<div align="center">

Indian Institute of Technology Madras

Department of Data Science and Artificial Intelligence

DA5000: Mathematical Foundations of Data Science

Tutorial III

</div>

Problem

1. If the variance of a Dataset is correctly computed with the formula using $(N-1)$ in the denominator, which of the following option is true? Also, explain the reason for using $(N-1)$ in the denominator.

   (a) Dataset is a sample

   (b) Dataset is a population

   (c) Dataset could be either a sample or a population

2. Consider a continuous random variable X with probability density function $f(x)$ over the interval $[a, b]$. The variance σ^2 of X can be expressed using the following two formulas:

   1. $\sigma^2 = \int_a^b (x - \mu)^2 f(x) dx$
   2. $\sigma^2 = \int_a^b x^2 f(x) dx - \mu^2$

   Explain how the two formulas are equivalent.

3. Let X and Y be continuous random variables with joint probability density function $f_{X,Y}(x, y)$. Prove that:
$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

   , where a and b are constants, and

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

4. Let (X, Y) be continuous random variables with joint PDF:

$$f_{X,Y}(x, y) = \begin{cases} \frac{2}{x^2} & \text{if } 1 < y < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Find the marginal densities $f_X(x)$ and $f_Y(y)$.

   (b) Determine if X and Y are independent.

   (c) Calculate the expected value $E[X \cdot Y]$.

5. Suppose a company is analyzing the relationship between the number of hours a person spends exercising weekly (denoted by E) and their weight (denoted by W). The joint probability density function (PDF) of the number of hours exercised and weight is given by $f_{E,W}(e, w)$.

$$f_{E,W}(e, w) = \begin{cases} c_1 \cdot e \cdot w & \text{if } 0 \le e \le 5 \text{ and } 30 \le w \le 50 \\ c_2 \cdot e \cdot w^2 & \text{if } 0 \le e \le 5 \text{ and } 50 \le w \le 100 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

   (a) Find the values of c_1 and c_2

   (b) Determine the PDF which describes the distribution of exercise hours among the population.

<div align="center">

1

</div>

(c) Calculate the conditional PDF which represents the weight distribution for people who exercise a given number of hours e.

(d) Find the expected number of hours a person in the range 75-80 kgs exercises.

6. Suppose two persons, A and B have a meeting at a given time and each will arrive at the meeting place with a delay between 0 and 2 hours, with all pairs of delays being equally likely. The first to arrive will wait for 20 minutes and leave after that if the other hasnt arrived. What is the probability that they will meet?