INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination

M. Tech (CS), 2024-2025 (Semester – I)

Probability and Stochastic Processes

Date: 17.09.2024 | Maximum Marks: 80 | Duration: 3.0 Hours

Note: Answer as much as you can but the maximum you can score in each of the groups is 40 for a total of 80. Please mention the questions attempted on the back of the cover page of your answer script.

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

Group A

(QA1) (i) If there are m persons and n possible birthdays, find out an expression for the probability that m persons have distinct birthdays.

(ii) Now, consider the persons sequentially, one after another.

(a) Find out an expression, in terms of i and n, for the probability that the first i persons fail to have distinct birthdays.

(b) Find out a lower bound on i so that the probability of all birthdays being distinct is at least $1/2$.

[4+(4+2)=10]

Ans (i): The probability that all m persons have different birthdays is

$$\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{m-1}{n}\right)=\prod_{i=1}^{m-1}\left(1-\frac{i}{n}\right)$$

Ans (ii): Let us consider each person one at a time. Let E_j be the event that the j-th person's birthday does not match any of the birthdays of the first $j-1$ persons. Then, the probability that the first i persons do not have distinct birthdays is Pr $\left(\overline{E_1}\cup\overline{E_2}\cup\cdots\overline{E_i}\right)$

$$\begin{aligned}\Pr\left(\bigcup_{j=1}^{i}\overline{E_j}\right) &\leq \sum_{j=1}^{i}\Pr\left(\overline{E_j}\right)\\ &\leq \sum_{j=1}^{i}\tfrac{j-1}{n}\\ &= \tfrac{i(i-1)}{2n}\end{aligned}$$

If $i\leq\sqrt{n}$, $\Pr\left(\bigcup_{j=1}^{i}\overline{E_j}\right)\leq\frac{1}{2}$. So, with $\lfloor\sqrt{n}\rfloor$ persons, the probability of all birthdays being distinct is at least $1/2$.

(QA2) 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

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

[10]

Ans: Let $C_n = \overline{A_n}$ and $C = \overline{A}$ where $A = \bigcap\limits_{n=1}^{\infty} A_n$. Since $A_{n+1} \subseteq A_n$, $C_{n+1} \supseteq C_n$. So events C_n are increasing monotonically.

$$C = \overline{A} = \overline{\left(\bigcap_{n=1}^{\infty} A_n\right)} = \bigcup_{n=1}^{\infty} \overline{A_n} = \bigcup_{n=1}^{\infty} C_n.$$

Now use the results on increasing sequences, to obtain

$$1 - \text{Pr}(A) = \text{Pr}(\overline{A}) = \text{Pr}(C) = \lim_{n\to\infty} \text{Pr}\left(C_n\right) = \lim_{n\to\infty}\left(1 - \text{Pr}\left(A_n\right)\right).$$

So $1 - \text{Pr}(A) = 1 - \lim\limits_{n\to\infty} \text{Pr}\left(A_n\right)$. Therefore, $\text{Pr}(A) = \lim\limits_{n\to\infty} \text{Pr}\left(A_n\right)$.

(QA3) Consider $m+1$ boxes with the i-th box containing i red balls and $m-i$ white balls, where $i = 0, \ldots, m$. We choose a box at random, where all boxes are equally likely to be chosen. Then, we choose a ball at random from that box, n successive times with replacement (i.e., the ball drawn is replaced each time, and a new ball is selected independently). Suppose, a white ball is chosen each of the n times. What is the probability that if we draw a ball one more time, it will be white? Estimate this probability for large m. [10]

Ans: Let W_n be the event that a white ball was chosen for n preceding draws and E be the event that a white ball was drawn at the $n+1$-th draw.

We want to find $\text{Pr}(E \mid W_n)$. Now $\text{Pr}(E \mid W_n) = \frac{\text{Pr}(E \cap W_n)}{\text{Pr}(W_n)}$. Applying total probability theorem, we have

$$\begin{aligned}
\text{Pr}(W_n) &= \sum_{i=0}^{m} \text{Pr}\left(i\text{-th box chosen}\right) \cdot \text{Pr}\left(n \text{ white balls chosen with replacement from } i\text{-th box}\right) \\
&= \tfrac{1}{m+1} \sum_{i=0}^{m} \left(\tfrac{m-i}{m}\right)^n
\end{aligned}$$

Similarly, $\text{Pr}\left(E \cap W_n\right) = \text{Pr}\left(W_{n+1}\right) = \frac{1}{m+1} \sum\limits_{i=0}^{m} \left(\frac{m-i}{m}\right)^{n+1}$.

Now, for large m, we have

$$\begin{aligned}
\Pr(W_n) &= \tfrac{1}{m+1}\sum_{i=0}^{m}\left(\tfrac{m-i}{m}\right)^n \\
&\approx \tfrac{1}{(m+1)m^n}\int_0^m (m-x)^n dx \\
&= \tfrac{-1}{(m+1)m^n}\int_m^0 y^n dy \\
&= \tfrac{1}{(m+1)m^n}\int_0^m y^n dy \\
&= \tfrac{1}{(m+1)m^n}\cdot\tfrac{m^{n+1}}{n+1} \\
&\approx \tfrac{1}{n+1}
\end{aligned}$$

Similarly, $\Pr\left(E \cap W_n\right) = \Pr\left(W_{n+1}\right) \approx \frac{1}{n+2}$.

So, $\Pr(E \mid W_n) = \frac{\Pr(E\cap W_n)}{\Pr(W_n)} \approx \frac{1/(n+2)}{1/(n+1)} = \frac{n+1}{n+2}$.

(QA4) (i) Let X be a random variable with PMF p_X, and let $g(X)$ be a function of X. Then, show that the expected value of the random variable $g(X)$ is given by

$$E[g(X)] = \sum_x g(x)p_X(x)$$

.

(ii) Let X be a random variable with $\text{var}[X] = 0$. What will be the PMF of X? Justify your result.

[6+4=10]

Ans (i): Let $Y = g(X)$ and $p_Y(y) = \sum_{\{x \mid g(x)=y\}} p_X(x)$, where $p_X(x)$ and $p_Y(y)$ denotes the pmf of X and Y, respectively.

$$\begin{aligned}
E[g(X)] &= E[Y] \\
&= \sum_y y\, p_Y(y) \\
&= \sum_y y \sum_{\{x \mid g(x)=y\}} p_X(x) && \text{(add the pmfs)} \\
&= \sum_y \sum_{\{x \mid g(x)=y\}} y\, p_X(x) \\
&= \sum_y \sum_{\{x \mid g(x)=y\}} g(x)\, p_X(x) \\
&= \sum_x g(x)\, p_X(x) && \text{(as } y = g(x) \text{ is a function.)}
\end{aligned}$$

(Ans (ii): Let $p_X(x)$ denote the pmf of X, i.e. $\Pr(X = x)$. As $\text{var}[X] = 0$, we have

$$\begin{aligned}
& E[(X - E[X])^2 &&= 0 \\
\text{or,}\quad & E\left[(f(X))^2\right] &&= 0 \quad \text{where } f(X) = X - E[X] \\
\text{or,}\quad & \sum_x (f(X))^2\, p_X(x) &&= 0
\end{aligned}$$

As both $(f(X))^2$ and $p_X(x)$ are non-negative, so $p_X(x) = 0$ for $f(X) \neq 0$. So, $p_X(x) = 1$ for $f(X) = 0$. This implies that the pmf of X is as follows:

$$\text{pmf}_X(x) = \begin{cases} 1 & \text{if } x = a; \\ 0 & \text{o/w.} \end{cases}$$

(QA5) (i) Let X be a random variable that assumes only nonnegative values. Show that for all $a > 0$, $\Pr(X \geq a) \leq \frac{E[X^m]}{a^m}$, where m is a positive integer ≥ 2.

(ii) Show that for a random variable X and any $t > 1$,

$$\Pr\left(|X - E[X]| \geq t \cdot \sqrt{\text{var}[X]}\right) \leq \frac{1}{t^2}.$$

Prove all results that you need. [5+5=10]

Ans (i): For $a > 0$, let

$$I = \begin{cases} 1 & \text{if } X^m \geq a^m; \\ 0 & \text{otherwise.} \end{cases}$$

Since $X \geq 0$, we have $I \leq 1 \leq \frac{X^m}{a^m}$. As I is a 0-1 r.v., $E[I] = \Pr(I = 1) = \Pr(X^m \geq a^m)$. Take expectations on both sides to get

$$\Pr(X^m \geq a^m) = E[I] \leq E\left[\frac{X^m}{a^m}\right] = \frac{E\left[X^m\right]}{a^m}.$$

Ans (ii): Done in the class.

Group B

(QB1) There are n urns of which the i-th urn contains $i-1$ red balls and $n-i$ blue balls. You pick an urn at random and remove two balls at random without replacement. Find the probability that

(i) the second ball is blue;

(ii) the second ball is blue given the first ball is blue.

[3+7=10]

(Ans:) Let C_i denote the color of the i-th ball picked. Each urn contains $n-1$ balls, so there are $n(n-1)$ balls in all, of which $\frac{n(n-1)}{2}$ are blue (B) and $\frac{n(n-1)}{2}$ are red (R).

(a) The answer is $\frac{1}{2}$. You can see it using total probability.

(b) We seek $\Pr(C_2 = B \mid C_1 = B)$ by conditioning on the choice of the urn. $\Pr(C_1 = B) = \frac{\frac{n(n-1)}{2}}{n(n-1)} = \frac{1}{2}$.

$$\Pr(C_2 = B \mid C_1 = B) = \frac{\Pr(C_1, C_2 = B)}{\Pr(C_1 = B)} = \frac{\Pr(C_1, C_2 = B)}{\frac{1}{2}}. \tag{1}$$

Now, to find $\Pr(C_1, C_2 = B)$, we condition on the choice of the urn as

$$\begin{aligned}\Pr(C_1, C_2 = B) &= \sum_{i=1}^{n} \Pr(C_1, C_2 = B \mid i\text{-th urn was chosen}) \cdot \Pr(i\text{-th urn was chosen}) \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{n-i}{n-1} \cdot \frac{n-i-1}{n-2} \quad [\text{as Pr(i-th urn was chosen)} = \frac{1}{n}]\end{aligned}$$

Replace the above in Equation (2), to get

$$\Pr(C_2 = B \mid C_1 = B) = \frac{\Pr(C_1, C_2 = B)}{\Pr(C_1 = B)} = \frac{\left(\sum_{i=1}^{n} \frac{(n-i)(n-i-1)}{n(n-1)(n-2)}\right)}{\frac{1}{2}} = \frac{\frac{2n(n-1)(n-2)}{6n(n-1)(n-2)}}{\frac{1}{2}} = \frac{2}{3}$$

(QB2) Let $X_1, \ldots, X_n$ be independent random variables such that $E[X_i] \neq 0$. Show that

$$\frac{\text{var}\left(\prod_{i=1}^{n} X_i\right)}{\prod_{i=1}^{n} E[X_i]^2} = \prod_{i=1}^{n}\left(\frac{\text{var}(X_i)}{E[X_i]^2}+1\right)-1$$

[10]

(Ans:) Use independence and the formula of variance, $\text{var}[X] = E[X^2] + E[X]^2$, to get

$$\begin{aligned}
\text{var}\left(\prod_{i=1}^{n} X_i\right) &= E\left[\left(\prod_{i=1}^{n} X_i\right)^2\right] - \left(E\left[\prod_{i=1}^{n} X_i\right]\right)^2 \\
&= E\left[\prod_{i=1}^{n} X_i^2\right] - \left(\prod_{i=1}^{n} E[X_i]\right)^2 \\
&= \prod_{i=1}^{n} (E[X_i^2]) - \prod_{i=1}^{n} (E[X_i])^2 \\
&= \prod_{i=1}^{n} \left(\text{var}[X_i] + E[X_i]^2\right) - \prod_{i=1}^{n} (E[X_i])^2
\end{aligned}$$

Now, divide both sides by $\prod_{i=1}^{n} E[X_i]^2$ to get the result.

(QB3) There are n letters marked for n envelopes. The letters are mixed up and put randomly inside the envelopes. A *match* occurs if a letter goes into the envelope it is marked for. What is the probability of exactly k matches? [10]

(Ans:) Let $P(i)$ denote the probability of *derangement* of i letters (which we know from our class). For the probability of exactly k matches, consider a fixed set of k letters. The probability $\mathcal{P}$ that this set of k letters (and only this set) match, and by implication the other $n-k$ letters *derange* is

$$\mathcal{P} = \frac{1}{n}\frac{1}{n-1}\cdots\frac{1}{n-(k-1)}P(n-k) = \frac{(n-k)!}{n!}P(n-k).$$

Now as there are $\binom{n}{k}$ choices of the set of k letters, the requisite probability is

$$\binom{n}{k}\mathcal{P} = \frac{n!}{k!(n-k)!}\frac{(n-k)!}{n!}P(n-k) = \frac{P(n-k)}{k!}$$

(QB4) Let X and Y be independent Poisson variables with parameters λ_1 and λ_2, respectively. Prove that:

(i) $X+Y$ is Poisson with parameter $\lambda_1+\lambda_2$;

(ii) the conditional distribution of X, given $X+Y=n$, is binomial; and find its parameters.

[4+6=10]

Ans (i): Let $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ denote the pdf of X_1 and X_2, respectively. Let $X = X_1 + X_2$. The pdf of X is

$$f_X(x) = f_{X_1+X_2}(x) = \Pr(X = x) = \sum_{x_1=0}^{x} f_{X_1}(x_1)\cdot f_{X_2}(x-x_1).$$

We prove this as follows:
For each x, the event $[X = x]$ is the union of the disjoint events $[X_1 = x_1 \text{ and } X_2 = x - x_1]$ for $x_1 = 0, 1, \dots, x$. Then,

$$\begin{aligned} f_X(x) &= \Pr(X = x) \\ &= \sum_{x_1=0}^{x} \Pr(X_1 = x_1 \text{ and } X_2 = x - x_1) \\ &= \sum_{x_1=0}^{x} \Pr(X_1 = x_1)\cdot\Pr(X_2 = x - x_1) (\text{ because } X_1 \text{ and } X_2 \text{ are independent.}) \\ &= \sum_{x_1=0}^{x} f_{X_1}(x_1)\cdot f_{X_2}(x-x_1) \end{aligned}$$

Applying this to sum of independent Poissons, we have: $f(X_i = x_i) = \Pr(X_i = x_i) = e^{-\lambda_i}\frac{\lambda_i^{x_i}}{x_i!}$, for $i = 1, 2$. We use the formula deduced above.

$$\begin{aligned} f_X(x) &= \sum_{x_1=0}^{x} e^{-\lambda_1}\frac{\lambda_1^{x_1}}{x_1!}\cdot e^{-\lambda_2}\frac{\lambda_2^{x_2}}{x_2!} \\ &= e^{-(\lambda_1+\lambda_2)}\sum_{x_1=0}^{x}\frac{\lambda_1^{x_1}}{x_1!}\cdot\frac{\lambda_2^{x_2}}{x_2!} \\ &= e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^x}{x!} \end{aligned}$$

To see the last deduction, consider

$$\begin{aligned} (\lambda_1+\lambda_2)^x &= \sum_{x_1=0}^{x}\binom{x}{x_1}\lambda_1^{x_1}\lambda_2^{x-x_1} \\ &= x!\sum_{x_1=0}^{x}\frac{\lambda_1^{x_1}}{x_1!}\cdot\frac{\lambda_2^{x_2}}{x_2!} \end{aligned}$$

Thus $X_1 + X_2 \sim \text{Poi}(\lambda_1+\lambda_2)$.

Ans (ii): We need to use the result in (i) as follows:

$$\begin{aligned}
\Pr(X=k \mid X+Y=n) &= \frac{\Pr(X=k \cap X+Y=n)}{\Pr(X+Y=n)} \\
&= \frac{\Pr(X=k \cap Y=n-k)}{\Pr(X+Y=n)} \\
&= \frac{\Pr(X=k)\cdot\Pr(Y=n-k)}{\Pr(X+Y=n)} \\
&= \frac{e^{-\lambda_1}\frac{\lambda_1^k}{k!}\cdot e^{-\lambda_2}\frac{\lambda_2^{(n-k)}}{(n-k)!}}{e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}} \\
&= \frac{n!}{k!(n-k)!}\frac{\lambda_1^k\lambda_2^{(n-k)}}{(\lambda_1+\lambda_2)^n} \\
&= \binom{n}{k}\frac{\lambda_1^k\lambda_2^{(n-k)}}{(\lambda_1+\lambda_2)^n} \\
&= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}
\end{aligned}$$

Thus, we see that the conditional distribution follows Bin $\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$.

(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.

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

Ans (i): 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 $\text{var}(X) = \text{var}(Y)$ and $\text{cov}(X,Y) = E[(X-E[X])(Y-E[Y])] = -E[(X-E[X])^2] = -\text{var}(X)$.

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

Ans (ii): 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 $E\left[\sum_{j=1}^m Y_j\right] = \sum_{j=1}^m \gamma_j$.

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

[5+5=10]