3^a Prova - FECD A - 03/06/2024

1. **4 PONTOS** A mouse is trapped in a room with three exits at the center of a maze.

    - Exit 1 leads directly outside the maze after 3 minutes.
    - Exit 2 leads back to the room after 5 minutes.
    - Exit 3 leads back to the room after 7 minutes.

    Every time the mouse makes a choice, it is equally likely to choose any of the three exits. Upon reentering the room, the process resets, and everything proceeds as at the experiment's start. Considering the probabilities shown in Figure 1, what is the expected time for the mouse to leave the maze?

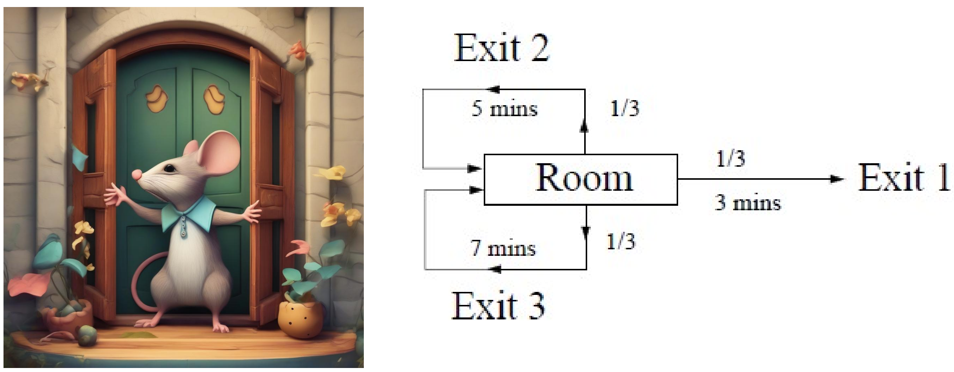

Figura 1: Mouse in a maze. Diagram with probabilities.

---

2. **4 PONTOS** In the previous problem, find the variance of the random time for the mouse to leave the maze.

---

3. **4 PONTOS** In the first problem, let N be the number of trials until the mouse leaves the maze. Find $\mathbb{E}(N)$ and $\mathbb{V}(N)$.

---

4. **2 PONTOS** Your task is to prove the law of iterated expectation:

$$\mathbb{E}(Y) = \mathbb{E}\left[\mathbb{E}(Y|X)\right]$$

    Assume that X and Y are both discrete random variables with possible values $\{1, 2, 3\}$ and $\{5, 7\}$. For this, remember that the random variable $\mathbb{E}(Y|X)$ is discrete with the possible values and associated probabilities given by:

$$\begin{aligned} &\mathbb{E}(Y|X=1) \text{ with probability } \mathbb{P}(X=1)\\ &\mathbb{E}(Y|X=2) \text{ with probability } \mathbb{P}(X=2)\\ &\mathbb{E}(Y|X=3) \text{ with probability } \mathbb{P}(X=3) \end{aligned}$$

    We also know how to obtain each of these expected values. For example,

$$\mathbb{E}(Y|X=2) = \sum_y y * \mathbb{P}(Y=y|X=2) = (5) * \mathbb{P}(Y=5|X=2) + (7) * \mathbb{P}(Y=7|X=2)$$

---

5. **4 PONTOS** A retail store has collected data on its customers to better understand their purchasing behavior. The data consists of four variables for each customer:

- X_1: Amount spent on electronics (in dollars)
- X_2: Amount spent on clothing (in dollars)
- X_3: Amount spent on groceries (in dollars)
- X_4: Amount spent on home goods (in dollars)

Assume the data follows a multivariate normal distribution with expected vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Given the following parameters: - Mean vector: $\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3, \mu_4] = [100, 150, 200, 250]$ - Covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 400 & 150 & 100 & 50 \\ 150 & 300 & 80 & 60 \\ 100 & 80 & 500 & 200 \\ 50 & 60 & 200 & 450 \end{bmatrix}$$

Define the subvectors $\mathbf{Y} = [X_1, X_2]$ and $\mathbf{Z} = [X_3, X_4]$. The conditional mean vector $\mu_{\mathbf{Y}|\mathbf{Z}}$ is given by:

$$\mu_{\mathbf{Y}|\mathbf{Z}} = \mu_{\mathbf{Y}} + \Sigma_{\mathbf{YZ}}\Sigma_{\mathbf{ZZ}}^{-1}(\mathbf{Z} - \mu_{\mathbf{Z}})$$

The conditional covariance matrix $\Sigma_{\mathbf{Y}|\mathbf{Z}}$ is given by:

$$\Sigma_{\mathbf{Y}|\mathbf{Z}} = \Sigma_{\mathbf{YY}} - \Sigma_{\mathbf{YZ}}\Sigma_{\mathbf{ZZ}}^{-1}\Sigma_{\mathbf{ZY}}$$

(a) Obtain the conditional distribution of $\mathbf{Y}$ given $\mathbf{Z} = [z_3, z_4] = [240, 250]$. Mostre explicitamente todas as matrizes envolvidas mas deixe as operações matriciais (multiplicação e inversa) apenas indicadas.

(b) Suponha que o sub-vetor $\mathbf{Z} = [z_3, z_4]$ tenha o seu valor exatamente igual à sua média marginal: $\mathbf{Z} = [z_3, z_4] = [\mu_3, \mu_4] = [200, 250]$. Pode-se afirmar que (V ou F e justifique):

- A média conditional $\mu_{\mathbf{Y}|\mathbf{Z}}$ é igual à média marginal $[\mu_1, \mu_2]$.
- A matriz de covariâcia conditional $\Sigma_{\mathbf{Y}|\mathbf{Z}}$ é igual à matriz de covariância conditional $\Sigma_{\mathbf{YY}}$.

---

6. **4 PONTOS** You are given a dataset X with 10 variables (as columns of a table) and N rows. We estimate the Σ covariance matrix of the dataset, which is a 10×10 symmetric and positive definite matrix. Its eigenvectors $bsv_1, \mathbf{v}_2, \ldots, \mathbf{v}_{10}$ are 10 vectors of dimension 10×1. Any item of the dataset (a row of the dataset seen as a 10×1 column vector) can be written as a linear combination of the 10 eigenvectors: $\mathbf{x} = \sum_k c_k \mathbf{v}_k$.

   - Explain why it is always possible to represent any $\mathbf{x}$ as a linear combination $\sum_k c_k \mathbf{v}_k$ of the eigenvectors. That is, what are the properties of the eigenvectors that allow this representation?
   - Is this representation unique or there is more than one way to represent it as a linear combination $\sum_k c_k \mathbf{v}_k$ of the eigenvectors?
   - The values of c_k in such representation is a function g of $\mathbf{v}_k$ and $\mathbf{x}$. What is this $c_k = g(\mathbf{v}_k, \mathbf{x})$ function?.
   - If we express one the eigenvector, say $\mathbf{v}_3$, as a linear combination of all eigenvectors (including $\mathbf{v}_3$), what will be the linear combination $\sum_k c_k \mathbf{v}_k$?

---

7. **4 PONTOS** Let X be a univariate continuous random variable. Based on X we want to decide if the item belongs to class 1 or class 2. The density of X in each class is given by $f_1(x)$ and $f_2(x)$. The proportion of items coming from class 1 is π. A very simple classification rule is selected: A threshold a is chosen and if $X > a$, classify it as 1. Otherwise, classify it as 2. The misclassification error cost is equal to the two possible errors.

   - Show that the probability of error for this rule is given by

   $$\mathbb{P}(\text{error}) = \pi \int_{-\infty}^{a} f_1(x)dx + (1-\pi)\int_{a}^{\infty} f_2(x)dx \ .$$

   DICA: $\mathbb{P}(A) = \mathbb{P}(A| \in 1)\mathbb{P}(\in 1) + \mathbb{P}(A| \in 2)\mathbb{P}(\in 2)$

   - Taking derivatives, show that to minimize $\mathbb{P}(\text{error})$ the threshold a should satisfy

   $$\pi f_1(a) = (1-\pi) f_2(a) \ .$$

8. **4 PONTOS** Assume that a factor analysis model can represent three correlated random variables X_1, X_2, and X_3 with one single latent factor F. More specifically, assume that

$$X_1 = 0.9F + \epsilon_1$$
$$X_2 = 0.7F + \epsilon_2$$
$$X_3 = 0.5F + \epsilon_3$$

F and the vector $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \epsilon_3)$ are independent. Additionally, $\mathbf{E}(F) = 0$ and $\mathbf{V}(F) = 1$. Also, $\mathbf{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ e $\mathbf{V}(\boldsymbol{\epsilon}) = \text{diag}(0.1, 0.2, 0.1)$.

- Obtain 3×1 expected vector $\mathbb{E}(\mathbf{X})$.
- Obtain the 3×3 covariance matrix $\mathbb{V}(\mathbf{X})$. You can only indicate which operations are necessary.