INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination

M. Tech (CS) - I Year, 2018-2019 (Semester - I)

Probability and Stochastic Processes

Date: 03.09.2018 Maximum Marks: 90 Duration: 3 Hours

Note: This is a 2-page question paper. The question paper is of 100 marks.

Answer as much as you can, but the maximum you can score is 90.

(Q1) Let there be n sticks each of which is broken into one long and one short part. The $2n$ parts are arranged into n pairs from which new sticks are formed. Find the probability that

(a) the parts will be joined in the original order.

(b) that all long parts are paired with short parts.

[6+6=12]

(Ans:) This problem has similarities with the problem of pairing couples. Think of the long part of the stick as female and the other part as male.

(Ans a:) The possible number of arrangements is $(2n)!$. Of them, fix a couple (a long and a short part of the same stick). There are n of them. They can be permuted in $n!$ ways. Now, in an arrangement, each couple can be arranged in 2 ways, giving 2^n for n couples. So, the said probability is $\frac{2^n n!}{(2n)!}$.

(Ans b:) Now permute n long sticks in $n!$ ways and n short sticks in $n!$ ways and pair them up in $(n!)^2$ ways. For each such arrangement, one can again order the long and short in 2 ways, leading to 2^n for n couples. Thus, the said probability is $\frac{2^n(n!)^2}{(2n)!}$ which for a better form is $\frac{2^n}{\binom{2n}{n}}$.

◀

(Q2) Airlines find that each passenger who reserves a seat fails to turn up with probability $\frac{1}{10}$ independently of the other passengers. So, Indigo Airlines always sells 10 tickets for their 9 seater aeroplane while Air India always sells 20 tickets for their 18 seater aeroplane. Which is more often over-booked? [5]

(Ans:) Let X and Y denote the number of people to whom Indigo and Air India sell tickets, respectively. So, X and Y take value in $[1, \ldots, 10]$ and $[1, \ldots, 20]$. The failure probability of a passenger turning up is $\frac{1}{10}$ and the success probability is $\frac{9}{10}$ for both airlines.

So, $\Pr(X = k) = \binom{10}{k} \left(\frac{9}{10}\right)^k \left(\frac{1}{10}\right)^{10-k}$ and $\Pr(Y = k) = \binom{20}{k} \left(\frac{9}{10}\right)^k \left(\frac{1}{10}\right)^{20-k}$.

Now, the probability of Indigo to be overbooked is

$$\begin{aligned} \Pr(X > 9) &= \Pr(X = 10) \\ &= \binom{10}{10} \left(\frac{9}{10}\right)^{10} \left(\frac{1}{10}\right)^{10-10} \\ &= \left(\frac{9}{10}\right)^{10} \\ &= a \quad \text{(say)} \end{aligned}$$

Now, the probability of Air India to be overbooked is

$$\begin{aligned} \Pr(Y > 18) &= \Pr(Y = 19) + \Pr(Y = 20) \\ &= \binom{20}{19} \left(\frac{9}{10}\right)^{19} \left(\frac{1}{10}\right) + \binom{20}{20} \left(\frac{9}{10}\right)^{20} \\ &= 2\left(\frac{9}{10}\right)^{19} + \left(\frac{9}{10}\right)^{20} \\ &= \left(\frac{9}{10}\right)^{10} \left(2\left(\frac{9}{10}\right)^{9} + \left(\frac{9}{10}\right)^{10}\right) \\ &= a\left(\frac{20}{9}a + a\right) \\ &= a\left(\frac{29}{9}a\right) \\ &> a \end{aligned}$$

As $\Pr(Y > 18) > \Pr(X > 9)$, Air India is more often overbooked. ◀

(Q3) For the following, let X_1 and X_2 be two independent random variables. $\text{Bin}(\cdot, \cdot)$ and $\text{Poi}(\cdot)$ denote the Binomial and Poisson distributions, respectively. "pdf" means the probability distribution function.

(i) Let $X_i \sim \text{Bin}(n_i, p), i = 1, 2$. Find out the pdf of $X_1 + X_2$.

(ii) Let $X_i \sim \text{Poi}(\lambda_i), i = 1, 2$. Find out the pdf of $X_1 + X_2$.

[6+6=12]

(Ans:) 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 $[X_1 = x_1$ and $X_2 = x - x_1]$ for $x_1 = 0, 1, \ldots, x$. Then,

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

(i) We have $f(X_i = x_i) = \Pr(X_i = x_i) = \binom{n_i}{x_i} p^{x_i} (1-p)^{(n_i - x_i)}$, for $i = 1, 2$. We use the formula deduced above.

$$\begin{aligned} f_X(x) &= \sum_{x_1=0}^{x} \binom{n_1}{x_1} p^{x_1} (1-p)^{(n_1 - x_1)} \binom{n_2}{x - x_1} p^{x - x_1} (1-p)^{(n_2 - (x - x_1))} \\ &= p^x (1-p)^{n_1 + n_2 - x} \sum_{x_1=0}^{x} \binom{n_1}{x_1} \binom{n_2}{x - x_1} \\ &= \binom{n_1 + n_2}{x} p^x (1-p)^{n_1 + n_2 - x} \end{aligned}$$

Consider the co-efficient of both LHS and RHS of any y^x in the identity $(1 + y)^{n_1}(1 + y)^{n_2} = (1 + y)^{n_1 + n_2}$ to get $\sum_{x_1=0}^{x} \binom{n_1}{x_1}\binom{n_2}{x - x_1} = \binom{n_1 + n_2}{x}$. Thus $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$

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

◀

(Q4) **(i)** Let X and Y be independent discrete random variables, and let $g, h : \mathbb{R} \to \mathbb{R}$. Show that $g(X)$ and $h(Y)$ are independent.

(ii) Let X and Y be independent Bernoulli random variables with parameter $\frac{1}{2}$. Show that $X + Y$ and $|X - Y|$ are dependent though uncorrelated.

[4+6=10]

(Ans (i):) Let $g', h' \in \mathbb{R}$. We have

$$\begin{aligned}
\Pr\left(g(X) = g', h(Y) = h'\right) &= \sum_{x,y:g(x)=g' \text{ and } h(y)=h'} \Pr\left(X = x, Y = y\right) \\
&= \sum_{x,y:g(x)=g' \text{ and } h(y)=h'} \Pr\left(X = x\right) \cdot \Pr\left(Y = y\right) \\
&\quad \text{since } X \text{ and } Y \text{ are independent.} \\
&= \sum_{x:g(x)=g'} \Pr\left(X = x\right) \sum_{y:h(y)=h'} \Pr\left(Y = y\right) \\
&= \Pr\left(g(X) = g'\right) \cdot \Pr\left(h(Y) = h'\right).
\end{aligned}$$

(Ans (ii):) Let us look at the covariance of the random variables $X + Y$, and $|X - Y|$.

$$\text{cov}\left(X + Y, |X - Y|\right) = E[(X + Y)(|X - Y|)] - E[X + Y] \cdot E[|X - Y|]$$

The values of the random variables and their corresponding probabilities are as follows:

$(X+Y)\,\lvert X-Y\rvert$	0	1	2
Probability	$\frac{5}{8}$	$\frac{1}{4}$	$\frac{1}{8}$

$(X+Y)$	0	1	2
Probability	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$(X-Y)$	0	1
Probability	$\frac{1}{2}$	$\frac{1}{2}$

$$\begin{aligned}
\text{cov}\left(X + Y, |X - Y|\right) &= E[(X + Y)(|X - Y|)] - E[X + Y] \cdot E[|X - Y|] \\
&= \frac{1}{4} + \frac{1}{4} - 1 \cdot \frac{1}{2} \\
&= 0
\end{aligned}$$

So, $X + Y$ and $|X - Y|$ are uncorrelated.

Looking at the probabilities $\Pr(X + Y = 0, |X - Y| = 0)$, $\Pr(X + Y = 0)$ and $\Pr(|X - Y| = 0)$, we see that $\Pr(X + Y = 0, |X - Y| = 0) \neq \Pr(X + Y = 0) \cdot \Pr(|X - Y| = 0)$. So, they are dependent.

(Q5) Let X be a random variable defined over a sample space Ω such that $\mathrm{E}[X] = \mu$. Show that $\Pr(X \geq \mu) > 0$ and $\Pr(X \leq \mu) > 0$. [4+4=8]

[Hints: Can you try to prove using contradiction?]

(Ans:) Assume, for a contradiction, $\Pr(X \geq \mu) = 0$. Then,

$$\mu = \mathrm{E}[X] = \sum_x x \Pr(X = x) = \sum_{x<\mu} x \Pr(X = x) < \sum_{x<\mu} \mu \Pr(X = x) = \mu$$

which cannot be. So, $\Pr(X \geq \mu) \neq 0$ and as $\Pr(\cdot) \geq 0$, we have the result.

Similarly, assume $\Pr(X \leq \mu) = 0$. Then,

$$\mu = \mathrm{E}[X] = \sum_x x \Pr(X = x) = \sum_{x>\mu} x \Pr(X = x) > \sum_{x>\mu} \mu \Pr(X = x) = \mu$$

which cannot be. So, $\Pr(X \leq \mu) \neq 0$ and as $\Pr(\cdot) \geq 0$, we have the result. ◀

(Q6) Independent trials, each resulting in a success with probability p or a failure with probability $q = 1 - p$, are performed. Compute the probability that a run of n consecutive successes occurs before a run of m consecutive failures. [15]

(Ans:) Notice a nature of this problem. Supposing we are having a sequence of successes. Once we reach n consecutive successes before m consecutive failures, we are done. But say before reaching n consecutive successes, we encounter a failure. This wipes out the run of successes and brings us back to square one. This characteristic of the problem suggests that there is a recursive nature to this problem. Thus the solution idea is to condition on some trials and then deduce a recursive formula.

Let E be the event that a run of n consecutive successes occur before a run of m consecutive failures. Let T_1 be the event that the first trial results in a success. Then,

$$\Pr(E) = p \Pr(E \mid T_1) + q \Pr(E \mid \overline{T_1}) \tag{1}$$

Let us determine $\Pr(E \mid T_1)$ and $\Pr(E \mid \overline{T_1})$. We condition again on $S_{2,n}$, the event that all trials between 2 and n result in successes. So, we have

$$\Pr(E \mid T_1) = \Pr(E \mid S_{2,n}T_1) \cdot \Pr(S_{2,n} \mid T_1) + \Pr(E \mid \overline{S_{2,n}}T_1) \cdot \Pr(\overline{S_{2,n}} \mid T_1) \tag{2}$$

Notice that $\Pr(E \mid S_{2,n}T_1) = 1$. Trials are independent, so events $S_{2,n}$ and T_1 are independent and hence, $\Pr(S_{2,n} \mid T_1) = \Pr(S_{2,n}) = p^{n-1}$ and $\Pr(\overline{S_{2,n}} \mid T_1) = \Pr(\overline{S_{2,n}}) = 1 - p^{n-1}$. We are left to tackle $\Pr(E \mid \overline{S_{2,n}}T_1)$. This event says that we started with a success and kept on having successes starting from the second trial onwards and had a failure before the n-th success. So, this wipes out all previous successes and brings us to a situation where we start with a failure, i.e. $\Pr(E \mid \overline{S_{2,n}} T_1) = \Pr(E \mid \overline{T_1})$. Replacing the above observations in Equation 2, we have

$$\begin{aligned} \Pr(E \mid T_1) &= \Pr(E \mid S_{2,n} T_1) \cdot \Pr(S_{2,n} \mid T_1) + \Pr(E \mid \overline{S_{2,n}} T_1) \cdot \Pr(\overline{S_{2,n}} \mid T_1) \\ \Pr(E \mid T_1) &= p^{n-1} + (1 - p^{n-1}) \Pr(E \mid \overline{T_1}) \qquad (3) \end{aligned}$$

So, now we need to determine $\Pr(E \mid \overline{T_1})$. We again condition on $F_{2,m}$, the event that all trials between 2 and m result in failures. We have

$$\Pr(E \mid \overline{T_1}) = \Pr(E \mid F_{2,m} \overline{T_1}) \cdot \Pr(F_{2,m} \mid \overline{T_1}) + \Pr(E \mid \overline{F_{2,m}}\, \overline{T_1}) \cdot \Pr(\overline{F_{2,m}} \mid \overline{T_1}) \qquad (4)$$

Notice that $\Pr(E \mid F_{2,m} \overline{T_1}) = 0$ as $F_{2,m} \overline{T_1}$ denotes the event that m failures have occurred before n successes. By similar logic as earlier, we would have $\Pr(E \mid \overline{F_{2,m}}\, \overline{T_1}) = \Pr(E \mid T_1)$ and $\Pr(\overline{F_{2,m}} \mid \overline{T_1}) = \Pr(\overline{F_{2,m}}) = 1 - q^{m-1}$. So, we have

$$\begin{aligned} \Pr(E \mid \overline{T_1}) &= \Pr(E \mid F_{2,m} \overline{T_1}) \cdot \Pr(F_{2,m} \mid \overline{T_1}) + \Pr(E \mid \overline{F_{2,m}}\, \overline{T_1}) \cdot \Pr(\overline{F_{2,m}} \mid \overline{T_1}) \\ \Pr(E \mid \overline{T_1}) &= (1 - q^{m-1}) \Pr(E \mid T_1) \qquad (5) \end{aligned}$$

Replacing Equation 5 in Equation 3, we have

$$\begin{aligned} \Pr(E \mid T_1) &= p^{n-1} + (1 - p^{n-1})(1 - q^{m-1}) \Pr(E \mid T_1) \\ \Pr(E \mid T_1) &= \frac{p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1} q^{m-1}} \qquad (6) \end{aligned}$$

Replacing Equation 6 in Equation 1, we have

$$\begin{aligned} \Pr(E) &= p \Pr(E \mid T_1) + q \Pr(E \mid \overline{T_1}) \\ &= p \Pr(E \mid T_1) + q\{(1 - q^{m-1}) \Pr(E \mid T_1)\} \\ &= \Pr(E \mid T_1)(p + q + q^{m-1}) \\ &= \Pr(E \mid T_1)(1 + q^{m-1}) \\ &= \frac{p^{n-1}(1 - q^m)}{p^{n-1} + q^{m-1} - p^{n-1} q^{m-1}} \qquad (7) \end{aligned}$$

◀

(Q7) A person has a matchbox in his left pocket and another one in his right pocket. Both the matchboxes initially contained N match sticks. Whenever the person needs a match stick, he is equally likely to take the match box from either pocket. Consider the moment when the person first finds that one of his matchboxes is empty. What is the probability that there are exactly i matchsticks in the other box, $i = 0, 1, \ldots, N$? [10]

(Ans:) Let E be the event that the person first discovers that the left hand matchbox (LHM) is empty and there are k match sticks in the right hand matchbox (RHM). E occurs if and only if the $(N+1)$-th choice from the LHM is made at the $(N+1) + (N-k)$ trial.

This is all about getting exactly r successes after a certain number of independent trials.

Let X be the random variable indicating the number of trials required to obtain r successes.

$$\Pr(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, \ldots$$

We have discussed how to obtain this in the class on September 20, 2018.

Coming back to our case, $p = \frac{1}{2}$, $r = N + 1$, $n = N + 1 + N - K = 2N - k + 1$. So,

$$\Pr(E) = \binom{2N-k}{N} \left(\frac{1}{2}\right)^{2N-k+1}$$

There is an equal probability of the *mirror* event (LHM replaced by RHM and vice-versa). So, the final answer is $2\Pr(E)$. ◀

(Q8) Let X and Y be independent random variables taking positive integer values and having the same mass function $f(x) = 2^{-x}$ for $x = 1, 2, \ldots$. Find (i) $\Pr(\min\{X, Y\} \leq x)$; and (ii) $\Pr(X \text{ divides } Y)$. [5+5=10]

(Ans:) (i) X and Y are independent and $x \geq 1$. The complement of $\Pr(\min\{X, Y\} \leq x)$ is that both X and Y are greater than x. Thus,

$$\begin{aligned}
\Pr(\min\{X, Y\} \leq x) &= 1 - \Pr(X > x \text{ and } Y > x) \\
&= 1 - \Pr(X > x) \cdot \Pr(Y > y) \\
&= 1 - (2^{-x} \cdot 2^{-x}) \\
&= 1 - 4^{-x}
\end{aligned}$$

(ii)

$$\begin{aligned}
\Pr(X \text{ divides } Y) &= \sum_{i=1}^{\infty} \Pr(Y = iX) \\
&= \sum_{i=1}^{\infty} \sum_{x=1}^{\infty} \Pr(Y = iX \text{ and } X = x) \\
&= \sum_{i=1}^{\infty} \sum_{x=1}^{\infty} \Pr(Y = iX) \cdot \Pr(X = x) \\
&= \sum_{i=1}^{\infty} \sum_{x=1}^{\infty} 2^{-ix} \cdot 2^{-x} \\
&= \sum_{i=1}^{\infty} \frac{1}{2^{i+1} - 1}
\end{aligned}$$

◀

(Q9) Suppose we roll a standard fair die 200 times. Let X be the sum of the numbers that appear over the 200 rolls. Use Chebyshev's inequality to bound $\Pr[X \geq 750]$. [8]

(Ans:) Let X_i be the random variable that denotes the value obtained in the i^{th} roll of the die, X_i can take any integral value in the range $[1, 6]$ with equal probability. So, $E[X_i] = \frac{7}{2}$ and $\text{Var}[X_i] = E[X_i^2] - (E[X_i])^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$. Since $X = \sum_{i=1}^{200} X_i$, $E[X] = 200 \times \frac{7}{2} = 700$ and as X_i's are independent, $\text{Var}[X] = 200 \times \frac{35}{12} = \frac{1750}{3}$. Now using Chebyshev's inequality, we have

$$\begin{aligned}
\Pr[X \geq 750] &\leq \Pr[|X - 700| \geq 50] \\
&\leq \frac{\text{Var}[X]}{50^2} \\
&= \frac{1750}{3 \times 2500} \\
&= \frac{7}{30}.
\end{aligned}$$

◀

(Q10) If X is a random variable with mean 0 and finite variance σ^2, then for any $a > 0$, show that $\Pr(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$. [10]

[Hints: The above is a different form of Chebyshev's inequality. Use Markov's inequality to prove it by observing $\Pr(X \geq a) = \Pr(X + b \geq a + b)$, for $b > 0$. You will obtain an expression involving σ, a and b. Now try to find a suitable b.]

(Ans:) The proof technique is almost same as the deduction we did for Chebyshev's inequality in the class. Let $b > 0$.

$$\begin{aligned}
\Pr(X \geq a) &= \Pr(X + b \geq a + b) \\
&\leq \Pr\left((X + b)^2 \geq (a + b)^2\right) \\
&\leq \frac{E[(X + b)^2]}{(a + b)^2} \\
&= \frac{\sigma^2 + b^2}{(a + b)^2}
\end{aligned}$$

as $E[(X + b)^2] = E[X^2 + 2b \cdot X + b^2] = E[X^2] + 2b \cdot E[X] + b^2 = \sigma^2 + b^2$. Minimizing the above expression, we have $b = \frac{\sigma^2}{a}$ and putting back into the above Equation, we get the final result. ◀