

2ª Prova - FECD A - 27/05/2024

30 pontos - Renato Assunção

1. **5 PONTOS** Consider two discrete random variables X and Y with the following joint probability mass function (pmf):

$$P(X = x, Y = y) = \begin{cases} 0.1 & \text{if } (x, y) = (0, 0) \\ 0.2 & \text{if } (x, y) = (0, 1) \\ 0.1 & \text{if } (x, y) = (1, 0) \\ 0.4 & \text{if } (x, y) = (1, 1) \\ 0.2 & \text{if } (x, y) = (1, 2) \\ 0 & \text{otherwise} \end{cases}$$

- (a) Compute the marginal pmf of X , $P(X = x)$, for $x = 0, 1$.
- (b) Compute the marginal pmf of Y , $P(Y = y)$, for $y = 0, 1, 2$.
- (c) Compute the conditional pmf $P(Y = y | X = 1)$ for $y = 0, 1, 2$.
- (d) Compute the conditional pmf $P(X = x | Y = 1)$ for $x = 0, 1$.
- (e) Are X and Y independent random variables? Justify your answer.

Solution:

- (a) The marginal pmf of X is computed by summing the joint pmf over all values of Y :

$$P(X = 0) = P(0, 0) + P(0, 1) = 0.1 + 0.2 = 0.3$$

$$P(X = 1) = P(1, 0) + P(1, 1) + P(1, 2) = 0.1 + 0.4 + 0.2 = 0.7$$

- (b) The marginal pmf of Y :

$$P(Y = 0) = P(0, 0) + P(1, 0) = 0.1 + 0.1 = 0.2$$

$$P(Y = 1) = P(0, 1) + P(1, 1) = 0.2 + 0.4 = 0.6$$

$$P(Y = 2) = P(1, 2) = 0.2$$

- (c) The conditional pmf $P(Y = y | X = 1)$ is computed using:

$$P(Y = y | X = 1) = \frac{P(X = 1, Y = y)}{P(X = 1)}$$

$$P(Y = 0 | X = 1) = \frac{0.1}{0.7} \approx 0.143$$

$$P(Y = 1 | X = 1) = \frac{0.4}{0.7} \approx 0.571$$

$$P(Y = 2 | X = 1) = \frac{0.2}{0.7} \approx 0.286$$

- (d) The conditional pmf $P(X = x | Y = 1)$ is computed using:

$$P(X = x | Y = 1) = \frac{P(X = x, Y = 1)}{P(Y = 1)}$$

$$P(X = 0 | Y = 1) = \frac{0.2}{0.6} = \frac{1}{3}$$

$$P(X = 1 | Y = 1) = \frac{0.4}{0.6} = \frac{2}{3}$$

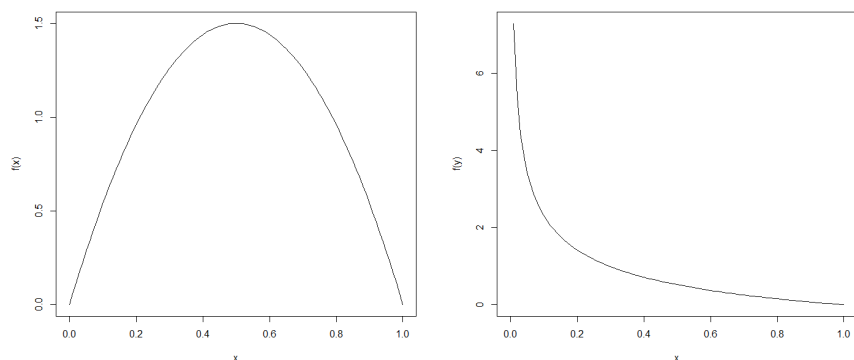


Figure 1: Left: Density $f_X(x) = 6(x - x^2)$. Density of Y : $f_Y(y) = 2(1/\sqrt[3]{y} - 1)$.

- (e) To check if X and Y are independent, verify if $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ for *all* pairs (x, y) . An equivalent alternative is to show that $P(X = x, Y = y) \neq P(X = x) \cdot P(Y = y)$ for one single pair of x and y . Since $P(X = 0, Y = 0) = 0.1 \neq 0.3 \times 0.2 = P(X = 0) \cdot P(Y = 0)$, X and Y are not independent. Indeed, we could have used any pair (x, y) because $P(X = x, Y = y) \neq P(X = x) \cdot P(Y = y)$, for all pairs (x, y) .

2. **5 PONTOS** X is a random variable with support in the interval $[0, 1]$ and density function given by $f_X(x) = 6(x - x^2)$ for $x \in [0, 1]$. The plot of this density function is on the left-hand side of Figure 1. A random value of X is taken and then a random cube is created with X as its edge. Let $Y = X^3$ be the volume of this random cube.

- a) **5%:** Which of the following four probabilities is the highest one: $\mathbb{P}(X \in [0.0, 0.2])$, $\mathbb{P}(X \in [0.8, 1.0])$, $\mathbb{P}(X \in [0.2, 0.4])$, or $\mathbb{P}(X \in [0.4, 0.6])$?
- b) **5%:** Give a rough estimate of the probability you selected in the previous item. It is not necessary to be a precise estimate.
- c) **80%:** Find the probability density of Y given by $f_Y(y) = 2\left(\frac{1}{\sqrt[3]{y}} - 1\right)$ with graph shown on the right-hand side of Figure 1.
- d) **10% :** Show that $\mathbb{E}(Y) = 1/5$.

Solution:

- a) **5%:** As the intervals have all the same length, it is easy to compare the areas under the curve for each interval. It is clear that $[0.4, 0.6]$ has the largest area and therefore $\mathbb{P}(X \in [0.4, 0.6])$ is the correct answer.
- b) **5%:** The exact answer is 0.296, by integrating the density on the interval $(0.4, 0.6)$. However, I was not expecting the exact answer. An answer based on a rough visual estimation of the proportion of area is enough. Any answer between 0.25 and 0.40 is considered correct in this item.
- c) **80%:** We have $X = \sqrt[3]{Y}$ and $dx/dy = 1/(3y^{2/3})$. Therefore, $f_Y(y) = f_X(\sqrt[3]{y}|dx/dy|) = 2(1/(\sqrt[3]{y}) - 1)$.
- d) **10% :**

$$\mathbb{E}(Y) = \int_0^1 y f_Y(y) dy = 2 \int_0^1 (y^{2/3} - y) dy = 1/5.$$

3. **5 PONTOS** O gráfico à esquerda na Figura 2 exibe uma amostra do vetor aleatório (X, Y) com certa densidade $f(x, y)$. Com base neste gráfico, identifique a opção correta:

- $\mathbb{E}(Y|X = 8) \approx ??$: (i) 8 (ii) -5 (iii) 33 (iv) 15 (v) 0
- Suponha que sabemos que $X < 5$. Ache $\mathbb{E}(X|Y = 10 \text{ e } X < 5) \approx ??$: (i) 0 (ii) 1 (iii) 8 (iv) -10 (v) 4
- $\sigma(Y|X = 0) = \sqrt{\mathbb{V}(Y|X = 0)}$ é aproximadamente igual a: (i) 1.5 (ii) 0 (iii) 0.1 (iv) 10.0 (v) -1.0
- $\sigma(Y|X = 2) = \sqrt{\mathbb{V}(Y|X = 2)}$ é aproximadamente igual a: (i) 1.5 (ii) 0 (iii) 0.1 (iv) 10.0 (v) -1.0
- $\sigma(Y|X = x) = \sqrt{\mathbb{V}(Y|X = x)}$ é uma função de x . Ela é:
 - (i) crescente em x ;

- (ii) constante com respeito a x ;
- (iii) decrescente em x ;
- (iv) parabólica em x .

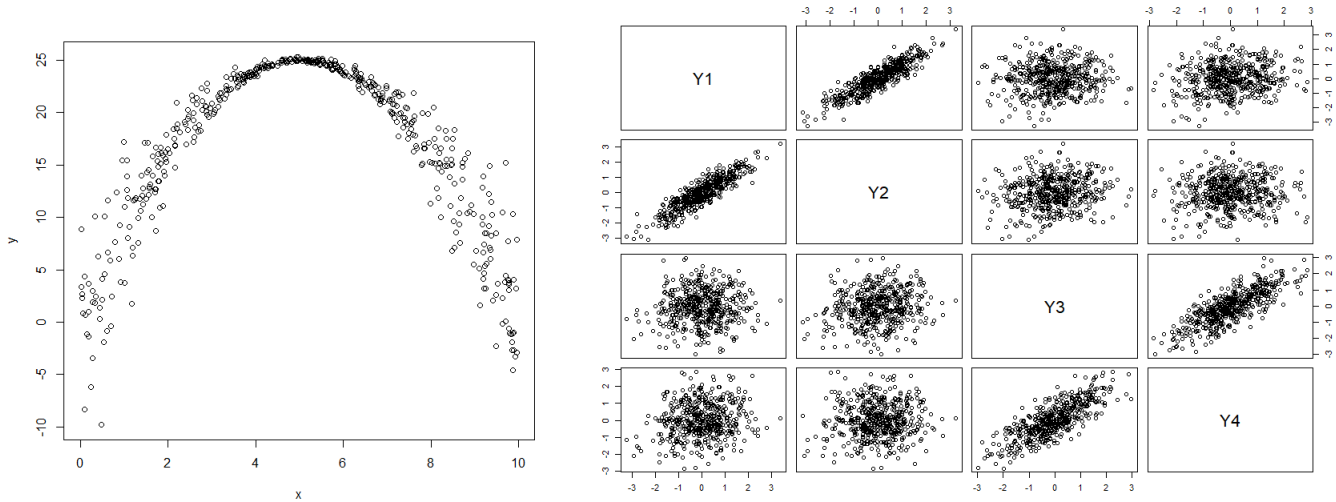


Figura 2: Left: Amostra de um vetor aleatório (X, Y) . Right: Scatterplot of sample of random vetor $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$.

Solution:

- $\mathbb{E}(Y|X = 8) \approx 15$.
- $\mathbb{E}(X|Y = 10 \text{ e } X < 5) \approx 1.5$.
- ANULADA (todos ganham 1 ponto). Deveria ser $\sigma(Y|X = 5) = \sqrt{\mathbb{V}(Y|X = 5)}$. Dois de vocês me avisaram!! $\sigma(Y|X = 0) \approx 3.5$. Nenhuma das opções é aproximadamente igual a este valor.
- $\sigma(Y|X = 2) = \sqrt{\mathbb{V}(Y|X = 2)} \approx 1.5$
- Claramente, o desvio-padrão condicional $\sigma(Y|X = x)$ decresce quando vamos de $x = 0$ até o meio do suporte ($x = 5$). A partir de $x = 5$, o desvio-padrão condicional cresce. Assim, dentre as opções, o correto é “parabólica em x ”.

4. **5 PONTOS** Temos interesse em gerar uma amostra pelo método Monte Carlo de uma v.a. X que possui densidade de probabilidade

$$f(x) = \begin{cases} x/50, & \text{se } x \in (0, 10) \\ 0, & \text{se } x < 0 \text{ ou } x > 10 \end{cases}$$

Assumindo que você consegue gerar variáveis aleatórias com distribuição $U(0, 1)$, escreva um pseudo-código para gerar uma amostra de tamanho B usando:

- o método da acumulada inversa de Stan Ulam.
- o método de aceitação e rejeição de von Neumann. Veja que você deve usar uma distribuição com densidade $g(x)$ que tenha um suporte $\mathcal{S}_g$ que **contenha** o suporte $\mathcal{S}_f$ de $f(x)$ (que é o intervalo $(0, 10)$). Os suportes $\mathcal{S}_g$ e $\mathcal{S}_f$ não precisam ser idênticos mas apenas $\mathcal{S}_g \subset \mathcal{S}_f$.

Solution:

- o método da acumulada inversa de Stan Ulam: para $x \in (0, 10)$, a distribuição acumulada é dada por

$$\mathbb{F}(x) = \int_0^x \frac{t}{50} dt = \frac{x^2}{100}$$

e portanto a sua inversa $\mathbb{F}^{-1}$ é encontrada fazendo $u = \mathbb{F}(x)$ e invertendo $x = 10\sqrt{u} = \mathbb{F}^{-1}(u)$. Assim, para o método da acumulada inversa de Stan Ulam, o comando

```
x = ( 10 * sqrt( runif(B) ) )
```

gera o vetor x com a densidade desejada.

- Para o método de aceitação-rejeição de von Neumann, vamos usar a densidade de uma uniforme em $(0, 10)$. Para obter uma amostra de $U(0, 10)$, basta gerar da $U(0, 1)$ e multiplicar os valores por 10. A densidade $g(x)$ da $U(0, 10)$ é constante e igual a $1/10$ para $x \in (0, 10)$. A densidade alvo possui o máximo em $x = 10$ e igual a $f(10) = 10/50 = 0.2$. Assim, podemos tomar $M = 2$ para garantir que $f(x)/(Mg(x)) \leq 1$ para todo $x \in (0, 10)$. Como $M = 2$, para gerar B valores devemos simular aproximadamente $M * B = 2 * B$ valores da densidade $g(x)$.

```
n = 0
xf = 0
while( n < B){
  xg = 10 * runif(1)
  prob = (xg/50)/(2*(1/10))
  aceita = rbinom(1, 1, prob)
  if(aceita == 1){
    xf = c(xf, xg)
    n = n+1
  }
}
```

-
5. **5 PONTOS** O gráfico à direita na Figura 2 exibe os gráficos de dispersão de uma amostra do vetor aleatório $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$. Se as variâncias de todas as quatro variáveis é igual a 1 e sabendo que a covariância entre as variáveis Y_i e Y_j é $\text{Cov}(Y_i, Y_j) = \rho_{ij}\sigma_i\sigma_j$, obtenha valores aproximados para os elementos $(2, 1)$, $(2, 3)$, e $(3, 4)$ da matriz 4×4 de covariância do vetor $\mathbf{Y}$.

Solution:

- $(2, 1) : \text{Cov}(Y_1, Y_2) \approx 1 * 1 * 0.9 = 0.9$. Aceitamos qualquer valor para $\rho_{ij} > 0.75$.
- $(2, 3) : \text{Cov}(Y_2, Y_3) \approx 1 * 1 * 0.1 = 0.1$. Aceitamos qualquer valor para $0 \leq \rho_{ij} < 0.25$.
- $(3, 4) : \text{Cov}(Y_3, Y_4) \approx 1 * 1 * 0.8 = 0.8$. Aceitamos qualquer valor para $\rho_{ij} > 0.75$. Para ser coerente com os gráficos, devemos ter a covariância aqui menor a resposta fornecida no primeiro item pois claramente as variáveis 1 e 2 tem correlação maior que as variáveis 3 e 4.

-
6. **5 PONTOS** O vetor aleatório (X, Y) tem suporte no quadrado $(0, 1)^2$ e possui densidade de probabilidade $f(x, y) = (6/5)(x^2 + 2xy + y^2)$ neste quadrado.

- Obtenha a densidade marginal $f_X(x)$ e obtenha o seu valor em $x = 1/2$ (isto é, avalie $f_X(1/2)$.)
- Obtenha a densidade condicional $f_{Y|X}(y|x = 1/2)$.

Solution: A densidade marginal $f_X(x)$:

$$\int_0^1 f(x, y) dy = \frac{6}{5} \left(x^2 + x + \frac{1}{3} \right)$$

e $f_X(1/2) = 6/5(1/4 + 1/2 + 1/3) = 13/10$.

A densidade condicional $f_{Y|X}(y|x = 1/2)$:

$$f_{Y|X}(y|x = 1/2) = \frac{f(x, y)}{f_X(1/2)} = \frac{(6/5)(1/4 + y + y^2)}{13/10} = (12/13)(1/4 + y + y^2)$$

OBS: na prova, escrevi $f(x, y) = (6/7)(x^2 + 2xy + y^2)$ quando deveria ser $f(x, y) = (6/5)(x^2 + 2xy + y^2)$. Um pequeno erro de digitação que não compromete a execução do exercício com os valores fornecidos.
