DS313: Statistical Foundations of Data Science
Assignment No: 3

Course Instructor: Siddhartha Sarma *Date:* 28 Feb 2025
Submission deadline: 19 Mar 2025, 10 AM

Related topics:

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

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. For a non-negative integer-valued random variable N, show that $\mathbf{E}[N] = \sum_{i=1}^{\infty} \Pr(N \geq i)$.

2. Let X be a binomial random variable with parameters n and p. Show that $\mathbf{E}\left[\frac{1}{X+1}\right] = \frac{1-(1-p)^{n+1}}{(n+1)p}$.

3. Show that X is a Poisson random variable with parameter λ, then $\mathbf{E}[X^n] = \lambda \mathbf{E}[(X+1)^{n-1}]$.

4. [PA] Suppose that X, Y, and Z are independent random variables that are each equally likely to be either 1 or 2. Find the probability mass function of (a) $W = XYZ$, (b) $V = XY + XZ + YZ$, and (c) $U = X^2 + YZ$. Find $\mathrm{Cov}(W, V), \mathrm{Cov}(W, U)$ and $\mathrm{Cov}(V, U)$.

5. The random variables X and Y are of discrete type, independent, with $\Pr(X = n) = \alpha_n$ and $\Pr(Y = n) = \beta_n, \quad n = 0, 1, \ldots$. Show that, if $Z = X + Y$, then

$$\Pr(Z = n) = \sum_{k=0}^{n} \alpha_k \beta_{n-k}, n = 0, 1, \ldots$$

   [PA] Verify the above expression by considering X and Y as Poisson and Geometric random variables, respectively.

6. [PA] If X and Y are independent, identically distributed (i.i.d) binomial random variables with parameters n and p, then show that $X+Y$ is also binomial random variable and then find its parameters.

7. Find the moment generating functions for random variables with (a) Binomial, (b) Geometric and (c) Poisson distributions and then calculate mean and variance for each of them.

8. If U, V are i.i.d. non-negative random variables with $\Pr(U = k) = \Pr(V = k) = p_k, \quad k = 0, 1, 2, \ldots$. If $\Pr(U = k | U + V = k) = \Pr(U = k - 1 | U + V = k) = 1/(k+1), k \geq 0$, then show that U and V are geometric random variables.

9. Consider the following probability space $(\Omega, \mathcal{E}, \Pr)$, where $\Omega = \{1, 2, \ldots\}$, $\mathcal{E}$ is the σ -algebra of all subsets of Ω and $\Pr(\{k\}) = 2^{-k}$ for each $k = 1, 2, \ldots$. Find $\mathbf{E}[X|Y]$ if $X(k) = k$ and $Y(k) = (-1)^k$ for each $k = 1, 2, \ldots$.

10. If a random variable X has the following CDF,

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 2x^2/3, & 0 \leq x < 0.5 \\ 1 - \exp(-0.75x) & 0.5 \leq x < 1.5 \\ x/2 - 0.05, & 1.5 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

(a) Find $\Pr(X = i), i = 0.5, 1, 1.5, 2$

(b) Find $\Pr(\frac{1}{2} < X < \frac{3}{2})$.

(c) Find the expression for the pdf of X.

[PA] Write a program that generates samples from the above distribution and verify its output by plotting a histogram of the cumulative function of the samples generated by your program.

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