DS313: Statistical Foundations of Data Science
Assignment No: 2

Course Instructor: Siddhartha Sarma *Date:* 17 Feb 2025
Submission deadline: 6 Mar 2025, 10 AM (Bring hardcopies to the exam hall)

Related topics:

- **Discrete random variables and their distributions**
- **Joint PMF of discrete random variables**
- **Functions of random variables**
- **Conditional distribution**
- **Expectation, variance and moments**

Notations:

- $F_X(x), p_X(x)$ denote the cumulative distribution function (CDF) and probability mass function (PMF) of a random variable X, respectively.
- $p_{XY}(x, y), p_{X|Y}(x|y)$ denote joint and conditional PMFs, respectively.
- $\mathbf{E}[.]$, Cov(., .) denote the expectation and covariance, respectively.
- $Bin(N, p)$ denotes the binomial distribution with parameters N and p.

Instructions & Information:

- For the questions prefixed with [PA], you need to solve them on paper and write code to simulate the underlying experiment as well. You are free to choose the programming language for writing your code. As discussed in the second tutorial session, you should plot the normalised frequencies of the events under consideration as bar plots and compare them to their corresponding probabilities calculated analytically. Please refer to the course moodle page for examples.
- *The programming assignments should be submitted through moodle only. You also need to indicate in moodle quiz corresponding to the Assignment#1 the questions you have submitted in hard copy.*
- You must include a table similar to the one below on the front page of your submission to indicate the page numbers of the answer to the assignment questions.

Question no.	1	2	3	4	5	6	7	8	9	10
Page no.										

---

1. If the CDF of X is given by

$$F_X(x) = \begin{cases} 0, & x < -1 \\ 1/2, & -1 \le x < 1 \\ 4/5, & 1 \le x < 3 \\ 9/10, & 3 \le x < 4 \\ 1, & x \ge 4 \end{cases}$$

   (i) Find the PMF of X; (ii) Find $\Pr(1.5 \le X \le 3.95)$; (iii) Find $\Pr(X \ge 0)$

2. [PA] (i) A biased coin with probability of heads p and tails $1-p$ is tossed until k heads are obtained. Find the distribution of the number of flips. (*This is called the negative binomial distribution.*)

   (ii) If X represents the difference between the number of heads and the number of tails obtained when the coin mention in (i) is tossed n times. Find the PMF of X for a given n.

3. A communication network is shown Fig. 1, where S, D and R represent the source, destination and intermediate routers, respectively. The link capacities in megabits per second (Mbps) are given by $C_1 = C_3 = 5, C_2 = C_5 = 10$ and $C_4 = 8$, and are the same in each direction. Information flow from the source (S) to the destination (D) can be split among multiple paths. For example, if all links are working, then the maximum communication rate is 10 Mbps: 5 Mbps can be routed over links 1 and 2, and 5 Mbps can be routed over links 3 and 5. Let F_i be the event that link i fails. Suppose that F_1, F_2, F_3, F_4 and F_5 are independent and $\Pr(F_i) = 0.1$ for each i. Let X be defined as the maximum rate (in Mbits per second) at which data can be sent from the source node to the destination node. Find the PMF p_X.

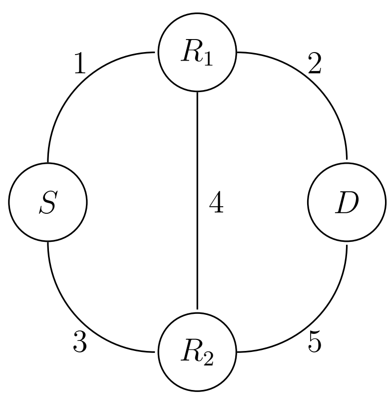

Figure 1: The communication network mentioned in Q. 3.

4. Suppose a discrete random variable X takes on one of the values from the set $\{0, 1, \dots, N\}$. If for some constant c, $\Pr(X = i) = c\Pr(X = i-1), i = 1, 2, \dots, N$, express the PMF and CDF of X in terms of c and N and find $\mathbf{E}[X]$. If another random variable Y is defined as $Y = -\log(p_X(X))$, find $\mathbf{E}[Y]$.

5. [PA]Write algorithms for generating samples from the following distributions using a uniform random number generator that produces samples from the interval $[0, 1]$: (a) Binomial with parameters N and p, (b) Poisson with parameter $\lambda > 0$, (c) Geometric with parameter p. Validate your proposed algorithms by comparing the normalised histograms generated using them with the corresponding PMFs.

6. Let $q_N(k) := \binom{N}{k} p^k (1-p)^{N-k}$, where $0 \leq p \leq 1$; $k, N \in \mathbb{Z}$, and $k \leq N$. Show that (a) $q_N(k+1) > q_N(k)$, if $k < Np-(1-p)$. (b) if $Np-(1-p)$ is an integer, $q_N(k)$ assumes its maximum value for two values of k, namely, $k_0 = Np-(1-p)$ and $k_1 = Np - (1-p) + 1$.

7. Let $X \sim Bin(n, p)$ and $Y \sim Bin(n, 1-p)$ represent binomial random variables. Prove that $\Pr(X \leq i) = 1 - \Pr(Y \leq n - i - 1)$

8. [PA]Two fair dice are rolled. Find the joint probability mass function of X and Y when (a) X is the largest value obtained on any die and Y is the sum of the values; (b) X is the value on the first die and Y is the larger of the two values; (c) X is the smallest and Y is the largest value obtained on the dice.

9. [PA] Refer to the *Solved Problem.*23 of chapter 1 of [MIT_BT]. Assume that input and output of the channel are represented by random variables X and Y, respectively. Now instead of a single binary communication channel, we consider a 2-stage cascaded binary communication channel, i.e., two such binary communication channels are connected in series. If (ϵ_0, ϵ_1) values for the first and the second stage are $(0.25, 0.3)$ and $(0.6, 0.45)$, respectively, then find the conditional probability $p_{Y|X}(y|x)$ for the 2-stage cascaded channel.

10. Let Y be a Poisson random variable with mean $\mu > 0$ and let Z be a geometrically distributed random variable with parameter p with $0 < p < 1$. Assume Y and Z are independent.

    (a) Find $\Pr(Y < Z)$. Express your answer as a simple function of μ and p.

    (b) Find $\Pr(Y < Z|Z = i)$ for $i \geq 1$.

    (c) Find $\Pr(Y = i|Y < Z)$ for $i \geq 0$. Express your answer as a simple function of p, μ and i.

A few questions were adapted from the following references

- Papoulis & Pillai, "Probability, Random Variables and Stochastic Processes", McGraw Hill, 4th Ed
- Sheldon Ross, 'A first course in Probability", 8th Ed
- Lecture notes of Prof. Hajek, University of Illinois Urbana-Champaign
- Capiński & Zastwniak, "Probability Through Problems", First Ed, Springer