MAE 326 - Lista de Exercícios 4

Prof. Antonio Galves

28 de outubro de 2013

Algoritmo para seleção do alcance de Cadeias de Markov:
Dada uma amostra $X_0^n = (X_0, \ldots, X_n)$ de símbolos do alfabeto A finito, definimos

$$\hat{k}_n = \inf\{k \geq 0 : \hat{p}_n^{(k)} \approx \hat{p}_n^{(j)}, \forall j \geq k\},$$

como um critério para selecionarmos o alcance do mecanismo aleatório Markoviano que gerou a amostra. Na definição acima, $\hat{p}_n^{(k)} \approx \hat{p}_n^{(j)}$, para $j \geq k$ se e somente se

$$\max_{a \in A} \max_{a_{-j}^{-1} \in A^j} \{|\hat{p}_n^{(k)}(a|a_{-k}^{-1}) - \hat{p}_n^{(j)}(a|a_{-j}^{-1})|\} \leq \delta,$$

onde $\delta \in (0,1)$ é o limiar fixado. Sempre que $\hat{p}_n^{(k)} \approx \hat{p}_n^{(j)}$, diremos que essas duas medidadas de probabilidades são estatisticamentes iguais.

1. Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias iid assumindo valores em $A = \{0, 1\}$ e com distribuição dada por

   $$\mathbb{P}(X_i = 1) = p \text{ e } \mathbb{P}(X_i = 0) = 1 - p, \text{ para } \ i = 1, 2, \ldots$$

   Para $n = 1000$ e $p = 0.5$, para quais valores de $\delta > 0$ o majorante da desigualdade de Chebyshev é menor ou igual a 0.05?

2. Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias iid assumindo valores em $A = \{0, 1\}$ e com $\mathbb{P}(X_i = 1) = p$, para $i = 1, \ldots, n$. Se $p = 0.5$ e $\delta = 0.02$, encontre $\bar{n}$ tal que para todo $n \geq \bar{n}$ temos

   $$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\Big| \geq \delta\Big) \leq 0.05.$$

3. Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias iid assumindo valores em $A = \{1, \cdots, |A|\}$ com distribuição dada por

$$\mathbb{P}(X_i = a) = p(a), \ \forall a \in A$$

onde $p(a)$ é conhecido para todo $a \in A$. Definimos para cada $a \in A$ e $n \geq 0$,

$$\hat{p}_n(a) = \sum_{t=1}^{n} 1_{\{X_t = a\}},$$

o número de vezes que a cadeia $(X_t)_{t \geq 0}$ visita o símbolo a até o instante n. Usando a desigualdade de Chebyshev, sabemos que para todo $\delta > 0$ e $n \geq 0$

$$\mathbb{P}(|\hat{p}_n(a) - p(a)| > \delta) \leq \frac{(a(1 - p(a))^2}{n\delta^2}.$$

Mostre, a partir da desigualde acima, que para qualquer $\epsilon > 0$ e $a \in A$, existe $\bar{n} = \bar{n}(\epsilon, a)$ tal que,

$$\mathbb{P}(\hat{p}_n(a) - \delta \leq p(a) \leq \hat{p}_n(a) + \delta) \geq 1 - \epsilon.$$

4. Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias iid assumindo valores em $A = \{1, \cdots, |A|\}$ com distribuição dada por

$$\mathbb{P}(X_i = a) = p(a), \ \forall a \in A$$

onde $p(a)$ é desconhecido para todo $a \in A$. Usando a desigualdade de Chebyshev, sabemos que para todo $\delta > 0$ e $n \geq 0$

$$\mathbb{P}(|\hat{p}_n(a) - p(a)| > \delta) \leq \frac{1}{4n\delta^2}.$$

Mostre, a partir da desigualde acima, que para qualquer $\epsilon > 0$, existe $\bar{n} = \bar{n}(\epsilon)$ tal que para todo $a \in A$,

$$\mathbb{P}(\hat{p}_n(a) - \delta \leq p(a) \leq \hat{p}_n(a) + \delta) \geq 1 - \epsilon.$$

5. Dada uma amostra $X = (X_0, \cdots, X_{10000})$ gerada por um algoritmo aleatório assumindo valores em $A = \{0, 1\}$, obteve-se os seguintes valores das funções de contagem:

a_0	a_1	a_2	$N_{10000}(a_0a_1a_2)$
0	0	0	3211
0	0	1	790
0	1	0	195
1	0	0	805
0	1	1	810
1	0	1	190
1	1	0	780
1	1	1	3220

a_0	a_1	$N_{10000}(a_0a_1)$
0	0	3940
0	1	1036
1	0	4010
1	1	1015

a_0	$N_{10000}(a_0)$
0	4981

Utilizando $\hat{k}_n$ como critério de estimação, teste a hipótese nula

$$H_0 = \{X \text{ foi gerada por uma Cadeia de Markov de alcance } 1\}$$

assumindo $\delta = 0.01$.

6. Dada uma amostra $X = (X_0, \cdots, X_{100000})$ gerada por um algoritmo aleatório assumindo valores em $A = \{0, 1\}$, obteve-se os seguintes valores das funções de contagem:

a_0	a_1	a_2	a_3	$N_{100000}(a_0a_1a_2a_3)$
0	0	0	0	24149
0	0	0	1	5850
0	0	1	0	3199
0	1	0	0	2800
1	0	0	0	6250
0	0	1	1	4250
0	1	0	1	2100
1	0	0	1	1400
0	1	1	0	1599
1	0	1	0	1900
1	1	0	0	4600
0	1	1	1	6000
1	1	0	1	3000
1	0	1	1	3050
1	1	1	0	5950
1	1	1	1	23901

a_0	a_1	a_2	$N_{100000}(a_0a_1a_2)$
0	0	0	30000
0	0	1	7449
0	1	0	4900
1	0	0	7650
0	1	1	7599
1	0	1	4950
1	1	0	7600
1	1	1	29851

a_0	a_1	$N_{100000}(a_0a_1)$
0	0	37450
0	1	12499
1	0	12600
1	1	37451

a_0	$N_{100000}(a_0)$
0	49950

Utilizando $\hat{k}_n$ como critério de estimação, teste a hipótese nula

$$H_0 = \{X \text{ foi gerada por uma Cadeia de Markov de alcance } 2\}$$

assumindo $\delta = 0.01$