Algoritmos de Monte Carlo e Cadeias de Markov - CPS767 2018/1

Prof. Daniel R. Figueiredo

Gabriel Matos Cardoso Leite

Segunda Lista de Exercícios

March 29, 2018

Questão 1: Cauda do dado em ação

Considere um icosaedro (um sólido Platônico de 20 faces) honesto, tal que a probabilidade associada a cada face é $\frac{1}{20}$. Considere que o dado será lançado até que um número primo seja obtido, e seja Z a variável aleatória que denota este número. Responda às perguntas abaixo:

1. Determine a distribuição de Z, ou seja $P[Z = k], k = 1, 2, \ldots$. Que distribuição é esta?
2. Utilize a desigualdade de Markov para calcular um limitante para $P[Z \geq 10]$.
3. Utilize a desigualdade de Chebyshev para calcular um limitante para $P[Z \geq 10]$.
4. Calcule o valor exato de $P[Z > 10]$ (dica: use probabilidade complementar). Compare os valores obtidos.

Solução:

1. Estamos interessados em executar o experimento de lançar o dado por um número de vezes até que se obtenha um sucesso (um número primo). Os primos entre 1 e 20 são: $2, 3, 5, 7, 11, 13, 17, 19$. Com isso, a distribuição da variável Z é Geométrica com $p = \frac{8}{20}$
2. A desigualdade de Markov é dada pela equação:

   $$P[Z \geq a) \leq \frac{E[Z]}{a}$$

   O valor esperado da variável Z é dado por:

   $$E[Z] = \mu = \frac{1}{p} = \frac{20}{8} = 2.5$$

   Com isso temos,

   $$P[Z \geq 10] \leq \frac{2.5}{10}$$
   $$P[Z \geq 10] \leq 0.25$$

3. Primeiro precisamos calcular a variância de Z.

$$Var[Z] = \sigma^2 = \frac{1-p}{p^2} = \frac{1-\frac{8}{20}}{\frac{8}{20}^2} = 3.75$$

Agora precisamos colocar o que temos na forma da desigualdade de Chebyshev.

$$P[|Z-\mu| \geq k\sigma] \leq \frac{1}{k^2}$$

$$\begin{aligned} Z &\geq 10 \\ Z - 2.5 &\geq 7.5 \\ k\sigma &= 7.5 \\ k &= \frac{7.5}{\sqrt{3.75}} \approx 3.87 \end{aligned}$$

Com isso podemos calcular o valor do limitante.

$$P[|Z-2.5| \geq 7.5] \leq \frac{1}{3.87^2} \approx 0.067$$

4.

$$\begin{aligned} P[Z > 10] &= 1 - P[Z \leq 10] \\ 1 - (1-(1-p)^{10}) &= \frac{12}{20}^{10} \\ P[Z > 10] &\approx 0.006 \end{aligned}$$

O limitante calculado pela desigualdade de Chebyshev foi bem próximo do valor real enquanto o obtido pela desigualdade de Markov ficou bem acima do valor real.

Questão 2: Cauda do dado em ação 2

Considere um icosaedro (um sólido Platônico de 20 faces) honesto, tal que a probabilidade associada a cada face é $\frac{1}{20}$. Considere que o dado será lançado n vezes, e seja Z a variável aleatória que denota o número de vezes que o resultado do dado foi um múltiplo de seis. Responda às perguntas abaixo:

1. Determine a distribuição de Z, ou seja $P[Z = k], k = 1, 2, \ldots, n$. Que distribuição é esta?

2. Seja $n = 1000$, utilize a desigualdade de Chebyshev para calcular um limitante para $P[Z > 300]$.

3. Seja $n = 1000$, utilize a desigualdade de Chernoff para calcular um limitante para $P[Z > 300]$.

4. Seja $n = 1000$, calcule o valor exato para $P[Z > 300]$. Compare os valores obtidos.

5. Determine o valor z em função de n tal que $Z \leq z$ whp (*with high probability*).

Solução:

1. Estamos interessados em executar o experimento de lançar o dado por um número de vezes e contar o número de sucessos (múltiplos de 6). Os múltiplos de 6 entre 1 e 20 são: $6, 12, 18$. Com isso, a distribuição da variável Z é Binomial com $p = \frac{3}{20}$

2. O valor esperado de Z é:

$$E[Z] = \mu = np = 1000 \cdot \frac{3}{20} = 150$$

   E a sua variância é:

$$Var[Z] = \sigma^2 = np(1-p) = 150 \cdot \frac{17}{20} = 127.5$$

   Agora colocamos na forma da desigualdade

$$P[|Z - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

$$\begin{aligned} Z &\geq 300 \\ Z - 150 &\geq 150 \\ k\sigma &= 150 \\ k &= \frac{150}{\sqrt{127.5}} \approx 13.28 \end{aligned}$$

   Com isso podemos calcular o valor do limitante.

$$P[|Z - 150| \geq 150] \leq \frac{1}{13.28^2} \approx 0.0057$$

3. Precisamos primeiro calcular o valor de δ.

$$\begin{aligned} (1 + \delta) &= 300 \\ \delta &= 1 \end{aligned}$$

   Com isso podemos calcular o valor do limitante.

$$P[Z > 300] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu}$$

$$P[Z > 300] \leq \left(\frac{e^{1}}{4} \right)^{150} \approx 6.8 \times 10^{-26}$$

4.

$$P[Z > 300] = 1 - P[Z \leq 300]$$
$$P[Z > 300] \approx 1.11 \times 10^{-16}$$

5. Pela definição da desigualdade de Chernoff temos:

$$P[Z > (1+\delta)\mu] \leq e^{-\delta^2 \frac{\mu}{3}}$$

Como Z tem distribuição Binomial, $\mu = np$, $p = 0.15$ e $z = (1-\delta)0.15n$,

$$P[Z > z] \leq e^{-\delta^2 \frac{0.15n}{3}}$$
$$P[Z > z] - 1 \leq e^{-\delta^2 \frac{0.15n}{3}} - 1$$
$$1 - P[Z > z] \geq 1 - \frac{1}{e^{\delta^2 \frac{0.15n}{3}}}$$
$$P[Z < z] \geq 1 - \frac{1}{e^{\delta^2 \frac{0.15n}{3}}}$$

Quando $n \to \infty$, $\frac{1}{e^{\delta^2 \frac{0.15n}{3}}} \to 0$ e $P[Z < (1-\delta)0.15n] \to 1$ *with high probability.*

Questão 3: Pesquisa

Você leu no jornal que uma pesquisa com 1500 pessoas indicou que 40% dos entrevistados preferem o candidato A enquanto 60% preferem o candidato B. Estime a margem de erro desta pesquisa usando uma confiança de 90%. O que você precisa assumir? (dica: use a lei dos grandes números).

Solução:

Seja

$$X_n = \begin{cases} 1, & \text{se o n-ésimo voto for no candidato A} \\ 0, & \text{caso contrário} \end{cases}$$

Para $n = 1, 2, 3, \ldots, 1500$. Então os $X_1, X_2, \ldots, X_{1500}$ são iid's e Bernoullis com probabilidade $p = 0.4$. Com isso temos,

$$E[X_n] = \mu = 0.4 \text{ e } Var[X_n] = \sigma^2 = 0.4 \cdot 0.6 = 0.24$$

Pela Lei dos Grandes Números temos

$$P\left[\left|\frac{X_1 + X_2 + \cdots + X_{1500}}{1500} - \mu\right| < \epsilon\right] \geq 1 - \frac{\sigma^2}{\epsilon^2 n}$$

e

$$\beta = 1 - \frac{\sigma^2}{\epsilon^2 n}$$

Para uma confiança de 90%

$$\begin{aligned}\beta &= 1 - \frac{\sigma^2}{\epsilon^2 n}\\ \epsilon &= \sqrt{\frac{0.24}{1500 \cdot 0.1}}\\ \epsilon &= 0.04\end{aligned}$$

Para isso foi preciso assumir que as X_n eram iid's e sua probabilidade.

Questão 4: Moedas

Você tem duas moedas: uma honesta e outra enviesada que produz cara com probabilidade $\frac{3}{4}$. Uma das duas moedas é escolhida aleatoriamente e lançada n vezes. Seja S_n o número de caras que foram observadas nas n jogadas. Responda às perguntas abaixo:

1. A lei dos grandes números pode ser aplicada para prever a fração de caras que será observada?
2. Podemos determinar qual moeda foi escolhida, depois da mesma ser lançada um número n grande?
3. Determine o valor de n tal que tenhamos 95% de chance de acertar qual moeda foi escolhida.

Solução:

1. Sim, temos uma sequência de variáveis iid's, μ e σ^2
2. Podemos determinar a moeda com alguma confiança β escolhendo um epsilon pequeno e lançando a moeda um número n suficiente para que a média amostral se aproxime do valor esperado que é 0.5 e 0.75 para as moedas honesta e enviesada, respectivamente.
3. Assumindo $\epsilon = 0.01$ e testando a moeda honesta, pela Lei dos Grandes Números temos

$$P\left[\left|\frac{X_1 + X_2 + \cdots + X_n}{1500} - 0.5\right| < 0.01\right] \geq 1 - \frac{0.25}{0.0001n}$$

Para uma confiança de 95%

$$\begin{aligned}\beta &= 1 - \frac{\sigma^2}{\epsilon^2 n}\\ n &= \frac{0.25}{0.0001 \cdot 0.05}\\ n &= 50000\end{aligned}$$

Isso quer dizer que se depois de 50000 lançamentos a fração de caras obtidas não estiver no intervalo $[0.49, 0.51]$, a moeda escolhida foi a moeda enviesada com 95% de confiança, se positivo, a moeda escolhida foi a moeda honesta com 95% de confiança.

Questão 5: Sanduíches

Você convidou 64 pessoas para uma festa e agora precisa preparar sanduíches para os convidados. Você acredita que cada convidado irá comer $0, 1$ ou 2 sanduíches com probabilidades $\frac{1}{4}, \frac{1}{2}$ e $\frac{1}{4}$, respectivamente. Assuma que o número de sanduíches que cada convidado irá comer é independente de qualquer outro convidado. Quantos sanduíches você deve preparar para ter uma confiança de 95% de que não vai faltar sanduíches para os convidados?

Solução:

Seja C_i a variável do número de sanduíches que o convidade i vai comer, $C_i = 0, 1$ ou 2. E $C_1, C_2, \ldots, C_{64}$ são iid's. Para aplicar a Lei dos Grandes Números precisamos calcular μ e σ^2.

$$E[C_i] = \mu = \sum_{j=0}^{2} j \cdot p_j = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

$$E[C_i^2] = \sum_{j=0}^{2} j^2 \cdot p_j = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} = 1.5$$

$$\sigma^2 = E[C_i^2] - \mu^2 = 1.5 - 1^2 = 0.5$$

Temos $n = 64$ e precisamos calcular o valor de epsilon para um grau de confiança de 95%

$$P\left[\left|\frac{C_1 + C_2 + \cdots + C_{36}}{36} - 1\right| < \epsilon\right] \geq 1 - \frac{0.5}{\epsilon^2 64}$$

$$\beta = 1 - \frac{\sigma^2}{\epsilon^2 n}$$

$$\epsilon = \sqrt{\frac{0.5}{64 \cdot 0.05}}$$

$$\epsilon = 0.40$$

Ou seja, com 95% de confiança o número de sanduíches que uma pessoa irá comer está no intervalo $[0.60, 1.40]$. Para garantir que não faltará nenhum sanduíche para os 64 convidados, multiplicamos o limite superior do intervalo pelo número de pessoas, obtendo $1.40 \times 64 = 89.6$. Portanto serão necessários 90 sanduíches para que não falte nenhum sanduíche com 95% de confiança.

Questão 6: Vértices Isolados

Considere o modelo de grafo aleatório de Erdös-Rényi (também conhecido por $G(n, p)$), onde cada possível aresta de um grafo rotulado com n vértices ocorre com probabilidade p, independentemente. Responda às perguntas abaixo:

1. Determine a distribuição do grau do vértice 1 (em função de n e p)?

2. Determine a probabilidade do vértice 1 não ter arestas incidentes, ou seja, estar isolado.
3. Determine o valor esperado do número de vértices isolados no grafo (dica: use v.a. indicadora).
4. Mostre que se $p = (1+\epsilon)logn/n$, para qualquer $\epsilon > 0$, o modelo $G(n,p)$ não possui vértices isolados, whp. (dica: use método do primeiro momento)

Solução:

1. Seja A_{ij} uma variável aleatória tal que,

$$A_{ij} = \begin{cases} 1, & \text{se a aresta } j \text{ está presente no vértice } i \\ 0, & \text{caso contrário} \end{cases}$$

   O grau do vértice 1 G_{n1} é

$$G_{n1} = \sum_{j=1}^{n-1} A_{1j}$$

   e G_{n1} tem distribuição Binomial com probabilidade p.

2.

$$\begin{aligned} P[G_{n1} = 0] &= \binom{n-1}{0} p^0 (1-p)^{n-1} \\ &= (1-p)^{n-1} \end{aligned}$$

3. Seja

$$I_i = \begin{cases} 1, & \text{se o vértice } i \text{ é isolado} \\ 0, & \text{caso contrário} \end{cases}$$

   A variável $T_n = \sum_i I_i$ representa o total de vértices isolados no grafo.

$$\begin{aligned} E[T_n] &= E[\sum_i^n I_i] \\ &= \sum_i^n E[I_i] \\ &= \sum_i^n (1-p)^{n-1} \\ &= n(1-p)^{n-1} \end{aligned}$$

4. Queremos $\lim_{n\to\infty} P[T_n > 0] = 0$. Pelo método do primeiro momento, se $\lim_{n\to\infty} E[T_n] = 0$ a probabilidade também vai pra zero.

$$\begin{aligned}
\lim_{n\to\infty} E[T_n] &= \lim_{n\to\infty} n(1-p)^{n-1} \\
&= \lim_{n\to\infty} n\left(1-\left((1+\epsilon)\frac{\log n}{n}\right)\right)^{n-1} \\
&= \lim_{n\to\infty} e^{\log\left(n\left(1-\left((1+\epsilon)\frac{\log n}{n}\right)\right)^{n-1}\right)} \\
&= \lim_{n\to\infty} e^{(n-1)\log\left(n\left(1-\left((1+\epsilon)\frac{\log n}{n}\right)\right)\right)} \\
&= \lim_{n\to\infty} e^{(n-1)\log(n-((1+\epsilon)\log n))} \\
&= e^{\lim_{n\to\infty}(n-1)\log(n-((1+\epsilon)\log n))} \\
&= e^{\lim_{n\to\infty}\frac{\log(n-((1+\epsilon)\log n))}{\frac{1}{(n-1)}}}
\end{aligned}$$

Aplicando l'Hôpital ao limite

$$\begin{aligned}
\lim_{n\to\infty}\frac{\log\left(n-((1+\epsilon)\log n)\right)}{\frac{1}{(n-1)}} &= \lim_{n\to\infty} -\frac{\frac{n-1+\epsilon}{n^2-(1+\epsilon)n\log n}}{\frac{1}{(n-1)^2}} \\
&= \lim_{n\to\infty} -\frac{(n-1-\epsilon)(n-1)^2}{n^2-(1+\epsilon)n\log n}
\end{aligned}$$

Como o numerador tem um fator da ordem de n^3 e o denomidador da ordem de n^2, o

$$\lim_{n\to\infty}\frac{\log\left(n-((1+\epsilon)\log n)\right)}{\frac{1}{(n-1)}} = -\infty$$

Então,

$$\lim_{n\to\infty} E[T_n] = e^{-\infty} = 0$$

Portanto provamos que,

$$\lim_{n\to\infty} P[T_n > 0] = 0$$

ou

$$\lim_{n\to\infty} P[T_n = 0] = 1 \text{ ,whp}$$