LISTA DE EXERCÍCIOS– DISCIPLINA DE ANÁLISE DE DADOS

Sandro Dias Pinto Vitenti

Esses exercícios se referem aos capítulos 2, 3 e parte do 5 do livro Statistics, A Guide to the Use of Statistical Methods in the Physical Sciences - R.J. Barlow. Os exercícios devem ser feitos com programação em qualquer linguagem desejada.

---

1. Utilizando a distribuição uniforme, a distribuição normal e o sampler desenvolvido nas primeiras aulas,

   (a) gere amostras de N pontos $\{x_i\}$ das seguintes distribuições:

$$\{x_i\} \sim \mathrm{Unif}(0,1), \tag{1}$$

$$\{x_i\} \sim \mathrm{Norm}(0,1), \tag{2}$$

   e

$$\{x_i\} \sim \left[\frac{e^{-\frac{(x-x_0)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} - \frac{e^{-\frac{(x+x_0)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\right]\frac{1}{\mathrm{Erf}\left(\frac{x_0}{\sqrt{2}\sigma}\right)}, \quad x \in (0,\infty) \text{ e } x_0 > 0, \tag{3}$$

   tal que a primeira é a distribuição uniforme no intervalo $[0,1]$, a segunda é a distribuição normal com média 0 e variância 1 e Erf é a função erro (escolher um desvio padrão para a última distribuição).

   (b) Dados os n pontos de cada distribuição, faça uma binagem nos intervalo de $(0,1)$, $(-3,3)$, $(x_0-3\sigma, x_o+3\sigma)$ respectivamente para cada distribuição (cuidado para o intervalo não ultrapassar o domínio das distribuições), i.e, divida o intervalo em intervalos menores disjuntos (x_i, x_{i+1}) e veja quantos pontos n_i existem em cada subintervalo. Calcule o valor esperado de pontos em cada subintervalo usando:

$$\bar{n}_i = n\int_{x_i}^{x_{i+1}} \mathrm{P}(x)\mathrm{d}x, \tag{4}$$

   e calcule a diferença relativa com o valor real encontrado n_i. Faça esse procedimento para $n = 10, 100, 1000, 10000$.

(c) Plote os valores encontrados para diferença relativa $(n_i/\bar{n}_i - 1)$ de cada subintervalo de cada distribuição. Plote também as diferenças relativas como função do número total N.

2. Para a distribuição Gaussiana na Eq. (2), calcule

$$\bar{x} = \sum_{i=1}^{N} \frac{x_i}{N}, \quad \text{e} \quad \bar{V} = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N}, \tag{5}$$

para $N = 10$. Repita o procedimento $t = 1000$ vezes. Para cada vez, calcule a média dos valores encontrados, i.e, na segunda vez, some o primeiro e o segundo resultado e divida por 2. Na terceira vez, some os três resultados e calcule a média, e assim por diante até a milésima vez. Plote esses dados em função do numero de cálculos t. A variância está convergindo para o valor esperado?

3. Para a distribuição na Eq. (3), calcule as relações dadas pelas equações 2.13 e 2.15 do livro *Statistics, A Guide to the Use of Statiscal Methods in the Physical Sciences - R.J. Barlow* para diferentes valores de N. Veja para quais valores de N o resultado se estabiliza em até duas casas decimais.

4. Usando as técnicas usadas nos exercícios anteriores

(a) gere um conjunto de dados $\{x_i, y_i\}$ tal que x_i é distribuído pela Eq. (1) e y_i pela Eq. (2). Calcule a covariância $\mathrm{Cov}(x, y)$, a correlação $\mathrm{Cor}(x, y)$ e analise qual valor de N é necessário para que a covariância chegue perto de 0 com uma precisão de duas casas decimais.

(b) Repita os passos da letra (a) para os conjuntos $\{x_i, y_i + x_i\}$ e $\{x_i, y_i - x_i\}$.

(c) Faça um scatter plot para os 3 casos e compare com os valores de correlação obtidos anteriormente.

5. Gere n pontos da distribuição binomial

$$P(r) = p^r (1-p)^{(n-r)} \binom{n}{r} \tag{6}$$

para valores de $n = 10, 100, 1000, 10000$, de forma que o produto $\lambda = np$ seja constante (atenção: $p < 1$). Faça o mesmo para a distribuição de Poisson

$$P(r) = e^{-\lambda} \frac{\lambda^r}{r!}. \tag{7}$$

Plote os gráficos das distribuições e analise para qual valor de n a distribuição binomial pode ser aproximada por uma distribuição de Poisson.

6. Para a distribuição de Poisson em (7), gere dados com valores de $\lambda = 1, 5, 20, 100, 1000$ e compare os resultados com a distribuição Gaussiana com média λ e desvio padrão $\sqrt{\lambda}$.

7. Dada a Gaussiana multivariada

$$P(x,y) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{1}{2}\left(\frac{y-\mu_y}{2\sigma_y}\right)^2}}{2\pi\sigma_x\sigma_y}, \tag{8}$$

vamos renomear o expoente como

$$\chi^2 = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2. \tag{9}$$

Dado intervalos Ω_n definidos por $\chi^2 \leq n^2$, $n = 1, 2, 3$, calcule

$$P(n\sigma) \equiv \int_{\Omega_n} P(x,y)\mathrm{d}x\mathrm{d}y. \tag{10}$$

8. Repita o exercício 7 para a Eq. 3.24 do livro tal que χ^2 é definida pelo novo expoente, i.e., $\chi^2 = -2\ln(n_0 P)$ onde n_0 é a constante de normalização da distribuição $P = e^{-\frac{1}{2}\chi^2}/n_0$.

9. Dados os 8 estimadores da seção 5.1 do livro,

   (a) calcule as diferentes estimativas da média para as 3 distribuições do exercício 1.

   (b) Cheque a consistência, bias e eficiência para cada um deles numericamente.

10. Dada a distribuição uniforme em (1),

    (a) Gere conjuntos de 100 pontos para L variáveis aleatórias uniformes, isto é, $\{x_i^L\}_{N=100}$ para $L = 2, 8, 16, 64, 1024$. Gerados os pontos, defina uma nova variável como a soma das variáveis, i.e.,

    $$z_i \equiv \sum_{j=1}^{L} \frac{x_i^j}{L}. \tag{11}$$

    Faça a binagem de $\{z_i\}$ e plote um histograma para cada valor de L, plote também uma distribuição Gaussiana usando $\bar{z}$ e $V(z)$.

    (b) Repita a letra (a) para variáveis Gaussianas unidimensionais.

    (c) Repita a letra (a) para Gaussiana multivariada dada em (8).