INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination

M. Tech (CS), 2023-2024 (Semester – I)

Probability and Stochastic Processes

Date: 21.09.2023 Maximum Marks: 75 Duration: 3.0 Hours

Note:

Answer all questions in Group-A and answer as much as you can in Group-B but the maximum you can score in Group-B is 40.

$E[X]$ and var$[X]$ denote the expectation and variance of the random variable X, respectively.

Group A

(QA1) Let $A_1, A_2, \ldots$ be a decreasing sequence of events, so that $A_1 \supseteq A_2 \supseteq \ldots$. Now define a limiting event A as

$$A = \lim_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} A_n$$

Then, show that

$$\Pr(A) = \Pr\left(\lim_{n \to \infty} A_n\right) = \lim_{n \to \infty} \Pr\left(A_n\right)$$

[7]

(QA2) Two gamblers, G_1 and G_2 bet on the outcomes of successive flips of a coin. On each flip, if the coin comes up heads, G_1 collects Rs. 1 from G_2; and if the coin comes up tails, G_2 collects Rs. 1 from G_1. The gamblers continue to do this until one of them runs out of money. Assume that the successive flips of coins are independent and the probability of a coin flip resulting in a head is p, and resulting in a tail is $1-p$. Show that with probability 1, either G_1 or G_2 will wind up with all the money. [8]

(QA3) (i) State and prove Bayes' theorem.

(ii) Show that for any n events $A_1, A_2, \ldots, A_n$, $\Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr\left(A_i\right)$

[(1+3)+4=8]

(QA4) Let $Z = X_1 + X_2 + \cdots + X_N$ where N is a random variable that takes non-negative integer values and $X_1, X_2, \ldots$ are identically distributed random variables. If $N = 0$, we let $Z = 0$. Assume that $N, X_1, X_2, \ldots$ are mutually independent. Deduce an expression for $E[Z]$. [4]

(QA5) (i) State and prove Chebyshev's inequality.

(ii) Let X denote a random variable indicating the number of heads in a sequence of n independent fair coin flips. Compute a lower bound on $\Pr(X \leq 2n/3)$.

(Ans:) $\Pr(X \leq 2n/3) = 1 - \Pr(X > 2n/3)$. Obtain an upper bound on $\Pr(X > 2n/3)$ from Chebyshev. Chebyshev gives $\Pr(X > 2n/3) \leq \frac{9}{n}$. Thus $\Pr(X \leq 2n/3) > 1 - \frac{9}{n}$.

[(1+3)+4=8]

Group B

(QB1) (i) Let $\mathcal{F}$ be a σ-field of subsets of Ω and suppose that $B \in \mathcal{F}$. Show that $\mathcal{G} = \{A \cap B : A \in \mathcal{F}\}$ is a σ-field of subsets of B.

(Ans:) Notice that we are to prove $\mathcal{G}$ is a σ-field of subsets of B and not a σ-field of subsets of Ω. We need to check that $\mathcal{G}$ staisfies the definition ofσ-field:

(i) $\emptyset \in \mathcal{F}$. So, $\emptyset = \emptyset \cap B \in \mathcal{G}$.

(ii) If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_i (A_i \cap B) = (\cup_i A_i) \cap B \in \mathcal{G}$.

(iii) If $A \in \mathcal{F}$, then $\bar{A} \in \mathcal{F}$. So, $B \setminus (A \cap B) = \bar{A} \cap B \in \mathcal{G}$.

(ii) A fair coin is tossed repeatedly. Show that, with probability 1, a head turns up sooner or later.

(Ans:) Pr(a head turns up sooner or later) $= 1 -$ Pr(no head ever turns up). As we want Pr(a head turns up sooner or later) $= 1$, we will want to show

$$\Pr(\text{ no head ever turns up }) = 0$$

.

$$\begin{aligned}
\Pr(\text{ no head ever turns up }) &= \lim_{n\to\infty} \Pr(\text{ no head turns up in first } n \text{ tosses }) \\
&= \lim_{n\to\infty} \left(\frac{1}{2}\right)^n \\
&= 0
\end{aligned}$$

[5+5=10]

(QB2) A coin that has probability of heads equal to p is tossed successively and independently until a head comes twice in a row or a tail comes twice in a row. Find the expected value of the number of tosses. [10]

(Ans:) Argue using total expectation theorem.

Let X be the random variable denoting the number of tosses until the process is over. Let H_i and T_i be the event that a head and tail occurs, respectively at the i-th toss. Let $p = \Pr(H_i)$ and $q = \Pr(T_i)$ and $p + q = 1$. Let us look at the total expectationin terms of H_1 and T_1, where $p = \Pr(H_1)$ and $q = \Pr(T_1)$:

$$E[X] = p\, E[X \mid H_1] + q\, E[X \mid T_1].$$

Again, use total expectation,

$$E[X \mid H_1] = p\, E[X \mid H_1 \cap H_2] + q\, E[X \mid H_1 \cap T_2] = 2p + q(1 + E[X \mid T_1]),$$

as the conditional event $X \mid H_1 \cap H_2$ indicates the game is over with two tosses and the conditional event $X \mid H_1 \cap T_2$ indicates that the influence of the first toss being a head is gone and the next toss being a tail is of importance, i.e., the first toss just adds 1 to the count and only the second toss matters in determining the number of additional tosses till the end. Thus, $E[X \mid H_1 \cap H_2] = 2$ and $E[X \mid H_1 \cap T_2] = 1 + E[X \mid T_1]$. Similarly, we have

$$E[X \mid T_1] = 2q + p(1 + E[X \mid H_1]).$$

Solve the expressions for $E[X \mid H_1]$ and $E[X \mid T_1]$ to obtain, $E[X \mid H_1] = \frac{2+q^2}{1-pq}$ and $E[X \mid T_1] = \frac{2+p^2}{1-pq}$.

Replace it in the first expression for $E[X]$, to obtain $E[X] = p \cdot \frac{2+q^2}{1-pq} + q \cdot \frac{2+p^2}{1-pq}$. Use the fact $p + q = 1$, to simplify the above expression to $E[X] = \frac{2+pq}{1-pq}$ which is the final answer.

(QB3) Let $X_1, \dots X_n$ be independent random variables and let $Z = X_1 + \cdots + X_n$. Suppose that each X_i, $i = 1, \dots, n$ is a Bernoulli with parameter p_i, and that $p_1, \dots, p_n$ are chosen so that the mean of Z is a given $\mu > 0$. Deduce what p_i's should be so that the variance of Z is maximized. [10]

(Ans:) $Z = X_1 + \cdots + X_n$ where $X_1, \dots X_n$ are independent random variables and $E[Z] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n p_i = \mu$. As each X_i is a Bernoulli with parameter p_i, var$[X_i] = p(1 - p)$. So,

$$\text{var}[Z] = \sum_{i=1}^n \text{var}[X_i] = \sum_{i=1}^n p_i(1 - p_i) = \mu - \sum_{i=1}^n p_i^2.$$

So maximizing the variance of Z is equivalent to minimizing $\sum_{i=1}^n p_i^2$ subject to the constraint $\sum_{i=1}^n p_i = \mu$. Now, show that $p_i = \mu/n\ \forall i$ by solving the above constrained minimization problem as follows:

$$\sum_{i=1}^n p_i^2 = \sum_{i=1}^n (\mu/n)^2 + \sum_{i=1}^n (p_i - \mu/n)^2.$$

This is minimized when $p_i = \mu/n\ \forall i$.

(QB4) Suppose that X and Y are independent, identically distributed geometric random variables with parameter p. Show that

$$\Pr\left(X = i \mid X + Y = n\right) = \frac{1}{n-1}, \quad i = 1, \dots, n-1.$$

[10]

(Ans:) $X, Y \sim$ Geo(p).

$$\Pr\left(X = i \mid X + Y = n\right) = \frac{\Pr\left(X = i, X + Y = n\right)}{\Pr\left(X + Y = n\right)} = \frac{\Pr\left(X = i\right)\Pr\left(Y = n - i\right)}{\Pr\left(X + Y = n\right)}$$

The last step could be written because X and Y are independent.

$\Pr\left(X = i\right) = p(1-p)^{i-1}$, for $i \geq 1$, and $\Pr\left(Y = n - i\right) = p(1-p)^{n-i-1}$, for $n - i \geq 1$.

So, it follows that

$$\Pr\left(X = i\right)\ \Pr\left(Y = n - i\right) = \begin{cases} p^2(1-p)^{n-2} & \text{if } i = 1, \dots, n-1 \\ 0, & \text{otherwise} \end{cases}$$

Note that the product $\Pr\left(X = i\right)\ \Pr\left(Y = n - i\right)$ is independent of i. So, for any i and j such that $i, j \in [1, n-1]$, we have

$$\Pr\left(X = i \mid X + Y = n\right) = \Pr\left(X = j \mid X + Y = n\right).$$

Hence, $\Pr\left(X = i \mid X + Y = n\right) = \frac{1}{n-1}, \quad i = 1, \dots, n-1.$

(QB5) (i) Consider n independent tosses of a coin with probability of a head equal to p. Let X and Y be the number of heads and of tails, respectively. Compute the correlation coefficient of X and Y.

(Ans:) The correlation coefficient of X and Y is $\rho(X, Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\ \text{var}(Y)}}$.

Now, we have $X + Y = n$. So, $E[X] + E[Y] = n$.

Using these two, we have $(X - E[X]) + (Y - E[Y]) = 0$, or $(X - E[X]) = -(Y - E[Y])$. Thus, we have var$(X) =$ var(Y) and cov$(X, Y) = E[(X - E[X])(Y - E[Y])] = -E[(X - E[X])^2] = -var(X)$.

So, $\rho(X, Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\ \text{var}(Y)}} = -1$.

(ii) Show that $\text{Cov}\left(\sum\limits_{i=1}^{n} X_i, \sum\limits_{j=1}^{m} Y_j\right) = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} \text{Cov}(X_i, Y_j)$.

(Ans:) Let $E[X_i] = \mu_i,\ i = 1, \dots, n$ and $E[Y_j] = \gamma_j,\ j = 1, \dots, m$. By linearity of expectations, $E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mu_i$ and $E\left[\sum_{j=1}^{m} Y_j\right] = \sum_{j=1}^{m} \gamma_j$.

$$\begin{aligned}
\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) &= E\left[\left(\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right]\right)\left(\sum_{j=1}^{m} Y_j - E\left[\sum_{j=1}^{m} Y_j\right]\right)\right] \\
&= E\left[\left(\sum_i X_i - \sum_i \mu_i\right)\left(\sum_j Y_j - \sum_j \gamma_j\right)\right] \\
&= E\left[\left(\sum_i (X_i - \mu_i)\right)\left(\sum_j (Y_j - \gamma_j)\right)\right] \\
&= E\left[\sum_i \sum_j \left(X_i - \mu_i\right)\left(Y_j - \gamma_j\right)\right] \\
&= \sum_i \sum_j E\left[\left(X_i - \mu_i\right)\left(Y_j - \gamma_j\right)\right] \\
&= \sum_i \sum_j \text{Cov}(X_i, Y_j)
\end{aligned}$$

[5+5=10]