1. A Tabela abaixo mostra a distribuição conjunta do vetor aleatório discreto (X, Y). Obtenha: (a) a distribuição marginal da variável X, (b) a distribuição condicional $(Y|X = 1)$, (c) a distribuição condicional $(X|Y = 0)$ e (d) a distribuição condicional $(X|Y = 1)$.

$y\|x$	$x = 0$	$x = 1$	$x = 2$
$y = 0$	0.2	0.2	0.30
$y = 1$	0.2	0	0.05
$y = 2$	0	0	0.05

Solução: (a) $\mathbb{P}(X = x)$: Temos $\mathbb{P}(X = 0) = 0.2 + 0.2 + 0 = 0.4$, $\mathbb{P}(X = 1) = 0.2 + 0 + 0 = 0.2$, $\mathbb{P}(X = 2) = 0.3 + 0.05 + 0.05 = 0.4$.

(b) $\mathbb{P}(Y = y|X = 1)$: Temos $\mathbb{P}(Y = 0|X = 1) = \mathbb{P}(Y = 0, X = 1)/\mathbb{P}(X = 1) = 0.2/0.2 = 1$. Analogamente, $\mathbb{P}(Y = 1|X = 1) = \mathbb{P}(Y = 1, X = 1)/\mathbb{P}(X = 1) = 0/0.2 = 0$ e $\mathbb{P}(Y = 2|X = 1) = 0/0.2 = 0$.

(c) $\mathbb{P}(X = x|Y = 0)$: Temos $\mathbb{P}(X = 0|Y = 0) = \mathbb{P}(X = 0, Y = 0)/\mathbb{P}(Y = 0) = 0.2/\mathbb{P}(Y = 0)$. Analogamente, $\mathbb{P}(X = 1|Y = 0) = \mathbb{P}(X = 1, Y = 0)/\mathbb{P}(Y = 0) = 0.2/\mathbb{P}(Y = 0)$ e $\mathbb{P}(X = 2|Y = 0) = 0.3/\mathbb{P}(Y = 0)$. Como estas probabilidades devem somar, temos que $1 = (0.2+0.2+0.3)/\mathbb{P}(Y = 0)$, o que implica que $\mathbb{P}(Y = 0) = 0.7$. Concluímos que $\mathbb{P}(X = 0|Y = 0) = 2/7$, $\mathbb{P}(X = 1|Y = 0) = 2/7$ e $\mathbb{P}(X = 2|Y = 0) = 3/7$.

---

2. O gráfico na Figura 1 exibe uma amostra do vetor aletório (X, Y) com certa densidade normal bivariada $f(x, y)$. Com base neste gráfico, identifique a opção correta para o vetor esperado $\boldsymbol{\mu}$ e para a matriz de covariâncias $\boldsymbol{\Sigma}$:

- (a) $\boldsymbol{\mu} = (0, 0)$, (b) $\boldsymbol{\mu} = (10, 10)$, (c) $\boldsymbol{\mu} = (10, 35)$, (d) $\boldsymbol{\mu} = (35, 10)$ (e) $\boldsymbol{\mu} = (35, 35)$
- (a) $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 2.5 \end{bmatrix}$, (b) $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.4 \\ -0.4 & 2.5^2 \end{bmatrix}$, (c) $\boldsymbol{\Sigma} = \begin{bmatrix} 2.5^2 & 0.4 \\ 0.4 & 1 \end{bmatrix}$, (d) $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.9 * 2.5 \\ 0.9 * 2.5 & 2.5^2 \end{bmatrix}$, e (e) $\boldsymbol{\Sigma} = \begin{bmatrix} 0.9 * 2.5 & 1 \\ 2.5^2 & 0.9 * 2.5 \end{bmatrix}$,

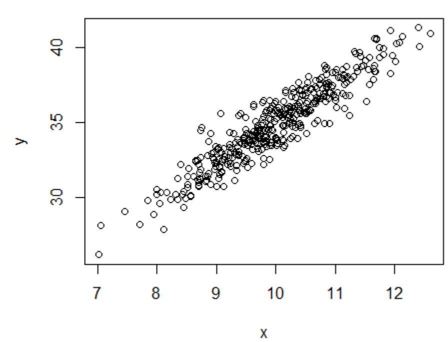

Figura 1: Amostra de um vetor aleatório (X, Y).

Solução: $\boldsymbol{\mu} = (10, 35)$ e $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.9 * 2.5 \\ 0.9 * 2.5 & 2.5^2 \end{bmatrix}$.

---

3. Considerando a densidade normal bivariada $f(x, y)$ que gerou a amostra de dados na Figura 1, diga, sem fazer nenhuma conta e usando apenas a visualização da amostra, qual é aproximadamente a distribuição condicional de $(X|Y = 40)$.

Solução: $(X|Y = 40) \approx N(12, 0.25^2)$.

---

4. Suponha que o vetor aleatório contínuo e positivo $\mathbf{X} = (X_1, X_2)$ possui a densidade $f_1(\mathbf{x}) = 6 \exp(-(3x_1 + 2x_2))$ quando o indivíduo pertence à população 1. Quando ele pertence à população 2, temos $f_2(\mathbf{x}) = \exp(-(x_1 + x_2))$. O custo $c(1|2)$ do erro de classificar erradamente no grupo 1 um indivíduo do grupo 2 é 3 vezes maior que o custo contrário $c(2|1)$ de colocar no grupo 2 alguém do grupo 1. Se o grupo 1 constitui 90% da população total, mostre que a região ótima R_1 de classificação no grupo 1 é dada pelo semi-plano $2x_1 + x_2 \leq \log(18)$.

Solução: A regra de Bayes fornece a região ótima para classificar todos os pontos $\mathbf{x} = (x_1, x_2)$ do plano. Devemos classificar como classe 1 aqueles pontos em R_1: isto é, aqueles pontos $\mathbf{x} = (x_1, x_2)$ tais que

$$\begin{aligned}
\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &\geq \frac{\pi_2}{\pi_1} \frac{c(1| \in \pi_2)}{c(2| \in \pi_1)} \\
\frac{6\exp(-(3x_1 + 2x_2))}{\exp(-(x_1 + x_2))} &\geq \frac{0.1}{0.9}\, 3 \\
e^{-3x_1 - 2x_2 + x_1 + x_2} &\geq 18^{-1} \\
-2x_1 - x_2 &\geq -\log(18) \\
2x_1 + x_2 &\leq \log(18)
\end{aligned}$$

Esta região é o semi-plano determinado pela linha reta $x_2 = \log(18) - 2x_1$. Como a origem $(0,0)$ tem $2 \times 0 + 0 < \log(18) \approx 2.89$, a origem está na região R_1 bem como todos os pontos do plano que estão no mesmo semi-plano que ela.

---

5. Um provedor de serviços de telefonia está analisando a força do sinal de sua rede. Seja (X, Y) as coordenadas da posição de um certo cliente em relação à torre de celular mais próxima (isto é, assuma que a torre está na origem $(0,0)$). Analisando os dados, a companhia descobriu que a distribuição de probabilidade da posição (X, Y) deste cliente segue uma densidade de probabilidade conjunta dada por

$$f(x, y) = \begin{cases} 3(x^2 + y^2)/8, & \text{se } -1 < x < 1 \text{ e } -1 < y < 1 \\ 0, & \text{caso contrário} \end{cases}$$

- Encontre a densidade marginal $f(x)$.
- Encontre a probabilidade $\mathbb{P}(X \in (0.5, 1) \text{ e } Y \in (0.5, 1))$

Solução: Para a densidade marginal $f(x)$, calculamos

$$f_X(x) = \int_{-1}^{1} f(x, y) dx\ dy = \int_{-1}^{1} \frac{3}{8}(x^2 + y^2) dy = \frac{1}{4} + \frac{3}{4} x^2$$

se $x \in (-1, 1)$.

Para a probabilidade $\mathbb{P}(X \in (0.5, 1) \text{ e } Y \in (0.5, 1))$, temos

$$\mathbb{P}((X, Y) \in (0.5, 1) \times (0.5, 1)) = \int_{1/2}^{1} \int_{1/2}^{1} \frac{3}{8}(x^2 + y^2) dx\ dy = \frac{7}{64}$$
