**Inferência em Cadeias de Markov
com Dados Incompletos**

Juan Carlos Raúl Soto Sotelo

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

Área de concentração: Probabilidade
Orientador: Pablo Ferrari

Resumo

Neste trabalho estudamos o problema de estimar uma cadeia de Markov quando as observações do processo estão incompletas. Propomos uma construção regenerativa para construir outro processo com a mesma matriz de transição. mostramos também que os métodos de reamostragem por blocos para cadeias de Markov podem ser usados com observações incompletas.

Finalmente, propomos modelos para o processo de omissão de dados e um método de imputação para os valores não observados.

Sumário

Introdução	1
1 Inferência em cadeias de Markov	3
1.1 Definições e resultados da literatura	3
1.2 Determinação da ordem	4
1.2.1 Criterio de Determinação Eficiente	4
1.2.2 Fator de Bayes	5
1.2.3 Comparação dos métodos via simulação	6
1.3 Cadeias de ordem infinita	6
2 Estimação com dados faltantes	8
2.1 Introdução	8
2.2 Construção Regenerativa	8
3 Reamostragem em Cadeias de Markov	10
3.1 Introdução	10
3.2 Reamostragem uniforme em blocos	11
3.3 Reamostragem em Blocos Balanceados	12
3.4 Reamostragem Importante em Blocos	13
4 Modelos para o processo de omissão de dados	17
4.1 Introdução	17
4.2 Modelos para a omissão de dados	17
5 Discussão e Conclusões	19
Referências Bibliográficas	21

Introdução

A modelagem markoviana foi introduzida por Andrej Andreivich Markov no começo do século XX, para representar situações que apresentam dependência entre observações sucessivas. Um modelo markoviano se caracteriza por ter memória finita.

A inferência estatística em modelos markovianos teve seu início nos trabalhos de Bartlett da década de 50, em Billingsley (1961) encontramos os principais resultados obtidos até aquela data. Posteriormente Tong (1975), Schwarz (1978) e Zhao, Dorea e Gonçalves (2001) estudaram o problema da determinação da ordem da cadeia.

Nos últimos anos encontramos trabalhos encaminhados a duas abordagens: aproximar o processo de memória infinita por processos de memória finita Lalley (1986), Bressau, Fernández e Galves (1999), Comets, Fernández e Ferrari (2000), e Fernández e Galves (2002); e usar métodos de amostragem em modelos markovianos Kulperger e Prakasa Rao (1989), Athreya e Fu (1992), Fuh, Fan e Hung (1998), Fuh, Fan e Wung (2000)

Neste trabalho de doutorado estudamos a inferência para cadeias de Markov quando a sequência observada é incompleta, extendendo as abordagens de Little e Rubin no caso de observações independentes.

Este trabalho está estruturado da seguinte forma:

No Capítulo 1 apresentamos os principais resultados sobre inferência estatística em cadeias de Markov.

No Capítulo 2 introduzimos o problema de estimação com dados faltantes e apresentamos uma construção regenerativa para construir uma sequência completa com a mesma matriz de transição dos dados originais.

No Capítulo 3 introduzimos as técnicas de abordagem como outra alternativa para estimar a cadeia de Markov na presença de dados incompletos.

No capítulo 4 fazemos um estudo dos tempos de omissão para modelar a distribuição do processo que gerou a omissão dos dados. Proponhemos 2 modelos gerais de dependência e no caso de modelos do tipo 2, um método de imputação para os valores omitidos.

Finalmente, no Capítulo 5 discutiremos as conclusões e possíveis extensões deste trabalho.

Capítulo 1

Inferência em cadeias de Markov

Neste capítulo faremos uma revisão dos principais resultados inferenciais em cadeias de Markov.

1.1 Definições e resultados da literatura

$\mathcal{A}$: alfabeto finito (por simplicidade tomamos $\mathcal{A} = \{1, \dots, s\}$).

$\mathcal{A}^{\mathbb{Z}}$: o espaço das seqüências infinitas $x = (x_i) \in \mathcal{A}^{\mathbb{Z}}$.

x_m^n : denotará a seqüência $x_m, \dots, x_n$, $-\infty \leq m \leq n \leq \infty$.

x_m^n, x : denotará a seqüência $x_m, \dots, x_n, x$, $-\infty \leq m \leq n < \infty$.

Seja $X = (X_n)_{n \in \mathbb{Z}}$ um processo estacionário tomando valores no alfabeto finito $\mathcal{A}$ e definido no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$.

Definição 1.1.1 *Se existe $k \geq 1$ tal que para qualquer $a_{-\infty}^0 \in \mathcal{A}^{\mathbb{N}}$*

$$\mathbb{P}(X_0 = a_0 | X_{-k-m}^{-1} = a_{-k-m}^{-1}) = \mathbb{P}(X_0 = a_0 | X_{-k}^{-1} = a_{-k}^{-1}) \quad (1.1)$$

para todo $m \geq 1$ então chamaremos $X = (X_n)_{n \in \mathbb{Z}}$ de cadeia de Markov de ordem k . Se não existe nenhum k satisfazendo (??), dizemos que é uma cadeia de ordem infinita (c.o.i.)

Por comodidade denotaremos $p(a_0 | a_{-k}^{-1}) = \mathbb{P}(X_0 = a_0 | X_{-k}^{-1} = a_{-k}^{-1})$ e $\mathbf{P} = [p(a_0 | a_{-k}^{-1})]$ a matriz de transição.

Seja $x_1^N \in \mathcal{A}^N$ uma realização de uma cadeia de Markov de ordem k $(X_n)_{n \in \mathbb{Z}}$. Defina

$$n(a_{-k}^0) = \sum_{j=1}^N \mathbb{1}_{\{X_j^{j+k-1} = a_{-k}^0\}}, \text{ e } n(a_{-k}^{-1}, \cdot) = \sum_{a_0 \in \mathcal{A}} n(a_{-k}^0).$$

O estimador de máxima verossimilhança de $p(a_0 | a_{-k}^{-1})$ é dado por

$$\hat{p}_{MV}(a_0 | a_{-k}^{-1}) = \frac{n(a_{-k}^0)}{n(a_{-k}^{-1})}$$

Este estimador é consistente com distribuição assintótica normal Billingsley (1961).

Na estimação no enfoque bayesiano, usam-se prioris Dirichlet para cada linha da matriz de transição. A idéia baseia-se no fato que fixando k estados a_{-k}^{-1} , o número de transições $n(a_{-k}^0)$ tem distribuição multinomial.

$$\underbrace{f(\mathbf{P}|x_1^N)}_{\text{posteriori}} \propto \underbrace{f(\mathbf{P})}_{\text{priori}} \times \underbrace{L(x_1^N|\mathbf{P})}_{\text{verossimilhança}} \quad (1.2)$$

Se $p(\cdot|a_{-k}^{-1}) \sim \text{Dirichlet}(b_{a_{-k}^{-1},1}, \dots, b_{a_{-k}^{-1},s})$, b : parâmetros.

Temos que

$$p(\cdot|a_{-k}^{-1})|x_1^N \sim \text{Dirichlet}(n(a_{-k}^{-1}, 1) + b_{a_{-k}^{-1},1}, \dots, n(a_{-k}^{-1}, s) + b_{a_{-k}^{-1},s})$$

Neste caso as expressões para a média, moda e variância da distribuição posteriori são dadas por:

$$\begin{aligned} \mathbb{E} p(a_0|a_{-k}^{-1})|x_1^N &= \frac{b_{a_{-k}^0}^*}{b_{a_{-k}^0, \bullet}^*} \\ \text{moda} [p(a_0|a_{-k}^{-1})|x_1^N] &= \frac{b_{a_{-k}^0}^* - 1}{b_{a_{-k}^0, \bullet}^* - s} \\ \text{var} [p(a_0|a_{-k}^{-1})|x_1^N] &= \frac{b_{a_{-k}^0}^* (b_{a_{-k}^0, \bullet}^* - b_{a_{-k}^0}^*)}{b_{a_{-k}^0, \bullet}^{*2} (b_{a_{-k}^0, \bullet}^* + 1)} \end{aligned}$$

sendo $b_{a_{-k}^0}^* = n(a_{-k}^0) + b_{a_{-k}^0}$, e $b_{a_{-k}^0, \bullet}^* = \sum_{a_0 \in \mathcal{A}} n(a_{-k}^0) + b_{a_{-k}^0}$.

Próva-se que os estimadores de máxima verossimilhança e bayesiano são assintoticamente equivalentes com a mesma distribuição assintótica com taxa de convergência dada por :

$$\hat{p}(a_0|a_{-k}^{-1}) - p(a_0|a_{-k}^{-1}) = O_{q.c.} \left(\sqrt{\frac{\log \log n}{n}} \right)$$

Zhao, Dorea e Gonçalves (2001)

1.2 Determinação da ordem

1.2.1 Criterio de Determinação Eficiente

Um problema freqüente quando temos uma amostra de uma cadeia de Markov é determinar a ordem da cadeia. Diferentes abordagens incluem o teste de razão de verossimilhança (LRT), o critério de informação de Akaike (AIC), e o critério de informação bayesiana (BIC).

Zhao, Dorea e Gonçalves (2001) propõem o critério de determinação eficiente (EDC) para estimar a ordem de uma cadeia de Markov. A ordem r é estimada por $\hat{r}$ tal que

$$EDC(\hat{r}) = \min\{EDC(t), t = 0, 1, \dots, R\} \quad (1.3)$$

$$EDC(t) = -2 \log L_t + \gamma(t) c_n \quad (1.4)$$

- L_t : O valor da verossimilhança estimada com a amostra.
- $\gamma(t)$: O número de parâmetros livres a estimar, se a cadeia é de ordem t , e
- C_n : Uma seqüência de números positivos dependendo de n .

Inclui casos particulares como o AIC e BIC. Zhao, Dorea e Gonçalves (2001) dão condições para ter um estimador consistente da ordem da cadeia :

- Consistência forte

$$\frac{C_n}{n} \rightarrow +\infty \text{ e } \frac{C_n}{\log \log n} \rightarrow +\infty$$

- Consistência fraca

$$\frac{C_n}{n} \rightarrow \infty \text{ e } C_n \rightarrow +\infty$$

1.2.2 Fator de Bayes

A ideia do fator de Bayes é a seguinte:

Suponha que desejamos escolher do conjunto de modelos $\{M_1, \dots, M_r\}$ um modelo para o conjunto de observações Y e que os dados tem densidade $f_i(y_i|\theta_i)$ se o modelo for M_i . Sejam $p_1, \dots, p_r$ as prioris para cada modelo e $\pi_i(\theta_i)$ as prioris dos parâmetros dado o modelo.

$$P(M_i|Y) = \frac{m_i(y)p_i}{\sum m_j(y)p_j} \quad (1.5)$$

onde $m_i(y) = \int f_i(y|\theta_i)\pi_i(\theta_i)d\theta_i$ é a distribuição marginal dos dados dado o modelo M_i , a razão

$$B_{ij} = \frac{P(M_i|Y)/p_i}{P(M_j|Y)/p_j} = \frac{m_i(y)}{m_j(y)} \quad (1.6)$$

é chamada o fator de Bayes do M_i a M_j .

No caso de uma cadeia de Markov

$$m_k = \frac{1}{s^k} \left(\frac{\Gamma(sb)}{\Gamma(b)^s} \right)^{s^k} \times \prod_{a_{-k}^{-1} \in \mathcal{A}^k} \frac{\prod_{a_0 \in \mathcal{A}} \Gamma(n(a_{-k}^0) + b_{a_{-k}^0})}{\Gamma(n(a_{-k}^{-1}, \bullet) + b_{a_{-k}^{-1}, \bullet})} \quad (1.7)$$

1.2.3 Comparação dos métodos via simulação

Soto (2002), apresenta um estudo de simulação comparando os diferentes critérios para cadeias de Markov de varias ordens, número de estados e tamanhos de amostra. Obtivemos os seguinte:

- O teste de razão de verossimilhança só tinha bom performance no caso independente.
- O AIC em geral, tende a super-estimar a ordem da cadeia.
- O BIC tem boa performance, melhorando conforme aumenta o tamanho da amostra.
- O Fator de Bayes, consegue determinar em maior porcentagem a verdadeira ordem da cadeia independentemente da priori.

Consistência do fator de Bayes

O fator de Bayes conseguiu estimar consistentemente a ordem de uma cadeia de Markov, isto é, se a ordem é r se cumpre:

$$B_{rr'} \rightarrow +\infty, \forall r' \neq r, \text{ e } \log B_{rr'} = O(\log n)$$

Teorema 1.2.1 (Super-estimação e sub-estimação da ordem) *Seja r a ordem de uma cadeia de Markov $(X_n)_{n \in \mathbb{Z}}$ e suponha que temos uma realização $x_1^n \in \mathcal{A}^n$ de esta cadeia. Se $k \neq r$ é a ordem estimada, temos que*

- se $k > r$ (super-estimação)

$$\hat{p}(a_0|a_{-k}^{-1}) \xrightarrow{q.c.} p(a_0|a_{-r}^{-1})$$

$$\sqrt{n(a_{-k}^{-1})} (\hat{p}(a_0|a_{-k}^{-1}) - p(a_0|a_{-r}^{-1})) \rightarrow N(0, \Sigma_1)$$

- se $k < r$ (sub-estimação)

$$\hat{p}(a_0|a_{-k}^{-1}) \xrightarrow{q.c.} p(a_0|a_{-k}^{-1})$$

$$\sqrt{n(a_{-k}^{-1})} (\hat{p}(a_0|a_{-k}^{-1}) - p(a_0|a_{-k}^{-1})) \rightarrow N(0, \Sigma_2)$$

1.3 Cadeias de ordem infinita

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

$$p^{[k]}(a_0|a_{-k}^{-1}) := \mathbb{P}(X_0 = a_0 | X_{-k}^{-1} = a_{-k}^{-1})$$

Bressaud, Fernández e Galves (1999) dan resultados sobre a velocidade de convergência de aproximações markovianas de cadeias de ordem infinita. Guiol (2002), estuda o comportamento assintótico de estimadores da entropia para cadeias de ordem infinita com perda de memória exponencial.

Usando o teorema ergódico, prova-se que se x_1^n é uma amostra de uma cadeia de ordem infinita, para k fixo,

$$\hat{p}(a_0|a_{-k}^{-1}) \xrightarrow{q.c.} p^{[k]}(a_0|a_{-k}^{-1})$$

Isto é, podemos estimar consistentemente a aproximação markoviana de uma cadeia de ordem infinita.

Capítulo 2

Estimação com dados faltantes

2.1 Introdução

Um problema bastante comum quando coletamos dados é a falta de algumas observações. Rubin (1976) classifica o mecanismo de dados ausentes de acordo com a probabilidade de não esposta. Assim, quando a probabilidade de não resposta depende dos dados presentes mas não dos ausentes, os dados são ditos perdidos por um processo aleatório (MAR - missing at random). Se o mecanismo de não esposta não depende de dados presentes ou ausentes, dizemos que os dados são perdidos por um processo totalmente aleatório (MCAR - missing completely at random).

2.2 Construção Regenerativa

Seja $\{X_n\}$ uma cadeia de Markov com matriz de transição P e espaço de estados E , e $\{M_n\}$ um processo estocástico com valores em $\{0, 1\}$ onde a seguinte regra é imposta: sempre que $M_n = 1$ o valor de X_n não é observado.

Definição 2.2.1 A sequência $T_0 = 0, T_1 = \inf\{n \geq 1 : M_n = 1\}$ e para $n > 1$, $T_n = \inf\{n > T_{n-1} : M_n = 1\}$ é chamada *tempos de omissão* do processo $\{X_n\}$ e M_n o *processo de omissão*.

Baseados nas idéias de Rubin(1977) podemos estabelecer relações entre os processos X_n e M_n que nos permitam estimar o modelo com dados faltantes:

- Modelo 1: M_n e X_n são processos independentes
- Modelo 2: M_n depende de $X_{i=1}^n$, $n = 1, 2, \dots$
- Modelo 3: O processo M_n, X_n é uma cadeia de Markov bivariada

Proposição 2.2.1 Se $M_n \in \sigma(X_1, \dots, X_n)$ para todo $n = 1, 2, \dots$, dado $\{T_1^\infty = t_1^\infty\}$, as sequências $\{X_{T_{i-1}+1}^{T_i-1}\}, i = 1, 2, \dots$ são independentes.

Prova:

Por construção os $\{T_n\}$ são tempos de parada e da propriedade markoviana segue o resultado. $\square$.

Este teorema nos auxiliará para construir um novo processo sem valores ausentes. Defina o processo W_n da seguinte maneira:

- Para $1 \leq n \leq T_1$, $W_n = X_n$
- Para $n \geq T_1$, $Z_1 = \inf\{n > T_1 : X_n = X_{T_1}\}$, $W_{T_1+1} = X_{Z_1+1}$

Teorema 2.2.1 Se $M_n \in \sigma(X_1, \dots, X_n)$ para todo $n = 1, 2, \dots$, dado $\{T_1^\infty = t_1^\infty\}$, o processo W_n é uma cadeia de Markov com matriz de transição P .

Prova: Por construção é imediata $\square$.

Exemplo 2.2.1 Considere 21 realizações de uma cadeia de Markov em $S = 1, 2, 3$ e matriz de transição

$$P = \begin{pmatrix} 0,3 & 0,4 & 0,3 \\ 0,5 & 0,3 & 0,2 \\ 0,2 & 0,4 & 0,4 \end{pmatrix}$$

$X : 2, 1, *, 3, 1, *, *, 1, 2, *, 3, *, 3, 2, 1, 1, 3, 2, 2, 3, 3$.

Aquí, $*$ indica que o valor não é observado.

Note-se que $M_3, M_6, M_7, M_{10}, e M_{12}$ são iguais a zero e $T_1 = 3, T_2 = 6, T_3 = 7, T_4 = 10, T_5 = 12$ são os tempos de omissão dos dados.

$$W = x_1^2, x_9, x_{15}^{21} = 2, 1, 2, 1, 1, 3, 2, 2, 3, 3$$

Como consequencia do ultimo teorema temos o seguinte resultado:

Teorema 2.2.2 Seja $\{x_1^n\}$ uma realização de uma cadeia de Markov $\{X_n\}$ ergódica com matriz de transição P e $\{M_n\}$ um processo de omissão associado a $\{X_n\}$. Os estimadores obtidos do processo W_n são consistentes e assintoticamente normais.

Exemplo 2.2.2 (Continuação) Com os valores obtidos de w_n a matriz de transição estimada é dada por

$$\hat{P} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

Os resultados anteriores continuam sendo válidos se considermos uma cadeia de Markov de ordem $K > 1$.

Capítulo 3

Reamostragem em Cadeias de Markov

3.1 Introdução

A utilização de técnicas de reamostragem tem tido um importante papel na estatística, especialmente quando o cálculo de estimadores por métodos analíticos é complicado. Uma delas a técnica *bootstrap*, foi introduzida por Efron(1979) e teve grande desenvolvimento nos últimos 20 anos, mostrando-se uma interessante ferramenta de estimação.

O método de reamostragem (*bootstrap*) proposto por Efron (1979), consiste em : Considere uma amostra aleatória $z = \{z_1, \dots, z_n\}$ de uma função distribuição F e a estatística de interesse $s(z)$. A estimativa *bootstrap* para a esperança de $s(z)$ é

$$\hat{e} = E_{\hat{F}}s(z^*), \quad (3.1)$$

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

$$\hat{e}_B = \frac{1}{B} \sum_{b=1}^B s(z^{*b}), \quad (3.2)$$

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

Athreya e Fu(1992) estendem as ideias de Efron para o caso de cadeias de Markov, introduzindo a reamostragem uniforme em blocos.

Neste capítulo mostraremos que as técnicas de reamostragem podem ser aplicadas no caso de ter uma realização incompleta de uma cadeia de Markov, mostraremos que podem ser usadas tanto na sequência original como no processo construído $\{W_n\}$.

3.2 Reamostragem uniforme em blocos

Suponha $x = \{x_0, \dots, x_n\}$ uma realização de uma cadeia de Markov ergódica (recorrente positiva, aperiódica e irredutível) $\{X_j; j = 0, \dots, n\}$, em um espaço de estado finito $S = \{1, \dots, s\}$, com probabilidade de transição $P = (p_{ij})$. A propriedade da ergodicidade implica a existência de uma medida invariante $\pi = (\pi_j), \pi_j > 0, \sum_{j=1}^s \pi_j = 1, \pi_j = \sum_i \pi_i p_{ij}, j = 1, \dots, s$ e para todo $i \in S, p_{ij}^{(n)} \rightarrow \pi_j$, quando $n \rightarrow \infty$, onde $p_{ij}^{(n)} = P\{X_n = j \mid X_0 = i\}$. Seja n_{ij} o número de transido estado i para o j em x , e n_i o número de visitas do estado i em x . O estimador de máxima verossimilhança da probabilidade de transição P é $\hat{P}_n \equiv (\hat{p}_n(i, j))$, onde $\hat{p}_n(i, j) = \frac{n_{ij}}{n_i}$.

Este estimador pode ser usado como uma aproximação da distribuição amostral, mas a dificuldade de calcular a matriz de variância-covariância faz este procedimento pouco apropriado para aplicações. Assim o método de reamostragem é uma técnica usada para estudar este tipo de problema.

Seja a realizada cadeia de Markov x , esta é particionada em blocos, onde o critério de parada é o tempo até que ocorra pelo menos duas transições para todos os estados em cada um dos blocos.

Determinando um critério 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 ficar o número de blocos k , e observar a cadeia até um n aleatório. A segunda maneira é fixar o tamanho da amostra n , assim o número de blocos k é uma variável aleatória.

Considerando a segunda alternativa, observamos a cadeia até o tempo fixado n e obtemos k blocos. Seja $T^{(l)}$ o tempo até que ocorra pelo menos duas transições cada dos estados no bloco l , sendo $l = 1, \dots, k$, temos a seguinte decomposta amostra original: $x = \{\eta_1, \dots, \eta_k, \eta_{k+1}\}$ onde $\eta_1 = \{x_0, \dots, x_{T^{(1)}-1}\}$ e $\eta_{k+1} = \{x_{T^{(k+1)}}, \dots, x_n\}$.

Seja $g(\eta_l)$ o número de visitas do estado i no bloco η_l , e $h(\eta_l)$ indica o número de transições ij no bloco η_l , então $p_l(i, j) = \frac{h(\eta_l)}{g(\eta_l)}$ é o estimador de p_{ij} dentro do l -ésimo bloco.

Seja

$$\hat{p}_k(i, j) \equiv \frac{\sum_{\alpha=1}^k h(\eta_\alpha)}{\sum_{\alpha=1}^k g(\eta_\alpha)}, \quad (3.3)$$

o estimador de P .

Pela lei forte P_k é consistente. Então para estimar a distribuição de $P_k - P$, é proposto o seguinte método:

1. Decomponha a amostra original da seguinte maneira: $\{\eta_0, \eta_1, \dots, \eta_k\}$, onde $\eta_0 \equiv \{X_0, X_1, \dots, X_{T_{\Delta}^{(1)}-1}\}$.
Seja $\hat{F}_k$ a medida de probabilidade uniforme dos ciclos $\{\eta_\alpha; \alpha = 1, \dots, k\}$.
2. Com a amostra original fixada, retire uma amostra *bootstrap* de tamanho k de acordo com $\hat{F}_k$. Denotando a amostra por $\eta_1^*, \dots, \eta_k^*$. Definimos:

$$\tilde{p}_k(i, j) \equiv \frac{\sum_{\alpha=1}^k h(\eta_\alpha^*)}{\sum_{\alpha=1}^k g(\eta_\alpha^*)}, \quad (3.4)$$

3. Aproxima a distribuição de $\sqrt{k}(\hat{p}_k(i, j) - p_{ij})$ pela distribuição condicional de $\sqrt{k}(\tilde{p}_k(i, j) - \hat{p}_k(i, j))$ dado x .

Uma maneira diferente deste método é fixar o tamanho da amostra original n . Seja k_n o número aleatório de ciclos, então é feita seleção uniforme destes k_n ciclos.

3.3 Reamostragem em Blocos Balanceados

Este método consiste em uma reamostragem balanceada. Seja um estado recorrente Δ fixado, pela propriedade forte de Markov, os ciclos $\{X_j; j = T_{\Delta}^{(n)}, \dots, T_{\Delta}^{(n+1)} - 1\}$ são i.i.d para $n = 1, 2, \dots$, onde $T_{\Delta}^{(n)}$ é o tempo do n -ésimo retorno para Δ .

Observe a cadeia até o tempo fixo n , o número de ciclos k_n é uma variável aleatória. Seja $\{x_0, \dots, x_n\}$ a realização do processo, para quaisquer dois estados i, j , que são diferentes de Δ , seja

$$\eta_\alpha \equiv \{X_j; j = T_{\Delta}^{(\alpha)}, \dots, T_{\Delta}^{(\alpha+1)} - 1\},$$

o α -ésimo ciclo, então a amostra original pode ser decomposta da seguinte maneira :

$$\{\eta_0, \eta_1, \dots, \eta_{k_n}, \eta_{k_n+1}\},$$

onde $\eta_0 = \{x_0, x_1, \dots, x_{T_{\Delta}^{(1)}-1}\}$ e $\eta_{k_n+1} = \{x_{T_{\Delta}^{k_n+1}}, \dots, x_n\}$. Seja $g(\eta_\alpha)$ indica o número de visitas do estado i durante o ciclo η_α , e $h(\eta_\alpha)$ indica o número de transições $i \rightarrow j$ durante o ciclo η_α . Seja

$$\hat{p}_{k_n}(i, j) \equiv \frac{\sum_{\alpha=1}^{k_n} h(\eta_\alpha)}{\sum_{\alpha=1}^{k_n} g(\eta_\alpha)},$$

$$\hat{\pi}_k(i, j) \equiv \frac{\sum_{\alpha=1}^k g(\eta_\alpha)}{\sum_{\alpha=1}^k T_\alpha}$$

os estimadores de π e P , onde T_α é o estimador de p_{ij} .

Pela lei forte $\hat{p}_{k_n}(i, j)$ é consistente. Para a aproximação *bootstrap* da distribuição amostral de $\hat{p}_{k_n}(i, j) - p_{ij}$, selecione B amostras aleatórias uniformemente, com reposição, da amostra decomposta $\chi = \{\eta_1, \dots, \eta_{k_n}\}$, e aplique a técnica de reamostragem balanceada. O objetivo deste procedimento é garantir que cada bloco apareça o mesmo número de vezes em todas as reamostras. Defina

$$\hat{\sigma}^2 = \frac{1}{(\sum_{\alpha=1}^{k_n} g(\eta_\alpha))^2} \sum_{\alpha=1}^{k_n} (g(\eta_\alpha)(p_\alpha(i, j) - \hat{p}_{k_n}(i, j)))^2$$

onde $p_\alpha(i, j) = h(\eta_\alpha)/g(\eta_\alpha)$ é o estimador de p_{ij} dentro do α -ésimo bloco. Então o esquema de reamostragem balanceada consiste em:

1. Repete cada bloco B vezes, obtendo um total de Bk_n blocos:

$$\eta_1, \dots, \eta_1, \eta_2, \dots, \eta_2, \dots, \eta_{k_n}, \dots, \eta_{k_n}$$

2. Permuta a sequência acima de tal forma que cada grupo contenha os k_n blocos:

$$\eta_{11}^\dagger, \dots, \eta_{1k_n}^\dagger, \eta_{21}^\dagger, \dots, \eta_{2k_n}^\dagger, \dots, \eta_{B1}^\dagger, \dots, \eta_{Bk_n}^\dagger$$

3. As reamostras balanceadas são denotadas por:

Para cada reamostra balanceada, $\chi_b^\dagger, b = 1, \dots, B$, a estimativa *bootstrap* de $\hat{p}_{k_n}(i, j)$ é

$$\hat{p}_{k_nb}^\dagger = \frac{\sum_{\alpha=1}^{k_n} h(\eta_{b\alpha}^\dagger)}{\sum_{\alpha=1}^{k_n} g(\eta_{b\alpha}^\dagger)}$$

A distribuição de $(\hat{p}_{k_n}(i, j) - p_{ij})/\hat{\sigma}$ pode ser aproximada pela distribuição empírica de $(\hat{p}_{k_nb}^\dagger(i, j) - \hat{p}_{k_n}(i, j))/\hat{\sigma}_b^\dagger, b = 1, \dots, B$

3.4 Reamostragem Importante em Blocos

Com a mesma notação de reamostragem balanceada, o algoritmo de reamostragem em blocos para estimar $P\{(\hat{p}_{k_n}(i, j) - p_{ij})/\hat{\sigma} \leq x\}$ sobre a reamostragem importante consiste em:

1. Seja

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

e seja a distribuição de probabilidade da reamostragem $\{q_\alpha; \alpha = 1, \dots, k_n\}$ de $\{\eta_\alpha; \alpha = 1, \dots, k_n\}$ dada por

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

onde A é escolhido para minimizar $\exp(A^2)\Phi(x - A)$ com $\Phi(\cdot)$ sendo a função distribuição normal padrão.

2. Com a mesma amostra $\chi = \{\eta_1, \dots, \eta_{k_n}\}$ fixada, retire uma amostra de tamanho $k_n, \{\eta_1^\dagger, \dots, \eta_{k_n}^\dagger\}$, sobre a reamostragem importante com distribuição de probabilidade $\{q_\alpha; \alpha = 1, \dots, k_n\}$. Seja l_α o número de vezes que η_α aparece em $\{\eta_1^\dagger, \dots, \eta_{k_n}^\dagger\}$. Então a estimativa de *bootstrap* de $\hat{p}_{k_n}(i, j)$ é

$$\tilde{p}_{k_n}(i, j) = \frac{\sum_{\alpha=1}^{k_n} h(\eta_\alpha^\dagger)}{\sum_{\alpha=1}^{k_n} g(\eta_\alpha^\dagger)} \prod_{\alpha=1}^{k_n} (k_n q_\alpha)^{-l_\alpha}.$$

Então é possível aproximar a distribuição de $\hat{p}_{k_n}(i, j) - p_{ij}$ pela distribuição condicional de $\tilde{p}_{k_n}(i, j) - \hat{p}_{k_n}(i, j)$, dado x .

Teorema 3.4.1 *Se $E_\Delta T_\Delta^2 < \infty$, então para quase todas as realizado processo $\{X_n; n \geq 0\}$, nós temos para cada i, j ,*

$$\sqrt{k}(\tilde{p}_k(i, j) - \hat{p}_k(i, j)) \rightarrow N(0, \sigma^2)$$

em distribuição, onde σ^2 será definido na prova.

Prova:

Seja

$$\begin{aligned}
\sqrt{k}(\tilde{p}_k(i, j) - \hat{p}_k(i, j)) &= \sqrt{k} \left(\frac{\sum h(\eta_t^*)}{\sum g(\eta_t^*)} - \frac{\sum h(\eta_t)}{\sum g(\eta_t)} \right) \\
&= \sqrt{k} \left[\frac{\sum h(\eta_t^*) \sum g(\eta_t)}{\sum g(\eta_t^*) \sum g(\eta_t)} - \frac{\sum h(\eta_t)}{\sum g(\eta_t)} \right] \\
&= \sqrt{k} \frac{\sum h(\eta_t^*)}{\sum g(\eta_t)} \left(\frac{\sum g(\eta_t)}{\sum g(\eta_t^*)} - \frac{\sum h(\eta_t)}{\sum h(\eta_t^*)} \right) \\
&= \sqrt{k} \frac{\sum h(\eta_t^*) \sum g(\eta_t)}{\sum g(\eta_t^*) \sum g(\eta_t)} - \sqrt{k} \frac{\sum h(\eta_t^*)}{\sum g(\eta_t)} + \sqrt{k} \frac{\sum h(\eta_t^*)}{\sum g(\eta_t)} - \sqrt{k} \frac{\sum h(\eta_t)}{\sum g(\eta_t)} \\
&= \frac{\sqrt{k}}{\sum g(\eta_t)} \sum h(\eta_t^*) \left(\frac{\sum g(\eta_t)}{\sum g(\eta_t^*)} - 1 \right) + \frac{\sqrt{k}}{\sum g(\eta_t)} (\sum h(\eta_t^*) - \sum h(\eta_t)) \\
&= \sqrt{k} \frac{\sum h(\eta_t^*)}{\sum g(\eta_t^*) \sum g(\eta_t)} (\sum g(\eta_t) - \sum g(\eta_t^*)) \\
&\quad - \frac{E(h(\eta_1))}{\sqrt{k}(E(g(\eta_1)))^2} (\sum g(\eta_t) - \sum g(\eta_t^*)) \\
&\quad + \frac{\sqrt{k}}{\sum g(\eta_t)} (\sum h(\eta_t^*) - \sum h(\eta_t)) - \frac{1}{\sqrt{k}E(g(\eta_1))} (\sum h(\eta_t^*) - \sum h(\eta_t)) \\
&\quad + \frac{1}{\sqrt{k}E(g(\eta_1))} (\sum h(\eta_t^*) - \sum h(\eta_t)) + \frac{E(h(\eta_1))}{\sqrt{k}(E(g(\eta_1)))^2} (\sum g(\eta_t) \\
&= \left[\frac{k^{-1} \sum h(\eta_t^*)}{k^{-1} \sum g(\eta_t^*) k^{-1} \sum g(\eta_t)} - \frac{Eh(\eta_1)}{(Eg(\eta_1))^2} \right] \frac{1}{\sqrt{k}} (\sum g(\eta_t) - \sum g(\eta_t^*)) \\
&\quad + \left(\frac{1}{\sum g(\eta_t)/k} - \frac{1}{Eg(\eta_1)} \right) \frac{1}{\sqrt{k}} (\sum h(\eta_t^*) - \sum h(\eta_t)) \\
&\quad + \left(\frac{1}{Eg(\eta_1)} \frac{1}{\sqrt{k}} (\sum h(\eta_t^*) - \sum h(\eta_t)) - \frac{Eh(\eta_1)}{(Eg(\eta_1))^2} \frac{1}{\sqrt{k}} (\sum g(\eta_t^*) - \sum g(\eta_t)) \right) \\
&= W_{k1} + W_{k2} + W_{k3}
\end{aligned}$$

W_{k1}, W_{k2} convergem para 0 com probabilidade 1. Seja $c \equiv \frac{1}{Eg(\eta_1)}$ e $d \equiv \frac{Eh(\eta_1)}{Eg(\eta_1)^2}$. Então W_{k3} é dado por

$$\frac{1}{\sqrt{k}} \sum_{t=1}^k (c(h(\eta_t^*) - h(\eta_t)) + d(g(\eta_t^*) - g(\eta_t)))$$

$$\begin{aligned}
&= \frac{1}{\sqrt{k}} \sum_{t=1}^k ((ch(\eta_t^*) + dg(\eta_t^*)) - (ch(\eta_t) + dg(\eta_t))) \\
&\rightarrow N(0, c^2 var(h(\eta_1)) + d^2 var(g(\eta_1)) + 2cd cov(h(\eta_1), g(\eta_1))),
\end{aligned}$$

em distribuição. $\square$

Capítulo 4

Modelos para o processo de omissão de dados

4.1 Introdução

Neste capítulo nosso interesse é modelar a distribuição do processo de omissão de dados $\{M_n\}$. Apresentaremos várias estruturas de dependência baseadas na ideia de processos MAR e MCAR.

4.2 Modelos para a omissão de dados

Segundo Rubin(1976) quando a probabilidade de não resposta depende dos dados presentes mas não dos ausentes, os dados são ditos perdidos por um processo aleatório (MAR - missing at random). Se o mecanismo de não resposta não depende de dados presentes ou ausentes, dizemos que os dados são perdidos por um processo totalmente aleatório (MCAR - missing completely at random).

No capítulo 2, vimos que se $M_n \in \sigma(X_1, \dots, X_n, \forall n = 1, 2, \dots)$ podemos obter estimativas consistentes para a matriz de transição. A ideia é propor modelos de probabilidade que se encaixem como um processo MAR.

- Modelo 1: $P(M_n | M_1^{n-1}, X_1^n) = P(M_n | M_{n-k}^{n-1}), \forall n = 1, 2, \dots$
- Modelo 2: $P(M_n | M_1^{n-1}, X_1^n) = P(M_n | X_{n-k}^{n-1}), \forall n = 1, 2, \dots$

Modelos do tipo 1 podem ser tratados como cadeias de Markov de ordem k , e podemos usar as técnicas habituais de estimação. Neste modelo todos os valores têm igual probabilidade de ser omitidos.

Modelos do tipo 2, podem ser tratados como uma generalização dos modelos de Markov onde o processo $\{X_n\}$ não é mais oculto se não parcialmente observado e o

processo observado corresponderia a $\{M_n\}$. As estimativas podem ser obtidas construindo um algoritmo do tipo EM que generalize as estimativas Baum-Welch. As vantagens de modelos do tipo 2 é que podemos imputar os valores não observados por os valores mais possíveis mediante uma variante do algoritmo de Viterbi.

Capítulo 5

Discussão e Conclusões

Neste trabalho estudamos o problema estimação de dados incompletos no caso de uma cadeia de Markov.

Usando uma construção regenerativa construímos um novo processo completo com a mesma matriz de transição, que pode ser utilizado para estimar as probabilidades de transição. vimos também que podemos estender as ideias de reamostragem em blocos de Athreya e Fu para ter estimativas bootstrap das transições.

Quando é de interesse modelar o processo de omissão dos dados sugerimos 2 modelos, um dos quais permite imputar as observações faltantes por os valores mais probáveis.

Embora tenhamos obtido resultados satisfatórios na análise de dados incompletos em cadeias de Markov ainda tem muito por ser feito, possíveis extensões deste trabalho seriam:

- Propor outras estrutura de dependência para a omissão dos dados.
- Estender alguns resultados para processos de Markov a tempo contínuo
- Incluir o problema da determinação da ordem da cadeia dentro da abordagem do bootstrap.

Referências Bibliográficas

- Adke, S. R. & Manjunath, S. M. (1984). *An introduction to finite Markov processes*. John Wiley & Sons. New York.
- Adnan, A. (1996). *Properties of the Akaike information criterion*. *Microelectron. Reliab.*, Vol. 36, Nº 4, 457-46.
- Anderson, T. W. and Goodman L. A. (1957). *Statistical inference about Markov chains*, *Ann. Math. Statist.*, 28, 89-110.
- Akaike, H. (1973) *Information theory and an extension of the maximum likelihood principle*. 2nd Inter. Symp. on Information Theory (Petrov. B. N. and Csaki F. eds,) Akademiai Kiado, Budapest, 267 - 281.
- Basawa, I. and Prakasa Rao, B. (1980). *Statistical inference for stochastic processes*. Academic Press. London.
- BHAT, N. (1984). *Elements of applied stochastic processes*. John Wiley & Sons, Inc. USA.
- BILLINGSLEY, P. (1961). Statistical methods in Markov chains *Annals of Mathematical Statistics*, 32, 12-40.
- BICKEL, P.; DOKSUM, K. (1977). *Mathematical Statistics* Holden-Day, Inc. USA.
- KARLIN, S.; TAYLOR H. (1975). *A first course in stochastic processes*. Academic Press, Inc. New York.
- DOREA, C.; GALVES A.; KIRA, E. and PEREIRA ALENCAR A. (1997). Markovian modeling of the stress contours of Brazilian and European Portuguese. *REBRAPE*, Vol. 11, Nº 2, 161-173.
- FERRARI, P. e GALVES, A. (1997). *Acoplamento e processos estocásticos*. 21º Colóquio Brasileiro de Matemática. IMPA. Brasil.
- GARTHWAITE, P.; JOLLIFFE, I.; JONES, B. (1995). *Statistical inference*. Prentice Hall International. UK.
- GRAY, R. (1990). *Entropy and information theory*. Springer-Verlag. New York.

- GRAYBILL, F. (1976). *Theory and application of the linear model*. Wadsworth Publishing Company, Inc.
- GUTTORP, P. (1995). *Stochastic modeling of scientific data*. Chapman and Hall. London.
- HOEL, P.; PORT, S. and STONE, C. (1972). *Introduction to stochastic processes*. Houghton Mifflin Company. Boston.
- KATZ, R. W. (1981). On some criteria for estimating the order of a Markov chain. *Technometrics*, **23**, 243-249.
- KULLBACK, S.; LEIBLER, R. (1951). On information and sufficiency *Ann. Math. Statist.*, **22**, 79 - 86.
- KULLBACK, S. (1959). *Information theory and statistics*. John Wiley & Sons, Inc. USA.
- MACKAY, D. (2001). *Information theory, inference, and learning algorithms*. Preprint, 62pp.
- RISSANEN, J. (2002). Lectures on statistical modeling theory. www.cs.tut.fi/~rissanen/papers/lectures2.ps, 69pp.
- ROSS, S. (1996). *Stochastic processes*. 2nd ed. John Wiley & Sons. USA.
- SCHWARZ, G. (1978). Estimating the dimension of a model *Annals of Statist.*, **6**, 461-464.
- SAKAMOTO, Y.; ISHIGURO, M. and KITAGAWA, G. (1986). *Akaike information criterion statistics*. Dordrecht: Kluwer Academic Publishers. Translated from Japanese.
- TONG, H. (1975). Determination of the order of a Markov chain by Akaike's information criterion *J.Appl. Prob.*, **12**, 488-497.