Indian Institute of Technology Mandi
भारतीय प्रौद्योगिकी संस्थान मण्डी

IC252-Data Science 2

Assignment– 04

General instructions:

- Utilize Python programming language for implementation.
- Ensure the program is well-documented to enhance comprehension.
- Employ functions and loops for efficient code organization.
- Implement error handling to manage invalid inputs or unexpected scenarios.
- Optimize the code for performance and readability where applicable.

Question 1: Imagine you live in a place with three weather conditions: sunny, rainy or cloudy. You also have three clothing choices: T-shirt, sweater or Jacket.

(a) Create a joint probability table representing the likelihood of experiencing a specific weather condition alongside your clothing choice (Hint: Simulate a large number of samples of weather conditions with clothing choices)

(b) Calculate the marginal probabilities of the weather and the clothing choices based on the joint probability table. Plot these probabilities (two plots one for weather and the other for clothing choice).

(c) Find the conditional probability of each clothing choice given a specific weather condition.

Question 2: A factory produces lightbulbs of two types: incandescent and LED. The probability of an incandescent bulb being defective is 0.1, while for LED bulbs it's 0.05. A box contains 2 incandescent and 3 LED bulbs. You pick two bulbs at random without replacement

(a) Define the random variables. Find and plot the joint distribution.

(b) Calculate the marginal probability distributions.

(c) Plot the PMF of the random variable X (probability mass function, as it deals with discrete values).

(d) Calculate the conditional probability of getting one defective bulb given the first chosen is incandescent.

(e) Are X and Y independent events? Explain your answer using the joint probability distribution. If Yes, would X and Y still be independent events if the bulb were drawn with replacement. Why or why not?

(f) How would the joint probability distribution and the marginal probabilities change if the bulbs were drawn with replacement after each pick?

(g) How would the PMF of X be affected by sampling with replacement? Would it remain the same, or would the probabilities change? Explain why.

Question 3: Two professors teach the same Statistics course. Professor A's exams follow a normal distribution with a mean (μ) of 78 and a standard deviation (σ) of 5. Professor B's exams are also normally distributed, with a mean (μ) of 85 and a standard deviation (σ) of 3. A student randomly picks a professor's course to enroll in (let Professor = "A" or "B"). They are also interested in knowing the difficulty level (Easy, Medium, Hard) assigned to their exam by the professor. The difficulty is assigned independently with the following probabilities:

- Easy: 0.3
- Medium: 0.5
- Hard: 0.2

(a) Find the joint probability distribution $P(X, Y)$. This will represent the probability of getting a specific score (X) along with a particular difficulty level (Y).

(b) Calculate the marginal probability distributions.

(c) Calculate the conditional probability $P(X > 80 | Y = \text{“}Easy\text{”})$ - the probability of getting a score higher than 80 given the exam is Easy.

(d) Plot the probability density functions (PDFs) of both Professor A's and Professor B's exams on the same graph. Clearly label the axes and curves.

(e) How would the joint probability distribution change if the difficulty level was not assigned independently and instead, professors were more likely to assign harder exams to students with higher expected scores (based on professor)?

Question 4: A small radio repair shop repairs two types of radios: AM and FM. The time to repair an AM radio follows a normal distribution with a mean of 1 hour ($\mu_{AM} = 1$) and a standard deviation of 0.5 hours ($\sigma_{AM} = 0.5$). The time to repair an FM radio follows a normal distribution with a mean of 1.5 hours ($\mu_{FM} = 1.5$) and a standard deviation of 0.75 hours ($\sigma_{FM} = 0.75$). Simulate the repair times for 100 AM radios and 100 FM radios using their respective normal distributions. You can use any random number generation library or statistical software to achieve this.

(a) Plot the joint probability distribution of the repair times for AM and FM radios.

(b) What is the probability that the repair time for the FM radio will be less than 1 hour given the AM radio repair takes 2 hours?

(c) Define a new random variable, T, representing the total repair time for both AM and FM radios. $T = RepairTimeAM + RepairTimeFM$. Simulate the total repair time (T) for 100 pairs of AM and FM radios.

(d) Plot the distribution (histogram) of the total repair time (T). Calculate the mean and standard deviation of the total repair time (T) obtained through simulation

(e) Assume a customer arrives at the shop, and there's already one repair in progress (either AM or FM). Let Y be a random variable indicating the remaining repair time (from the ongoing repair). Derive the probability density function (PDF) of Y if Y originated from an AM radio repair (YAM) and another PDF if Y originated from an FM radio repair (YFM). Utilize the change of variable formula for this.

Question 5: Joint Distribution and independence

(a) Generate two independent random variables, X and Y, with uniform distributions over the intervals $[0, 1]$ and $[1, 2]$, respectively and Calculate their joint probability distribution.

(b) Verify independence by checking if $P(X = x \cap Y = y) = P(X = x)P(Y = y)$ for all x and y.

(c) Based on the joint distribution, calculate $P(X > 0.5|Y = 1.5)$ (conditional probability).

(d) Plot the conditional probability distribution of X given $Y = 1.5$.

(e) Define a new random variable $Z = X + Y$ and utilize the change of variable formula to determine the probability density function (pdf) of Z.

(f) Validate the result by generating random samples of Z and comparing the empirical distribution with the theoretical pdf.