estes motores funcionam independentes mas o processo é dependente visto que se qualquer motor estragar o processo é interrompido.

Figura 5.5 apresenta, de forma reduzida, os dois primeiros blocos de uma máquina de trefilar.

Nas próximas figuras apresentamos as fotos de vários componentes da máquina de trefilar. Na Figura 5.6 podemos visualizar uma máquina de trefilar desde o abastecimento, rolo de arame ao fundo, o primeiro bloco até o último bloco.

Na Figura 5.7 podemos visualizar o rebobinador.

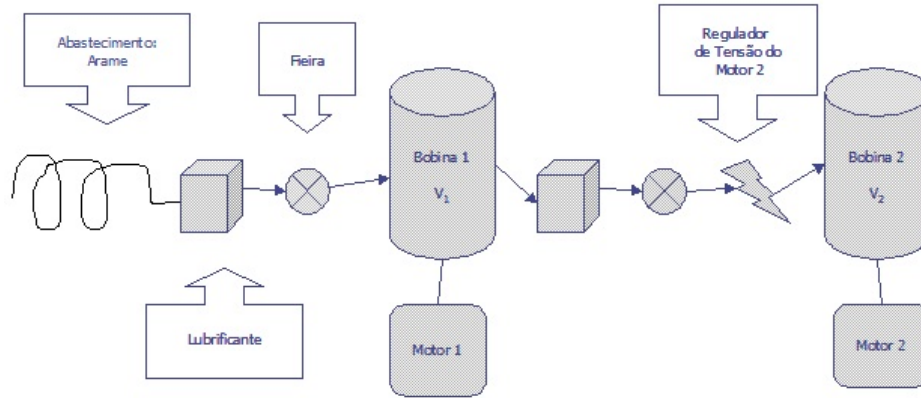


Figura 5.5: Modelo de uma Máquina de trefilar



(a)



(b)

Figura 5.6: Foto de uma Máquina de trefilar



(a)



(b)

Figura 5.7: Foto de um rebobinador de uma Máquina de trefilar

A seguir a função dos componentes básicos de uma máquina de trefilar é descrita.

Bobina: responsável pelo arraste do arame. Quando há acúmulo de arame em uma bobina, a máquina é programada para que somente as bobinas seguintes a este continuem em funcionamento, para que haja uma redução na quantidade de arame naquela bobina.

Lubrificante: responsável pela lubrificação do arame, aumentando o tempo de vida da fieira. Normalmente é apenas sabão.

Na figura 5.8 temos uma foto real de uma caixa de fieira juntamente com o sabão lubrificante.



Figura 5.8: Caixa de lubrificante e de fieira de uma Máquina de trefilar

Fieira: responsável pela redução da bitola do arame. Seu interior é feita de vídia e pode ser vista em corte longitudinal, parte (a), na Figura 5.9 e foto real, parte (b).

A cada estágio o diâmetro transversal do arame é reduzido de modo a não danificá-lo. Este arame é utilizado para fabricação de eletrodos para solda, grampos para grampeador etc. Em uma máquina de trefilar temos: $V_1 < V_2 < \dots < V_n$ sendo que n é a quantidade de bobinas. A velocidade final da última bobina é, em média, de 11 a 15 m/s.

Neste trabalho foram analisados os dados dos rolamentos dianteiro de cada um



Figura 5.9: (a) Vista em corte longitudinal e (b) foto real de uma fieira uma Máquina de trefilar

dos motores dos 7 blocos que compõem a máquina de trefilar 16032. Neste caso não foi considerado o rebobinador. A Figura 5.10 mostra os pontos onde são medidos as vibrações referentes aos rolamentos dianteiro e trazeiro do motor. Neste trabalho os dados são referentes ao rolamento definido pela posição 2.

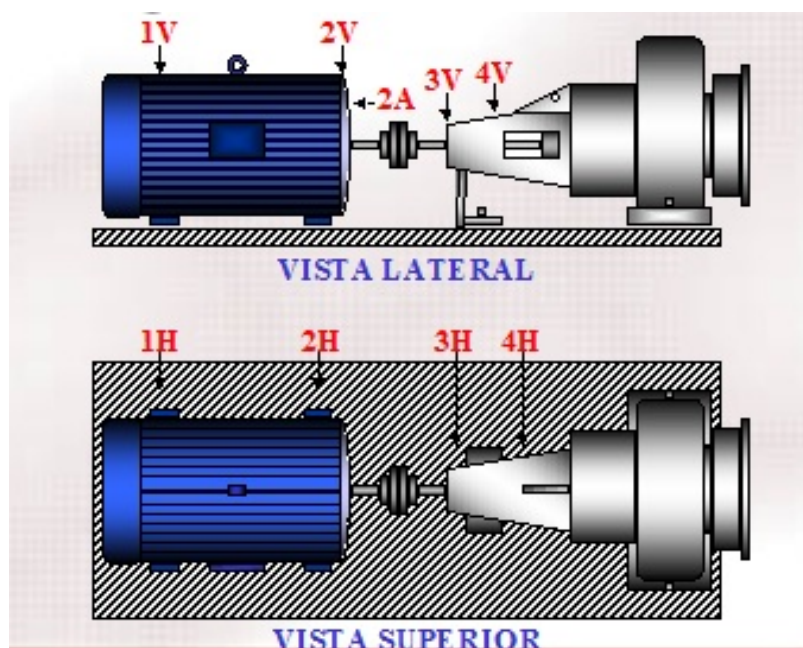


Figura 5.10: Pontos de medição de vibração do motor de uma Máquina de trefilar

5.9.4 Análise dos dados da máquina de Trefilar 16032

A seguir apresentamos os espectros de aceleração global de cada um dos rolamentos dianteiros dos motores da máquina de trefilar 16032.

A Figuras 5.11 a 5.17 apresentam os espectros dos roloamentos dianteiros dos blocos 1 a 7.

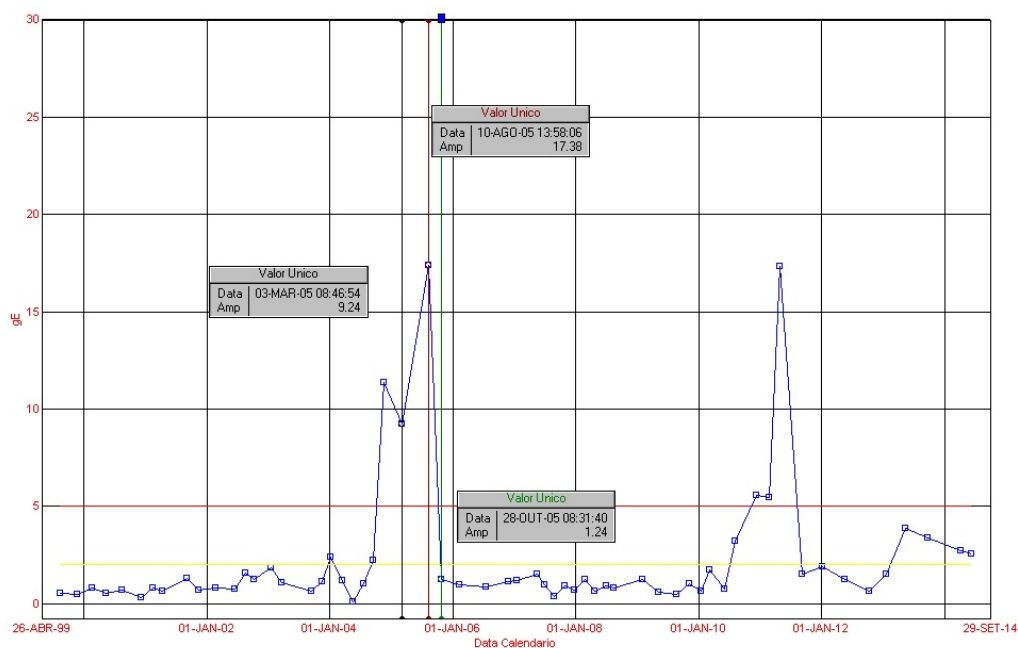


Figura 5.11: Espectro do rolamento dianteiro do bloco 1 da máquina de trefilar 16032

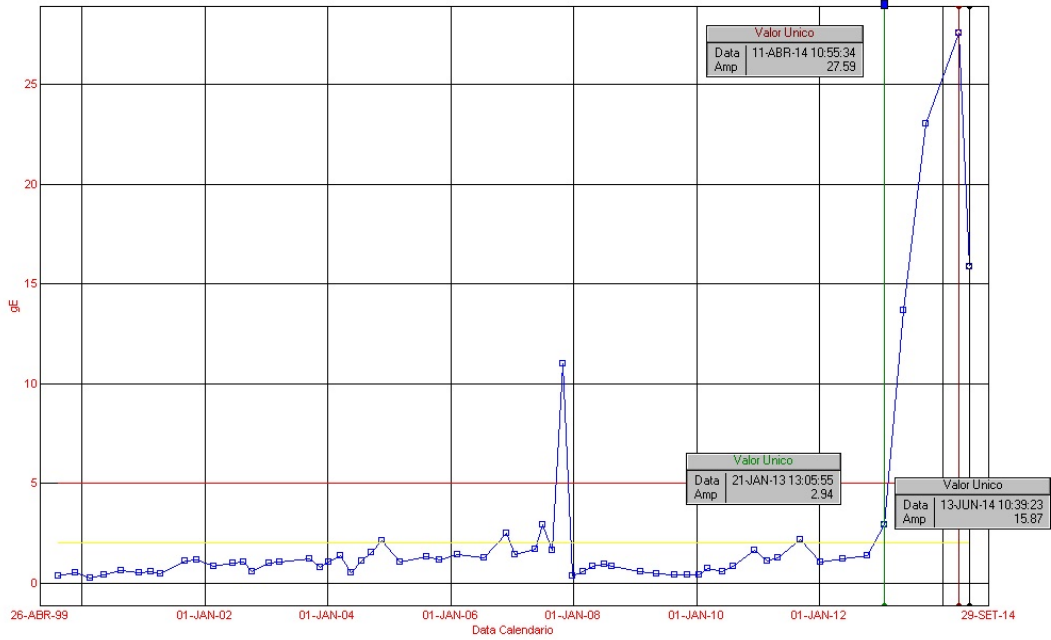


Figura 5.12: Espectro do rolamento dianteiro do bloco 2 da máquina de trefilar 16032

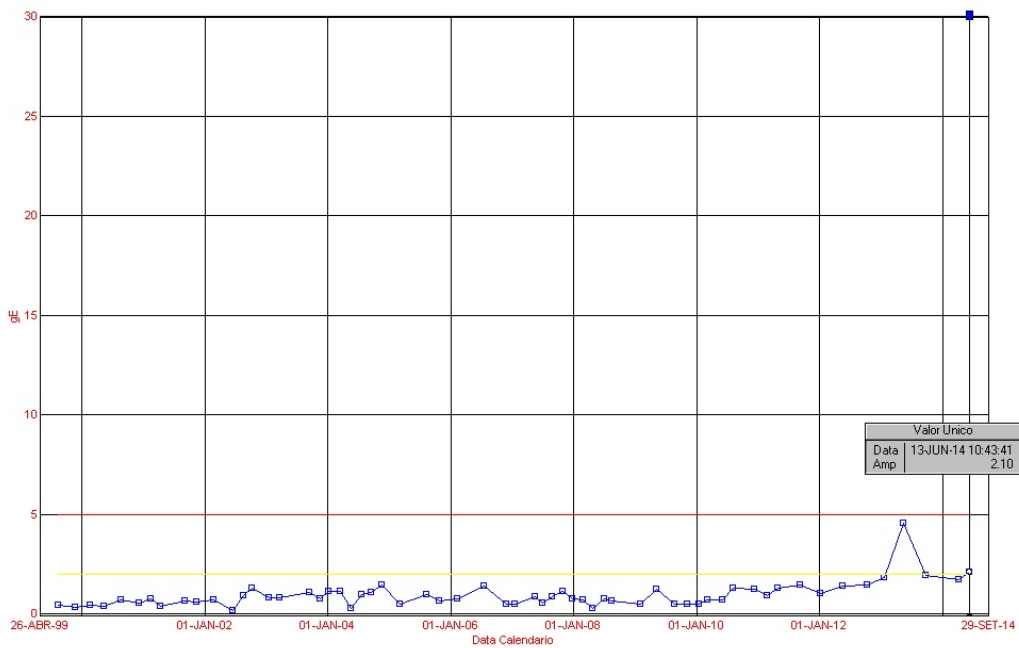


Figura 5.13: Espectro do rolamento dianteiro do bloco 3 da máquina de trefilar 16032

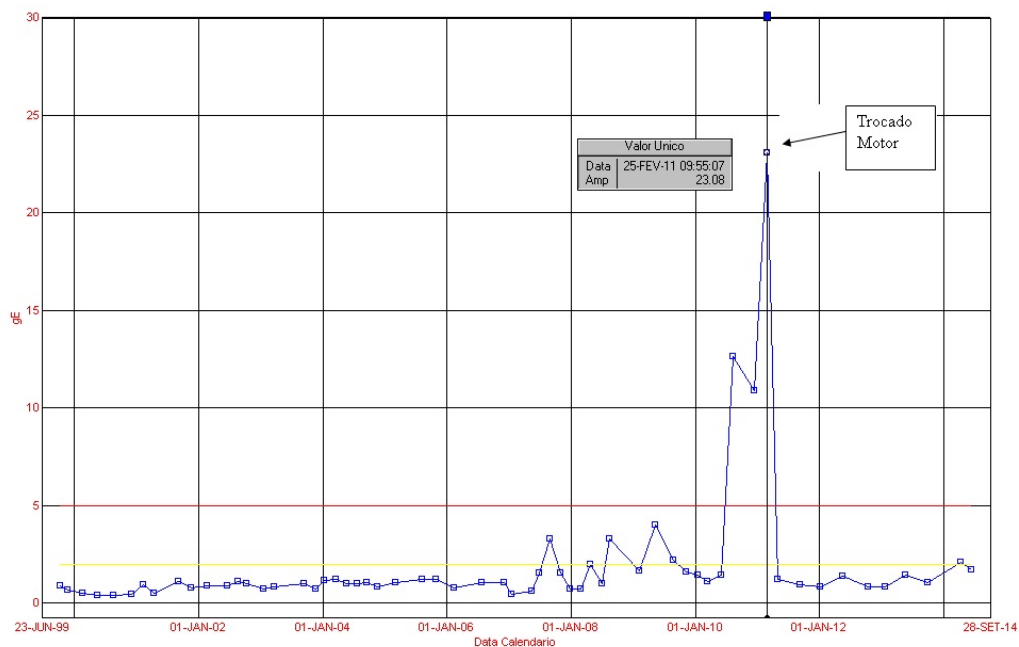


Figura 5.14: Espectro do rolamento dianteiro do bloco 4 da máquina de trefilar 16032

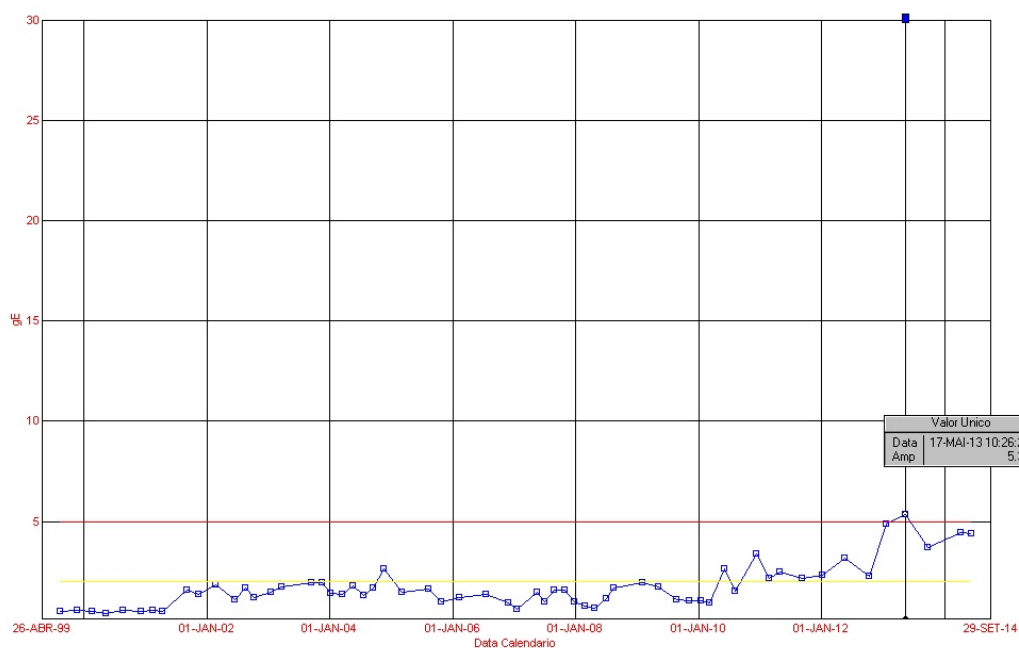


Figura 5.15: Espectro do rolamento dianteiro do bloco 5 da máquina de trefilar 16032

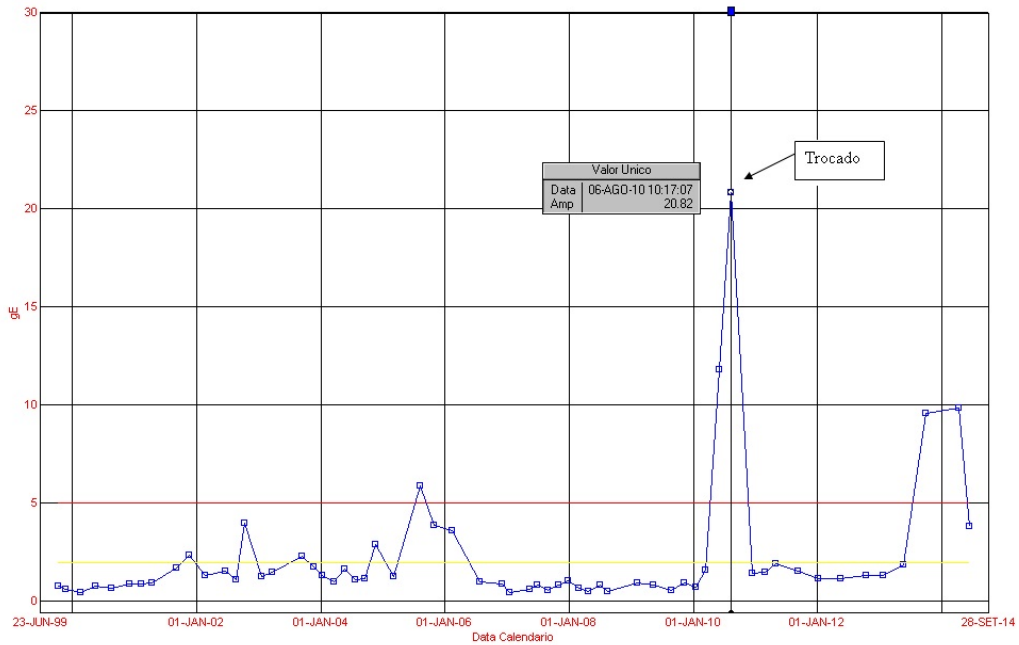


Figura 5.16: Espectro do rolamento dianteiro do bloco 6 da máquina de trefilar 16032

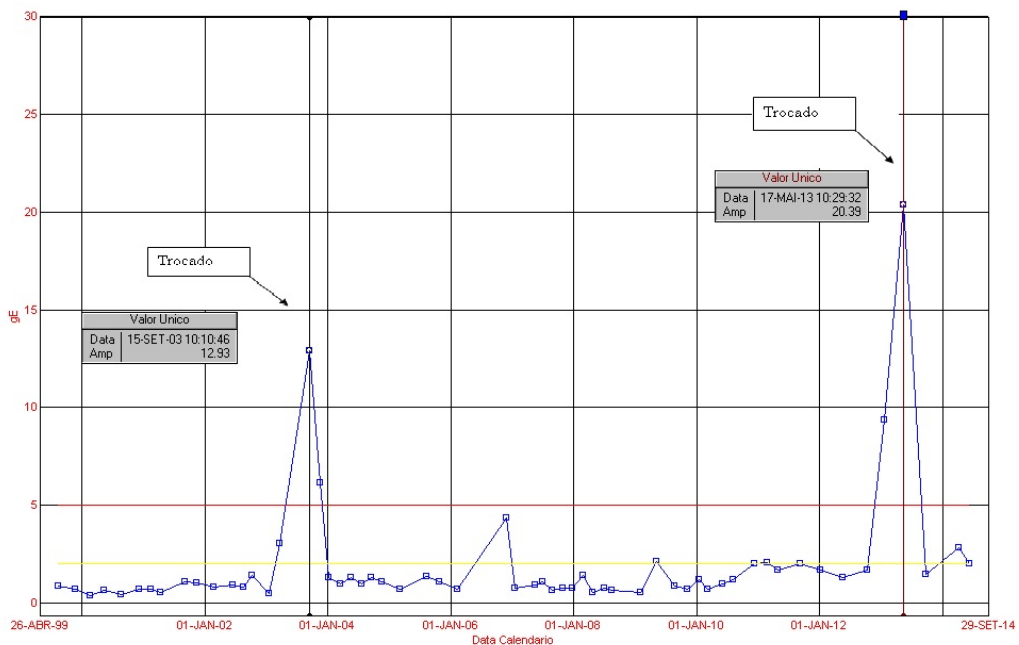


Figura 5.17: Espectro do rolamento dianteiro do bloco 7 da máquina de trefilar 16032

Nosso objetivo é encontrar a contribuição da função de densidade marginal entre cada função de densidade marginal e as demais. Esta contribuição é expressa pelos coeficientes de contribuição. Assim pode-se verificar o comportamento do nível de aceleração de cada bloco com os demais.

Os coeficientes de contribuição devem satisfazer a equação matricial

$$\begin{pmatrix} t_{n+1}^{(11)} \\ t_{n+1}^{(12)} \\ \vdots \\ t_{n+1}^{(17)} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{17} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{27} \\ \vdots & & & \\ \lambda_{71} & \lambda_{72} & \dots & \lambda_{77} \end{pmatrix} \begin{pmatrix} t_n^{(11)} \\ t_n^{(12)} \\ \vdots \\ t_n^{(17)} \end{pmatrix}$$

Os resultados dos coeficientes de contribuição estimados, λ 's, encontrados são:

$$\begin{pmatrix} 0.0000 & 0.1213 & 0.0699 & 0.0001 & 0.0232 & 0.0396 & 0.0015 \\ 0.0156 & 0.0108 & 0.4205 & 0.0078 & 0.0000 & 0.0154 & 0.0032 \\ 0.0290 & 0.5695 & 0.0974 & 0.1150 & 0.0604 & 0.3382 & 0.1655 \\ 0.0192 & 0.0109 & 0.0830 & 0.0117 & 0.0609 & 0.0135 & 0.0026 \\ 0.0245 & 0.2728 & 0.0765 & 0.8516 & 0.7406 & 0.5526 & 0.8246 \\ 0.8903 & 0.0052 & 0.1587 & 0.0080 & 0.0585 & 0.0172 & 0.0025 \\ 0.0214 & 0.0095 & 0.0942 & 0.0059 & 0.0564 & 0.0236 & 0.0000 \end{pmatrix}$$

De acordo com os coeficientes estimados podemos escrever

$$\left\{ \begin{array}{l}
 t_{n+1}^{(11)}(y|x) = 0,0156t_n^{(12)}(y|x) + 0,0290t_n^{(13)}(y|x) + 0,0192t_n^{(14)}(y|x) + 0,0245t_n^{(15)}(y|x) + \\
 \quad + 0,8903t_n^{(16)}(y|x) + 0,0214t_n^{(17)}(y|x) \\
 t_{n+1}^{(22)}(y|x) = 0,1213t_n^{(21)}(y|x) + 0,0108t_n^{(22)}(y|x) + 0,5695t_n^{(23)}(y|x) + 0,0109t_n^{(24)}(y|x) + \\
 \quad + 0,2728t_n^{(25)}(y|x) + 0,0052t_n^{(26)}(y|x) + 0,0095t_n^{(27)}(y|x) \\
 t_{n+1}^{(33)}(y|x) = 0,0699t_n^{(31)}(y|x) + 0,4205t_n^{(32)}(y|x) + 0,0974t_n^{(33)}(y|x) + 0,0830t_n^{(34)}(y|x) + \\
 \quad + 0,0765t_n^{(35)}(y|x) + 0,1587t_n^{(36)}(y|x) + 0,0942t_n^{(37)}(y|x) \\
 t_{n+1}^{(44)}(y|x) = 0,0001t_n^{(41)}(y|x) + 0,0078t_n^{(42)}(y|x) + 0,1150t_n^{(43)}(y|x) + 0,0117t_n^{(44)}(y|x) + \\
 \quad + 0,8516t_n^{(45)}(y|x) + 0,0080t_n^{(46)}(y|x) + 0,0059t_n^{(47)}(y|x) \\
 t_{n+1}^{(55)}(y|x) = 0,0232t_n^{(51)}(y|x) + 0,0604t_n^{(53)}(y|x) + 0,0609t_n^{(54)}(y|x) + \\
 \quad + 0,7406t_n^{(55)}(y|x) + 0,0585t_n^{(56)}(y|x) + 0,0564t_n^{(57)}(y|x) \\
 t_{n+1}^{(66)}(y|x) = 0,0396t_n^{(61)}(y|x) + 0,0154t_n^{(62)}(y|x) + 0,3382t_n^{(63)}(y|x) + 0,0135t_n^{(64)}(y|x) + \\
 \quad + 0,5526t_n^{(65)}(y|x) + 0,0172t_n^{(66)}(y|x) + 0,0236t_n^{(67)}(y|x) \\
 t_{n+1}^{(77)}(y|x) = 0,0015t_n^{(71)}(y|x) + 0,0032t_n^{(72)}(y|x) + 0,1655t_n^{(73)}(y|x) + 0,0026t_n^{(74)}(y|x) + \\
 \quad + 0,8246t_n^{(75)}(y|x) + 0,0025t_n^{(76)}(y|x)
 \end{array} \right.$$

Através do sistema de equações acima pode-se concluir, considerando os maiores coeficientes de contribuição, que as densidades $t_{n+1}^{(ii)}(y|x)$ com $i = 1, \dots, 7$ são determinadas considerando as densidades $t_n^{(16)}(y|x)$, $t_n^{(23)}(y|x)$, $t_n^{(32)}(y|x)$, $t_n^{(45)}(y|x)$, $t_n^{(55)}(y|x)$, $t_n^{(65)}(y|x)$ e $t_n^{(75)}(y|x)$.

Considerando uma política de manutenção mais eficaz pode-se dizer que os desgastes de todos os blocos da máquina podem ser explicados apenas pelos desgastes dos blocos 2, 3, 5 e 6. Assim basta monitorar apenas estes blocos reduzindo-se portanto o tempo de monitoramento bem como os custos.

Os dados utilizados na estimação dos coeficientes de contribuição são apresentados no anexo 9.

Perspectivas de Continuidade do Trabalho de Pesquisa

Como proposta de continuidade de trabalho propomos:

1. Encontrar o coeficiente de contribuição de duas ou mais cadeias de Markov com espaço de estados contínuo e discreto em regime estacionário.

Conclusões

Neste trabalho propusemos uma metodologia para verificar o quanto cada sequência, dado um conjunto de sequências, através de sua distribuição de probabilidade marginal ou densidade de transição marginal, contribui para a estimativa da distribuição de probabilidade marginal ou densidade de transição marginal de determinadas sequências. Esta contribuição é medida através dos coeficientes de contribuição.

Algumas observações foram feitas, considerando caso discreto:

1. Quando modelamos uma sequência utilizando PCT obtemos os contextos que realmente influenciam na estimativa do próximo símbolo, evitando assim, de fixar uma certa ordem no modelo.
2. A metodologia proposta foi suficiente para verificar a contribuição entre algumas distribuições:
 - (a) Entre os dois textos escritos por Almeida Garret em 1799, G003 e G004, percebemos que a forma rítmica do texto G003 influencia moderadamente a forma rítmica do texto G004. Já a forma rítmica do texto G004 tem pouca influência na forma rítmica do texto G003.
 - (b) Entre os dois textos escritos por Antônio Vieira em 1608, V002 e V004, percebemos que a forma rítmica do texto V002 influencia fortemente a forma rítmica do texto V004. A forma rítmica do texto V004 também influencia fortemente a forma rítmica do texto V002.
 - (c) Entre os textos escritos pelo Marquês D'Alorna em 1802, A003, e pela Marquesa D'Alorna escrito em 1750, A004, percebemos que há uma forte influência entre as formas rítmicas dos dois textos.

(d) Entre os textos escritos por Gandavo em 1502, G008, e por Ramalho Urtigão escrito em 1836, O001, percebemos que a forma rítmica do texto G008 não é influenciada pela forma rítmica do texto O001. Já a forma rítmica do texto O001 tem pouca influência da forma rítmica do texto G008.

As observações feitas, considerando caso contínuo, foram:

1. O modelo proposto relaciona as funções de densidade de transição marginais.
2. Considerando duas cadeias distintas e independentes, como esperado, não houve contribuição.
3. Já o caso em que as duas cadeias eram distintas e dependentes houve contribuição. De acordo com as simulações os parâmetros λ 's sofreram alteração quando foram variados os parâmetros N^* e w . Já os parâmetros A e ϵ_h não alteraram os valores dos λ 's.
4. O modelo mostrou-se satisfatório quando utilizado como metodologia de política de manutenção.

Anexo 1

8.1 Matrizes de Frequências de Transições Estimadas da Aplicação do Capítulo 4

As matrizes a seguir referem-se às matrizes de frequência de transição entre os textos das aplicações da seção 4.7. Para encontrar as matrizes de probabilidade de transição basta calcular

$$P^{ij} = \frac{F^{ij}}{\sum_{i=1}^{11} F^{ij}}$$

onde $j = 1, 2, \dots, 11$. Neste caso estamos adotando soma de coluna igual a 1.

8.1.1 Matriz de Frequência de Transição dos Textos G003 e G004

$$F_{(G003,G003)} = \begin{pmatrix} 146 & 0 & 492 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 212 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1517 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 104 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4831 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3833 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2146 & 0 \\ 492 & 0 & 1025 & 0 & 0 & 2314 & 0 & 0 & 3575 & 0 & 0 \\ 0 & 147 & 0 & 66 & 3179 & 0 & 1315 & 1691 & 0 & 681 & 333 \\ 0 & 48 & 0 & 33 & 1133 & 0 & 641 & 734 & 0 & 438 & 269 \\ 0 & 17 & 0 & 5 & 307 & 2 & 86 & 149 & 4 & 32 & 2 \end{pmatrix}$$

$$F_{(G003,G004)} = \begin{pmatrix} 6 & 1 & 15 & 4 & 76 & 39 & 55 & 137 & 143 & 121 & 41 \\ 1 & 1 & 9 & 1 & 22 & 20 & 20 & 39 & 37 & 37 & 25 \\ 10 & 4 & 43 & 12 & 191 & 114 & 165 & 336 & 294 & 253 & 95 \\ 1 & 0 & 4 & 1 & 13 & 9 & 8 & 18 & 21 & 20 & 9 \\ 32 & 24 & 116 & 20 & 579 & 417 & 449 & 1048 & 1001 & 834 & 311 \\ 25 & 17 & 116 & 15 & 432 & 309 & 328 & 860 & 809 & 671 & 251 \\ 21 & 12 & 79 & 9 & 258 & 191 & 177 & 480 & 467 & 329 & 123 \\ 43 & 36 & 162 & 26 & 947 & 572 & 672 & 1564 & 1675 & 1269 & 440 \\ 59 & 27 & 207 & 27 & 925 & 613 & 693 & 1574 & 1560 & 1271 & 456 \\ 28 & 9 & 102 & 12 & 403 & 281 & 302 & 684 & 756 & 508 & 211 \\ 3 & 1 & 12 & 1 & 85 & 53 & 52 & 117 & 124 & 117 & 39 \end{pmatrix}$$

$$F_{(G004,G003)} = \begin{pmatrix} 5 & 3 & 10 & 0 & 50 & 20 & 13 & 42 & 45 & 35 & 6 \\ 5 & 2 & 6 & 0 & 17 & 14 & 9 & 33 & 32 & 10 & 4 \\ 13 & 3 & 33 & 0 & 131 & 90 & 70 & 247 & 164 & 98 & 16 \\ 2 & 0 & 4 & 2 & 20 & 16 & 5 & 29 & 35 & 13 & 2 \\ 90 & 22 & 171 & 14 & 562 & 487 & 283 & 882 & 920 & 438 & 62 \\ 60 & 8 & 124 & 7 & 393 & 293 & 173 & 590 & 640 & 287 & 43 \\ 61 & 25 & 147 & 8 & 423 & 333 & 192 & 696 & 718 & 260 & 58 \\ 113 & 61 & 330 & 28 & 1024 & 806 & 451 & 1562 & 1615 & 712 & 155 \\ 133 & 44 & 336 & 20 & 1086 & 851 & 466 & 1592 & 1533 & 700 & 127 \\ 109 & 35 & 259 & 17 & 815 & 676 & 349 & 1267 & 1261 & 551 & 90 \\ 47 & 9 & 97 & 8 & 310 & 247 & 135 & 465 & 449 & 193 & 41 \end{pmatrix}$$

$$F_{(G004,G004)} = \begin{pmatrix} 33 & 0 & 196 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 132 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 865 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 128 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3931 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2618 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2921 & 0 \\ 196 & 0 & 669 & 0 & 0 & 1749 & 0 & 0 & 4243 & 0 & 0 \\ 0 & 71 & 0 & 72 & 1844 & 0 & 1343 & 1480 & 0 & 1186 & 892 \\ 0 & 34 & 0 & 44 & 1125 & 0 & 1063 & 974 & 0 & 1084 & 1105 \\ 0 & 27 & 0 & 12 & 830 & 4 & 387 & 472 & 26 & 239 & 4 \end{pmatrix}$$

8.1.2 Matriz de Frequência de Transição dos Textos V002 e V004

$$F_{(V002,V004)} = \begin{pmatrix} 8 & 2 & 21 & 2 & 94 & 67 & 50 & 141 & 139 & 69 & 10 \\ 0 & 0 & 6 & 0 & 21 & 15 & 8 & 23 & 26 & 7 & 4 \\ 27 & 13 & 57 & 4 & 224 & 149 & 110 & 320 & 347 & 160 & 23 \\ 0 & 0 & 1 & 0 & 6 & 0 & 1 & 14 & 11 & 3 & 0 \\ 72 & 46 & 207 & 9 & 812 & 555 & 363 & 1196 & 1167 & 578 & 70 \\ 58 & 20 & 174 & 3 & 632 & 418 & 277 & 919 & 893 & 440 & 69 \\ 34 & 21 & 86 & 7 & 364 & 262 & 146 & 454 & 498 & 243 & 30 \\ 117 & 58 & 305 & 13 & 1178 & 846 & 541 & 1798 & 1802 & 821 & 114 \\ 112 & 72 & 315 & 21 & 1203 & 842 & 546 & 1743 & 1802 & 820 & 120 \\ 38 & 25 & 137 & 7 & 478 & 319 & 235 & 800 & 723 & 338 & 49 \\ 6 & 1 & 10 & 0 & 58 & 36 & 29 & 77 & 82 & 51 & 5 \end{pmatrix}$$

$$F_{(V004,V002)} = \begin{pmatrix} 9 & 2 & 20 & 0 & 67 & 60 & 34 & 119 & 120 & 38 & 3 \\ 7 & 2 & 8 & 0 & 37 & 34 & 11 & 66 & 67 & 26 & 0 \\ 19 & 5 & 51 & 1 & 196 & 161 & 99 & 324 & 320 & 129 & 14 \\ 0 & 0 & 7 & 0 & 7 & 11 & 3 & 16 & 13 & 8 & 1 \\ 104 & 18 & 219 & 7 & 811 & 613 & 356 & 1185 & 1223 & 472 & 62 \\ 63 & 16 & 135 & 3 & 563 & 434 & 235 & 810 & 856 & 353 & 41 \\ 54 & 6 & 110 & 1 & 350 & 272 & 170 & 542 & 547 & 233 & 21 \\ 130 & 31 & 356 & 9 & 1190 & 927 & 474 & 1761 & 1766 & 749 & 92 \\ 138 & 20 & 333 & 9 & 1214 & 891 & 508 & 1803 & 1773 & 726 & 76 \\ 75 & 9 & 180 & 6 & 564 & 435 & 215 & 830 & 806 & 369 & 40 \\ 4 & 1 & 15 & 0 & 76 & 64 & 40 & 137 & 106 & 46 & 5 \end{pmatrix}$$

8.1.3 Matriz de Frequência de Transição dos Textos A003 e A004

$$F_{(A003,A004)} = \begin{pmatrix} 14 & 3 & 40 & 2 & 102 & 100 & 43 & 184 & 182 & 84 & 14 \\ 2 & 2 & 10 & 0 & 38 & 30 & 15 & 46 & 61 & 20 & 7 \\ 41 & 11 & 83 & 0 & 268 & 202 & 137 & 387 & 420 & 177 & 47 \\ 1 & 1 & 3 & 0 & 9 & 11 & 1 & 14 & 21 & 8 & 1 \\ 98 & 32 & 207 & 6 & 829 & 566 & 354 & 1186 & 1199 & 541 & 98 \\ 69 & 14 & 210 & 5 & 681 & 495 & 324 & 939 & 945 & 469 & 81 \\ 36 & 10 & 105 & 2 & 318 & 261 & 134 & 457 & 449 & 215 & 43 \\ 143 & 45 & 357 & 13 & 1187 & 808 & 511 & 1757 & 1662 & 741 & 154 \\ 141 & 38 & 308 & 14 & 1127 & 868 & 531 & 1729 & 1719 & 765 & 144 \\ 52 & 15 & 117 & 9 & 409 & 300 & 173 & 584 & 627 & 257 & 61 \\ 7 & 5 & 21 & 1 & 57 & 48 & 27 & 94 & 104 & 42 & 7 \end{pmatrix}$$

$$F_{(A004,A003)} = \begin{pmatrix} 17 & 4 & 33 & 0 & 82 & 77 & 42 & 141 & 150 & 51 & 7 \\ 1 & 2 & 15 & 0 & 20 & 22 & 9 & 37 & 47 & 20 & 3 \\ 31 & 15 & 76 & 5 & 267 & 161 & 101 & 326 & 348 & 110 & 21 \\ 2 & 1 & 4 & 0 & 10 & 5 & 2 & 14 & 6 & 7 & 1 \\ 121 & 39 & 312 & 12 & 791 & 667 & 329 & 1140 & 1146 & 401 & 67 \\ 77 & 29 & 190 & 4 & 619 & 470 & 226 & 911 & 825 & 289 & 49 \\ 51 & 14 & 123 & 3 & 357 & 290 & 131 & 523 & 542 & 190 & 26 \\ 185 & 49 & 407 & 17 & 1186 & 1012 & 467 & 1623 & 1712 & 631 & 89 \\ 192 & 52 & 422 & 18 & 1131 & 969 & 468 & 1724 & 1708 & 604 & 100 \\ 76 & 25 & 155 & 9 & 552 & 474 & 205 & 777 & 741 & 262 & 43 \\ 15 & 1 & 35 & 2 & 101 & 85 & 50 & 162 & 160 & 39 & 7 \end{pmatrix}$$

$$F_{(A004,A004)} = \begin{pmatrix} 134 & 0 & 470 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 176 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1461 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 52 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5025 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3689 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2250 & 0 \\ 470 & 0 & 991 & 0 & 0 & 2228 & 0 & 0 & 3689 & 0 & 0 \\ 0 & 120 & 0 & 36 & 3255 & 0 & 1421 & 1562 & 0 & 674 & 320 \\ 0 & 45 & 0 & 11 & 1255 & 0 & 677 & 662 & 0 & 352 & 317 \\ 0 & 11 & 0 & 5 & 339 & 0 & 100 & 128 & 11 & 43 & 20 \end{pmatrix}$$

8.1.4 Matriz de Frequência de Transição dos textos G008 e O001

$$F_{(G008,G008)} = \begin{pmatrix} 65 & 0 & 259 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 81 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1079 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 48 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4995 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3101 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3188 & 0 \\ 259 & 0 & 820 & 0 & 0 & 2022 & 0 & 0 & 4117 & 0 & 0 \\ 0 & 50 & 0 & 32 & 3131 & 0 & 1795 & 1344 & 0 & 604 & 263 \\ 0 & 30 & 0 & 15 & 1512 & 0 & 1268 & 794 & 0 & 500 & 181 \\ 0 & 1 & 0 & 1 & 271 & 0 & 77 & 84 & 1 & 9 & 2 \end{pmatrix}$$

$$F_{(O001,G008)} = \begin{pmatrix} 4 & 1 & 14 & 2 & 72 & 53 & 46 & 110 & 102 & 59 & 8 \\ 1 & 0 & 2 & 0 & 25 & 14 & 26 & 43 & 31 & 27 & 2 \\ 10 & 4 & 40 & 2 & 193 & 114 & 125 & 288 & 285 & 179 & 16 \\ 1 & 0 & 2 & 0 & 10 & 3 & 3 & 11 & 20 & 10 & 1 \\ 59 & 6 & 174 & 6 & 757 & 463 & 507 & 1067 & 1129 & 659 & 71 \\ 33 & 8 & 118 & 6 & 571 & 353 & 358 & 808 & 803 & 443 & 43 \\ 20 & 9 & 60 & 3 & 379 & 240 & 223 & 571 & 531 & 343 & 21 \\ 67 & 18 & 276 & 15 & 1170 & 711 & 710 & 1611 & 1608 & 1021 & 109 \\ 77 & 22 & 233 & 5 & 1094 & 710 & 746 & 1679 & 1666 & 990 & 100 \\ 42 & 11 & 128 & 7 & 593 & 356 & 359 & 828 & 847 & 462 & 67 \\ 10 & 2 & 32 & 2 & 131 & 84 & 85 & 201 & 197 & 108 & 8 \end{pmatrix}$$

$$F_{(0001,0001)} = \begin{pmatrix} 98 & 0 & 373 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 171 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1256 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 61 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4898 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3544 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2400 & 0 \\ 373 & 0 & 883 & 0 & 0 & 2287 & 0 & 0 & 3773 & 0 & 0 \\ 0 & 102 & 0 & 30 & 3064 & 0 & 1427 & 1454 & 0 & 757 & 488 \\ 0 & 42 & 0 & 17 & 1224 & 0 & 797 & 761 & 0 & 488 & 371 \\ 0 & 27 & 0 & 14 & 438 & 1 & 115 & 203 & 6 & 55 & 1 \end{pmatrix}$$

Anexo 2

9.1 Dados Rolamentos Máquina de Trefilar

Tabela 9.1: Valores dos níveis de aceleração, em G, do rolamento dianteiro dos blocos da máquina de trefilar 16032,

Bloco 1	Bloco 2	Bloco 3	Bloco 4	Bloco 5	Bloco 6	Bloco 7
0,5650	0,3690	0,4330	0,8800	0,5440	0,7670	0,8450
0,4680	0,5410	0,3040	0,6580	0,5970	0,6350	0,7120
0,8020	0,2530	0,4400	0,5220	0,5140	0,4790	0,3600
0,5330	0,4730	0,3650	0,4090	0,4220	0,8000	0,6560
0,7280	0,6360	0,6770	0,4170	0,5730	0,6880	0,4390
0,3200	0,5020	0,5500	0,4480	0,5220	0,8890	0,6690
0,8250	0,5920	0,7350	0,9300	0,5650	0,8760	0,6720
0,6440	0,4910	0,3600	0,4800	0,5100	0,9400	0,5360
1,2940	1,1120	0,6690	1,1060	1,5970	1,7110	1,0890
0,7250	1,1690	0,5920	0,8050	1,3660	2,3470	1,0430
0,8180	0,8310	0,6760	0,8970	1,8700	1,3070	0,7940
0,7560	0,9880	0,1660	0,8920	1,1230	1,5570	0,9240
1,5990	1,0460	0,8910	1,1040	1,7090	1,1040	0,8280
1,2410	0,5700	1,3020	1,0220	1,2180	3,9840	1,4230
1,8740	0,9970	0,8240	0,7500	1,4780	1,2800	0,4680
1,1070	1,0390	0,8130	0,8100	1,7410	1,4680	3,0650
0,6340	1,1870	1,0770	0,9730	1,9330	2,3290	12,9280

1,1590	0,7930	0,7420	0,7480	1,9510	1,7440	6,1560
2,3890	1,0800	1,1120	1,1570	1,4450	1,3290	1,2960
1,1970	1,3900	1,1040	1,2430	1,3850	0,9870	0,9860
0,1180	0,5380	0,2520	1,0240	1,7790	1,6370	1,2840
1,0430	1,1090	0,9500	1,0150	1,2920	1,1310	0,9450
2,2460	1,5400	1,0970	1,0610	1,7130	1,1740	1,3070
11,3700	2,1250	1,4300	0,8160	2,6670	2,9050	1,0570
9,2420	1,0710	0,5110	1,0340	1,4530	1,2660	0,7030
17,3840	1,3420	0,9470	1,2260	1,6430	5,8800	1,3520
1,2420	1,1410	0,6530	1,2030	1,0030	3,8860	1,3520
0,9850	1,4210	0,7730	0,7650	1,2100	3,6280	1,0940
0,8850	1,2710	1,3900	1,0410	1,3750	0,9800	0,7210
1,1250	2,5130	0,4670	1,0750	0,9250	0,8930	4,3450
1,2120	1,4190	0,5100	0,4750	0,6440	0,4720	0,7610
1,5450	1,7070	0,8460	0,6010	1,4770	0,6190	0,9160
0,9850	2,9350	0,5230	1,5510	0,9720	0,8140	1,0780
0,3940	1,6260	0,8770	3,3170	1,6030	0,5470	0,6260
0,9420	10,9970	1,1420	1,5470	1,5570	0,8200	0,7540
0,6990	0,3660	0,7480	0,7310	1,0110	1,0580	0,7610
1,2710	0,5950	0,7050	0,7060	0,8080	0,6770	1,4300
0,6560	0,8220	0,2900	1,9930	0,6700	0,5170	0,5330
0,9330	0,9540	0,7760	0,9980	1,1570	0,8260	0,7660
0,8280	0,8490	0,6710	3,2690	1,6820	0,5190	0,6600
1,2410	0,5460	0,4690	1,6520	1,9750	0,9520	0,5560
0,5940	0,4630	1,2460	4,0130	1,7630	0,8450	2,1290
0,4780	0,4020	0,5030	2,2140	1,0860	0,5490	0,8670
1,0360	0,4080	0,4780	1,5950	1,0580	0,9250	0,6710
0,6460	0,4020	0,5030	1,4150	1,0250	0,7440	1,1670
1,7260	0,7560	0,7120	1,1140	0,9220	1,5840	0,6890

0,7650	0,5600	0,6930	1,4620	2,6390	11,8020	0,9930
3,2220	0,8470	1,2870	12,6490	1,5250	20,8240	1,1880
5,5600	1,6230	1,2510	10,8980	3,4080	1,4330	2,0220
5,4590	1,1110	0,9370	23,0760	2,1750	1,4820	2,0670
17,3310	1,2720	1,2650	1,2260	2,4710	1,9230	1,6860
1,5470	2,1890	1,4230	0,9670	2,1890	1,5470	2,0250
1,8840	1,0520	1,0320	0,8400	2,3300	1,1840	1,6930
1,2490	1,2370	1,4110	1,3630	3,1800	1,1930	1,2780
0,6700	1,3960	1,4580	0,8490	2,2940	1,3260	1,6960
1,5140	2,9370	1,8480	0,8260	4,8700	1,3220	9,3680
3,8670	13,6550	4,5650	1,4410	5,3610	1,8650	20,3860
3,4090	23,0470	1,9310	1,0260	3,6880	9,5840	1,4490
2,7180	27,5890	1,7430	2,0680	4,4310	9,8780	2,8370
2,5800	15,8680	2,1040	1,7030	4,4010	3,8400	2,0060

Anexo 3

10.1 Reflexões Sobre Processo de Otimização Minimax

Nesta seção iremos apresentar algumas notas em relação ao processo de otimização.

A Figura 10.1 itens (a) e (b) representam graficamente o modelo discreto, considerando $r = 2$ seqüências, dado pela equação 4.10 e descrito a seguir

$$\mu_{n+1}^{(j)} = \sum_{k=1}^r \lambda_{jk} Q_M^{(jk)} \mu_n^{(k)}$$

onde $j = \{1, 2, \dots, r\}$, $n = \{0, 1, \dots\}$, $\lambda_{jk} \geq 0$ e $\sum_{k=1}^r \lambda_{jk} = 1$. Nesta equação $\mu_0^{(j)}$ é a distribuição de probabilidade inicial da j -ésima seqüência.

Uma das restrições utilizadas neste trabalho é que: dadas r seqüências então $\sum_1^r \lambda_{ij} = 1$ sendo $\lambda_{ij} \geq 0$ e $i, j \in \{1, 2, \dots, r\}$. Nesta seção iremos discutir alguns resultados considerando, por simplicidade, apenas $r = 2$ seqüências discretas. O caso contínuo é semelhante.

Na Figura 10.1 item (a) o vetor U representa o vetor $Q_M^{(11)} \mu_n^{(1)}$ onde $\mu_n^{(1)}$ é o vetor de distribuição de probabilidade da seqüência 1 e $Q_M^{(11)}$ é a matriz de probabilidade de transição. O vetor V representa o vetor $Q_M^{(12)} \mu_n^{(2)}$ onde $\mu_n^{(2)}$ é o vetor de distribuição de probabilidade da seqüência 2 e $Q_M^{(12)}$ é a matriz de probabilidade de transição da seqüência 2 para a seqüência 1. Assim, a figura apresenta o vetor $\mu_n^{(1)}$ como combinação linear dos vetores U e V . Os coeficientes da combinação linear são os coeficientes de contribuição λ_{11} e λ_{12} .

Na Figura 10.1 item (b) o vetor U representa o vetor $Q_M^{(21)} \mu_n^{(2)}$ onde $\mu_n^{(2)}$ é o vetor de distribuição de probabilidade da seqüência 2 e $Q_M^{(21)}$ é a matriz de probabilidade de transição da seqüência 1 para a seqüência 2. O vetor V representa o vetor $Q_M^{(22)} \mu_n^{(2)}$ onde $\mu_n^{(2)}$ é o vetor de distribuição de probabilidade da seqüência 2 e $Q_M^{(22)}$ é a matriz de probabilidade de transição. Assim, a figura apresenta o vetor $\mu_n^{(2)}$ como combinação

linear dos vetores U e V . Os coeficientes da combinação linear são os coeficientes de contribuição λ_{21} e λ_{22} .

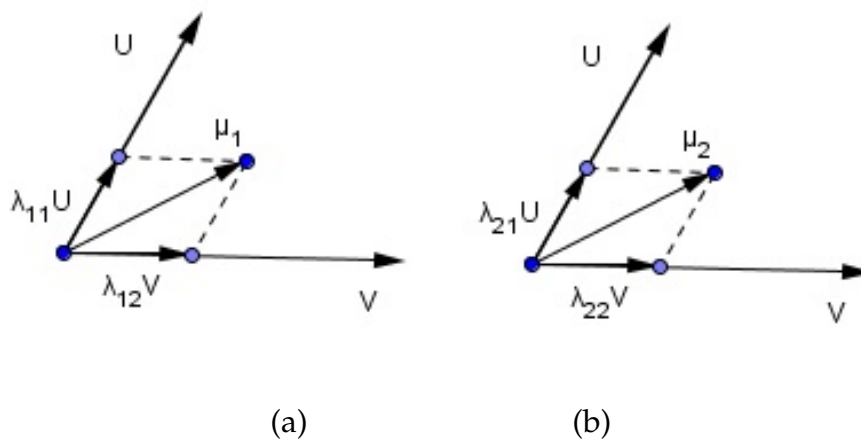


Figura 10.1: Representação Gráfica do modelo proposto.

Podemos observar que, caso estejamos em regime estacionário, então $Q_M^{(11)} \mu_n^{(1)} = \mu_n^{(1)}$ o que implica em $\lambda_{11} = 1$ e $\lambda_{12} = 0$. Analogamente se $Q_M^{(21)} \mu_n^{(2)} = \mu_n^{(2)}$ então $\lambda_{21} = 0$ e $\lambda_{22} = 1$. Fora do regime estacionário $Q_M^{(11)} \mu_n^{(1)} \neq \mu_n^{(1)}$ o que implica em $0 \leq \lambda_{11} < 1$ e $0 < \lambda_{12} \leq 1$. Analogamente $Q_M^{(21)} \mu_n^{(2)} \neq \mu_n^{(2)}$ então $0 < \lambda_{21} \leq 1$ e $0 \leq \lambda_{22} < 1$.

Nas seções seguintes iremos descrever alguns resultados possíveis dos coeficientes de contribuição.

10.1.1 Valores dos λ 's Iguais a 0,5

No processo de otimização com restrições os valores dos λ 's são determinados de acordo com o conjunto das restrições. Observe a figura 10.2. Nela apresentamos apenas duas restrições apresentadas no sistema de equações a seguir:

$$\begin{cases} \lambda_{11} + \lambda_{12} = 1 \\ 284\lambda_{11} + 284\lambda_{12} = 263 \end{cases}$$

A reta que passa pelos pontos (1,0) e (0,1) representa a primeira restrição. A segunda restrição foi tomada aleatoriamente como exemplo. O sistema acima pode ser escrito

como

$$\begin{cases} \lambda_{11} + \lambda_{12} = 1 \\ \lambda_{11} + \lambda_{12} = 263/284 \end{cases}$$

Observa-se que as retas que representam as restrições são paralelas pois os coeficientes dos λ 's são todos iguais a 1. Isto implica em resultado $\lambda_{11} = \lambda_{12} = 0,5$.

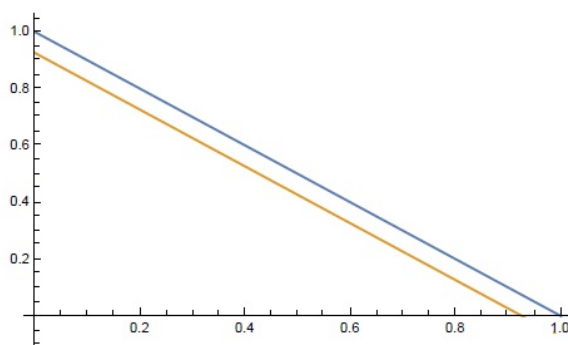


Figura 10.2: Gráfico das restrições.

Assim, se alguma das retas que representam as restrições for paralela à reta que representa a restrição $\lambda_{11} + \lambda_{12} = 1$ então os resultados dos coeficientes de contribuição relativos a esta reta serão iguais a 0,5. Caso contrário os coeficientes de contribuição serão diferentes de 0,5.

Um dos casos em que temos resultado 0,5 para os coeficientes é quando as duas sequências são iguais.

10.1.2 Valores dos λ 's Diferentes de 0,5

O conjunto das soluções viáveis dos coeficientes de contribuição são o intervalo $[0,1]$. No processo de otimização utilizamos o comando *minimax*. Este comando considera o menor valor entre um conjunto de máximos. Assim se tivermos r restrições teremos r máximos. Considerando o conjunto de máximos tomamos o menor valor como sendo a solução do coeficiente de contribuição.

Caso nenhuma das retas que representam as restrições for paralela à reta que re-

presenta a restrição $\lambda_{11} + \lambda_{12} = 1$ então os coeficientes terão resultados diferentes de 0,5.

Observe a Figura 10.3. Nela são apresentadas duas restrições. A reta que passa pelos pontos (1,0) e (0,1) representa a restrição de igualdade $\lambda_{11} + \lambda_{12} = 1$ a outra reta representa a restrição $400\lambda_{11} + 284\lambda_{12} = 263$, criada como exemplo. Estas restrições formam retas não paralelas no intervalo [0,1]. Isto implica em resultado $\lambda_{11} \neq \lambda_{12}$.

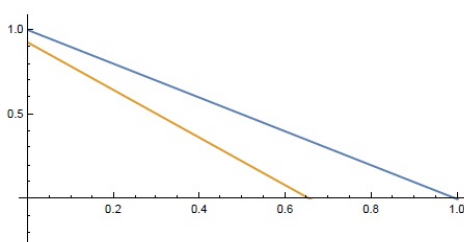


Figura 10.3: Gráfico das restrições.

O ponto de máximo da reta $400\lambda_{11} + 284\lambda_{12} = 263$ no intervalo [0,1] é aproximadamente o ponto (0;0,92). Como os coeficientes devem satisfazer à restrição $\lambda_{11} + \lambda_{12} = 1$ então obtém-se $\lambda_{11} = 0$ e $\lambda_{12} = 1$.

Na Figura 10.4 A reta que passa pelos pontos (1,0) e (0,1) representa a restrição de igualdade $\lambda_{11} + \lambda_{12} = 1$ e a outra representa a restrição (II): $\lambda_{12} = \frac{400}{284}\lambda_{11} + \frac{200}{284}$, criada como exemplo. De acordo com a figura os valores dos coeficientes são $\lambda_{11} = 1$ e $\lambda_{12} = 0$.

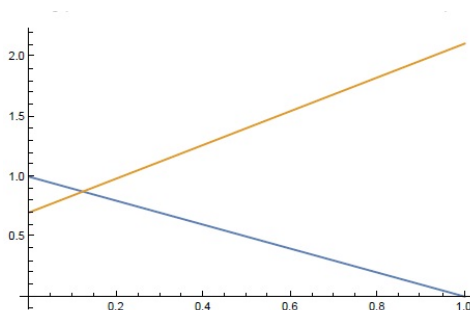


Figura 10.4: Gráfico das restrições.

A Figura 10.5 apresenta três restrições: a reta que passa pelos pontos (1,0) e (0,1) representa a restrição $\lambda_{11} + \lambda_{12} = 1$ e as restrições (I): $\lambda_{12} = \frac{400}{284}\lambda_{11} + \frac{400}{284}$ e (II): $\lambda_{12} =$

$-\frac{400}{284}\lambda_{11} + \frac{500}{284}$, ambas criadas como exemplo, sendo a primeira crescente e a segunda decrescente. De acordo com a figura os valores dos coeficientes são $\lambda_{11} = 1$ e $\lambda_{12} = 0$.

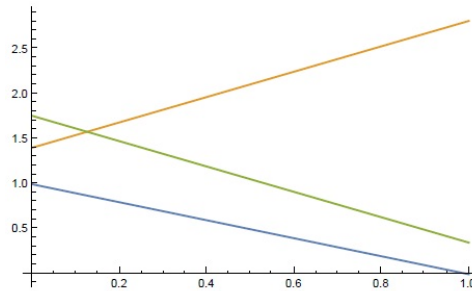


Figura 10.5: Gráfico das restrições.

Pode-se observar que as restrições (I) e (II) possuem um ponto de interseção. Este ponto representa a solução desejada. Este ponto é dado por $(\frac{1}{8}, \frac{7}{8})$. Assim os coeficientes são iguais a $\lambda_{11} = \frac{1}{8}$ e $\lambda_{12} = \frac{7}{8}$.

A figura 10.6 apresenta três restrições: a reta que passa pelos pontos (1,0) e (0,1) representa a restrição $\lambda_{11} + \lambda_{12} = 1$ e as restrições (I): $\lambda_{12} = \frac{400}{284}\lambda_{11} + \frac{400}{284}$ e (II): $\lambda_{12} = -\frac{400}{284}\lambda_{11} + \frac{1250}{284}$, ambas criadas como exemplo. De acordo com a figura os valores dos coeficientes são $\lambda_{11} = 1$ e $\lambda_{12} = 0$.

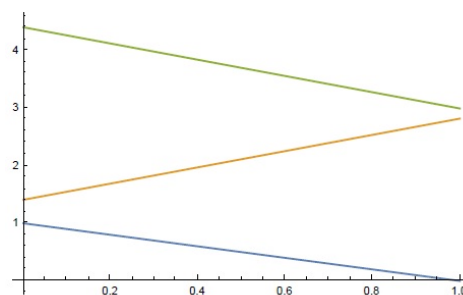


Figura 10.6: Gráfico das restrições.

Observamos que quando as retas que representam as restrições, exceto a restrição $\lambda_{11} + \lambda_{12} = 1$, não são concorrentes e nem paralelas no intervalo $[0,1]$, então os coeficientes de contribuição tem seus valores extremos, ou seja, são 0 ou 1. Caso as retas sejam concorrentes então os coeficientes de contribuição tem seus valores entre $]0,1[$.

Anexo 4

11.1 Cartas, Almeida Garrett (Garrett, A.) - Texto G003

Neste anexo apresentamos uma parte da carta escrita por Almeida Garrett.

"Deus guarde a Vossa Senhoria Tenho a satisfação de responder hoje á carta que Vossa Senhoria me dirigiu em 13 de Maio ultimo, com a copia authentica do Officio da Secretaria de Estado dos Negocios do Reino em data de 9 do corrente mez pelo qual Sua Magestade foi servida mandar-me participar que houvera porbem attender a minha proposta e aggraciar a Vossa Senhoria conforme o meu rogo,com a mercê do Habito de Cristo, que lhe conferiu por honra deste Conservatorio Real ao qual Vossa Senhoria mereceu pertencer pelo seu talento e applicação".

Referências Bibliográficas

- J. Abfalq, H. P. Kriegel, P. Kroger, P. Kunath, A. Pryakhin, and M. Renz. Similarity search on time series based on threshold queries. 2006.
- A. Akimov, A. Kolesnikov, and P. Franti. Lossless compression of map contours by context tree modeling of chain codes. *Pattern Recognition*, 40(3):944–952, March 2007.
- P. D. Andersen, A. Iversen, H. Madsen, and C. Rode. Dynamic modeling of presence of occupants using inhomogeneous Markov chains. *Energy and Buildings*, 69:213–223, February 2014.
- K. B. Athreya and G. S. Atuncar. Kernel estimation for real-valued Markov chains. *Sankhya: The Indian Journal of Statistics, Series*, 60(January 1995):1–17, 1998.
- G. S. Atuncar, L. H. Duczmal, and L. F. Bessegato. Rotinas em R para técnicas de suavização por núcleos estimadores. *17.º Simpósio Nacional de Estatística e Probabilidade*, pages 1 – 74, 2006.
- P. J. Avery and D. A. Henderson. Fitting Markov chain models to discrete state series such as DNA sequences. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(1):53–61, February 1999.
- G. Bejerano. Variations on probabilistic suffix trees: statistical modeling and prediction of protein families. *Bioinformatics*, 17(1):23–43, 2001.
- G. Bejerano. Algorithms for variable length Markov chain modeling. *Bioinformatics*, 20(5):788–789, January 2004.
- D. J Bemdt. Using dynamic time warping to find patterns in time series. pages 359–370, 1994.
- Abraham Berman and Robert J. Plemmons. *Nonnegative matrices in the mathematical sciences*. Academic Press, New York, 1979.
- L. F. Bertuccelli and J. P. How. Estimation of non-stationary Markov chain transition models. *2008 47th IEEE Conference on Decision and Control*, pages 55–60, 2008.

- S. Bhattacharya. Markov chain model to explain the dynamics of human depression. 2014, 2014.
- S. R. Browning. Multilocus association mapping using variable-length Markov chains. *The American Journal of Human Genetics*, 78(June):903–913, 2006.
- P. Bühlmann and A. J. Wyner. Variable length markov chains. *The Annals of Statistics*, 27(2):480–513, April 1999.
- V. S. M. Campos and C. C. Y. Dorea. Kernel estimation for stationary density of Markov chains with general state space. 57(3):443–453, 2005.
- Y. Chen, M. A. Nascimento, B. C. Ooi, and A. K. H. Tung. SpADe: On shape-based pattern detection in streaming time series. *2007 IEEE 23rd International Conference on Data Engineering*, pages 786–795, 2007.
- W Ching, M Ng, and E Fung. Higher-order multivariate Markov chains and their applications. *Linear Algebra and its Applications*, 428(2-3):492–507, January 2008.
- W. K; Ching, E. S.; Fung, and M. K. NG. A multivariate Markov chain model for categorical data sequences and its applications in demand predictions. *IMA Journal of Management Mathematics*, 13(3):187–199, July 2002.
- W. K. Ching, E. S. Fung, and M. K. Ng. Higher-order Markov chain models for categorical data sequences. *Naval Research Logistics*, 51(4):557–574, June 2004.
- S. T. Chiu. Bandwidth selection for kernel density estimation. *The Annals of Statistics*, 19:1883 – 1905, 1991.
- H. C. Cho, M. S. Fadali, and K. S. Lee. Online probability density estimation of nonstationary random signal using dynamic bayesian networks. *International Journal of Control, Automation and Systems*, 6(1):109 – 118, 2008.
- I. Csiszar and Z. Talata. Context tree estimation for not necessarily finite memory processes, via BIC and MDL. *IEEE Transactions on Information Theory*, 52(3):1007–1016, March 2006.
- C. C. Y Dorea. Strong consistency of kernel estimators for Markov transition densities. *Bulletin of the Brazilian Mathematical Society*, 33(June):409–418, 2002.
- D. Duarte, A. Galves, and N. L. Garcia. Markov approximation and consistent estimation of unbounded probabilistic suffix trees. *Bulletin of the Brazilian Mathematical*, 37 (April):581–592, 2006.
- R. M. Evaristo. Métodos de reamostragem de séries temporais baseados em wavelets. 2010.

- C. Faloutsos. Fast subsequence matching in time-series databases Matching in Time-Series Databases. pages 419–429, 1994.
- A. Galves, C. Galves, J. E. García, N. L Garcia, and F. Leonardi. Context tree selection and linguistic rhythm retrieval from written texts. *The Annals of Applied Statistics*, 6 (1):186–209, March 2012.
- A. Garivier and F. Leonardi. Context tree selection: A unifying view. *Stochastic Processes and their Applications*, 121(11):2488–2506, November 2011.
- J. Gill. *Bayesian methods: a social and behavioral sciences approach*. Chapman & Hall/CRC Statistics in the Social and Behavioral Science, 2nd editio edition, 2007.
- J. D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 57(2):357–384, 1989.
- E. Keogh and C. A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3):358–386, May 2004.
- F. G. Leonardi. A generalization of the PST algorithm: modeling the sparse nature of protein sequences. *Bioinformatics*, 22(11):1302–1307, March 2006.
- M. E. Lladser, M. D. Betterton, and R. Knight. *Multiple pattern matching : A Markov chain approach*. 2013.
- M. D. Morse and J. M. Patel. An efficient and accurate method for evaluating time series similarity. *Proceedings of the 2007 ACM SIGMOD . . .*, 2007.
- L. Page, S. Brin, R. Motwani, and T. Winograd. The pageRank citation ranking: bringing order to the web, 1999.
- P. C. B Phillips and J. Y. Park. Nonstationary density dstimation and kernel autoregression. (1181), 1998.
- H. Pranevicius, M. Pranevicius, O. Pranevicius, M. Snipas, N. Paulauskas, and F. Bukauskas. Continuous time Markov chain models of voltage gating of gap. (2):133–142, 2014.
- AE Raftery. A model for high-order Markov chains. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(3):528–539, 1985.
- J. Rissanen. A universal data compression system. *IEEE Transactions on Information Theory*, 29(5):656–664, September 1983.
- C. Robert and G. Casella. *Monte carlo statistical methods*. Springer Texts in Statistics, New York, 2nd editio edition, 2004.

- D. Ron, Y. Singer, and N. Tishby. The power of amnesia: Learning probabilistic automata with variable memory length. *Machine learning*, 25(2):117–149, 1996.
- M Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pages 832–837, 1956.
- GG Roussas. Nonparametric estimation in Markov processes. *Annals of the Institute of Statistical Mathematics*, pages 73 – 87, 1969.
- Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461–464, March 1978.
- M. Vlachos, G. Kollios, and D. Gunopulos. Discovering similar multidimensional trajectories. *Proceedings 18th International Conference on Data Engineering*, pages 673–684, 2002.
- J. A. Vrugt, C. J. F. T. Braak, C. G. H. Diks, and G. Schoups. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water Resources*, 51:457–478, January 2013.
- X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26(2):275–309, February 2012.