INDIAN STATISTICAL INSTITUTE

Class Test II

M Tech (CS), 2024 (Semester – I) *Probability and Stochastic Processes*

Date: 21.09.2024 Maximum Marks : 30 Duration : 1.0 hours

Note: The question paper is of 40 marks. Answer as much as you can, but the maximum you can score is 30. Answer a question within its allotted box.

Course:(MTech/JRF/PLP) __________ **Name:** ______________________________ **Roll Number:** __________

(Q1) A coin shows heads with probability p. Let X_n be the number of flips required to obtain a run of n consecutive heads. Show that $E(X_n) = \sum_{k=1}^{n} p^{-k}$.

[Hints: Try to form a recurrence using conditional expectation.] [10]

(Ans:) Let X_k be the random variable that denotes the number of tosses to get k consecutive heads. Note that $E[X_1] = \frac{1}{p}$. Observe that, for $n \geq 2$,

$$\begin{aligned} E\left(X_n \,|X_{n-1}\right) &= p(X_{n-1}+1) + (1-p)(X_{n-1}+1+E(X_n)) \\ E\left(E\left(X_n \,|X_{n-1}\right)\right) &= E\left(p(X_{n-1}+1) + (1-p)(X_{n-1}+1+E(X_n))\right) \\ E(X_n) &= p(E(X_{n-1})+1) + (1-p)(E(X_{n-1})+1+E(X_n)) \\ E(X_n) &= \frac{1}{p}(E(X_{n-1})+1) \end{aligned}$$

One can get the desired result after solving the above recurrence.

(Q2) Consider the coupon collector's problem studied in class and let X be the random variable denoting the time to collect n coupons. We know that $E[X] = nH_n$, where $H_n = \log n + O(1)$.

(i) Use Chebyshev's inequality to bound the deviation of X from $E[X]$ by an amount of nH_n. You can use the fact that the variance of a geometric r.v. with parameter p is $(1-p)/p^2$ and $\sum_{i=1}^{\infty}\left(\frac{1}{i}\right)^2 = \frac{\pi^2}{6}$.

(ii) Using union bound or otherwise, find out the probability that all coupons are not collected even after $2n\ln n$ steps. Which of the two bounds is better? Give justifications. [4+6=10]

(Ans (i):) We know that we can write $X = X_1 + X_2 + \cdots + X_n$, where $X_i \sim \text{Geo}(\frac{n-i+1}{n})$, i.e., X_i follows geometric with parameter $\frac{n-i+1}{n}$. We also know that $E[X] = nH_n$. So, $\text{var}[X_i] = (1-p)/p^2 \leq 1/p^2 = \left(\frac{n}{n-i+1}\right)^2$.

Now,

$$\text{var}[X] = \sum_{i=1}^{n} \text{var}[X_i] = \sum_{i=1}^{n}\left(\frac{n}{n-i+1}\right)^2 = n^2\sum_{i=1}^{n}\left(\frac{1}{i}\right)^2 \leq \frac{\pi^2 n^2}{6}.$$

By Chebyshev's inequality,

$$\Pr\left(|X - nH_n| \geq nH_n\right) \leq \frac{n^2\pi^2/6}{(nH_n)^2} = \frac{\pi^2}{6\,(H_n)^2} = O\left(\frac{1}{(\ln n)^2}\right).$$

(Ans (ii):) Let E_i be the event that the i-th coupon is not obtained after $2n\ln$ steps.

$$\Pr(E_i) = \left(1 - \frac{1}{p}\right)^{2n\ln n} < e^{-2n\ln n} = \frac{1}{n^2}.$$

Using the union bound, we have that the probability that all coupons are not collected even after $2n\ln n$ steps, is

$$\Pr\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n}\Pr\left(E_i\right) = \frac{1}{n}.$$

Now, coming to which of the two bounds is better, we have

$$\Pr\left(X \geq 2n\ln n\right) = \Pr\left(X - n\ln n \geq n\ln\right) \leq \Pr\left(|X - nH_n| \geq nH_n\right).$$

By comparing the two functions obtained in (i) $\left(O\left(\frac{1}{(\ln n)^2}\right)\right)$ and (ii) $\frac{1}{n}$, we see that $\frac{1}{n} \leq \left(O\left(\frac{1}{(\ln n)^2}\right)\right)$. So, the union bound based argument gives much better tail bound than Chebyshev.

(Q3) A rectangular piece of paper is coloured with three colours – red, blue and green. 12% of the surface is coloured red, another 12% is coloured blue and the rest is coloured green. Show that, irrespective of the manner in which the colours are distributed, it is possible to draw a square with all its vertices green, on the piece of paper. [10]

(Ans:) Observe that 24% of the surface is colored non-green, i.e., either red or blue.
Let E_i be the event that the i-th vertex of a randomly selected square is either red or blue, i.e. non-green. Notice that $\Pr(E_i) = 0.24$.
Now the probability that the square has its vertices coloured either red or blue is $\Pr\left(\bigcup_{i=1}^{4} E_i\right)$. By the union bound, we have

$$\Pr\left(\bigcup_{i=1}^{4} E_i\right) \leq \sum_{i=1}^{4} \Pr\left(E_i\right) = 4 \times 0.24 = 0.96 < 1.$$

So, there exists squares with all vertices coloured green. At least 4% of the squares have only green vertices.

(Q4) An electronic device transmits a 1 with probability p and a 0 with probability $1 - p$, independent of earlier transmissions. The number of transmissions within a given time interval T follows a Poisson PMF with parameter λ. Show that

(i) the PMF of the number of 1s transmitted during the same time interval T follows Poisson.

(ii) Find the parameter of the Poisson also. [8+2=10]

(Ans (i):) Let X be the number of 1's transmitted and Y be the number of 0's transmitted within a time interval T. Let $Z = X + Y$. By the question statement, $Z \sim \text{Poi}(\lambda)$, i.e., $\Pr(Z = z) = \frac{e^{-\lambda}\lambda^z}{z!}$.

We want to find $\Pr(X = n)$. We know $\Pr(X = n) = \sum_{m=0}^{\infty} \Pr(X = n \cap Y = m)$. So, we need to find $\Pr(X = n \cap Y = m)$.

$$
\begin{aligned}
\Pr(X = n \cap Y = m) &= \Pr(Z = n + m) \cdot \Pr(X = n \cap Y = m \mid Z = n + m) \\
&= \binom{n+m}{n} p^n (1-p)^m \cdot e^{-\lambda} \cdot \frac{\lambda^{(n+m)}}{(n+m)!} \\
&= \frac{(n+m)!}{n!\, m!} p^n (1-p)^m \frac{e^{-\lambda}\lambda^n \lambda^m}{(n+m)!} \\
&= \frac{e^{-\lambda}(\lambda p)^n}{n!} \frac{(1-p)^m \lambda^m}{m!} \\
&= \frac{\left(e^{-\lambda p}\right)(\lambda p)^n}{n!} \frac{e^{-\lambda} e^{\lambda p} (1-p)^m \lambda^m}{m!} \\
&= \frac{\left(e^{-\lambda p}\right)(\lambda p)^n}{n!} \frac{e^{-\lambda(1-p)} (\lambda(1-p))^m}{m!}
\end{aligned}
$$

Now coming back to the finding $\Pr(X = n)$.

$$
\begin{aligned}
\Pr(X = n) &= \sum_{m=0}^{\infty} \Pr(X = n \cap Y = m) \\
&= \sum_{m=0}^{\infty} \frac{\left(e^{-\lambda p}\right)(\lambda p)^n}{n!} \frac{e^{-\lambda(1-p)} (\lambda(1-p))^m}{m!} \\
&= \frac{e^{-\lambda p}(\lambda p)^n e^{-\lambda(1-p)}}{n!} \sum_{m=0}^{\infty} \frac{\lambda^m (1-p)^m}{m!} \\
&= \frac{e^{-\lambda p}(\lambda p)^n e^{-\lambda(1-p)}}{n!} \cdot e^{\lambda(1-p)} \\
&= \frac{e^{(-\lambda p)}(\lambda p)^n}{n!}
\end{aligned}
$$

Thus, $X \sim \text{Poi}(\lambda p)$.

(Ans (ii):) The parameter is as found λp.