DS303: Statistical Foundations of Data Science
Assignment No: 2

Course Instructor: Siddhartha Sarma *Date:* 25 Feb 2024

Submission deadline: 10 Mar 2024, 11 AM (Bring hardcopies to the exam hall)

Related topics:

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

Notations:

- $\mathbf{E}[.]$, $\mathrm{Cov}(.,.)$ denote the expectation and covariance, respectively.
- $p_X(.)$ denotes the probability mass function (PMF) of a random variable X.
- $\Pr(.|.)$ represents conditional probability.

For the questions prefixed with [PA], in addition to solving them analytically, you should write a program in any programming language of your choice to simulate the underlying experiment. As discussed in the last 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 discussion section of the course moodle page for examples.

A separate submission link for the programming assignments will be provided on the course moodle page.

---

1. Suppose a discrete random variable X takes on one of the values from the set $\{0, 1, \ldots, N\}$. If for some constant c, $\Pr(X = i) = c\Pr(X = i - 1), i = 1, 2, \ldots, N$, 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 $\mathbf{E}[Y]$.

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

3. [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.

4. Let X have a Poisson distribution with parameter λ. Find the values of k for which $\Pr(X = k)$ is largest.

5. [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.

6. [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 $\text{Cov}(W, V), \text{Cov}(W, U)$ and $\text{Cov}(V, U)$.

7. 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

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

8. 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.

9. Two coins with probabilities of heads p_1 and p_2, respectively, are tossed together. A random variable X corresponds to the number of heads that show up when these coins are tossed. The mapping for another random variable Y is as follows: 1 if heads appears for the second coin, otherwise 0. Can p_1 and p_2 be chosen such that X and Y are statistically independent?

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.

    (d) Find $\mathbf{E}[Y | Y < Z]$, which is the expected value computed according to the conditional distribution found in part (c). Express your answer as a simple function of μ and p.

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