Reamostragem Ponderada em Blocos para Cadeias de Markov

Adriana Petrielli

Tese apresentada ao

Instituto de Matemática e Estatística

da Universidade de São Paulo

para obtenção do grau

de Mestrado em Estatística

Orientador: Prof. Dr. Adilson Simonis

São Paulo - Novembro - 2004

Reamostragem Ponderada em Blocos para Cadeias de Markov

Este exemplar corresponde à redação

final da dissertação devidamente

corrigida e defendida por Adriana Petrielli

e aprovada pela comissão julgadora.

São Paulo, novembro de 2004.

Banca Examinadora:

- Prof. Dr. Adilson Simonis (Orientador) - IME-USP
- Prof. Dr. Heleno Bolfarine - IME-USP
- Prof. Dr. Miguel Abadi - IMECC-UNICAMP

Agradecimentos

A elaboração dessa dissertação somente foi possível graças à contribuição direta ou indireta de pessoas que me apoiaram no âmbito pessoal e profissional.

Agradeço a Deus que me disponibilizou as condições necessárias à realização deste trabalho.

Aos meus pais Ademir e Terezinha pelos esforços dedicados à educação, saúde, felicidade e caráter.

Aos meus irmãos André, Márcia e Mauricio, que sempre me deram força e incentivaram meus estudos.

Ao meu namorado, Fábio, que esteve ao meu lado e confiou em mim.

À Paula, ao meus sobrinhos Gabriel e Giovanna.À Cristina, Pedro, Beatriz e Rafael. Agradeço, também, aos meus padrinhos Cecília e Durval pelo carinho. Ao meus tios Marizilda e Fernando que mesmo longe sempre estiveram presentes na minha vida.

Agradeço ao professor Adilson pela orientação, compreensão e apoio.

A todos os amigos que de alguma forma me ajudaram, em especial a: Valéria, Diana, Letícia, Ricardo, Mário, Cinthya e Denise.

À CAPES pelo apoio financeiro.

Sumário

Dada uma seqüência de realizações de uma cadeia de Markov em tempo discreto e espaço de estados finito, estimamos a distribuição de probabilidade amostral das transições dos estados e da medida invariante usando o método de reamostragem. Discutimos alguns métodos existentes na literatura e propomos dois novos métodos de reamostragem com o objetivo de reduzir o esforço computacional e garantir a redução da variância. Para tanto, consideramos algumas recentes idéias no contexto de classificação, tais como o *Adaboost*, na estimação desta distribuição. Alguns resultados númericos ilustram e nos fornecem evidências para os métodos propostos.

Abstract

Given a sequence of realizations of a discrete time Markov chain on a finite state space, we estimated transition probabilities and invariant distributions using a bootstrap method. We discussed some existing methods in literature and consider two new models of bootstrap with the objective to reduce the computational effort and to guarantee the reduction of the variance. For in such a way, we consider some recent ideas in the classification context, such as the *Adaboost*, to the estimation of this distribution. Some numerical results illustrate and in they supply evidences to them.

Índice

Introdução 1

1.Métodos de Reamostragem para Cadeias de Markov 5

1.1.Uniforme em Blocos 8

1.2.Balanceada em Blocos 10

1.3.Importância em Blocos 12

1.4.Balanceada de Importância em Blocos 16

2.Classificadores 20

2.1.Árvore de Decisão 23

2.2.Modelo de Regressão Logística 25

2.3.Combinação de Classificadores 26

2.3.1.*Bagging* 27

2.3.2.*Boosting* 28

2.3.3.Uma análise generalizada do *Adaboost* 29

2.3.4.Escolhendo α_b 31

2.3.5.Exemplo dos procedimentos *Bagging* e *Boosting* 35

3.Duas Novas Propostas de Reamostragem em Blocos para Cadeias de Markov 37

3.1.Ponderada em Blocos 38

3.2.Balanceada Ponderada em Blocos 41

4.Simulações dos Métodos de Reamostragem para Cadeias de Markov 44

4.1.Probabilidade de Transição 45

4.2.Medida Invariante 52

5.Considerações Finais 59

Referências Bibliográficas 61

Introdução

Atualmente, a capacidade dos sistemas de computação para gerar e armazenar dados é grande em relação ao seu poder de análise. Nesse contexto, a descoberta de conhecimento surge como uma forma de buscar informações úteis desconhecidas, que podem estar ocultas em grandes bases de dados. Esse é dito um processo de mineração, que tem como objetivo os métodos de extração de informações a partir dos dados.

Existem muitas estratégias no processo de mineração de grandes bancos de dados que produzem previsões, padrões e relações. Uma tarefa da mineração é predizer um valor de um atributo de uma nova observação quando os valores dos outros atributos são conhecidos. Uma possibilidade é usar um algoritmo de aprendizagem, também denominado algoritmo de predição ou método de classificação.

Uma coleção de observações com valores conhecidos do atributo é tratado como um conjunto de treinamento para um algoritmo de aprendizagem que pode produzir uma regra de classificação. Aplicando diversos algoritmos de aprendizagem a um mesmo conjunto de treinamento, obtém-se para cada um desses uma descrição do conceito (um classificador), que descreve esses exemplos de forma mais compacta. Um classificador é aquele que fornece uma decisão para um dado problema. Diferentes classificadores normalmente possuem previsões e precisões diferentes para o mesmo conjunto de dados. A abordagem tradicional para determinar qual é o melhor classificador para uma aplicação é selecionar aquele que fornece a maior precisão (ou erro de predição esperado menor) para essa aplicação. Porém, essa abordagem pode não ser a mais adequada. Informações descobertas pelos demais classificadores podem estar sendo ignoradas quando se opta por um único classificador. Assim, uma abordagem mais apropriada pode ser a combinação das saídas de diversos classificadores. Pesquisas recentes, Friedman, J., Hastie, T., Tibshirani, R.(2000), Dietterich, T.G.(2000), Breiman, L.(1996), mostram que utilizando alguma combinação

de classificadores é possível obter resultados com maior precisão comparados com os resultados obtidos por um único classificador.

Existem diversos problemas que sugerem a modelagem de cadeias de Markov para sua solução e é de fundamental importância a estimação da matriz de probabilidade de transição e da medida invariante (ou estacionária). O objetivo deste trabalho é apresentar alguns métodos de reamostragem existentes para a estimação da distribuição de probabilidades das transições e da medida invariante e considerar o procedimento de combinação de classificadores no contexto de cadeias de Markov, definindo como classificador a probabilidade de transição estimada assim como sua medida invariante estimada a partir de realizações do processo estocástico.

Dada uma seqüência de realizações de uma Cadeia de Markov em tempo discreto com espaço de estado finito, estimamos as probabilidades de transição usando o método de reamostragem (*Bootstrap*).

Um problema central na estatística é determinar o estimador para um determinado parâmetro de interesse e avaliar a acurácia deste estimador, através das estimativas do erro padrão e consequentemente encontrar intervalos de confiança. Efron (1979), introduziu a versão *bootstrap*, motivado por esse problema.

Este método tem por base a idéia de que a amostra original é tratada como se fosse a população que deu origem aos dados, ou seja, a idéia é repor a distribuição desconhecida da população pela distribuição empírica, da amostra original.

Propriedades dos estimadores podem ser baseados na distribuição empírica. Algumas vezes estas propriedades são determinadas analiticamente, mas na maioria das vezes podem ser aproximadas, por exemplo, pelos métodos de Monte Carlo.

O *bootstrap* de Efron pode ser definido da seguinte maneira: Considere uma amostra de n vetores aleatórios idependentes e identicamente distribuídos $X_1, \ldots, X_n$ e o estimador $\hat{\theta}(X_1, \ldots, X_n)$ (denotado por $\hat{\theta}$). O procedimento para estimar a acurácia de $\hat{\theta}$ é definido em termos da distribuição empírica F_n. Esta função tem para cada valor observado dos vetores aleatórios X_i a probabilidade $1/n$, para $i = 1, \ldots, n$.

A função distribuição empírica é o estimador de máxima verossimilhança da distribuição para as observações. A distribuição de $\hat{\theta}-\theta$ é aproximada pela distribuição obtida por gerar valores de $\hat{\theta}$ pela amostragem independente com reposição da distribuição empírica F_n. A estimativa *bootstrap* do erro padrão de $\hat{\theta}$ é então o desvio padrão da distribuição de *bootstrap* para $\hat{\theta} - \theta$.

A aplicação da técnica requer a geração de reamostras, isto é, amostras obtidas pela amostragem independente com reposição da distribuição empírica. O procedimento consiste em:

1. Gerar uma amostra de tamanho n (onde n é a amostra de tamanho original) com reposição, da distribuição empírica.
2. Calcular θ^*, o valor de $\hat{\theta}$ calculado na amostra.
3. Repete os passos (1) e (2) B vezes.

Pela repetição dos passos 1 e 2, B vezes, obtemos a aproximação de Monte Carlo para a distribuição de θ^*. O desvio padrão de θ^* é a aproximação de Monte Carlo para a estimativa do erro padrão para $\hat{\theta}$.

Desejamos conhecer a distribuição de $\hat{\theta} - \theta$, mas o que temos é a aproximação de Monte Carlo para a distribuição de $\theta^* - \hat{\theta}$. A idéia principal do *bootstrap* é que para n suficientemente grande esperamos que as duas distribuições sejam aproximadamente a mesma.

A idéia do *bootstrap* é que a variabilidade de θ^* será similiar a variabilidade de $\hat{\theta}$. Existe uma boa razão para acreditar que isto é verdade desde que n seja grande, pois assim F_n fica próxima de F, e amostrar com reposição de F_n é quase amostrar aleatoriamente F.

Um resultado importante é o Teorema Glivenko-Cantelli, Chung (1974), que afirma que a distribuição empírica converge uniformemente com probabilidade 1 para a distribuição F quando as observações são independente e identicamente distribuídas, ou seja,

$$\sup_{-\infty<x<\infty} \mid F_n(x) - F(x) \mid \longrightarrow 0,$$

quando $n \longrightarrow \infty$, com probabilidade 1.

O procedimento para encontrar intervalos de confiança (IC's) dos parâmetros consiste em obter B estimativas através da reamostragem, definir a função distribuição a partir destas estimativas e então encontrar o intervalo de confiança desejado.

A seqüência de Markov deve ser adaptada para ser utilizada neste contexto, e é introduzida a reamostragem em blocos. Com o objetivo de garantir a redução da variância é proposto a reamostragem ponderada, que utiliza idéias do método de combinação de classificadores.

No capítulo 1 discutimos alguns dos métodos de reamostragem em blocos bem conhecidos para cadeias de Markov; Reamostragem Uniforme em Blocos, Reamostragem Balanceada em Blocos, Reamostragem de Importância em Blocos e Reamostragem Balanceada de Importância em Blocos. No capítulo 2 apresentamos o conceito de classificadores dando como exemplos a Árvore de Decisão e a Regressão Logística, e descrevemos o procedimento de combinação de classificadores, onde são estudados os métodos de *Bagging* e *Boosting*. No capítulo 3 apresentamos duas novas propostas de reamostragem para Cadeias de Markov, que consistem em uma adaptação do procedimento de combinação de classificadores, a Reamostragem Ponderada em Blocos e a Reamostragem Balanceada Ponderada em Blocos. No capítulo 4 discutimos os métodos de reamostragem apresentando algumas evidências numéricas.

Capítulo 1

Métodos de Reamostragem para Cadeias de Markov

Dada uma seqüência de realizações de uma cadeia de Markov com espaço de estados finito, em tempo discreto, desejamos estimar as distribuições de probabilidade da estatística de máxima verossimilhança das transições e da medida invariante.

Considere $\underline{x} = \{x_0, ..., x_n\}$ uma realização de um cadeia de Markov ergódica (recorrente positiva, aperiódica e irredutível) $\{X_j; j = 0, ..., n\}$, em um espaço de estado finito $S = \{1, ..., s\}$, com probabilidade de transição $P = (p_{ij})$. A ergodicidade implica a existência de uma medida invariante $\Pi = (\pi_1, \ldots, \pi_s)$, $\pi_j > 0$, $\sum_{j=1}^{s} \pi_j = 1$, $\pi_j = \sum_i \pi_i p_{ij}$, $j = 1, ..., s$ e para todo $i \in S$, $p_{ij}^{(n)} \to \pi_j$ exponencialmente rápido quando $n \to \infty$, onde $p_{ij}^{(n)} = P\{X_n = j \mid X_0 = i\}$.

Seja n_{ij} o número de transições do estado i para o j em $\underline{x}$, e n_i o número de visitas ao estado i em $\underline{x}$. O estimador de máxima verossimilhança da probabilidade de transição P é dado por $\hat{P}_n \equiv (\hat{p}_n(i,j))$, onde $\hat{p}_n(i,j) = \begin{cases} \frac{n_{ij}}{n_i}, n_i > 0 \\ \delta_{ij}, cc \end{cases}$ com $\delta_{ij} = 1$ para $i = j$, e 0 caso contrário, e podemos estimar Π por $\hat{\Pi}_n \equiv (\hat{\pi}_n(i))$, onde $\hat{\pi}_n(i) = n_i/n$.

Seja a matriz de variância-covariância de $\hat{P}_n$ definido por

$$\Sigma_{ij,kl} = \begin{cases} \frac{p_{ij}(1-p_{ij})}{n_i}, (i,j) = (k,l) \\ \frac{-p_{ij}p_{il}}{n_i}, (i = k), (j \neq l) \\ 0, cc \end{cases} \tag{1.1}$$

A dificuldade em calcular a matriz de variância-covariância assintótica faz este procedimento pouco apropriado. Assim o método de reamostragem é uma técnica útil para estudar este tipo de problema.

O método de reamostragem (*bootstrap*) proposto por Efron(1979), consiste em :

Considere uma amostra independente e identicamente distribuída (i.i.d) $\underline{z} = \{z_1, ..., z_n\}$ de uma função distribuição F e estatística de interesse $s(\underline{z})$. A estimativa de *bootstrap* para a Esperança de $s(\underline{z})$ é dada por

$$\hat{e} = E_{\hat{F}} s(\underline{z}^*), \tag{1.2}$$

onde $\hat{F}$ é a função distribuição empírica, $E_{\hat{F}}$ é a esperança sob $\hat{F}$, e $\underline{z}^* = \{z_1^*, ..., z_n^*\}$ é retirada com distribuição uniforme de $\underline{z}$, com reposição. A menos que $s(\underline{z})$ seja a média ou outra estatística simples, não é fácil calcular exatamente $\hat{e}$, então aproximamos $\hat{e}$ por

$$\hat{e}_B = \frac{1}{B} \sum_{b=1}^{B} s(\underline{z}^{*b}), \tag{1.3}$$

onde cada $\underline{z}^{*b}$ é uma amostra de tamanho n retirada com reposição de $\underline{z}$, B é o número de simulações de Monte-Carlo, e $s(\underline{z}^{*b})$ é o valor da estatística s avaliado para $\underline{z}^{*b}$.

A fórmula (1.3) é um exemplo de uma estimativa de Monte Carlo da esperança $E_{\hat{F}} s(\underline{z}^*)$. De acordo com a lei dos Grandes Números $\hat{e}_B \to \hat{e}$ quando $B \to \infty$, e $var(\hat{e}_B - \hat{e}) = c/B$, onde $c \geq 0$ é uma constante, assim a variância vai para zero quando $B \to \infty$. A variância amostral de $\hat{e}_B$ pode ser estimada por,

$$\hat{var}(e_B) = \frac{1}{B-1} \sum_{b=1}^{B} (s(\underline{z}^{*b}) - \hat{e}_B)^2. \tag{1.4}$$

Iremos explorar esta idéia da reamostragem para construir um método para estimar a distribuição de probabilidade das transições e da medida invariante de uma Cadeia de

Markov de ordem um. Primeiro discutimos a reamostragem uniforme em blocos baseado no processo introduzido por Athreya e Fuh (1992). Davison (1986) propôs a técnica da reamostragem balanceada para reduzir o esforço computacional. Hall (1990) discutiu a eficiência assintótica da reamostragem balanceada em relação a reamostragem uniforme, em problemas de estimação de funções de distribuição e quantis. Johns (1998) e Davison (1988) apresentaram a técnica da reamostragem de importância, quando o tamanho da amostra é fixo e esta técnica mostrou-se um esquema amostral para simular um evento raro. É de fato uma técnica de redução de variância. Booth, Hall e Wood (1993) discutiram a técnica de reamostragem de importância balanceada, que combina os aspectos da reamostragem balanceada e da reamostragem de importância.

Athreya e Fuh (1992) ao estudarem a realização de uma cadeia de Markov $\underline{x}$, particionaram em blocos $\{x_l; l = T_\triangle^{(\alpha)}, \ldots, T_\triangle^{(\alpha+1)} - 1\}$, onde $T_\triangle^{(\alpha)}$ é o tempo do α-ésimo retorno a um estado recorrente $\triangle$. A partir de um critério apropriado de parada temos pela propriedade forte de Markov que esses blocos são i.i.d. Assim, o método de reamostragem pode ser aplicado nestes blocos. Existem duas alternativas para realizar esta amostragem. A primeira é fixar o número de blocos k, e observar a cadeia até um n aleatório. A segunda é fixar o tamanho da amostra n, assim o número de blocos k é uma variável aleatória, onde o último bloco pode ser descartado.

Considerando a segunda alternativa, observamos a cadeia até o tempo fixado n e obtemos k blocos. Para um estado recorrente $\triangle$ fixado, seja $T_\triangle^{(\alpha)}$ o tempo do α-ésimo retorno ao estado $\triangle$, então os blocos $\eta_\alpha = \{x_l; l = T_\triangle^{(\alpha)}, \ldots, T_\triangle^{(\alpha+1)} - 1\}$ são i.i.d. para $\alpha = 1, 2, \ldots$ A amostra original pode ser decomposta da seguinte maneira:

$$\chi = \{\eta_0, \eta_1, \ldots, \eta_k, \eta_{k+1}\} \text{ onde } \eta_0 = \{x_0, \ldots, x_{T_\triangle^{(1)}-1}\} \text{ e } \eta_{k+1} = \{x_{T_\triangle^{(k+1)}}, \ldots, x_n\}.$$

Seja $g(\eta_\alpha)$ o número de visitas do estado i no bloco η_α, $h(\eta_\alpha)$ indica o número de transições do estado i para o j no bloco η_α, e T_α é o tamanho do α-ésimo bloco, então $p_\alpha(i,j) = \frac{h(\eta_\alpha)}{g(\eta_\alpha)}$ e $\pi_\alpha(i) = \frac{g(\eta_\alpha)}{T_\alpha}$ são os estimadores de máxima verossimilhança de p_{ij} e π_i, respectivamente, dentro do α-ésimo bloco.

Os estimadores de p_{ij} e π_i de máxima verossimilhança são dados por

$$\hat{p}_n(i,j) = \frac{\sum_{\alpha=1}^{k} h(\eta_\alpha)}{\sum_{\alpha=1}^{k} g(\eta_\alpha)}, \tag{1.5}$$

$$\hat{\pi}_n(i) = \frac{\sum_{\alpha=1}^{k} g(\eta_\alpha)}{\sum_{\alpha=1}^{k} T_\alpha}. \tag{1.6}$$

As variâncias estimadas desses dois estimadores são dadas por

$$\hat{\sigma}^2 = \frac{1}{(\sum_{\alpha=1}^{k} g(\eta_\alpha))^2} \sum_{\alpha=1}^{k} (g(\eta_\alpha)(p_\alpha(i,j) - \hat{p}_n(i,j)))^2, \tag{1.7}$$

$$\tilde{\sigma}^2 = \frac{1}{(\sum_{\alpha=1}^{k} T_\alpha)^2} \sum_{\alpha=1}^{k} (T_\alpha(\pi_\alpha(i) - \hat{\pi}_n(i)))^2. \tag{1.8}$$

Pela lei forte $\hat{p}_n(i,j)$ e $\hat{\pi}_n(i)$ são consistentes. Para estimar as distribuições de $(\hat{\pi}_n(i) - \pi_i)$ e $(\hat{p}_n(i,j) - p_{ij})$, iremos discutir alguns métodos de reamostragem e propor dois novos métodos.

Iremos considerar a seguinte notação: $\{1\}$ Método de Reamostragem Uniforme em Blocos, $\{2\}$ Método de Reamostragem Balanceada em Blocos, $\{3\}$ Método de Reamostragem de Importância em Blocos e $\{4\}$ Método de Reamostragem Balanceada de Importância em Blocos.

1.1 Uniforme em Blocos

Nesta seção discutimos dois procedimentos, baseados na reamostragem uniforme em blocos, para estimar a distribuição das estimativas das probabilidades de transições e da distribuição invariante.

Para um dado conjunto de dados $\{x_0, x_1, \ldots, x_n\}$, particionado em k blocos, o esquema de reamostragem uniforme em estimar a $P\{(\hat{p}_n(i,j)-p_{ij})/\hat{\sigma} \leq x\}$ e a $P\{(\hat{\pi}_n(i)-\pi_i)/\tilde{\sigma} \leq x\}$ consiste nos seguintes passos:

1. *Para a transição (i,j) fixa, decomponha a amostra original em k blocos: $\{\eta_1, \eta_2, ..., \eta_k\}$. Seja $\hat{F}$ a medida de probabilidade uniforme dos blocos $\{\eta_\alpha; \alpha = 1, ..., k\}$.*

2. *Para $b = 1, \ldots, B$, com a amostra original fixada, retire uma amostra de tamanho k de acordo com $\hat{F}$. Denotando esta amostra por $\chi_b^{\{1\}} = \{\eta_{b1}^{\{1\}}, ..., \eta_{bk}^{\{1\}}\}$, definimos as estimativas da probabilidade de transição do estado i para o j e a medida invariante do estado i, na b-ésima iteração , respectivamente por :*

$$\tilde{p}_b^{\{1\}}(i,j) = \frac{\sum_{\alpha=1}^{k} h(\eta_{b\alpha}^{\{1\}})}{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{1\}})}, \tag{1.9}$$

$$\tilde{\pi}_b^{\{1\}}(i) = \frac{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{1\}})}{\sum_{\alpha=1}^{k} T_{b\alpha}^{\{1\}}}. \tag{1.10}$$

Como sugerido por Fan,T.H.e Hung,W.L. (1997), *as variâncias amostrais das estimativas de probabilidade e das medidas invariantes são dadas por:*

$$\hat{\sigma}_b^{2\{1\}} = \frac{1}{(\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{1\}}))^2} \sum_{\alpha=1}^{k} (g(\eta_{b\alpha}^{\{1\}})(p_{b\alpha}^{\{1\}}(i,j) - \hat{p}_n(i,j)))^2, \tag{1.11}$$

$$\tilde{\sigma}_b^{2\{1\}} = \frac{1}{(\sum_{\alpha=1}^{k} T_{b\alpha}^{\{1\}})^2} \sum_{\alpha=1}^{k} (T_{b\alpha}^{\{1\}})(\pi_{b\alpha}^{\{1\}}(i) - \hat{\pi}_n(i)))^2, \tag{1.12}$$

onde $g(\eta_{b\alpha}^{\{1\}})$ é o número de visitas ao estado i do bloco reamostrado $\eta_{b\alpha}^{\{1\}}$, $h(\eta_{b\alpha}^{\{1\}})$ é o número de transições (i,j) do bloco reamostrado $\eta_{b\alpha}^{\{1\}}$, $T_{b\alpha}^{\{1\}}$ é o tamanho do bloco reamostrado $\eta_{b\alpha}^{\{1\}}$, $p_{b\alpha}^{\{1\}}(i,j) = h(\eta_{b\alpha}^{\{1\}})/g(\eta_{b\alpha}^{\{1\}})$ e $\pi_{b\alpha}^{\{1\}}(i) = g(\eta_{b\alpha}^{\{1\}})/T_{b\alpha}^{\{1\}}$.

3. *As distribuições de* $(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma}$ *e* $(\hat{\pi}_n(i) - \pi_i)/\tilde{\sigma}$ *podem ser aproximadas, respectivamente, pela distribuições empíricas de* $(\tilde{p}_b^{\{1\}}(i,j) - \hat{p}_n(i,j))/\hat{\sigma}_b^{\{1\}}$ *e* $(\tilde{\pi}_b^{\{1\}}(i) - \hat{\pi}_n(i))/\tilde{\sigma}_b^{\{1\}}$, $b = 1, \ldots, B$.

Para estimar a função distribuição F de $M = (\hat{p}_n(i,j) - p_{ij})/\hat{\sigma}$, seja $\chi^{\{1\}} = \{\eta_1^{\{1\}}, \ldots, \eta_k^{\{1\}}\}$ a reamostra retirada aleatoriamente e com reposição de $\chi = \{\eta_1, \ldots, \eta_k\}$.

Seja $\tilde{p}^{\{1\}}(i,j)$ e $\hat{\sigma}^{2\{1\}}$, calculados na reamostra $\chi^{\{1\}}$, para a versão de $\hat{p}_n(i,j)$ e $\hat{\sigma}^2$, respectivamente. Então a estimativa de *bootstrap* de F é dada pela seguinte probabilidade condicional $\hat{F}(x) = P(M^{\{1\}} \leq x \mid \chi^{\{1\}})$, com $M^{\{1\}} = (\tilde{p}^{\{1\}}(i,j) - \hat{p}_n(i,j))/\hat{\sigma}^{\{1\}}$.

A partir da reamostragem uniforme, temos B reamostras independentes, $\chi_1^{\{1\}}, \ldots, \chi_B^{\{1\}}$, de χ, e seja $M_b^{\{1\}}$ a versão de M calculada na reamostra $\chi_b^{\{1\}}$. Assim, a aproximação de $\hat{F}$ pela reamostragem uniforme é dada por:

$$\hat{F}_u(x) = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}(M_b^{\{1\}} \leq x). \tag{1.13}$$

Note que condicionada a χ, $\hat{F}_u(x) \longrightarrow \hat{F}(x)$, com probabilidade 1, quando $B \longrightarrow \infty$, e que $E(\hat{F}_u(x) \mid \chi) = \hat{F}(x)$ e $Var(\hat{F}_u(x) \mid \chi) = \frac{1}{B}(1 - \hat{F}(x))\hat{F}(x)$.

Pelo Teorema 8 de Athreya e Fuh (1992), o $\sup_{-\infty<x<\infty} \mid \hat{F}(x) - \Phi(x) \mid \longrightarrow 0$, com probabilidade 1, quando $n \longrightarrow \infty$, onde $\Phi(.)$ é a função distribuição normal padrão.

Segue que com probabilidade 1, quando $n \longrightarrow \infty$, $\sup_{-\infty<x<\infty} \mid BVar\{\hat{F}_u(x) \mid \chi\} - \{1 - \Phi(x)\}\Phi(x) \mid \longrightarrow 0$, quando $B \longrightarrow \infty$.

Analogamente, para estimar a função distribuição F de $M = (\hat{\pi}_n(i) - \pi_i)/\tilde{\sigma}$, tem-se o Teorema 6 de Athreya e Fuh (1992).

1.2 Balanceada em Blocos

A idéia desta reamostragem é garantir que cada bloco apareça o mesmo número em todas as reamostras. Então, para uma realização $\{x_0, x_1, \ldots, x_n\}$ particionado, em k blo-

cos, o esquema de reamostragem balanceado em estimar a $P\{(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma} \leq x\}$ e a $P\{(\hat{\pi}_n(i) - \pi_i)/\tilde{\sigma} \leq x\}$ consiste nos seguintes passos:

1. *Fixe a transição (i,j) e decomponha a amostra original em k blocos: $\{\eta_1, \eta_2, ..., \eta_k\}$. Repita cada bloco B vezes, obtendo um total de Bk blocos.*

2. *Permute os Bk blocos aleatoriamente, e cada k blocos fornecem uma reamostra balanceada, num total de B reamostras, $\chi_b^{\{2\}} = \{\eta_{b1}^{\{2\}}, \ldots, \eta_{bk}^{\{2\}}\}$, $b = 1, \ldots, B$.*

3. *Para cada uma das reamostras definimos as estimativas da probabilidade de transição (i,j) e da medida invariante do estado i, na b-ésima iteração, respectivamente por :*

$$\tilde{p}_b^{\{2\}}(i,j) = \frac{\sum_{\alpha=1}^{k} h(\eta_{b\alpha}^{\{2\}})}{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{2\}})}, \tag{1.14}$$

$$\tilde{\pi}_b^{\{2\}}(i) = \frac{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{2\}})}{\sum_{\alpha=1}^{k} T_{b\alpha}^{\{2\}}}. \tag{1.15}$$

As variâncias amostrais das estimativas de probabilidade e da medida invariante são dadas por:

$$\hat{\sigma}_b^{2\{2\}} = \frac{1}{(\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{2\}}))^2} \sum_{\alpha=1}^{k} (g(\eta_{b\alpha}^{\{2\}})(p_{b\alpha}^{\{2\}}(i,j) - \hat{p}_n(i,j)))^2, \tag{1.16}$$

$$\tilde{\sigma}_b^{2\{2\}} = \frac{1}{(\sum_{\alpha=1}^{k} T_{b\alpha}^{\{2\}})^2} \sum_{\alpha=1}^{k} (T_{b\alpha}^{\{2\}})(\pi_{b\alpha}^{\{2\}}(i) - \hat{\pi}_n(i)))^2, \tag{1.17}$$

onde $g(\eta_{b\alpha}^{\{2\}})$ é o número de visitas ao estado i do bloco reamostrado $\eta_{b\alpha}^{\{2\}}$, $h(\eta_{b\alpha}^{\{2\}})$ é o número de transições (i,j) do bloco reamostrado $\eta_{b\alpha}^{\{2\}}$, $T_{b\alpha}^{\{2\}}$ é o tamanho do bloco reamostrado $\eta_{b\alpha}^{\{2\}}$, $p_{b\alpha}^{\{2\}}(i,j) = h(\eta_{b\alpha}^{\{2\}})/g(\eta_{b\alpha}^{\{2\}})$ e $\pi_{b\alpha}^{\{2\}}(i) = g(\eta_{b\alpha}^{\{2\}})/T_{b\alpha}^{\{2\}}$.

4. *A distribuição de* $(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma}$ *e* $(\hat{\pi}_n(i) - \pi_i)/\tilde{\sigma}$ *pode ser aproximada pela distribuição empírica de* $(\tilde{p}_b^{\{2\}}(i,j) - \hat{p}_n(i,j))/\hat{\sigma}_b^{\{2\}}$ *e* $(\tilde{\pi}_b^{\{2\}}(i) - \hat{\pi}_n(i))/\tilde{\sigma}_b^{\{2\}}$, $b = 1, \ldots, B$, *respectivamente.*

Para construir a aproximação da reamostragem balanceada de $\hat{F}$, primeiro obtemos B reamostras balanceadas $\chi_b^{\{2\}}$,$1 \leq b \leq B$. Seja $\tilde{p}^{\{2\}}(i,j)$ e $\hat{\sigma}^{2\{2\}}$, calculados na reamostra $\chi^{\{2\}}$, para a versão de $\hat{p}_n(i,j)$ e $\hat{\sigma}^{\{2\}}$, respectivamente. Então a estimativa de *bootstrap* de F é dada por $\hat{F}(x) = P(M^{\{2\}} \leq x \mid \chi)$, com $M^{\{2\}} = (\tilde{p}^{\{2\}}(i,j) - \hat{p}_n(i,j))/\hat{\sigma}^{\{2\}}$.

A aproximação de $\hat{F}$ pela reamostragem balanceada é dada por:

$$\hat{F}_b(x) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}(M_b^{\{2\}} \leq x). \tag{1.18}$$

Para obter a eficiência assintótica de $\hat{F}_b(x)$ em relação a $\hat{F}_u(x)$, é necessário verificar a variância assintótica de $\hat{F}_b(x)$, que é dado pelo teorema 1, página 228, de Fuh,C.D., Fan,T.H. e Hung,W.L. (1998), onde $\sup_{-\infty<x<\infty} \mid BVar\{\hat{F}_b(x) \mid \chi\} - \{(1-\Phi(x))\Phi(x) - \phi(x)^2\} \mid \longrightarrow 0$, com probabilidade 1, quando $n \longrightarrow \infty$, sendo $\Phi(.)$ a função distribuição normal padrão, e $\phi(.)$ a função densidade normal padrão.

De maneira análoga, obtemos a variância assintótica da distribuição de $\hat{\pi}(i)$. Comparando os resultados das variâncias da reamostragem balanceada e uniforme, é possível verificar que as estimativas da reamostragem de importância são mais eficientes em relação a reamostragem uniforme, quando $n \to \infty$, atraveés da seguinte relação:

$$\frac{Var(\hat{F}_u(x))}{Var(\hat{F}_b(x))} \sim \frac{\Phi(x)\{1-\Phi(x)\}}{\Phi(x)\{1-\Phi(x)\} - \phi(x)^2}. \tag{1.19}$$

1.3 Importância em Blocos

Com a notação da seção anterior, o algoritmo para estimar a $P\{(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma} \leq x\}$ consiste em:

1. *Para*

$$\epsilon_\alpha(i,j) = \hat{\sigma}^{-1}\frac{1}{\sum_{\alpha=1}^{k} g(\eta_\alpha)} g(\eta_\alpha)(p_\alpha(i,j) - \hat{p}_n(i,j)), \tag{1.20}$$

seja a distribuição de probabilidade de reamostragem $\{q_\alpha; \alpha = 1, \ldots, k\}$ *de* $\{\eta_\alpha; \alpha = 1, \ldots, k\}$ *dada por*

$$q_\alpha = \frac{exp(-A\epsilon_\alpha(i,j))}{\sum_{\alpha=1}^{k} exp(-A\epsilon_\alpha(i,j))},$$

onde A *é escolhido de tal forma que minimize* $exp(A^2)\Phi(x - A)$, *com* $\Phi(.)$ *sendo a função distribuição normal padrão.*

2. *Para* $b = 1, \ldots, B$, *com a amostra fixada, retire uma sub-amostra de tamanho* k, $\chi_b^{\{3\}} = \{\eta_{b1}^{\{3\}}, \ldots, \eta_{bk}^{\{3\}}\}$, *através da distribuição de probabilidade dos blocos* $\{q_\alpha; \alpha = 1, \ldots, k\}$. *Seja* $l_{b\alpha}$ *o número de vezes que* $\eta_{b\alpha}$ *aparece na reamostra* $\chi_b^{\{3\}}$. *A estimativa de* $\hat{p}_n(i,j)$ *é dada por :*

$$\tilde{p}_b^{\{3\}}(i,j) = \frac{\sum_{\alpha=1}^{k} h(\eta_{b\alpha}^{\{3\}})}{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{3\}})} \prod_{\alpha=1}^{k} (kq_\alpha)^{-l_{b\alpha}} \tag{1.21}$$

A variância amostral da estimativa de probabilidade $\tilde{p}_b^{\{3\}}$ *é dada por:*

$$\hat{\sigma}_b^{2\{3\}} = \frac{1}{(\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{3\}}))^2} \sum_{\alpha=1}^{k} (g(\eta_{b\alpha}^{\{3\}})(p_{b\alpha}^{\{3\}}(i,j) - \hat{p}_n(i,j)))^2, \tag{1.22}$$

onde $g(\eta_{b\alpha}^{\{3\}})$ *é o número de visitas ao estado* i *do bloco reamostrado* $\eta_{b\alpha}^{\{3\}}$, $h(\eta_{b\alpha}^{\{3\}})$ *é o número de transições* (i,j) *do bloco reamostrado* $\eta_{b\alpha}^{\{3\}}$ *e* $p_{b\alpha}^{\{3\}}(i,j) = h(\eta_{b\alpha}^{\{3\}})/g(\eta_{b\alpha}^{\{3\}})$.

3. *Da mesma forma que nas seções anteriores, a distribuição de* $(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma}$ *pode ser aproximada pela distribuição empírica de* $(\tilde{p}_b^{\{3\}}(i,j) - \hat{p}_n(i,j))/\hat{\sigma}_b^{\{3\}}$, $b = 1, \ldots, B$.

O procedimento para estimar $P\{(\hat{\pi}_n(i) - \pi_i)/\tilde{\sigma} \leq x\}$ pelo método de reamostragem de importância em blocos consiste em:

1. *Para*

$$\epsilon_\alpha(i) = \tilde{\sigma}^{-1} \frac{1}{\sum_{\alpha=1}^{k} T_\alpha} T_\alpha(\pi_\alpha(i) - \hat{\pi}_n(i)), \tag{1.23}$$

seja a distribuição de probabilidade de reamostragem $\{q_\alpha; \alpha = 1, \ldots, k\}$ *de* $\{\eta_\alpha; \alpha = 1, \ldots, k\}$ *dada por*

$$q_\alpha = \frac{exp(-A\epsilon_\alpha(i))}{\sum_{\alpha=1}^{k} exp(-A\epsilon_\alpha(i))},$$

onde A *é escolhido de tal forma que minimize* $exp(A^2)\Phi(x - A)$, *com* $\Phi(.)$ *sendo a função distribuição normal padrão.*

2. *Para* $b = 1, \ldots, B$, *com a amostra fixada, retire uma amostra de tamanho* k, $\chi_b^{\{3\}} = \{\eta_{b1}^{\{3\}}, \ldots, \eta_{bk}^{\{3\}}\}$, *através da distribuição de probabilidade dos blocos* $\{q_\alpha; \alpha = 1, \ldots, k\}$. *Então a estimativa de* $\tilde{\pi}_n(i)$ *é dada por :*

$$\tilde{\pi}_b^{\{3\}}(i) = \frac{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{3\}})}{\sum_{\alpha=1}^{k} T_{b\alpha}^{\{3\}}} \prod_{\alpha=1}^{k} (kq_\alpha)^{-l_{b\alpha}}. \tag{1.24}$$

A variância amostral da estimativa da medida invariante é dada por:

$$\tilde{\sigma}_b^{2\{3\}} = \frac{1}{(\sum_{\alpha=1}^{k} T_{b\alpha}^{\{3\}})^2} \sum_{\alpha=1}^{k} (T_{b\alpha}^{\{3\}})(\pi_{b\alpha}^{\{3\}}(i) - \hat{\pi}_n(i)))^2, \tag{1.25}$$

onde $T_{b\alpha}^{\{3\}}$ *é o tamanho do bloco reamostrado* $\eta_{b\alpha}^{\{3\}}$ *e* $\pi_{b\alpha}^{\{3\}}(i) = g(\eta_{b\alpha}^{\{3\}})/T_{b\alpha}^{\{3\}}$.

3. *Assim, a distribuição de* $(\hat{\pi}_n(i) - \pi_i)/\tilde{\sigma}$ *pode ser aproximada pela distribuição empírica de* $(\tilde{\pi}_b^{\{3\}}(i) - \hat{\pi}_n(i))/\tilde{\sigma}_b^{\{3\}}$, $b = 1, \ldots, B$.

Na reamostragem de importância, cada bloco η_α tem uma probabilidade de ser selecionado $p(\eta_\alpha)$, com $p(\eta_\alpha) > 0$ e $\sum_{\alpha=1}^{k} p(\eta_\alpha) = 1$. Retiradas B reamostras $\chi_b^{\{3\}}$, $1 \leq b \leq B$, então a aproximação de $\hat{F}(x)$, pela reamostragem de importância é

$$\hat{F}_i(x) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}(M_b^{\{3\}} \leq x) \prod_{\alpha=1}^{k} (kp(\eta_\alpha))^{-Q_{b\alpha}^{\{3\}}}, \tag{1.26}$$

onde $Q_{b\alpha}^{\{3\}}$ é o número de vezes que η_α aparece na reamostra $\chi_b^{\{3\}}$. O objetivo é encontrar $p(\eta_\alpha)$ que minimize a variância assintótica de $\hat{F}_i$, que é equivalente a minimizar

$$\hat{v} = E'\{\mathbf{1}(M^{\{1\}} \leq x) \prod_{\alpha=1}^{k} (kp(\eta_\alpha))^{-Q_\alpha^{\{1\}}}\},$$

onde $Q_\alpha^{\{1\}}$ é o número de vezes que η_α aparece na reamostra $\chi^{\{1\}}$, sob a reamostragem uniforme. Então,

$$\begin{aligned} M^{\{1\}} &= \frac{\tilde{p}^{\{1\}}(i,j) - \hat{p}_n(i,j)}{\hat{\sigma}^{\{1\}}} \\ &= \frac{\tilde{p}^{\{1\}}(i,j) - \hat{p}_n(i,j)}{\hat{\sigma}} \frac{\hat{\sigma}}{\hat{\sigma}^{\{1\}}} \\ &\approx \sum_{\alpha=1}^{k} Q_\alpha^{\{1\}} \epsilon_\alpha(i,j), \end{aligned}$$

onde

$$\epsilon_\alpha(i,j) = \frac{1}{\hat{\sigma}} \frac{g(\eta_\alpha)(p_\alpha(i,j) - \hat{p}_n(i,j))}{\sum_{\alpha=1}^{k} g(\eta_\alpha)},$$

e

$$\hat{v} \approx E'\{\mathbf{1}(\sum_{\alpha=1}^{k} Q_\alpha^{\{1\}} \epsilon_\alpha(i,j) \leq x) \prod_{\alpha=1}^{k} (kp(\eta_\alpha))^{-Q_\alpha^{\{1\}}}\} = E\{\mathbf{1}(V_1 \leq x) exp(V_2) \mid \chi\},$$

onde $V_1 = \sum_{\alpha=1}^{k} Q_\alpha^{\{1\}} \epsilon_\alpha$, $V_2 = \sum_{\alpha=1}^{k} Q_\alpha^{\{1\}} \delta_\alpha$ e $\delta_\alpha = -log(k(p(\eta_\alpha))$. Com k aleatório, as variáveis aleatórias V_1 e V_2 são somas aleatórias multinomiais ponderadas. De acordo com o Teorema de Athreya, K.B. e Fuh,C.D.(1992), página 236, condicionada a χ, (V_1, V_2) é assintóticamente normal bivariada, com média $(0, E(\sum_{\alpha=1}^{k} \delta_\alpha))$, variância $(1, E(\sum_{\alpha=1}^{k} \delta_\alpha^2))$ e

covariância
$E(\sum_{\alpha=1}^{k} \epsilon_\alpha(i,j)\delta_\alpha)$. Então quando $n \longrightarrow \infty$,

$$(\hat{v}) \approx exp(s^2)\Phi(x - s\rho),$$

onde $s^2 = E(\sum_{\alpha=1}^{k} \delta_\alpha^2)$ e $\rho = \frac{1}{s}E(\sum_{\alpha=1}^{k} \epsilon_\alpha(i,j)\delta_\alpha)$.

Os valores de s e ρ que minimizam $exp(s^2)\Phi(x - s\rho)$ são $(s,\rho) = (A,1)$, onde $A = A(x) > 0$ é tal que minimiza $exp(A^2)\Phi(x - A)$. Seja $\delta_\alpha = A\epsilon_\alpha + c$, onde c é escolhido tal que $\sum_{\alpha=1}^{k} p(\eta_\alpha) = 1$. Então, a variância assintótica da reamostragem de importância é mínima quando

$$p(\eta_\alpha) = \frac{exp(-A\epsilon_\alpha(i,j))}{\sum_{\alpha=1}^{k} exp(-A\epsilon_\alpha(i,j))},$$

para $\alpha = 1, \ldots, k$. Por um argumento similar, obtemos a variância assintótica da distribuição de $\hat{\pi}(i)$. Comparando os resultados das variâncias da reamostragem de importância e uniforme, é possível verificar que as estimativas da reamostragem de importância são mais eficientes em relação a reamostragem uniforme, quando $n \longrightarrow \infty$, através da seguinte relação,

$$\frac{Var(\hat{F}_u(x))}{Var(\hat{F}_i(x))} \sim \frac{\Phi(x)\{1 - \Phi(x)\}}{exp(A^2)\Phi(x - A) - \Phi(x)^2}. \tag{1.27}$$

1.4 Balanceada de Importância em Blocos

A reamostragem balanceada de importância é a combinação da reamostragem balanceada e da reamostragem de importância. O balanceado nesta reamostragem significa que o número de vezes em que cada bloco aparece em qualquer reamostragem é proporcional a probabilidade de ser selecionado pela reamostragem de importância.

Com a mesma notação da seção anterior, o algoritmo para estimar a $P\{(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma} \leq x\}$ consiste em:

1. *Para*

$$\epsilon_\alpha(i,j) = \hat{\sigma}^{-1}\frac{1}{\sum_{\alpha=1}^{k} g(\eta_\alpha)} g(\eta_\alpha)(p_\alpha(i,j) - \hat{p}_n(i,j)). \tag{1.28}$$

Seja a distribuição de probabilidade de reamostragem $\{q'_\alpha; \alpha = 1, \ldots, k\}$ *de* $\{\eta_\alpha; \alpha = 1, \ldots, k\}$ *por*

$$q'_\alpha = \frac{exp(-A'\epsilon_\alpha(i,j))}{\sum_{\alpha=1}^{k} exp(-A'\epsilon_\alpha(i,j))},$$

onde A' *é escolhido de tal forma que minimize* $exp(A'^2)\Phi(x-A') - \{\phi(x) - A'\Phi(x)\}^2$ *com* $\Phi(.)$ *sendo a função distribuição normal padrão e* ϕ *a função densidade normal padrão.*

2. *Defina* $B'_\alpha = [kBq'_\alpha]$, *onde* $[.]$ *denota a parte inteira, obtenha* $r = kB - \sum_{\alpha=1}^{k} B'_\alpha$ *e* $d_\alpha = kBq'_\alpha - B'_\alpha$. *Ordene* $\{d_1, \ldots, d_k\}$ *do menor para o maior. Seja* $B_\alpha = B'_\alpha + 1$, *se a ordem de* d_α *é maior ou igual a* r, *ou,* $B_\alpha = B'_\alpha$, *caso contrário. Assim* B_α *é o número de vezes que o bloco* η_α *aparece na reamostra balanceada. Ou seja, repete o bloco* η_α, B_α *vezes, totalizando* Bk *blocos.*

3. *Permute os* Bk *blocos aleatoriamente, e a cada* k *blocos obtenha uma reamostra balanceada, num total de* B *reamostras,* $\chi_b^{\{4\}} = \{\eta_{b1}^{\{4\}}, \ldots, \eta_{bk}^{\{4\}}\}$, $b = 1, \ldots, B$.

4. *Para cada uma das reamostras a estimativa da probabilidade de transição* (i,j), *na* b*-ésima iteração é dada por :*

$$\tilde{p}_b^{\{4\}}(i,j) = \frac{\sum_{\alpha=1}^{k} h(\eta_{b\alpha}^{\{4\}})}{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{4\}})} \prod_{\alpha=1}^{k} (kq_\alpha)^{-l_{b\alpha}} \tag{1.29}$$

A variância amostral da estimativa de probabilidade é dada por:

$$\hat{\sigma}_b^{2\{4\}} = \frac{1}{(\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{4\}}))^2} \sum_{\alpha=1}^{k} (g(\eta_{b\alpha}^{\{4\}})(p_{b\alpha}^{\{4\}}(i,j) - \hat{p}_n(i,j)))^2, \tag{1.30}$$

onde $g(\eta_{b\alpha}^{\{4\}})$ *é o número de visitas ao estado* i *do bloco reamostrado* $\eta_{b\alpha}^{\{4\}}$, $h(\eta_{b\alpha}^{\{4\}})$ *é o número de transições* (i,j) *do bloco reamostrado* $\eta_{b\alpha}^{\{4\}}$, $p_{b\alpha}^{\{4\}}(i,j) = h(\eta_{b\alpha}^{\{4\}})/g(\eta_{b\alpha}^{\{4\}})$.

5. *A distribuição de* $(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma}$ *pode então ser aproximada pela distribuição empírica de* $(\tilde{p}_b^{\{4\}}(i,j) - \hat{p}_n(i,j))/\hat{\sigma}_b^{\{4\}}$, $b = 1, \ldots, B$.

Com a mesma notação, o algoritmo para estimar a $P\{(\tilde{\pi}_n(i,j) - \pi_i)/\tilde{\sigma} \leq x\}$ consiste em:

1. *Para*
$$\epsilon_\alpha(i) = \tilde{\sigma}^{-1} \frac{1}{\sum_{\alpha=1}^{k} T_\alpha} T_\alpha(\pi_\alpha(i) - \hat{\pi}_n(i)). \tag{1.31}$$
Seja a distribuição de probabilidade de reamostragem $\{q'_\alpha; \alpha = 1, \ldots, k\}$ *de* $\{\eta_\alpha; \alpha = 1, \ldots, k\}$ *dada por*
$$q'_\alpha = \frac{exp(-A'\epsilon_\alpha(i))}{\sum_{\alpha=1}^{k} exp(-A'\epsilon_\alpha(i))},$$
onde A' *é escolhido de tal forma que minimize* $exp(A'^2)\Phi(x - A') - \{\phi(x) - A'\Phi(x)\}^2$, *com* $\Phi(.)$ *sendo a função distribuição normal padrão e* ϕ *a função densidade normal padrão.*

2. *Defina* $B'_\alpha = [kBq'_\alpha]$, *onde* $[.]$ *denota a parte inteira, obtenha* $r = kB - \sum_{\alpha=1}^{k} B'_\alpha$ *e* $d_\alpha = kBq'_\alpha - B'_\alpha$. *Ordene* $\{d_1, \ldots, d_k\}$ *do menor para o maior. Seja* $B_\alpha = B'_\alpha + 1$, *se a ordem de* d_α *é maior ou igual a* r, *ou,* $B_\alpha = B'_\alpha$, *caso contrário. Assim* B_α *é o número de vezes que o bloco* η_α *aparece na reamostra balanceada. Ou seja, repete o bloco* η_α, B_α *vezes, totalizando* Bk *blocos.*

3. *Permute os* Bk *blocos aleatoriamente, e a cada* k *blocos obtemos uma reamostra balanceada, num total de* B *reamostras,* $\chi_b^{\{4\}} = \{\eta_{b1}^{\{4\}}, \ldots, \eta_{bk}^{\{4\}}\}$, $b = 1, \ldots, B$.

4. *Então a estimativa de* $\tilde{\pi}_n(i)$ *é dada por:*
$$\tilde{\pi}_b^{\{4\}}(i) = \frac{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{4\}})}{\sum_{\alpha=1}^{k} T_{b\alpha}^{\{4\}}} \prod_{\alpha=1}^{k} (kq_\alpha)^{-l_{b\alpha}}, \tag{1.32}$$

onde $T_{b\alpha}^{\{4\}}$ *é o comprimento de* $\eta_{b\alpha}^{\{4\}}$.

A variância amostral da estimativa de $\tilde{\pi}_b^{\{4\}}$ *é dada por:*

$$\tilde{\sigma}_b^{2\{4\}} = \frac{1}{(\sum_{\alpha=1}^{k} T_{b\alpha}^{\{4\}})^2} \sum_{\alpha=1}^{k} (T_{b\alpha}^{\{4\}}(\pi_{b\alpha}^{\{4\}}(i) - \hat{\pi}_n(i)))^2, \tag{1.33}$$

onde $T_{b\alpha}^{\{4\}}$ *é o tamanho do bloco reamostrado* $\eta_{b\alpha}^{\{4\}}$ *e* $\pi_{b\alpha}^{\{4\}} = g(\eta_{b\alpha}^{\{4\}})/T_{b\alpha}^{\{4\}}$.

5. *Assim, a distribuição de* $(\hat{\pi}_n(i) - \pi_i)/\tilde{\sigma}$ *pode ser aproximada pela distribuição empírica de* $(\tilde{\pi}_b^{\{4\}}(i) - \hat{\pi}_n(i))/\tilde{\sigma}_b^{\{4\}}$, $b = 1, \ldots, B$.

Obtenha B reamostras balanceadas de importância $\chi^{\{4\}}$, $1 \leq b \leq B$, a aproximação de reamostragem balanceada de importância de $\hat{F}$ é

$$\hat{F}_{bi}(x) = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}(M_b^{\{4\}} \leq x) \prod_{\alpha=1}^{k} (kp(\eta_\alpha))^{-Q_{b\alpha}^{\{4\}}} = \frac{1}{B} \sum_{b=1}^{B} U_b, \tag{1.34}$$

onde $Q_{b\alpha}^{\{4\}}$ é o número de vezes que η_α aparece em $\chi_b^{\{4\}}$. Temos por Athreya,K.B e Fuh,C.D.(1992),página 232, que

$$BVar(\hat{F}_{bi}(x)) \approx exp(s^2)\Phi(x-s) - \Phi(x)^2 - \{\phi(x) - s\Phi(x)\}^2.$$

A variância mínima assintótica da reamostragem balanceada de importância ocorre quando

$$p(\eta_\alpha) = \frac{exp(-A'\epsilon_\alpha(i,j))}{\sum_{\alpha=1}^{k} exp(-A'\epsilon_\alpha(i,j))},$$

para $\alpha = 1, \ldots, k$, onde A' minimiza $exp(A'^2)\Phi(x - A') - \{\phi(x) - A'\Phi(x)\}^2$. Por um argumento similar, obtemos variância assintótica da distribuição de $\hat{\pi}(i)$. Comparando os resultados das variâncias da reamostragem balanceada de importância e uniforme, é possível verificar que as estimativas da reamostragem balanceada de importância são mais eficientes em relação a reamostragem uniforme, quando $n \to \infty$, através da seguinte relação:

$$\frac{Var(\hat{F}_u(x))}{Var(\hat{F}_{bi}(x))} \sim \frac{\Phi(x)\{1 - \Phi(x)\}}{exp(A'^2)\Phi(x - A') - \Phi(x)^2 - \{\phi(x) - A'\Phi(x)\}^2}. \tag{1.35}$$

Capítulo 2

Classificadores

Na mineração de dados, o método de classificação é um dos mais utilizados. Pode ser utilizado na descoberta de características de clientes, na detecção de fraudes, em aplicações na análise de risco em créditos, entre outras. O objetivo fundamental deste método está em extrair conhecimento de um conjunto de dados tal que essas regras possam ser aplicadas a novos dados.

De certa forma, classificação pode ser resumida em encontrar uma função que mapeie objetos com características comuns, na mesma classe. A classificação pode possuir dois significados diferentes:

1. Dado um conjunto de observações, tentar estabelecer a existência de classes ou grupos nos dados;

2. Dado um conjunto de observações no qual as classes são conhecidas, o objetivo está em encontrar regras capazes de classificar novas observações entre as classes já existentes.

O primeiro tipo é conhecido como Aprendizado Não Supervisionado, o segundo como Aprendizado Supervisionado. Aqui daremos enfoque em aprendizado supervizionado.

Um aprendizado pode ser visto como um sistema que ajuda a construir uma tomada de decisões chamada de *classificador*. Esse sistema tem disponível um conjunto finito de exemplos de casos resolvidos. Os dados de cada caso consistem em um padrão de

observações e a correta classificação correspondente. O objetivo é moldar a estrutura do classificador para um problema específico, encontrando uma forma genérica de relatar qualquer padrão particular de observações em uma das classes especificadas.

Para definir um classificador considere $\underline{x}$ um vetor de medidas, então X é o espaço que contém todos os vetores possíveis de medidas. Supondo que os casos estejam em J classes, temos que $C = \{1, \ldots, J\}$ é o conjunto das classes. Assim para qualquer $\underline{x} \in X$, a regra prediz uma das classes em C para $\underline{x}$.

Definição 2.1: *Um classificador ou uma regra de classificação é uma função*

$$d : X \longrightarrow C.$$

Uma outra maneira de definir um classificador é definir A_j como um subconjunto de X tal que, $A_j = \{\underline{x}; d(\underline{x}) = j\}$. Os conjuntos $A_1, \ldots, A_J$ formam uma partição de X.

Definição 2.2: *Um classificador é uma partição de X, tal que para todo $\underline{x} \in A_j$ a classe predita é j.*

A construção de um classificador depende da amostra de aprendizado, ou de treinamento.

Definição 2.3: *A amostra de aprendizado denotado por $L = \{(x_1, j_1), \ldots, (x_N, j_N)\}$, onde $x_n \in X$ e $j_n \in \{1, \ldots, J\}$, $n = 1, \ldots, N$.*

O principal objetivo de um sistema de aprendizado está em classificar com sucesso casos nunca vistos. A taxa de erro é a forma mais utilizada para medir a performance de um classificador. Definimos, a seguir, três maneiras de estimar a taxa de erro de um classificador.

Definição 2.4: *Para $X \in X$, $Y \in C$, seja (X, Y) uma amostra da distribuição de probabilidade $P(A, j)$. Definimos*

$$R(d) = P(d(X) \neq Y). \tag{2.1}$$

Dado um classificador, denota-se por $R(d)$ a verdadeira taxa de classificação incorreta. Usando L constrói-se d, e com um outro conjunto de dados de L, observa-se a classificação correta para cada um dos casos deste novo conjunto de dados e então encontra-se a classificação predita usando $d(\underline{x})$, e temos que a proporção da classificação incorreta por d é o valor de $R^*(d)$.

Uma notação mais precisa seria: $P(d(X) \neq Y|L)$, ou seja, a probabilidade da classificação incorreta de uma nova amostra dada a amostra de aprendizado L.

Se L for usado para construir d e estimar $R(d)$ então tais estimativas são denominadas estimativas internas. Um tipo de estimativa possível é a resubstituição definida como,

Definição 2.5:

$$R(d) = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}(d(x_n) \neq j_n). \tag{2.2}$$

Uma outra maneira de obter uma estimativa interna é usar a estimação de uma amostra de teste, onde os casos de L são divididos em dois subconjuntos L_1 e L_2. O subconjunto L_1 é usado para construção de d, e o subconjunto L_2 para estimar $R(d)$.

Definição 2.6: *Seja N_2 o número de casos do subconjunto L_2. Então,*

$$R(d) = \frac{1}{N_2}\sum_{(x_n,j_n)\in L_2} \mathbf{1}(d(x_n) \neq j_n). \tag{2.3}$$

Um problema na classificação é o processo de *overfitting*, que surge quando um classificador tem um desempenho bom na amostra de treinamento mas nos dados novos tem desempenho ruim. Isso pode acontecer porque durante o processo de treinamento, existe um esforço muito grande do sistema para *acomodar* os dados de treinamento. Tal esforço pode levar a uma extração de regras pouco genéricas para casos novos.

Na literatura o classificador mais utilizado para combinar é o da Árvore de Decisão. Iremos na próxima seção discutir a construção de um classificador deste tipo.

2.1 Árvore de decisão

Árvores de Decisão são utilizadas em algoritmos de classificação, que são representações do conhecimento e, ao mesmo tempo, um eficiente meio de construir classificadores que predizem classes baseadas nos valores de atributos de um conjunto de dados. Tem a função de particionar recursivamente um conjunto de treinamento, até que cada subconjunto obtido contenha casos de uma única classe. Para isso, a técnica examina e compara a distribuição de classes durante a construção da árvore. Os resultados são os dados organizados que são utilizados para classificar novos casos.

A capacidade de discriminação de uma árvore de decisão vem da divisão do espaço em sub-espaços e a cada um destes é associado uma classe. Os mais conhecidos algoritmos de árvores de decisão são: CART(Classification and Regression Trees) de Breiman e Friedman (1984), ID3 de Quinlan (1986) e C4.5 de Quinlan (1993).

A estrutura de árvores de decisão é a seguinte:

1. Cada nó representa um atributo da base de dados.
2. Cada galho representa um valor do atributo.
3. Cada folha representa uma classe.
4. Cada percurso corresponde a uma regra de classificação.

Então um caso é classificado seguindo o caminho da raíz até a folha, enquanto suas características satisfazem as ligações. Existe apenas um caminho que é utilizado para classificar os dados.

Os passos para a construção de uma árvore de decisão são:

1. Escolher um atributo como sendo a raíz;
2. Criar galhos para todos os diferentes valores do atributo selecionado no passo 1;
3. Se todos os exemplos de treinamento sobre uma folha pertencerem a uma mesma classe, esta folha recebe o nome de classe. Se todas as folhas possuem uma classe, o algoritmo termina.

4. Caso contrário o nó é determinado com um atributo que não ocorra no trajeto da raíz, e galhos são criados para todos os valores, e retorna-se ao passo 3.

Um critério utilizado para a escolha do atributo pode ser a Entropia.

Um sistema completo de eventos $A_1, \ldots, A_J$ é um conjunto tal que um e somente um deles deve ocorrer a cada tentativa. Se tivermos os eventos $A_1, \ldots, A_J$, conjuntamente com suas probabilidades estimadas $p_1, \ldots, p_J$ ($p_i \geq 0, \sum_{i=1}^{J} p_i = 1$), então temos um esquema finito dado por

$$\begin{pmatrix} A_1 & A_2 & \ldots & A_J \\ p_1 & p_2 & \ldots & p_J \end{pmatrix}$$

Todo esquema finito descreve um estado de incerteza, onde se deseja prever o resultado de um experimento com base nas probabilidades de cada evento. O grau de incerteza é possivelmente diferente para esquemas diferentes.

A medida de entropia busca medir o grau de incerteza em cada esquema finito, e é definida como,

Definição 2.1.1:

$$Entropia(p_1, \ldots, p_J) = -\sum_{i=1}^{J} p_i log(p_i). \tag{2.4}$$

Na definição acima, os logaritmos são tomados numa base fixa qualquer e $p_i log(p_i) = 0$ sempre que $p_i = 0$.

O ganho da informação é dado pela diferença entre o valor da informação do atributo categórico e do atributo em questão, e é utilizado este conceito de ganho para construir as árvores de decisão. Em cada nó é colocado o atributo com maior ganho entre os atributos ainda não considerados no caminho desde a raíz até o nó.

Se o atributo for numérico utiliza-se um teste que produza uma partição do conjunto de exemplos de treinamento, utilizando para isso um ponto de referência, que pode ser qualquer ponto intermediário entre dois valores diferentes e consecutivos dos valores observados.

2.2 Modelo de Regressão Logística

Cox (1970) tornou o modelo de Regressão Logística popular entre os estatísticos. Este modelo fornece estimativas das probabilidades de cada observação pertencer a uma determinada classe, e a observação é classificada na classe com maior probabilidade.

Para a variável resposta $Y \in \{0, 1\}$, considere o modelo de regressão logístico linear simples em que $p(x)$, a probabilidade de $Y = 1$ dado o valor x de uma variável explicativa, e defina:

$$log\{\frac{p(x)}{1 - p(x)}\} = \alpha + \beta x, \tag{2.5}$$

onde α e β são parâmetros desconhecidos, obtendo assim a relação

$$p(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}, \tag{2.6}$$

sendo essas estimativas restritas no intervalo [0, 1].

O modelo estimado por Máxima Verossimilhança é dado por:

$$\begin{aligned} l(y, \alpha, \beta) &= \sum_{i=1}^{n} \{y_i log(p(x_i)) + (1 - y_i) log(1 - p(x_i))\} \\ &= \sum_{i=1}^{n} \{y_i x_i - log(1 + e^{\alpha+\beta x_i})\}, \end{aligned}$$

onde l é o logaritmo da função de verossimilhança.

Derivando em relação ao parâmetros desconhecidos obtemos as equações não-lineares, e assim pode ser utilizado o método iterativo de Newton-Raphson para obtenção da estimativa de máxima verossimilhanção dos parâmetros.

Estimando os parâmetros e as probabilidades, utilizamos estas para classificar as observações. Uma observação que apresenta alta probabilidade em pertencer a determinada classe é alocada para esta classe.

2.3 Combinação de Classificadores

Recentes resultados de Dietterich (2000), afirmam que é possível construir combinações de classificadores com maior precisão do que um único classificador e isso se deve a três razões.

A primeira razão é estatística. Um algoritmo de aprendizado faz uma busca no espaço H de hipóteses para identificar a melhor. O problema estatístico aparece quando o conjunto de treinamento é muito pequeno e com alta variabilidade comparado ao espaço de hipóteses. Sem uma quantidade suficiente de dados, o algoritmo de aprendizado pode encontrar diferentes hipóteses em H com a mesma precisão no conjunto de treinamento. Construindo uma combinação com todos esses classificadores, o algoritmo pode fazer uma combinação dos resultados dos mesmos e reduzir o risco de fornecer um resultado incorreto.

A segunda razão é computacional. Muitos algoritmos de aprendizado utilizam alguma forma de busca restrita que pode parar em um ótimo local. Nos casos em que o conjunto de treinamento é suficiente, ou seja o problema estatístico é ausente, ainda pode ser difícil para o algoritmo encontrar a melhor hipótese. Então uma combinação construída com a busca local partindo de pontos diferentes pode obter uma melhor aproximação da verdadeira função desconhecida do que qualquer um dos classificadores individuais.

A terceira razão é representacional. Na maioria das aplicações de aprendizado, a hipótese verdadeira pode não ser representada por nenhuma das hipóteses no espaço H. Desenvolvendo uma soma ponderada das hipóteses retiradas de H, pode ser possível ampliar o espaço das hipóteses. Há muitos algoritmos de aprendizado para os quais H é o espaço de todos os possíveis classificadores, mas com uma amostra de treinamento finita, esses algoritmos irão explorar apenas um conjunto finito de hipóteses e finalizarão a busca quando encontrarem uma hipótese que satisfaça o conjunto de treinamento. Deve-se considerar o espaço H como o das hipóteses efetivamente pesquisado pelo algoritmo de aprendizado para um dado conjunto de treinamento.

Um método de combinação de classificadores utilizado é o método do votante (*voting*). Ao utilizar este método, são gerados classificadores com diferentes conjuntos de treinamento e sua combinação é feita através da votação dos resultados de cada classificador.

Essa votação pode ou não ser ponderada e, a combinação, possivelmente, terá uma precisão melhor do que qualquer um dos classificadores que a gerou.

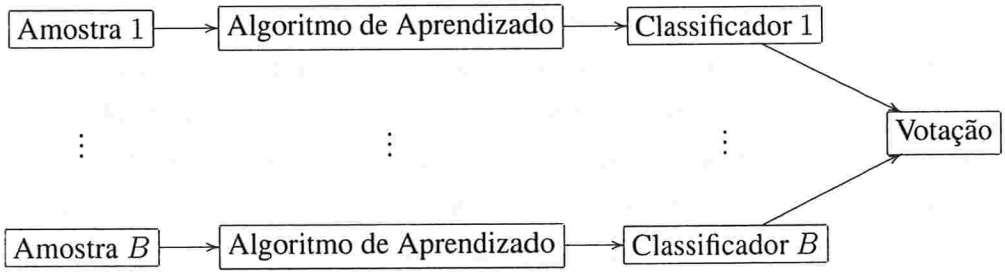
O método do votante (*voting*) pode ser dividido em dois tipos: os que mudam a distribuição do conjunto de treinamento baseado no desempenho dos classificadores anteriores, como o algoritmo *Boosting*, e os que não mudam a distribuição, como o algoritmo *Bagging*.

Nas próximas seções iremos descrever os algoritmos *Bagging* e *Boosting*.

2.3.1 *Bagging*

O algoritmo descrito por Breiman (1996) que recebeu o nome de *Bootstrap Aggregating*, ou *Bagging*, realiza a combinação de classificadores gerados pelo mesmo algoritmo de aprendizagem. De acordo com esse algoritmo, dado um conjunto de treinamento N, com n exemplos, são geradas B amostras desse conjunto, chamadas *Bootstrap*. Cada amostra é formada tomando aleatoriamente n exemplos de N, com substituição. Assim, eles possuem o mesmo número de exemplos de N, sendo que alguns podem aparecer mais de uma vez e outros podem não aparecer.

As amostras são usadas como novos conjuntos de treinamento, dando origem a B classificadores. Com o objetivo de gerar o classificador final os B classificadores são agregados, no caso de classificação a classe mais votada nas B amostras é escolhida.



Seja $S = \{(x_1, y_1), ..., (x_m, y_m)\}$ um conjunto de dados de treinamento, e $\hat{d}_b(x)$ um classificador construído com base na amostra S. Então o procedimento *Bagging* consiste em :

Para $b = 1, ..., B$,

1. *Construir uma amostra bootstrap* S_b^* *, de acordo com a distribuição uniforme.*
2. *Construir o classificador* $\hat{d}_b(x)$ *utilizando a amostra* S_b^*.

O classificador final é dado pela classe mais votada :

$$\hat{d}_B(x) = argmax_j\{\hat{d}_b(x) = j\}.$$

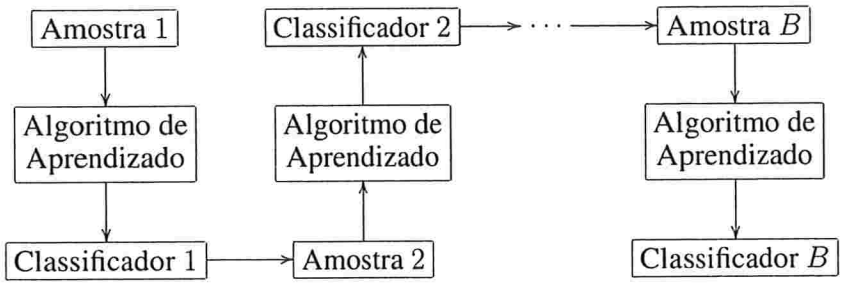
Existe vasta literatura deste procedimento no contexto do desempenho de classificadores instáveis, ou seja classificadores com alta variância.

2.3.2 *Boosting*

Boosting, descrito por Freund e Schapire (1996), é uma tentativa de produzir novos classificadores, que são melhores principalmente por classificar corretamente dados para os quais a eficiência dos classificadores comuns é ruim. Ao contrário de *Bagging*, a probabilidade de selecionar um exemplo não é igual para todos os exemplos do conjunto de treinamento. Isso depende da freqüência com que as amostras foram mal classificadas.

Um dos algoritmos mais utilizados do *Boosting* é o *AdaBoost* (*Adaptive Boosting*), que foi apresentado por Freund e Schapire (1997). Dado um conjunto de treinamento N, este algoritmo atribui um peso a cada um dos n exemplos, gerando um novo conjunto de treinamento N_1. Em cada iteração b, um novo conjunto N_b é gerado, com base em N_{b-1}, e a partir desses conjuntos vários classificadores são formados. Em cada iteração os pesos são ajustados de acordo com o desempenho do classificador anterior. Os exemplos que foram classificados de forma incorreta têm o peso modificado, se o erro de classificação é grande este peso é pequeno, se o erro é pequeno este peso é grande. Com isso, os classificadores modificam a distribuição do conjunto de treinamento.

Após as B iterações serem realizadas, a resposta do classificador final é formada através de uma votação ponderada. Nessa votação o peso do voto de cada classificador depende do seu desempenho no conjunto de treinamento usado para construí-lo. Note que o *Adaboost* gera os classificadores em série diferente do *Bagging*, que os gera em paralelo.



2.3.3 Uma análise generalizada do *Adaboost*.

Seja $S = \{(x_1, y_1), ..., (x_m, y_m)\}$ um conjunto de dados de treinamento, onde $y_i \in \{-1, +1\}$, $x_i \in X$, $i = 1, ..., m$, sendo $X \subset \mathbb{R}^p$, para algum $p \in \mathbb{N} = \{1, 2, \ldots\}$. Assumimos que um algoritmo de aprendizado base aceita como entrada um conjunto de treinamento S com distribuição $\mathcal{D}$ de probabilidade em $\{1, ..., m\}$.

Dada tal entrada, o algoritmo ajusta o classificador, em geral $h_b : X \longrightarrow \mathbb{R}$. O sinal de $h_b(x_i)$, onde $sinal[f] = \begin{cases} -1, & f < 0 \\ 1, & f \geq 0, \end{cases}$ é interpretado como o valor predito de x_i, e a $\mid h_b(x_i) \mid$ é a confiança desta predição. Então se $\mid h_b(x_i) \mid$ está próximo ou não de zero, isto é interpretado como uma baixa ou alta confiança.

A idéia do método é usar um aprendizado para formar uma regra de predição com alta acurácia usando, para isso, algoritmos base com distribuições diferentes sobre o mesmo conjunto de treinamento.

O efeito principal da regra de atualização do *Adaboost* é diminuir ou aumentar o peso dos casos, do conjunto de treinamento, classificados corretamente ou incorretamente por $h_b(x_i)$. O procedimento consiste em :

Seja a distribuição da primeira iteração $D_1(i) = \frac{1}{m}$, *para* $i = 1, ..., m$;

Para $b = 1, ..., B$

1. *Ajustar o classificador* $h_b(x_i)$ *usando a distribuição* D_b

2. *Define o erro de cada classificador como,*

$$\epsilon_b = P_{i \sim D_b}[h_b(x_i) \neq y_i],$$

isto é, para cada $b \geq 1$, *o erro de cada observação* i *é a probabilidade de* $h_b(x_i) \neq y_i$ *sob* D_b.

3. *Seja*

$$\alpha_b = \frac{1}{2} log(\frac{1 - \epsilon_b}{\epsilon_b})$$

4. *Modifica a distribuição da próxima iteração* $b + 1$,

$$D_{b+1}(i) = \frac{1}{Z_b} D_b(i) exp(-\alpha_b y_i h_b(x_i)),$$

para $i = 1, \ldots, m$, *tal que* $\sum_{i=1}^{m} D_{b+1}(i) = 1$.

Seja $f(x_i) = \sum_{b=1}^{B} \alpha_b h_b(x_i)$ o classificador final baseado na função discriminante que é uma combinação das saídas dos classificadores,

$$H(x_i) = sinal[f(x_i)]. \tag{2.7}$$

Uma propriedade do *Adaboost* consiste na redução do erro de treinamento. Schapire e Singer [23] (1999) provaram o seguinte limite do erro de treinamento do classificador final H.

Teorema 2.3.3.1: *Seja* E *a proporção de observações classificadas de forma incorreta no conjunto de treinamento, então* E *satisfaz a seguinte desigualdade,*

$$E = \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\{H(x_i) \neq y_i\}} \leq \prod_{b=1}^{B} Z_b \tag{2.8}$$

Prova: Se $H(x_i) \neq y_i$ então $y_i f(x_i) \leq 0$, assim $exp(-y_i f(x_i)) \geq 1$.

Logo, $\mathbf{1}_{\{H(x_i) \neq y_i\}} \leq exp(-y_i f(x_i))$, então

$$\frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\{H(x_i) \neq y_i\}} \leq \frac{1}{m} \sum_{i=1}^{m} exp(-y_i f(x_i)).$$

Sendo

$$D_{B+1}(i) = \frac{D_B(i)exp(-\alpha_b y_i h_b(x_i))}{Z_b},$$

assim

$$D_{B+1}(i) = \frac{(1/m)exp(-y_i f_(x_i))}{\prod_{b=1}^{B} Z_b}$$

e usando este fato temos que :

$$\frac{1}{m}\sum_{i=1}^{m} exp(-y_i f(x_i)) = \sum_{i=1}^{m} D_{B+1}(i) \prod_{n=1}^{B} Z_b = \prod_{b=1}^{T} Z_b \qquad \square$$

Uma conseqüência desse teorema é que com o objetivo de minimizar o erro de treinamento, uma aproximação seria minimizar Z_b em cada iteração do algoritmo, assim é possível aplicar esta idéia na escolha do α_b.

2.3.4 Escolhendo α_b

Seja $u_i = y_i h_b(x_i)$, e assumindo que $D_b(i) \neq 0$ para todo i, o objetivo é encontrar α_b que minimiza Z_b.

Para hipóteses $h_b \in [-1, 1]$ a escolha de α_b pode ser obtida por aproximar Z_b como segue.

Seja,

$$Z_b = \sum_{i=1}^{m} D_b(i)exp(-\alpha_n u_i) \leq \sum_{i=1}^{m} D_b(i) \left[\frac{1+u_i}{2}exp(-\alpha_b) + \frac{1-u_i}{2}exp(\alpha_b)\right]$$

Esta cota é válida se $u_i \in [-1, 1]$. Considerando $x = \alpha$, $y = -\alpha$, $\beta = \frac{1+u_i}{2}$ e uma $f(z) = exp(-z)$, então pela convexidade da função temos a seguinte desigualdade:

$$exp(-\alpha u_i) \leq \left(\frac{1+u_i}{2}\right) exp(-\alpha) + \left(\frac{1-u_i}{2}\right) exp(\alpha)$$

Assim é possível obter o seguinte corolário.

Corolário 2.3.4.1 : *Assumindo que cada* $h_b \in [-1, 1]$ *temos*

$$\alpha_b = \frac{1}{2} log \left(\frac{1 + r_b}{1 - r_b} \right) \tag{2.9}$$

onde $r_b = \sum_{i=1}^{m} D_b(i) y_i h_b(x_i)$. O erro de treinamento de H satisfaz

$$E \leq \prod_{b=1}^{B} \sqrt{1 - r_b^2}. \tag{2.10}$$

Prova: Pelo lado direito da desigualdade temos:

$$G = \sum_{i=1}^{m} D_b(i) \left[\frac{1 + u_i}{2} exp(-\alpha_b) + \frac{1 - u_i}{2} exp(\alpha_b) \right].$$

Derivando em relação a α_b:

$$\frac{\partial G}{\partial \alpha_b} = \sum_{i=1}^{m} \left[-\frac{1 + u_i}{2} exp(-\alpha_b) + \frac{1 - u_i}{2} exp(\alpha_b) \right]$$

E igualando a derivada a zero, temos:

$$\begin{aligned} \alpha_b &= \frac{1}{2} log \left[\frac{\sum_{i=1}^{m} D_b(i)(1 + u_i)}{\sum_{i=1}^{m} D_b(i)(1 - u_i)} \right] \\ &= \frac{1}{2} log \left[\frac{1 + r_b}{1 - r_b} \right] \end{aligned}$$

Segue que,

$$\begin{aligned} Z_b &\leq \sum_{i=1}^{m} D_b(i) \left[\frac{1 + u_i}{2} exp \left(-\frac{1}{2} log \left(\frac{1 + r_b}{1 - r_n} \right) \right) + \frac{1 - u_i}{2} exp \left(-\frac{1}{2} log \left(\frac{1 + r_b}{1 - r_b} \right) \right) \right] \\ &\leq \left(\frac{\sum_{i=1}^{m} D_b(i) + r_b}{2} \right) \left(\frac{1 + r_b}{1 - r_b} \right)^{-1/2} + \left(\frac{\sum_{i=1}^{m} D_b(i) - r_b}{2} \right) \left(\frac{1 + r_b}{1 - r_b} \right)^{1/2} \\ &\leq \left(\frac{1 + r_b}{2} \right) \left(\frac{1 - r_b}{1 + r_b} \right)^{1/2} + \left(\frac{1 - r_b}{2} \right) \left(\frac{1 + r_b}{1 - r_b} \right)^{1/2} \\ &\leq \sqrt{(1 + r_n)(1 - r_n)} \\ &\leq \sqrt{1 - r_b^2} \end{aligned}$$

Quando $h_b \in \{-1, 1\}$ temos:

$$\begin{aligned} r_b &= \sum_{i:y_i=h_b(x_i)} D_b(i)y_ih_b(x_i) + \sum_{i:y_i\neq h_b(x_i)} D_b(i)y_ih_b(x_i) \\ &= \sum_{i:y_i=h_b(x_i)} D_b(i) - \sum_{i:y_i\neq h_b(x_i)} D_b(i) \\ &= (1-\epsilon_b) - \epsilon_b = 1 - 2\epsilon_b. \end{aligned}$$

Substituindo $\alpha_b = \frac{1}{2}log\left[\frac{1-\epsilon_b}{\epsilon_b}\right]$ e $Z_b \leq 2\sqrt{\epsilon_b(1-\epsilon_b)}$, segue que,

é possível também determinar α_b da seguinte forma:

$$\begin{aligned} Z_b &= \sum_{i=1}^{m} D_n(i)exp(-y_i\alpha_b h_b(x_i)) \\ &= \sum_{i:y_i=h_b(x_i)} D_b(i)exp(-\alpha_b) + \sum_{i:y_i\neq h_b(x_i)} D_b(i)exp(\alpha_b) \\ &= (1-\epsilon_b)exp(-\alpha_b) + \epsilon_b exp(\alpha_b). \end{aligned}$$

Derivando em relação a α_b temos,

$$\frac{\partial Z_b}{\partial \alpha_b} = -(1-\epsilon_b)exp(-\alpha_b) + \epsilon_b exp(\alpha_b).$$

E igualando a derivada a zero temos,

$$\alpha_b = \frac{1}{2}log\left(\frac{1-\epsilon_b}{\epsilon_b}\right).$$

Assim Z_b pode ser escrito como:

$$\begin{aligned} Z_b &= (1-\epsilon_b)exp\left(\frac{-1}{2}log\left(\frac{1-\epsilon_b}{\epsilon_b}\right)\right) + \epsilon_b exp\left(\frac{1}{2}log\left(\frac{1-\epsilon_b}{\epsilon_b}\right)\right) \\ &= 2\sqrt{\epsilon_b(1-\epsilon_b)}. \end{aligned}$$

Portanto o erro do classificador final resulta em : $E \leq \prod_{b=1}^{B} 2\sqrt{(\epsilon_b(1-\epsilon_b))}$ □

O limite superior do erro, portanto, pode ser escrito na seguinte forma,

$$E \leq \prod_{b=1}^{B} 2\sqrt{1-4\gamma_b^2} = exp(-\sum_{b=1}^{B}(1/2)ln(\frac{1}{1-4\gamma_b^2})) \leq exp(-2\sum_{b=1}^{B}\gamma_b^2),$$

onde $\gamma_b = 1/2 - \epsilon_b$.

Assim, se cada classificador for ligeiramente melhor que uma escolha aleatória, o erro no conjunto de treinamento cai exponencialmente a zero.

Uma propriedade importante de um classificador é que ele tenha erro de generalização pequeno, ou seja, que a probabilidade do classificador errar em uma observação que não foi utilizada para sua construção seja pequena.

Dois métodos de analisar o erro de generalização do *Adaboost* têm sido propostos. O primeiro dado por Freund e Schapire (1997) usa o conceito de dimensão VC (esta dimensão é a medida de complexidade de um espaço de hipóteses) para limitar o erro de generalização da hipótese final em termos de erro de treinamento e um termo adicional que é uma função de dimensão VC da hipótese final e o número de exemplos de treinamento. O resultado deste primeiro teorema sugere que o *Adaboost* irá sofrer um desempenho ruim com dados novos, mesmo obtendo um desempenho bom no conjunto de treinamento, se o número de iterações crescer, com N fixo.

Schapire (1998) propôs uma análise alternativa para explicar a resistência do *Adaboost* em sofrer um desempenho ruim em dados novos, mesmo tendo obtido resultados bons na amostra. Este método é baseado no conceito de *margens* (a margem de uma observação é um número entre $[-1, 1]$ e é positivo se e somente se a hipótese final classifica corretamente uma observação. A magnitude da margem é uma medida de confiança na classificação). Nesta análise o erro de generalização é limitado pela probabilidade de as margens das observações da amostra de treinamento serem menores que um certo valor, mais um termo que vai a zero conforme N cresce. Então para que o erro de generalização seja pequeno é necessário que a proporção de exemplos que tenham margem menor que este valor seja pequeno.

2.3.5 Exemplo dos procedimentos *Bagging* e *Boosting*

Nesta seção, ilustraremos os métodos discutidos neste capítulo. O desempenho dos classificadores utilizados foi medido em relação às taxas de erro obtidas por simulações realizadas pelo *software* R, que consiste em um *software* estatítico gratuito.

Utilizando como referência o documento do pacote gbm, consideramos o conjunto de dados obtido através de uma simulação. Os dados gerados são de duas classes, e há 6 variáveis explicativas. Foram simuladas 6 variáveis com 10.000 observações i.i.d., para a amostra de treinamento, e aplicamos os classificadores *Bagging* e *Boosting*.

Para o método *Bagging* foi utilizado diretamento a função bagging do pacote ipred, do *software* R. Esta função cria classificadores via *Bagging* com o número desejado de amostras *bootstrap* combinando árvores de decisão. O método de *Boosting* foi testado utilizando o algoritmo *Adaboost* implementado no pacote gbm do *software* R.

Nos dois métodos o número de iterações dos algoritmos foi definido de 25 até 400 iterações e medido o respectivo desempenho. A tabela 2.1 apresenta as taxas de erro estimadas para os classificadores construídos pelos algoritmos *Bagging* e *Boosting*.

Iterações	Bagging	Boosting
25	0.286	0.412
50	0.264	0.349
100	0.273	0.258
150	0.263	0.195
200	0.260	0.186
250	0.264	0.175
300	0.259	0.174
350	0.258	0.174
400	0.261	0.173
450	0.262	0.173
500	0.260	0.173

Tabela 2.1: Taxas de erro estimadas dos algoritmos *Bagging* e *Boosting*, na amostra de treinamento. Simulações realizadas pelo *software* R.

É possível notar que com 100 iterações o algoritmo *Boosting* já apresenta evidências de um resultado superior ao do *Bagging*. A partir das taxas de erro da amostra de treinamento, verifica-se que para estes dados simulados o método *Bagging* não apresentou melhora ao aumentar o número de iterações, enquanto que o *Boosting* apresentou evidências numéricas de melhora na taxa de erro.

Capítulo 3

Duas novas propostas de Reamostragem em Blocos para Cadeias de Markov

Neste capítulo, duas propostas de reamostragem para cadeias de Markov, adaptadas e inspiradas pelo conceito do método *Boosting*, introduzido no capítulo anterior, são apresentadas.

A adaptação está na modificação da distribuição de probabilidade de reamostragem dos blocos. Como vimos, o método *Boosting* modifica a distribuição de probabilidade dos dados de acordo com a taxa de erro do ajuste do modelo. Na estimação da distribuição das probabilidades de transições e da medida invariante não temos uma taxa de erro, uma vez que não conhecemos a resposta verdadeira, assim utilizamos os estimadores de Máxima Verossimilhança para estimar uma distância e então propor os procedimentos.

A Reamostragem Ponderada em Blocos, para estimar a distribuição das probabilidades de transições, consiste em modificar a distribuição dos blocos de acordo com a distância destes blocos em relação as estimativas da matriz de Verossimilhança. Ou seja, os blocos mais distantes recebem pesos pequenos na distribuição da reamostragem e blocos com distâncias pequenas recebem pesos grandes. As amostras são retiradas de acordo com essa distribuição e são então calculadas as estimativas. O procedimento para a disribuição da medida invariante é análogo.

Já a Reamostragem Balanceada Ponderada em Blocos é uma adaptação da Reamostragem Balanceada de Importância em Blocos, que foi apresentada no Capíulo 1, mas com uma

modificação nos pesos de cada bloco, agora esses pesos levam em consideração a mesma idéia da Reamostragem Ponderada em Blocos, onde os pesos são funções das distâncias dos blocos em relação as estimativas da matriz de Verossimilhança. Ou seja, os blocos mais distantes recebem pesos pequenos na distribuição da reamostragem e blocos com distâncias pequenas recebem pesos grandes.

Considerando a seguinte notação: $\{5\}$ Método de Reamostragem Ponderada em Blocos e $\{6\}$ Método de Reamostragem Balanceada Ponderada em Blocos

3.1 Ponderada em Blocos

O algoritmo de Reamostragem Ponderada em Blocos para estimar a

$$P\{(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma} \leq x\}$$

consiste em:

1. *Seja $\triangle$ um estado recorrente fixado, seja $T_{\triangle}^{(\alpha)}$ o tempo do α-ésimo retorno ao estado $\triangle$, então os blocos $\{X_l; l = T_{\triangle}^{(\alpha)}, \ldots, T_{\triangle}^{(\alpha+1)} - 1\}$ são i.i.d. para $\alpha = 1, 2, \ldots$ Para quaisquer dois estados i e j, que são diferentes do estado $\triangle$, seja $\eta_\alpha = \{x_l; l = T_{\triangle}^{(\alpha)}, \ldots, T_{\triangle}^{(\alpha+1)} - 1\}$ o α-ésimo bloco. A amostra original pode ser decomposta em k blocos :$\{\eta_1, \eta_2, \ldots, \eta_k\}$.*

   Seja fixa a transição (i,j), calcula-se a distância de cada bloco em relação a $\hat{p}_n(i,j)$, (1.5):

   $$D_\alpha(i,j) = \mid p_\alpha(i,j) - \hat{p}_n(i,j) \mid, \qquad \alpha = 1, \ldots, k, \tag{3.1}$$

   e a partir desta distância é calculado um peso para cada bloco:

   $$W_\alpha(i,j) = \begin{cases} \frac{1}{D_\alpha(i,j)}, D_\alpha(i,j) \neq 0 \\ \frac{1}{min_l(D_l(i,j)}, \eta_l \neq \eta_\alpha, D_l(i,j) \neq 0 \end{cases} \tag{3.2}$$

   e $W_\alpha(i,j) = 1$ se todos os blocos tiverem $D_\alpha(i,j) = 0$, tal que $\sum_{\alpha=1}^{k} W_\alpha(i,j) = 1$.

   Seja a distribuição de probabilidade dos blocos na primeira iteração dada por: $\hat{F}^1(\eta_\alpha) = 1/k$, para $\alpha = 1, \ldots, k$.

2. *Para* $b = 1, ..., B$, *retire uma amostra de acordo com* $\hat{F}^b$: $\chi_b^{\{5\}} = \{\eta_{b1}^{\{5\}}, ..., \eta_{bk}^{\{5\}}\}$, *calcule as estimativas das probabilidades de transição* (i, j) *da seguinte forma:*

$$\tilde{p}_b^{\{5\}}(i,j) = \frac{\sum_{\alpha=1}^{k} h(\eta_{b\alpha}^{\{5\}})}{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{5\}})}. \tag{3.3}$$

A variância amostral da estimativa de probabilidade $\hat{p}_b^{\{5\}}$ *é dada por:*

$$\hat{\sigma}_b^{2\{5\}} = \frac{1}{(\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{5\}}))^2} \sum_{\alpha=1}^{k} (g(\eta_{b\alpha}^{\{5\}})(p_{b\alpha}^{\{5\}}(i,j) - \hat{p}_n(i,j)))^2, \tag{3.4}$$

onde $g(\eta_{b\alpha}^{\{5\}})$ *é o número de visitas ao estado* i *do bloco reamostrado* $\eta_{b\alpha}^{\{5\}}$, $h(\eta_{b\alpha}^{\{5\}})$ *é o número de transições* (i, j) *do bloco reamostrado* $\eta_{b\alpha}^{\{5\}}$ *e* $p_{b\alpha}^{\{5\}}(i,j) = h(\eta_{b\alpha}^{\{5\}})/g(\eta_{b\alpha}^{\{5\}})$.

Modifique a distribuição da próxima iteração $b+1$, *para* $\alpha = 1, \ldots, k$, *de acordo com os blocos reamostrados da* b*-ésima iteração :*

$$\hat{F}^{b+1}(\eta_\alpha) = \frac{\hat{F}^b(\eta_\alpha) W_\alpha(i,j) \mathbf{1}(\eta_\alpha \in \chi_b^{\{5\}}) + \hat{F}^b(\eta_\alpha) \mathbf{1}(\eta_\alpha \notin \chi_b^{\{5\}})}{\sum_{\alpha=1}^{k} \{\hat{F}^b(\eta_\alpha) W_\alpha(i,j) \mathbf{1}(\eta_\alpha \in \chi_b^{\{5\}}) + \hat{F}^b(\eta_\alpha) \mathbf{1}(\eta_\alpha \notin \chi_b^{\{5\}})\}}. \tag{3.5}$$

3. *A distribuição de* $(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma}$ *pode então ser aproximada pela distribuição empírica de* $(\hat{p}_b^{\{5\}}(i,j) - \hat{p}_n(i,j))/\hat{\sigma}_b^{\{5\}}$,$b = 1, \ldots, B$.

A idéia do procedimento para estimar a distribuição da medida invariante é a mesma do procedimento anterior, e também foi baseada no conceito do método *Boosting*. Este procedimento consiste em modificar a distribuição dos blocos de acordo com a distância desses blocos em relação as estimativas da medida invariante obtidas pela amostra original. Ou seja, os blocos mais distantes recebem pesos pequenos na distribuição da reamostragem e blocos com distâncias pequenas recebem pesos grandes. As amostras são retiradas de acordo com essa distribuição e são calculadas as estimativas obtidas por esta reamostragem.

O algoritmo para estimar a $P\{(\hat{\pi}_n(i) - \pi_i)/\tilde{\sigma} \leq x\}$ consiste em:

1. *Decomponha a amostra original em k blocos: $\{\eta_1, \eta_2, \ldots, \eta_k\}$, fixado o estado i. Calcule a distância de cada bloco em relação a $\hat{\pi}_n(i)$(1.6):*

$$D_\alpha(i) = \mid \pi_\alpha(i) - \hat{\pi}_n(i) \mid, \qquad \alpha = 1, \ldots, k, \tag{3.6}$$

a partir desta distância é calculado um peso para cada bloco:

$$W_\alpha(i) = \begin{cases} \frac{1}{D_\alpha(i)}, D_\alpha(i) \neq 0 \\ \frac{1}{min_l(D_l(i))}, \eta_l \neq \eta_\alpha, D_l(i) \neq 0 \end{cases} \tag{3.7}$$

e $W_\alpha(i) = 1$ se todos os blocos tiverem $D_\alpha(i) = 0$, tal que $\sum_{\alpha=1}^{k} W_\alpha(i) = 1$.

Seja a distribuição de probabilidade dos blocos na primeira iteração dada por : $\hat{F}^1(\eta_\alpha) = 1/k$, para $\alpha = 1, ..., k$.

2. *Para $b = 1, ..., B$, retire uma amostra de acordo com $\hat{F}^b$: $\chi_b^{\{5\}} = \{\eta_{b1}^{\{5\}}, ..., \eta_{bk}^{\{5\}}\}$, calcule a estimativa da medida invariante do estado i por:*

$$\tilde{\pi}_b^{\{5\}}(i) = \frac{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{5\}})}{\sum_{\alpha=1}^{k} T_{b\alpha}^{\{5\}}}. \tag{3.8}$$

A variância amostral da estimativa da medida invariante é dada por,

$$\tilde{\sigma}_b^{2\{5\}} = \frac{1}{(\sum_{\alpha=1}^{k} T_{b\alpha}^{\{5\}})^2} \sum_{\alpha=1}^{k} (T_{b\alpha}^{\{5\}}(\pi_{b\alpha}^{\{5\}}(i) - \hat{\pi}_n(i)))^2, \tag{3.9}$$

onde $g(\eta_{b\alpha}^{\{5\}})$ é o número de visitas ao estado i do bloco reamostrado $\eta_{b\alpha}^{\{5\}}$, $T_{b\alpha}^{\{5\}}$ é o tamanho do bloco reamostrado $\eta_{b\alpha}^{\{5\}}$ e $\pi_{b\alpha}^{\{5\}}(i) = g(\eta_{b\alpha}^{\{5\}})/T_{b\alpha}^{\{5\}}$.

Modifique a distribuição da próxima iteração $b+1$, para $\alpha = 1, \ldots, k$, de acordo com os blocos reamostrados da b-ésima iteração :

$$\hat{F}^{b+1}(\eta_\alpha) = \frac{\hat{F}^b(\eta_\alpha) W_\alpha(i) \mathbf{1}(\eta_\alpha \in \chi_b^{\{5\}}) + \hat{F}^b(\eta_\alpha) \mathbf{1}(\eta_\alpha \notin \chi_b^{\{5\}})}{\sum_{\alpha=1}^{k} \{\hat{F}^b(\eta_\alpha) W_\alpha(i) \mathbf{1}(\eta_\alpha \in \chi_b^{\{5\}}) + \hat{F}^b(\eta_\alpha) \mathbf{1}(\eta_\alpha \notin \chi_b^{\{5\}})\}}. \tag{3.10}$$

3. *A distribuição de $(\hat{\pi}_n(i) - \pi_i)/\tilde{\sigma}$ pode ser aproximada pela distribuição empírica de $(\tilde{\pi}_b^{\{5\}}(i) - \hat{\pi}_n(i))/\tilde{\sigma}_b^{\{5\}}$, $b = 1, \ldots, B$.*

3.2 Balanceada Ponderada em Blocos

Como vimos no Capítulo 1, o balanceado na reamostragem significa que o número de vezes em que cada bloco aparece em qualquer reamostragem é proporcional a probabilidade de ser selecionado pela reamostragem ponderada.

O algoritmo para estimar a $P\{(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma} \leq x\}$ consiste em:

1. *Fixe a transição* (i,j), *decomponha a amostra original em* k *ciclos* :$\{\eta_1, \eta_2, ..., \eta_k\}$. *Calcule a distância de cada bloco em relação a* $\hat{p}_n(i,j)$ (1.5):

$$D_\alpha(i,j) = \mid p_\alpha(i,j) - \hat{p}_n(i,j) \mid, \qquad \alpha = 1, \ldots, k, \tag{3.11}$$

e a partir desta distância calcule um peso para cada bloco:

$$W_\alpha(i,j) = \begin{cases} \frac{1}{D_\alpha(i,j)}, D_\alpha(i,j) \neq 0 \\ \frac{1}{min_l(D_l(i,j))}, \eta_l \neq \eta_\alpha, D_l(i,j) \neq 0 \end{cases} \tag{3.12}$$

e $W_\alpha(i,j) = 1$ *se todos os blocos tiverem* $D_\alpha(i,j) = 0$, *tal que* $\sum_{\alpha=1}^{k} W_\alpha(i,j) = 1$.

2. *A distribuição de probabilidade de reamostragem* $\{q'_\alpha; \alpha = 1, \ldots, k\}$ *de* $\{\eta_\alpha; \alpha = 1, \ldots, k\}$ *é*

$$q'_\alpha = W_\alpha(i,j).$$

3. *Defina* $B'_\alpha = [kBq'_\alpha]$, *onde* $[.]$ *denota a parte inteira, obtenha* $r = kB - \sum_{\alpha=1}^{k} B'_\alpha$ *e* $d_\alpha = kBq'_\alpha - B'_\alpha$. *Ordene* $\{d_1, \ldots, d_k\}$ *do menor para o maior. Seja* $B_\alpha = B'_\alpha + 1$, *se a ordem de* d_α *é maior ou igual a* r, *ou,* $B_\alpha = B'_\alpha$, *caso contrário. Assim* B_α *é o número de vezes que o bloco* η_α *aparece na reamostra balanceada. Ou seja, repete o bloco* η_α, B_α *vezes, totalizando* Bk *blocos.*

4. *Permute os* Bk *blocos aleatoriamente, e a cada* k *blocos obtenha uma reamostra balanceada, num total de* B *reamostras,* $\chi_b^{\{6\}} = \{\eta_{b1}^{\{6\}}, \ldots, \eta_{bk}^{\{6\}}\}$, $b = 1, \ldots, B$.

5. *Para cada uma das reamostras defina as estimativas da probabilidade de transição do estado i para o j e a medida invariante do estado i, na b-ésima iteração, respectivamente por :*

$$\tilde{p}_b^{\{6\}}(i,j) = \frac{\sum_{\alpha=1}^{k} h(\eta_{b\alpha}^{\{6\}})}{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{6\}})}. \tag{3.13}$$

A variância amostral das estimativa de probabilidade $\hat{p}_b^{\{6\}}$ é dada por,

$$\hat{\sigma}_b^{2\{6\}} = \frac{1}{(\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{6\}}))^2} \sum_{\alpha=1}^{k} (g(\eta_{b\alpha}^{\{6\}})(p_{b\alpha}^{\{6\}}(i,j) - \hat{p}_n(i,j)))^2, \tag{3.14}$$

onde $g(\eta_{b\alpha}^{\{6\}})$ é o número de visitas ao estado i do bloco reamostrado $\eta_{b\alpha}^{\{6\}}$, $h(\eta_{b\alpha}^{\{6\}})$ é o número de transições do estado i para o j do bloco reamostrado $\eta_{b\alpha}^{\{6\}}$, $p_{b\alpha}^{\{6\}}(i,j) = h(\eta_{b\alpha}^{\{6\}})/g(\eta_{b\alpha}^{\{6\}})$.

6. *A distribuição de $(\hat{p}_n(i,j) - p_{ij})/\hat{\sigma}$ pode então ser aproximada pela distribuição empírica de $(\tilde{p}_b^{\{6\}}(i,j) - \hat{p}_n(i,j))/\hat{\sigma}_b^{\{6\}}$, $b = 1, \ldots, B$.*

Com a mesma notação, o algoritmo para estimar a $P\{(\tilde{\pi}_n(i,j) - \pi_i)/\tilde{\sigma} \leq x\}$ consiste em:

1. *Fixe o estado i, decomponha a amostra original em k ciclos :$\{\eta_1, \eta_2, ..., \eta_k\}$. Calcule a distância de cada bloco em relação a $\tilde{\pi_n}(i)$:*

$$D_\alpha(i) = \mid \pi_\alpha(i) - \tilde{\pi}_n(i) \mid, \qquad \alpha = 1, \ldots, k, \tag{3.15}$$

e a partir desta distância calcule um peso para cada bloco:

$$W_\alpha(i) = \begin{cases} \frac{1}{D_\alpha(i)}, D_\alpha(i) \neq 0 \\ \frac{1}{min_l(D_l(i))}, \eta_l \neq \eta_\alpha, D_l(i) \neq 0 \end{cases} \tag{3.16}$$

e $W_\alpha(i) = 1$ se todos os blocos tiverem $D_\alpha(i) = 0$, tal que $\sum_{\alpha=1}^{k} W_\alpha(i) = 1$.

2. *A distribuição de probabilidade de reamostragem* $\{q'_\alpha; \alpha = 1, \ldots, k\}$ *de* $\{\eta_\alpha; \alpha = 1, \ldots, k\}$ *é*

$$q'_\alpha = W_\alpha(i).$$

3. *Defina* $B'_\alpha = [kBq'_\alpha]$, *onde* $[.]$ *denota a parte inteira, obtenha* $r = kB - \sum_{\alpha=1}^{k} B'_\alpha$ *e* $d_\alpha = kBq'_\alpha - B'_\alpha$. *Ordene* $\{d_1, \ldots, d_k\}$ *do menor para o maior. Seja* $B_\alpha = B'_\alpha + 1$, *se a ordem de* d_α *é maior ou igual a* r, *ou,* $B_\alpha = B'_\alpha$, *caso contrário. Assim* B_α *é o número de vezes que o bloco* η_α *aparece na reamostra balanceada. Ou seja, repete o bloco* η_α, B_α *vezes, totalizando* Bk *blocos.*

4. *Permute os* Bk *blocos aleatoriamente, e a cada* k *blocos obtenha uma reamostra balanceada, num total de* B *reamostras,* $\chi_b^{\{6\}} = \{\eta_{b1}^{\{6\}}, \ldots, \eta_{bk}^{\{6\}}\}$, $b = 1, \ldots, B$.

5. *Então a estimativa de* $\tilde{\pi}_n(i)$ *é dada por :*

$$\tilde{\pi}_b^{\{6\}}(i) = \frac{\sum_{\alpha=1}^{k} g(\eta_{b\alpha}^{\{6\}})}{\sum_{\alpha=1}^{k} T_{b\alpha}^{\{6\}}}. \tag{3.17}$$

A variância amostral da estimativa de $\tilde{\pi}_b^{\{6\}}$ *é dadas por,*

$$\tilde{\sigma}_b^{2\{6\}} = \frac{1}{(\sum_{\alpha=1}^{k} T_{b\alpha}^{\{6\}})^2} \sum_{\alpha=1}^{k} (T_{b\alpha}^{\{6\}}(\pi_{b\alpha}^{\{6\}}(i) - \hat{\pi}_n(i)))^2, \tag{3.18}$$

onde $T_{b\alpha}^{\{6\}}$ *é o tamanho do bloco reamostrado* $\eta_{b\alpha}^{\{6\}}$ *e* $\pi_{b\alpha}^{\{6\}} = g(\eta_{b\alpha}^{\{6\}})/T_{b\alpha}^{\{6\}}$.

6. *Assim, a distribuição de* $(\hat{\pi}_n(i) - \pi_i)/\tilde{\sigma}$ *pode ser aproximada pela distribuição empírica de* $(\tilde{\pi}_b^{\{6\}}(i) - \hat{\pi}_n(i))/\tilde{\sigma}_b^{\{6\}}$, $b = 1, \ldots, B$.

Capítulo 4

Simulações dos Métodos de Reamostragem para Cadeias de Markov

Neste capítulo iremos aplicar os métodos de reamostragem em cadeias de Markov estudados, com o intuito de ilustrar e comparar. Para uma cadeia de Markov ergódica de 3 estados, $S = \{0, 1, 2\}$, com estado inicial $X_0 = 1$, comparamos os métodos de reamostragem através dos intervalos de confiança estimados para a probabilidade de transição do estado 0 para o 0 e para a medida invariante do estado 0. A amostra original é gerada pela cadeia de Markov ergódica com matriz de probabilidade de transição dada por:

$$P = \begin{pmatrix} 0.33 & 0.33 & 0.34 \\ 0.5 & 0.25 & 0.25 \\ 0.5 & 0.35 & 0.15 \end{pmatrix},$$

e medida invariante dada por:

$$\pi = \begin{pmatrix} 0.429 & 0.255 & 0.316 \end{pmatrix}.$$

Seja o estado recorrente fixado $\triangle = 1$ e para o tamanho da amostra 350, obtemos 121 blocos. Fixamos a transição do estado 0 para o estado 0 e para estimar $P\{(\hat{p}_n(0,0) - p_{00})/\hat{\sigma} \leq x\}$ e $P\{(\hat{\pi}_n(0) - \pi_0)/\tilde{\sigma} \leq x\}$, implementamos, utilizando o *software* R , os

métodos de reamostragem Uniforme em Blocos, Balanceada em Blocos, Ponderada em Blocos e Balanceada Ponderada em Blocos. Utilizando diferentes números de simulações, com a mesma amostra original, comparamos os intervalos de confiança de 95% obtidos pela distribuição empírica, através do método percentil, e, também considerando que a distribuição limite é a Normal Padrão.

Temos um parâmetro θ e uma estimativa θ_h para θ. Seja θ^* uma estimativa não paramétrica de θ baseado na amostra de *bootstrap* e seja S^* uma estimativa do desvio padrão para θ baseado na amostra de *bootstrap*. Defina $M^* = (\theta^* - \theta_h)/S^*$. Para cada uma das B estimativas de θ^* existe um M^* correspondente. Seja a função distribuição dos valores observados

$$G_*(x) = \mathbb{P}\{M^* \leq x\}.$$

Para qualquer função distribuição F e $\tau \in (0,1)$, o τ-ésimo quantil de F é $F^{-1}(\tau) = inf\{x : F(x) \geq \tau\}$. O método percentil de *bootstrap* define $G_*^{-1}(\alpha)$ e $G_*^{-1}(1-\alpha)$ como os limites inferior e superior para o intervalo de confiança de $(1-2\alpha)$ para θ.

Apresentamos os resultados em duas seções. A primeira apresenta os resultados para a probabilidade de transição do estado 0 para o estado 0, p_{00}, e a segunda seção apresenta os resultados para a medida invariante do estado 0, π_0 .

4.1 Probabilidade de Transição

Aqui apresentamos os resultados obtidos pela simulação da comparação dos métodos de reamostragem Uniforme em Blocos, Balanceada em Blocos, Ponderada em Blocos e Balanceada Ponderada em Blocos, para estimar a $P\{(\hat{p}_n(0,0) - p_{00})/\hat{\sigma} \leq x\}$. A estimativa da probabilidade de transição p_{00} pelo método de Máxima Verossimilhança, é 0.3098, nesta amostra ($n = 350$).

Apresentamos para N fixo e diferentes números de simulações (B), os métodos de reamostragem utilizados para estimar os intervalos de confiança para a probabilidade de transição $(0,0)$. É possível verificar, através dos gráficos 4.1 a 4.18, a comparação das

variâncias das estimativas de cada um dos métodos, para B igual a 50, 100 e 150. Essas comparações nos fornecem evidências que o método de reamostragem Ponderada em Blocos apresentou as menores variâncias quando comparadas aos outros métodos, para todos os valores de B.

As tabelas 4.1 a 4.3 apresentam os intervalos de confiança de 95%, obtidos pela distribuição empírica, para a probabilidade de transição, e os respectivos comprimentos, para cada um dos métodos de reamostragem, para B igual 50, 100 e 150, respectivamente.

Apresentamos nas tabelas 4.4 a 4.6 ,para efeito de comparação, os intervalos de confiança de 95% considerando que $P\{(\hat{p}_n(0,0) - p_{00})/\hat{\sigma} \leq x\}$ tenha distribuição limite Normal Padrão.

Observando os resultados das estimativas das variâncias apresentados nos gráficos conjuntamente com os intervalos de confiança de 95% é possível comparar cada um dos métodos de reamostragem simulados. Nota-se, nos gráficos 4.1 ao 4.3, que as variâncias das estimativas da probabilidade de transição p_{00} do método de reamostragem Balanceada em Blocos são menores quando comparadas ao método de reamostragem Uniforme em Blocos, resultando em intervalos de confiança de 95%, para p_{00}, menores para todos os números de simulações (B) utilizados, verificados nas tabelas 4.1 a 4.3. Já as variâncias das estimativas da probabilidade de transição p_{00} do método de reamostragem Balanceada Ponderada em Blocos são menores quando comparadas ao método de reamostragem Balanceada em Blocos, como pode-se verificar nos gráficos 4.16 a 4.18, para todos os B's utilizados, refletindo nos intervalos de confiança e seus respectivos comprimentos, assim o método de reamostragem Balanceada Ponderada em Blocos se mostra, pelo menos de maneira empírica, mais eficiente do que o método de reamostragem Balanceada em Blocos.

O método que apresentou fortes evidências numéricas de eficiência, para esta amostra, foi o método de reamostragem Ponderada em Blocos, e isso pode ser verificado nas estimativas das variâncias apresentadas nos gráficos 4.4 a 4.6 e 4.10 a 4.15, nos resultados dos intervalos de confiança de 95%, obtidos pela distibuição empírica e supondo distribuição limite Normal Padrão, e seus respectivos comprimentos que se apresentaram menores quando comparados a todos os outros métodos de reamostragem simulados.

Método	IC de 95%	Comprimento
Uniforme	$(0.2051, 0.3856)$	0.1799
Balanceada	$(0.2416, 0.3803)$	0.1387
Balanceada Ponderada	$(0.2584, 0.3806)$	0.1222
Ponderada	$(0.3017, 0.3785)$	0.0768

Tabela 4.1: Estimativas para a probabilidade de transição $(0, 0)$, usando a Distribuição Empírica, com $B = 50$.

Método	IC de 95%	Comprimento
Uniforme	$(0.2500, 0.3701)$	0.1201
Balanceada	$(0.2460, 0.3630)$	0.1170
Balanceada Ponderada	$(0.2806, 0.3531)$	0.0725
Ponderada	$(0.2918, 0.3554)$	0.0636

Tabela 4.2: Estimativas para a probabilidade de transição $(0, 0)$, usando a Distribuição Empírica, com $B = 100$.

Método	IC de 95%	Comprimento
Uniforme	$(0.3035, 0.4113)$	0.1078
Balanceada	$(0.2914, 0.3729)$	0.0815
Balanceada Ponderada	$(0.3079, 0.3884)$	0.0805
Ponderada	$(0.2934, 0.3534)$	0.0600

Tabela 4.3: Estimativas para a probabilidade de transição $(0, 0)$, usando a Distribuição Empírica, com $B = 150$.

Método	IC de 95%	Comprimento
Uniforme	$(0.2500, 0.3700)$	0.1200
Balanceada	$(0.2460, 0.3654)$	0.1194
Balanceada Ponderada	$(0.2567, 0.3581)$	0.1014
Ponderada	$(0.2944, 0.3613)$	0.0670

Tabela 4.4: Estimativas para a probabilidade de transição $(0, 0)$, usando a Distribuição Normal Padrão, com $B = 50$.

Método	IC de 95%	Comprimento
Uniforme	$(0.2550, 0.3751)$	0.1200
Balanceada	$(0.2518, 0.3621)$	0.1103
Balanceada Ponderada	$(0.2954, 0.3629)$	0.0675
Ponderada	$(0.2985, 0.3601)$	0.0616

Tabela 4.5: Estimativas para a probabilidade de transição $(0, 0)$, usando a Distribuição Normal Padrão, com $B = 100$.

Método	IC de 95%	Comprimento
Uniforme	$(0.2913, 0.3914)$	0.1001
Balanceada	$(0.2994, 0.3875)$	0.0881
Balanceada Ponderada	$(0.3154, 0.3998)$	0.0844
Ponderada	$(0.3097, 0.3655)$	0.0558

Tabela 4.6: Estimativas para a probabilidade de transição $(0, 0)$, usando a Distribuição Normal Padrão, com $B = 150$.

- **Comparando os métodos de Reamostragem Uniforme (pontos mais claros) e Balanceada (pontos mais escuros) em Blocos**

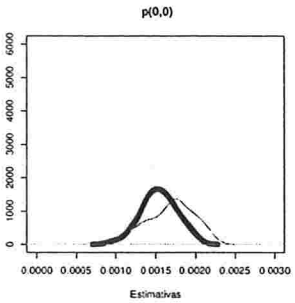


Figura 4.1: Estimativas da variância da probabilidade de transição p_{00} com $B = 50$.

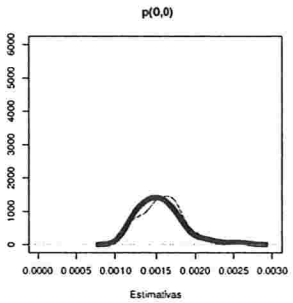


Figura 4.2: Estimativas da variância da probabilidade de transição p_{00} com $B = 100$.

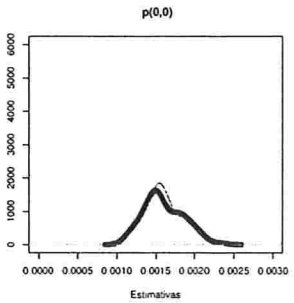


Figura 4.3: Estimativas da variância da probabilidade de transição p_{00} com $B = 150$.

- **Comparando os métodos de Reamostragem Uniforme (pontos mais claros) e Ponderada (pontos mais escuros) em Blocos**

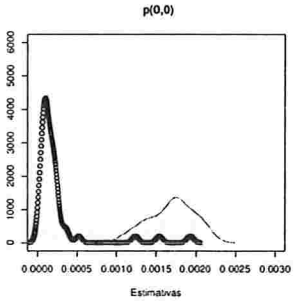


Figura 4.4: Estimativas da variância da probabilidade de transição p_{00} com $B = 50$.

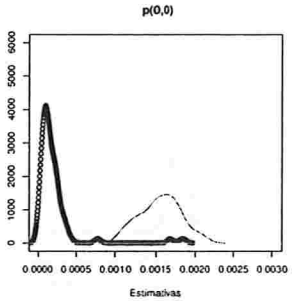


Figura 4.5: Estimativas da variância da probabilidade de transição p_{00} com $B = 100$.

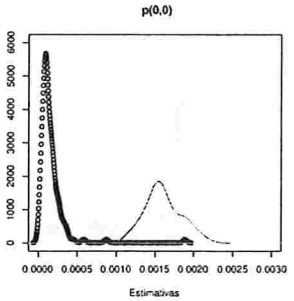


Figura 4.6: Estimativas da variância da probabilidade de transição p_{00} com $B = 150$.

- **Comparando os métodos de Reamostragem Uniforme (pontos mais claros) e Ponderada Balanceada (pontos mais escuros) em Blocos**

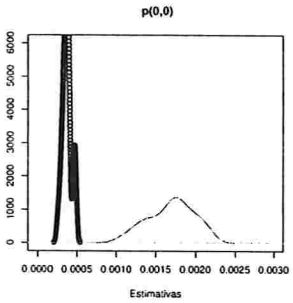


Figura 4.7: Estimativas da variância da probabilidade de transição p_{00} com $B = 50$.

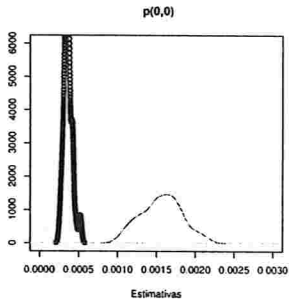


Figura 4.8: Estimativas da variância da probabilidade de transição p_{00} com $B = 100$.

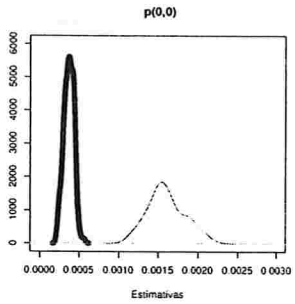


Figura 4.9: Estimativas da variância da probabilidade de transição p_{00} com $B = 150$.

- **Comparando os métodos de Reamostragem Ponderada (pontos mais escuros) e Balanceada (pontos mais claros) em Blocos**

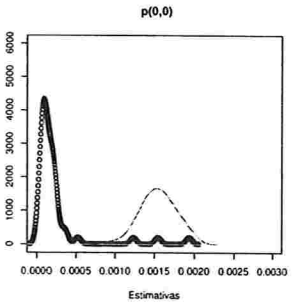


Figura 4.10: Estimativas da variância da probabilidade de transição p_{00} com $B = 50$.

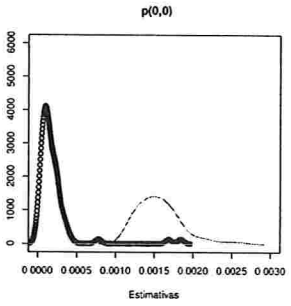


Figura 4.11: Estimativas da variância da probabilidade de transição p_{00} com $B = 100$.

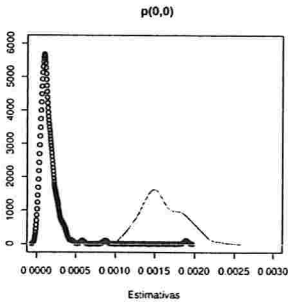


Figura 4.12: Estimativas da variância da probabilidade de transição p_{00} com $B = 150$.

- **Comparando os métodos de Reamostragem Ponderada (pontos mais escuros) e Balanceada Ponderada (pontos mais claros) em Blocos**

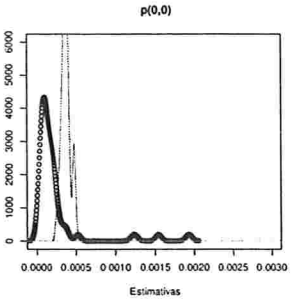


Figura 4.13: Estimativas da variância da probabilidade de transição p_{00} com $B = 50$.

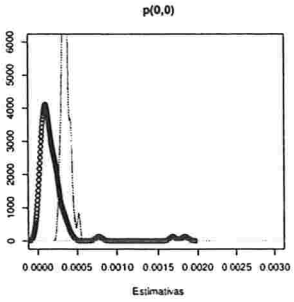


Figura 4.14: Estimativas da variância da probabilidade de transição p_{00} com $B = 100$.

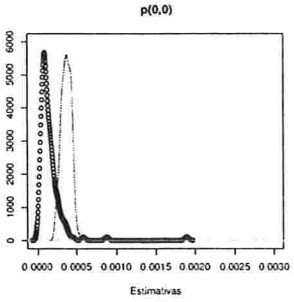


Figura 4.15: Estimativas da variância da probabilidade de transição p_{00} com $B = 150$.

- **Comparando os métodos de Reamostragem Balanceada (pontos mais claros) e Balanceada Ponderada (pontos mais escuros) em Blocos**

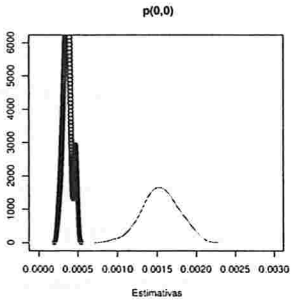


Figura 4.16: Estimativas da variância da probabilidade de transição p_{00} com $B = 50$.

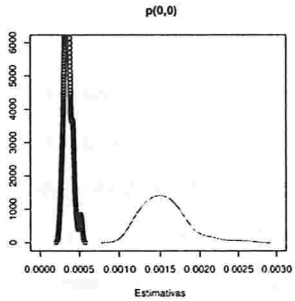


Figura 4.17: Estimativas da variância da probabilidade de transição p_{00} com $B = 100$.

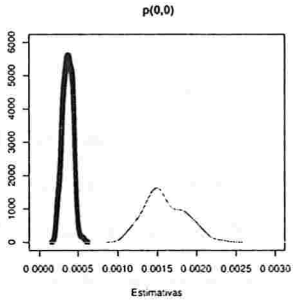


Figura 4.18: Estimativas da variância da probabilidade de transição p_{00} com $B = 150$.

4.2 Medida Invariante

Aqui apresentamos os resultados obtidos pela simulação da comparação dos métodos de reamostragem Uniforme em Blocos, Balanceada em Blocos, Ponderada em Blocos e Balanceada Ponderada em Blocos, para estimar a distribuição da medida invariante do estado 0, π_0. A estimativa de Máxima Verossimilhança da medida invariante do estado 0 é 0.441 (o valor verdadeiro é 0.429) .

Apresentamos para N fixo e diferentes números de simulações (B), os métodos de reamostragem utilizados para estimar os intervalos de confiança para a medida invariante. É possível verificar, através dos gráficos 4.19 a 4.36, a comparação das variâncias das estimativas de cada um dos métodos, para cada um dos B's.

As tabelas 4.7 a 4.9 apresentam os intervalos de confiança de 95% para a medida invariante do estado 0, pela distribuição empírica, e os respectivos comprimentos, para cada um dos métodos de reamostragem, para B igual 50, 100 e 150, respectivamente.

Já nas tabelas 4.10 a 4.12 estão os os intervalos de confiança de 95% supondo que a $P\{(\hat{\pi}_n(0) - \pi_0)/\tilde{\sigma} \leq x\}$ tem distribuição limite Normal Padrão.

Verificando os resultados das estimativas das variâncias apresentados nos gráficos conjuntamente com os intervalos de confiança de 95% é possível comparar cada um dos métodos de reamostragem simulados. Apesar das variâncias das estimativas da medida invariante π_0 do método de reamostragem Balanceada em Blocos se apresentarem próximas das variâncias do método de reamostragem Uniforme em Blocos, gráficos 4.19 a 4.21, os intervalos de confiança de 95%, para π_0, do método de reamostragem Balanceada em Blocos tiveram menores comprimentos quando comparados ao método de reamostragem Uniforme em Bloco, para todos os números de simulações (B) utilizados, verificados nas tabelas 4.7 a 4.9.

As variâncias das estimativas da medida invariante π_0 do método de reamostragem Balanceada Ponderada em Blocos são menores quando comparadas ao método de reamostragem Uniforme em Blocos e Balanceada em Blocos, como pode-se verificar nos gráficos 4.25 a 4.27 e 4.34 a 4.36, respectivamente, para todos os B's utilizados, refletindo nos intervalos de confiança e seus respectivos comprimentos. Assim o método de reamostragem

Balanceada Ponderada em Blocos se mostra mais eficiente, pelo menos empiricamente, do que os outros dois métodos.

O método que apresentou fortes evidências numéricas de eficiência, para esta amostra, foi o método de reamostragem Ponderada em Blocos, mesmo apresentando estimativas das variâncias próximas das estimativas obtidas pelo método de reamostragem Ponderada Balanceada, como pode ser visto nos gráficos a 4.31 a 4.33, mas quando comparados os resultados dos intervalos de confiança de 95% o método de reamostragem Ponderada em Blocos se mostrou mais eficiente, para todos os números de simulações utilizados, do que todos os outros métodos de reamostragem simulados para esta amostra.

Método	IC de 95%	Comprimento
Uniforme	$(0.3930, 0.4814)$	0.0884
Balanceada	$(0.3811, 0.4693)$	0.0882
Balanceada Ponderada	$(0.3933, 0.4815)$	0.0882
Ponderada	$(0.3888, 0.4762)$	0.0882

Tabela 4.7: Estimativas para a medida invariante do estado 0, usando a Distribuição Empírica, com $B = 50$.

Método	IC de 95%	Comprimento
Uniforme	$(0.4006, 0.4686)$	0.0680
Balanceada	$(0.4035, 0.4634)$	0.0599
Balanceada Ponderada	$(0.4021, 0.4617)$	0.0596
Ponderada	$(0.4075, 0.4558)$	0.0483

Tabela 4.8: Estimativas para a medida invariante do estado 0, usando a Distribuição Empírica, com $B = 100$.

Método	IC de 95%	Comprimento
Uniforme	$(0.3870, 0.4451)$	0.0581
Balanceada	$(0.4093, 0.4661)$	0.0568
Balanceada Ponderada	$(0.410, 0.466)$	0.0560
Ponderada	$(0.4037, 0.4545)$	0.0508

Tabela 4.9: Estimativas para a medida invariante do estado 0, usando a Distribuição Empírica, com $B = 150$.

Método	IC de 95%	Comprimento
Uniforme	$(0.3950, 0.4714)$	0.0764
Balanceada	$(0.3911, 0.4603)$	0.0692
Balanceada Ponderada	$(0.3933, 0.4595)$	0.0662
Ponderada	$(0.3980, 0.4602)$	0.0622

Tabela 4.10: Estimativas para a medida invariante do estado 0, usando a Distribuição Normal Padrão, com $B = 50$.

Método	IC de 95%	Comprimento
Uniforme	$(0.4107, 0.4688)$	0.0581
Balanceada	$(0.4165, 0.4697)$	0.0532
Balanceada Ponderada	$(0.4021, 0.4551)$	0.0530
Ponderada	$(0.4075, 0.4511)$	0.0436

Tabela 4.11: Estimativas para a medida invariante do estado 0, usando a Distribuição Normal Padrão, com $B = 100$.

Método	IC de 95%	Comprimento
Uniforme	$(0.4187, 0.4687)$	0.0500
Balanceada	$(0.4093, 0.4590)$	0.0497
Balanceada Ponderada	$(0.4121, 0.4571)$	0.0450
Ponderada	$(0.4140, 0.4534)$	0.0394

Tabela 4.12: Estimativas para a medida invariante do estado 0, usando a Distribuição Normal Padrão, com $B = 150$.

- **Comparando os métodos de Reamostragem Uniforme (pontos mais claros) e Balanceada (pontos mais escuros) em Blocos**

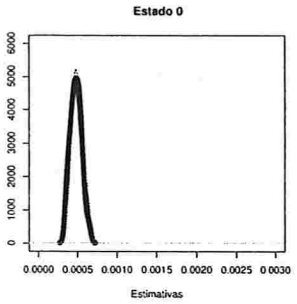


Figura 4.19: Estimativas da variância da medida invariante π_0 com $B = 50$.

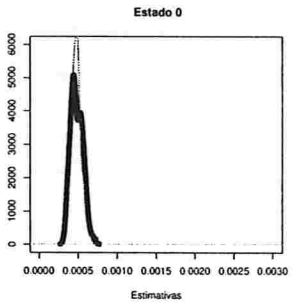


Figura 4.20: Estimativas da variância da medida invariante π_0 com $B = 100$.

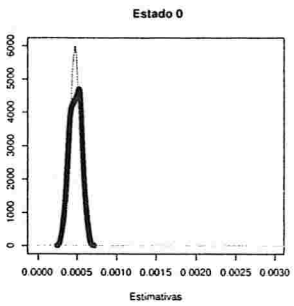


Figura 4.21: Estimativas da variância da medida invariante π_0 com $B = 150$.

- **Comparando os métodos de Reamostragem Uniforme (pontos mais claros) e Ponderada (pontos mais escuros) em Blocos**

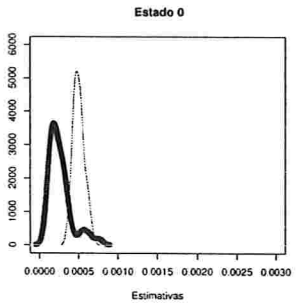


Figura 4.22: Estimativas da variância da medida invariante π_0 com $B = 50$.

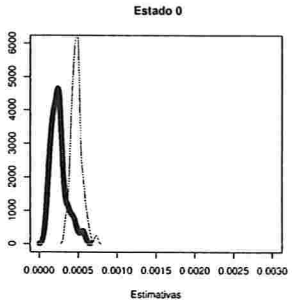


Figura 4.23: Estimativas da variância da medida invariante π_0 com $B = 100$.

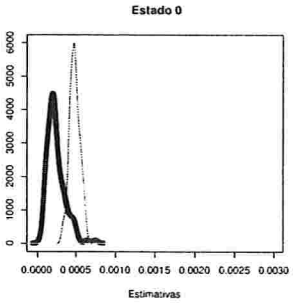


Figura 4.24: Estimativas da variância da medida invariante π_0 com $B = 150$.

- **Comparando os métodos de Reamostragem Uniforme (pontos mais claros) e Ponderada Balanceada (pontos mais escuros) em Blocos**

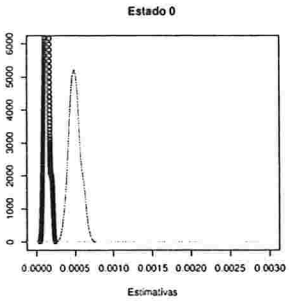


Figura 4.25: Estimativas da variância da medida invariante π_0 com $B = 50$.

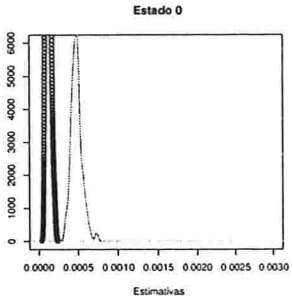


Figura 4.26: Estimativas da variância da medida invariante π_0 com $B = 100$.

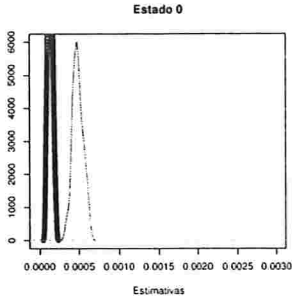


Figura 4.27: Estimativas da variância da medida invariante π_0 com $B = 150$.

- **Comparando os métodos de Reamostragem Ponderada (pontos mais escuros) e Balanceada (pontos mais claros) em Blocos**

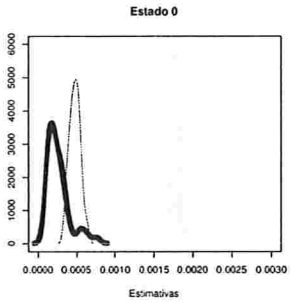


Figura 4.28: Estimativas da variância da medida invariante π_0 com $B = 50$.

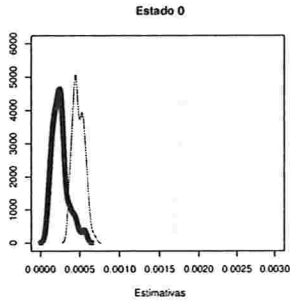


Figura 4.29: Estimativas da variância da medida invariante π_0 com $B = 100$.

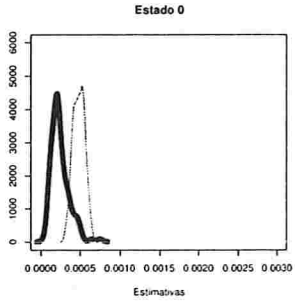


Figura 4.30: Estimativas da variância da medida invariante π_0 com $B = 150$.

- **Comparando os métodos de Reamostragem Ponderada (pontos mais escuros) e Balanceada Ponderada (pontos mais claros) em Blocos**

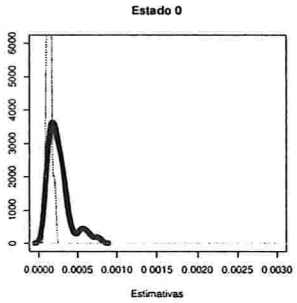


Figura 4.31: Estimativas da variância da medida invariante π_0 com $B = 50$.

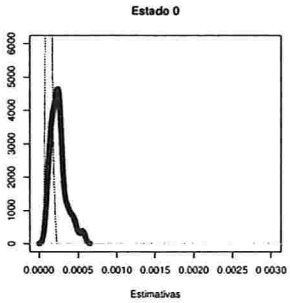


Figura 4.32: Estimativas da variância da medida invariante π_0 com $B = 100$.

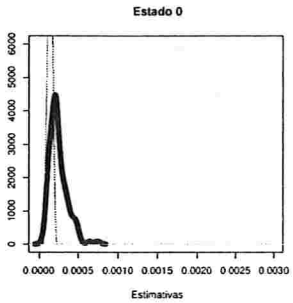


Figura 4.33: Estimativas da variância da medida invariante π_0 com $B = 150$.

- **Comparando os métodos de Reamostragem Balanceada (pontos mais claros) e Balanceada Ponderada (pontos mais escuros) em Blocos**

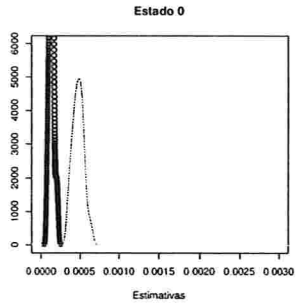


Figura 4.34: Estimativas da variância da medida invariante π_0 com $B = 50$.

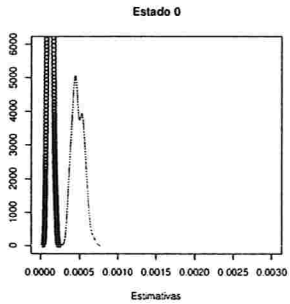


Figura 4.35: Estimativas da variância da medida invariante π_0 com $B = 100$.

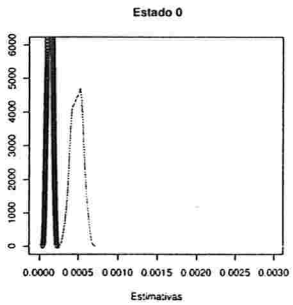


Figura 4.36: Estimativas da variância da medida invariante π_0 com $B = 150$.

Capítulo 5

Considerações Finais

Esse trabalho apresentou alguns resultados sobre reamostragem em blocos, para cadeias de Markov, no contexto da estimação da distribuição e probabilidade de transição de estados e da medida invariante. Discutimos a reamostragem Uniforme em Blocos, Balanceada em Blocos, de Importância em Blocos e de Importância Balanceada em Blocos. Apresentamos algumas idéias de cada método e também consideramos métodos de classificação tais como o *Bagging* e o *Boosting*.

Inspirados na idéia da mudança da distribuição dos dados ao longo do procedimento dado pelo método *Boosting* propusemos dois novos métodos de reamostragem em blocos para cadeia de Markov, a reamostragem Ponderada em Blocos e a reamostragem Ponderada Balanceada em Blocos. Como vimos, o método *Boosting* modifica a distribuição de probabilidade dos dados de acordo com a taxa de erro do ajuste do modelo. Na estimação da distribuição das probabilidades de transições e da medida invariante não temos uma taxa de erro conhecida da estimação, uma vez que não conhecemos a resposta verdadeira, assim utilizamos os estimadores de Máxima Verossimilhança para estimar uma distância e então propor os procedimentos.

Os métodos de reamostragem Uniforme em Blocos, Balanceada em Blocos, Ponderada em Blocos e Balanceada Ponderada em Blocos, foram aplicados em dados simulados e comparados através dos intervalos de confiança de 95% da distribuição empíricas e seus respectivos comprimentos. Visualizamos as variâncias de cada método de reamostragem simulado e observamos que para a amostra utilizada o método de reamostragem Ponderada

em Blocos se mostrou mais eficiente do que os outros métodos de reamostragem, para todos as simulações utilizadas, 50, 100 e 150. Também consideramos os intervalos de confiança segundo a lei limite conhecida dos métodos já tratados na literatura. Acreditamos que este trabalho contribui e propõe idéias para uma área de pesquisa relevante e ainda pouco explorada na área de reamostragem de cadeias de Markov.

Referências Bibliográficas

[1] James,B.R.(1942).*"Probabilidade: Um Curso em Nível Intermediário"*. Rio de Janeiro.

[2] Cox,D.R.(1970).*"The Analysis of Binary Data"*. Methuen,London.

[3] Chung,K.L.(1974).*"A Course in Probability Theory"*.2nd ed. Academic Press,New York, página 133.

[4] Breiman,L.,Friedman,J.,Olshen,R.,Stone,C.(1984).*"Classification and Regression Trees"*. Wadsworth,Belmont,CA.80.

[5] Efron,B.(1979).*"Bootstrap methods:another look at the jackknife"*. Annals of Statistics,7, 1 – 26.

[6] Davison,A.C.,Hinkley,D.V.,Schechtman,E.(1986).*"Efficient Bootstrap Simulation"*. Biometrika,73, 555 – 566.

[7] Quinlan,J.R.(1986).*"Induction of Decision Trees"*. Machine Learning,1, 81 – 106.

[8] Johns,M.V.Jr.(1988).*"Importance Sampling for Bootstrap Confidence Intervals'*. J.Amer.Statist.Assoc,83, 178 – 191.

[9] Davison,A.C.(1988).*"Discussion of Papers by D.V.Hinkley and by T.J.DiCiccio and J.P.Romano."*. J.Roy.Statist.Soc.Ser. B,50, 356 – 357.

[10] Kulberger,R.J.,Prakasa Rao,B.L.S.(1989).*"Bootstrapping a Finite State Markov Chain"*. Sankhya: The Indian Journal of Statisticas, $51(A)178 - 191$.

[11] Hall,P.(1990).*"Performance of Balanced Bootstrap Resampling in Distribution Function and Quantile Problems."* Probab.Theory Related Fields,85, 239 – 260.

[12] Athreya,K.B., Fuh,C.D.(1992). *"Bootstrapping Markov Chains:Countable Case."* Journal of Statistical Planning and Inference,33 : 311 – 331.

[13] Quinlan,J.R.(1993). *"C4.5:Programs for Machine Learning."* Morgan Kaufman,San Mateo,California.

[14] Booth,J.G.,Hall,P.,Wood,A.T.A.(1993). *"Balanced Importance Resampling for The Bootstrap"*. Ann.Statist. 21, 286 – 298.

[15] Guttorp,P.(1995). *"Stochastic Modeling of Scientific Data"*. Chapman and Hall.

[16] Breiman,L.(1996). *"Bagging Predictors"*. Machine Learning,24(2) : 123 – 140.

[17] Freund,Y.,Schapire,R.E.(1996). *"Experiments with a New Boosting Algorithm."* In 13 th International Conference on Machine Learning,148 – 156.

[18] Fan,T.H., Hung,W.L.(1997). *"Balanced Resampling For Bootstrapping Finite State Markov Chains."* Comm.Statist.:Simul.Comput,26(4), 1456 – 1475.

[19] Freund,Y.,Schapire,R.E.(1997). *"A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting."* Journal of Computer and System Sciences, 55, 119 – 139.

[20] Fuh,C.D.,Fan,T.H.,Hung,W.L.(1998). *"Importance Resampling for Markov Chains"*. Comm.Statist.Stochastic Models.,14, 1007 – 1026.

[21] Breiman,L.(1998). *"Arcing Classifiers"*. The Annals of Statistics,26(3), 801 – 849.

[22] Schapire,R.E.,Freund,Y.,Bartlett,P.,Lee,W.S.(1998). *"Boosting The Margin:A New Explanation For The Effectiveness Of Voting Methods"*. The Annals of Statistics. 26(5), 1651 – 1686.

[23] Schapire,R.E., Singer,Y.(1999). *"Improved Boosting Algorithms Using Confidence-rated Predictions"*. Machine Learning,37(3), 297 – 336.

[24] Schapire,R.E.(1999). *"Theoretical Views of Boosting and Applications"*. Proceedings of Algorithmic Learning Theory.

[25] Schapire,R.E.(1999). *"A Brief Introduction to Boosting"*. Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence.

[26] Dietterich,T.G.(2000). *"Ensemble Methods in Machine Learning"*. Multiple Classifier Systems,1 – 15.

[27] Friedman,J.,Hastie,T. ,Tibshirani,R.(2000). *"Additive Logistic Regression: A Statistical View of Boosting"*. The Annals of Statistics,28(2), 337 – 407.

[28] Fuh,C.D.,Fan,T.H.,Wung,W.L.(2000). *"Balanced Importance Resampling for Markov Chains"*. Journal of Statistical Planning and Inference,83, 221 – 224.

[29] W.Feller(1968). *"An Introduction to Probability Theory and Its Application"*. New York, London, Sydney, Third Edition, Vol I.

[30] www.r-project.org