Name:
Roll Number:
Department:
Program: BTech / MTech TA / MTech RA / PhD (Tick one)

CS6660: Mathematical Foundations of Data Science

Mid Term Exam 1

Date: 06 October 2024

Instructions:

- **This exam is for a total of 30 MARKS.**
- **You are allowed to keep ONE A4 sheet of written material containing formulae.**
- **Hints are provided for some questions.**
  However, it is NOT mandatory to solve the question using the approach in the hints.
  If you think you have a better approach in mind than the one given in the hint, feel free to present your approach.
- **Show all your working clearly.**
  We want to see your thought process, and possibly provide partial credit for the intermediate logical steps.
- **Plagiarism will NOT be entertained at any length.**
  If you are caught cheating during the exam, your answer script will NOT be evaluated.

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Assume that all random variables appearing in the questions below are defined with respect to $\mathscr{F}$.
Assume that all logarithms appearing in the questions below are natural logarithms, unless explicitly stated otherwise.

Name:
Roll Number:
Department:
Program: BTech / MTech TA / MTech RA / PhD (Tick one)

1. (a) **(4 Marks)**
   Fix $M \in \mathbb{N}$.
   Let $\Omega_1, \dots, \Omega_M$ be a *partition* of Ω, i.e., $\Omega_i \cap \Omega_j = \emptyset$ for all $i \neq j$, and $\bigcup_{i=1}^{M} \Omega_i = \Omega$.
   Let A and B be two events such that
   - A and B are conditionally independent conditioned on Ω_i, for all $i \in \{1, \dots, M\}$, and
   - B is independent of Ω_i for all $i \in \{1, \dots, M\}$.

   Prove that A and B are independent.

   (b) **(1 Mark)**
   A box contains three coins: two regular coins and one fake 2-headed coin.
   A friend of yours picks a coin uniformly at random, tosses it, and tells you that it landed heads.
   What is the probability that your friend chose the two-headed coin?

2. (a) **(2 Marks)**
   Two numbers are drawn independently and uniformly from the unit interval $[0, 1]$.
   The smaller of the two the numbers is known to be less than $1/3$.
   What is the probability that the larger one is greater than $3/4$?

   (b) **(3 Marks)**
   A miner is trapped in a mine containing three doors. The first door leads to a tunnel that takes him to safety after two hours of travel. The second door leads to a tunnel that returns him to the mine after three hours of travel. The third door leads to a tunnel that returns him to his mine after five hours. Assuming that the miner is at all times equally likely to choose any one of the three doors, what is the expected length of time until the miner reaches safety?
   Hint: Let D be a random variable that denotes the door chosen by the miner, and let T denote the time until safety. Compute $\mathbb{E}[T|\{D = 1\}]$, $\mathbb{E}[T|\{D = 2\}]$, and $\mathbb{E}[T|\{D = 3\}]$ separately, and use the law of iterated expectations to compute $\mathbb{E}[T]$.

3. Let X and Y be jointly continuous random variables with the joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 6xy, & 0 \le x \le 1,\ 0 \le y \le \sqrt{x}, \\ 0, & \text{otherwise.} \end{cases}$$

   (a) **(1 Mark)**
   Plot and shade the region of integration in 2-dimensions.

   (b) **(2 Marks)**
   Determine the marginal PDFs of X and Y.

   (c) **(1 Mark)**
   Are X and Y independent? Justify your answer.

   (d) **(3 Marks)**
   Evaluate $\mathbb{E}[X|Y]$.

   (e) **(3 Marks)**
   Evaluate Var$(X|Y = 1/2)$, defined as

$$\text{Var}(X|Y = 1/2) := \mathbb{E}[X^2|\{Y = 1/2\}] - \big(\mathbb{E}[X|\{Y = 1/2\}]\big)^2.$$

   You may leave the final answer in the form α/β, where α and β are co-prime (i.e., do not have any common factor).

Name:
Roll Number:
Department:
Program: BTech / MTech TA / MTech RA / PhD (Tick one)

4. On generating a random sample from the standard normal distribution on a computer.

Those of you who are familiar with Python programming language may be aware of the `numpy.random.normal()` module in Python for generating a random sample from the standard normal distribution.
In this question, we shall understand how this module works behind the scenes.
The basic premise here is that a computer first generates a random sample from Unif$(0, 1)$, and uses the uniform sample to generate the desired sample from standard normal distribution.

Let $U_1, U_2 \overset{\text{i.i.d.}}{\sim}$ Unif$(0, 1)$.

(a) **(2 Marks)**
Determine the PDF of $R = \sqrt{-2\,\log(1 - U_1)}$.

(b) **(1 Mark)**
Let $\Theta = 2\pi U_2$. What is the distribution of Θ?

(c) **(1 Mark)**
Argue that R and Θ are independent.

(d) **(3 Marks)**
Let $X = R\,\cos(\Theta)$ and $Y = R\,\sin(\Theta)$. Show that $X, Y \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.
The module `np.random.normal()` returns only one of the two samples X and Y, and masks the other.

(e) **(1 Mark)**
Determine the variance of $Z = 3X + 4Y$.

(f) **(2 Marks)**
Let $S = X + Y$ and $D = X - Y$.
Further, let W be independent of both X and Y, with

$$\mathbb{P}(\{W = w\}) = \begin{cases} \frac{1}{2}, & w = \pm 1, \\ 0, & \text{otherwise.} \end{cases}$$

Are S and WD jointly Gaussian? Justify.