ESTATÍSTICA COMPUTACIONAL E SIMULAÇÃO

Folha prática 2: Métodos de Monte Carlo em Inferência Estatística e Métodos de Reamostragem - bootstrap e jackniffe
Ano letivo 2021/22

1. Determine estimativas de $p = P(X > 2)$, com $X \sim C(0,1)$, i.e.,

$$f(x) = \frac{1}{\pi(1+x^2)}, -\infty < x < \infty,$$

usando

(a) a distribuição de Cauchy estandard

(b) a distribuição Uniforme definida em $(0,2)$

(c) Sabe-se que $p = 0.14758$. Compare os resultados anteriormente obtidos e as respetivas variâncias

2. Pretende-se calcular o valor do integral

$$I = \int_0^\infty log(x)x^5e^{-3x}dx.$$

Usando o método de integração de Monte Carlo, obtenha estimativas fazendo simulações de

(a) distribuição $X \sim Exp(3)$

(b) distribuição $X \sim Ga(6,3)$

(c) use ainda o método das réplicas pra estimar I com $X \sim Ga(6,3)$

(d) Sabe-se que $I = 0.1000009$. Compare os resultados anteriormente obtidos com o exato e com o valor estimado para o integral (e respetivo erro) no caso de usar o método das réplicas.

3. Pretende-se estimar o seguinte integral

$$\int_0^1 \frac{e^{-x}}{1+x^2}dx$$

usando a técnica da amostragem de importância. Considere as seguintes funções como candidatas a funções de importância:

$$f_1(x) = 1, 0 < x < 1, \quad f_2(x) = e^{-x}, 0 < x < \infty, \quad f_3(x) = \frac{1}{\pi(1+x^2)}, -\infty < x < \infty,$$

$$f_4(x) = \frac{e^{-x}}{1-e^{-1}}, 0 < x < 1, \quad f_5(x) = 4\frac{1}{\pi(1+x^2)}, 0 < x < 1.$$

Determine os erros associados e analise quais as que mais reduzem a variância da estimativa do integral.

4. Seja $X \sim N(1, \sigma^2)$. Sabe-se

$$\left]\frac{(n-1)S_c^2}{\chi_{1-\alpha/2,n-1}}, \frac{(n-1)S_c^2}{\chi_{\alpha/2,n-1}}\right[$$

é um intervalo de confiança para σ^2 a $100(1-\alpha)\%$. Usando o método de Monte Carlo determine um intervalo de confiança para σ^2.

5. Considere uma série temporal definida por $X_t = aX_{t-1} + Y_t$, onde $Y \sim N(0, \sigma^2)$. O estimador dos mínimos quadrados de a é dado por

$$T = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=2}^{n} X_{t-1}^2}.$$

Supondo o modelo com parâmetros $a = 0.3$ e $\sigma^2 = 1$, uma amostra com 100 observações e usando o método das réplicas (use 1000)

   (a) determine uma estimativa do parâmetro a;

   (b) calcule o desvio padrão, erro quadrático médio e viés associado à estimativa obtida; obtenha o boxplot da amostra dos enviezamentos.

   (c) determine um intervalo de confiança para a.

6. Suponha que se pretende

   (a) testar $H_0 : \sigma^2 = 4$ vs $H_1 : \sigma^2 > 4$, ao nível de significância $\alpha = 0.05$. Usando 100000 réplicas, determine estimativas do nível de significância do teste, considerando amostras de dimensão 100 de populações normais de valor médio unitário, tomando $\sigma^2 = 4$, $\sigma^2 = 10$ e $\sigma^2 = 1$. Comente os resultados.

   (b) determinar um intervalo de confiança para σ^2, considerando o método de Monte Carlo e usando dados provenientes de $N(1, 4)$.

7. Os dados que se encontram no ficheiro "tabaco.txt"dizem respeito a um estudo feito sobre a relação entre tabaco e cancro de pulmão. Foram estudados 25 grupos ocupacionais. A variável explicativa x é o número de cigarros fumados por dia por indivíduos do sexo masculino em cada grupo ocupacional. A variável resposta y é a taxa de mortalidade estandardizada devida a mortes por cancro do pulmão.

   (a) Represente graficamente os dados e obtenha o coeficiente de correlação empírico.

   (b) Gere 1000 amostras bootstrap dos dados, obtenha o respectivo coeficiente de correlação e calcule uma estimativa do erro padrão do coeficiente de correlação.

8. Considere os dados referentes a duração de erupções da nascente termal (geyser) Old Faithful - disponíveis na livraria MASS do R como "faithful$eruptions". Pretende-se determinar um intervalo de confiança para o valor médio da população subjacente aos dados. Considerando o estimador $\bar{X}$, determine esse intervalo de confiança usando o método de reamostragem de bootstrap.

9. Para uma amostra $X_1, \cdots, X_n$ iid a X, considere o estimador $\hat{\theta}_n = \bar{X}$ para o valor médio de X. Então sabemos que $se(\hat{\theta_n}) = \sqrt{Var(X)/n}$. Mostre que

   (a) $\hat{\theta}^{(j)} = \frac{n\bar{X} - X_j}{n-1}$.

   (b) $\bar{\hat{\theta}^{(\cdot)}} = \bar{X}$.

   (c) $\hat{se}_{jack}(\hat{\theta}_n) = \sqrt{S_n^2/n}$, com $S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2$.

10. Resolva o exercício 7 usando o método de reamostragem de Jackknife para estimar a correlação e calcule uma estimativa do seu viés.