MTH 201: Probability and Statistics

End Semester Exam (Section B)

26/04/2025

Sanjit K. Kaul

No books, notes, or devices are allowed. Just a pen/ pencil and eraser. Institute rules will apply with regards to cheating. Show your steps. Don't prioritize simplifying products and sums, if you are short on time. You have about 120 minutes.

Question 1. 5 **marks** Calculate $E[X^{1/Y} + XY|Y = 0.5]$, given that the joint PDF $f_{X,Y}(x, y) = 2$, $0 < x < y < 1$, and $f_{X,Y}(x, y) = 0$ otherwise.

Question 2. 25 **marks** You want to estimate the fraction of households in Delhi that have piped water. Assume that the availability of piped water at any household is modeled by a Bernoulli(p) random variable, where p is unknown. Answer the following questions.

(a) (5 marks) Calculate the minimum sample size required to ensure that the probability the sample mean is within ± 0.1 of p is at least 0.8.

(b) (10 marks) Suppose it is known that the variance of the Bernoulli random variable that models availability of piped water at any household is in the interval $(0, 0.25)$. Also, suppose we want to ensure that the probability the sample mean is within $(p - 0.05, p + 0.1)$ is at least 0.8. Calculate the minimum size of sample required.

(c) (10 marks) You randomly choose 100 households in Delhi and visit them to check availability of piped water at the chosen households. Forty percent of the chosen households have piped water. Provide a confidence interval with confidence 80%. Assume variance is unknown.

Question 3. 20 **marks** An experiment trial results in a pair of values drawn from the joint distribution of the pair of random variables X and Y. We perform such trials repeatedly till we observe $X < Y$ for the 10^{th} time. Assume X and Y are independent. X is exponential(1) and Y is uniform($0, 1$). Answer the following. Show your steps.

1) (10 marks) Derive the distribution of the total number of performed trials.
2) (10 marks) Derive the conditional distribution of X, given that $X < Y$.

Question 4. 20 **marks** A bin consists of 100 balls of which 20 are colored red, 30 are black, 10 are blue, 25 are green and 15 are purple. You draw balls randomly and with replacement and stop once you draw a green ball for the first time. Answer the following questions.

1) (5 marks) Derive the distribution of the number of draws from first principles.
2) (15 marks) Derive the distribution of the number of red balls drawn from the bin.

Question 5. 30 **marks** An experiment trial involves tossing two coins A and B. For each trial, before the toss, coin A chooses a probability of heads uniformly and randomly from $(0, 1)$. Similarly, coin B chooses a probability of heads uniformly and randomly from $(0, 0.5)$, independently of the choice made by coin A. We perform two such trails. Choices of probabilities made by the coins for the two trails are independent of each other. Coin A is declared winner in case it gives its first heads in an earlier trial than coin B. For example, Coin A gives its first heads in trial 2 and Coin B doesn't given any heads. In case both coins give their first heads in the same trial or if no heads are observed, we declare a draw. (10 **marks) Derive the probability that coin** A **wins**.

Now consider a slightly modified setting. As before, choices in different trials are independent of each other. However, for either trial, while coin A chooses a probability of heads uniformly and randomly from $(0, 1)$, coin B chooses a probability of heads such that for a choice of probability p by coin A, the conditional distribution from which coin b chooses its probability of heads is $1/p$ over the interval $(0, p)$ and is 0 otherwise. (20 **marks) As before, derive the probability that coin** A **wins**.

I. Useful Facts

An exponential RV X with expected value (mean) $1/\lambda$ has a PDF $f_X(x) = \lambda e^{-\lambda x}, x \geq 0$.

A Poisson random variable K with parameter α has PMF $P_K(k) = \alpha^k e^{-\alpha}/k!$. Its mean is α.

The Chebyshev's inequality states that for all $c > 0$, $P[|X - E[X]| \geq c] \leq \frac{\text{Var[X]}}{c^2}$.

Question 1. 5 **marks** Calculate $E[X^{1/Y} + XY|Y = 0.5]$, given that the joint PDF $f_{X,Y}(x, y) = 2$, $0 < x < y < 1$, and $f_{X,Y}(x, y) = 0$ otherwise.

$$E[X^{1/Y}+XY|Y=0.5]$$

$$= E\left[X^2 + \frac{X}{2} \middle| Y=0.5\right]$$

$$= E[X^2|Y=0.5] + \frac{1}{2}E[X|Y=0.5]$$

We require $f_{X|Y}(x|0.5)$.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Up to 1/5

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

$$= \int_0^y 2\,dx = 2y \qquad 0<y<1.$$

$$\therefore f_{X|Y}(x|0.5) = \frac{f_{X,Y}(x,0.5)}{f_Y(0.5)}$$

$$= \frac{2}{2(0.5)} = 2 \qquad 0<x<0.5$$

Correct value & range of x. Up to 1/5

$$E[X|Y=0.5] = \int_0^{0.5} x\, f_{X|Y}(x|0.5)\,dx$$

$$= 2\int_0^{0.5} x\,dx = 2\left(\frac{x^2}{2}\right)_0^{0.5}$$

$$= 2\left(\frac{1}{8}\right) = \frac{1}{4}.$$

1/5

$$E[X^2|Y=0.5] = \int_0^{0.5} x^2\, 2\,dx$$

$$= 2\left(\frac{x^3}{3}\right)_0^{0.5}$$

$$= \frac{2}{3}\left(\frac{1}{8}\right) = \frac{1}{12}.$$

1/5

$$\therefore E[X^{1/Y}+XY|Y=0.5]$$

$$= E[X^2|Y=0.5] + \frac{1}{2}E[X|Y=0.5]$$

$$= \frac{1}{12} + \frac{1}{2}\left(\frac{1}{4}\right) = \frac{1}{4}\left(\frac{1}{3}+\frac{1}{2}\right)$$

$$= \frac{1}{4}\cdot\frac{5}{6} = \frac{5}{24}.$$

1/5

Question 2. 25 **marks** You want to estimate the fraction of households in Delhi that have piped water. Assume that the availability of piped water at any household is modeled by a Bernoulli(p) random variable, where p is unknown. Answer the following questions.

(a) (5 marks) Calculate the minimum sample size required to ensure that the probability the sample mean is within ± 0.1 of p is at least 0.8.

(b) (10 marks) Suppose it is known that the variance of the Bernoulli random variable that models availability of piped water at any household is in the interval $(0, 0.25)$. Also, suppose we want to ensure that the probability the sample mean is within $(p - 0.05, p + 0.1)$ is at least 0.8. Calculate the minimum size of sample required.

(c) (10 marks) You randomly choose 100 households in Delhi and visit them to check availability of piped water at the chosen households. Forty percent of the chosen households have piped water. Provide a confidence interval with confidence 80%. Assume variance is unknown.

(a) Let the sample size be n.

Let X be Bernoulli(p). Note $E[X] = p$.

We require

$$P\left[p - 0.1 \le M_n(X) \le p + 0.1\right] \ge 0.8$$

That is

$$P\left[|M_n(X) - p| \le 0.1\right] \ge 0.8.$$

Correct application of Chebyshev's inequality 4/5

Since $M_n(X)$ is an unbiased estimator of p, $E[M_n(X)] = p$.

Thus, we require that

$$P\left[|M_n(X) - E[M_n(X)]| \le 0.1\right] \ge 0.8$$

Correct final answer 1/5

Application of Chebyshev's inequality gives

$$P\left[|M_n(X) - E[M_n(X)]| \le 0.1\right]$$

$$\ge 1 - \frac{Var[M_n(X)]}{(0.1)^2}$$

$$= 1 - \frac{Var[X]}{n(0.1)^2}$$

The minimum n can be obtained by setting

$$1 - \frac{Var[X]}{n(0.1)^2} \ge 0.8$$

The choice of n must work for any $Var[X]$. Also, it is largest when $Var[X]$ is the maximum.

maximum $Var[X] = (0.5)(0.5) = 0.25$.

$\therefore$ We require

$$1 - \frac{0.25}{n(0.1)^2} \ge 0.8$$

$$0.2 \ge \frac{1}{n\, 4 (0.1)^2}$$

$$n \ge \frac{1}{0.8\,(0.1)^2} = \frac{100}{0.8}$$

$$= \frac{1000}{8}$$

Choose $n = 125$.

(b)

We require

$$P\left[p - 0.05 \le M_n(X) \le p + 0.1\right] \ge 0.8.$$

We will use the Chebyshev's inequality as in (a).

Observe that

$$P\left[p - 0.05 \le M_n(X) \le p + 0.1\right]$$

$$\ge P\left[p - 0.05 \le M_n(X) \le p + 0.05\right]$$

$$= P\left[|M_n(X) - p| \le 0.05\right]$$

Choosing the correct interval 5/10

Choose n that satisfies

$$P\left[|M_n(X) - p| \le 0.05\right] \ge 0.8$$

Chebyshev's inequality gives

$$P\left[|M_n(X) - p| \le 0.05\right]$$

$$\ge 1 - \frac{Var[M_n(X)]}{(0.05)^2}$$

$$= 1 - \frac{Var[X]}{n(0.05)^2}$$

Correct Application of Chebyshev's 4/10

We require

$$1 - \frac{Var[X]}{n(0.05)^2} \ge 0.8$$ for any $Var[X]$, where X is Bern(p).

To choose n that satisfies the above, set $Var[X]$ to its max possible value. It is known that $Var[X]$ is in $(0, 0.25)$.

Correct answer 1/10

$\therefore$ We require

$$1 - \frac{0.25}{n(0.05)^2} \ge 0.8$$

$$0.2 \ge \frac{0.25}{n(0.05)^2}$$

$$n \ge \frac{0.25}{(0.2)(0.05)^2}$$

$$= \frac{0.25}{(0.2)(25)} 10^4$$

$$= \frac{100}{0.2} = \frac{1000}{2} = 500.$$

(c) Given $n = 100$.

We require to find c such that

$$P\left[|M_n(X) - p| \ge c\right] \le 1 - 0.8$$
$$= 0.2.$$

From Chebyshev's inequality,

$$P\left[|M_n(X) - p| \ge c\right] \le \frac{Var[X]}{nc^2}.$$

Application of Chebyshev's 5/10

Find c that satisfies

$$\frac{Var[X]}{nc^2} = 0.2$$

$$c^2 = \frac{Var[X]}{(0.2)(100)}$$

$Var[X]$ is unknown. Choose the maximum possible value to guarantee the required confidence of 80%.

$$c^2 = \frac{0.25}{(0.2)(100)}$$

$$= \frac{1}{80}$$

Correct c 2/10

$$c = \frac{1}{\sqrt{80}}$$

The confidence interval is

$$\left(0.4 - \frac{1}{\sqrt{80}},\ 0.4 + \frac{1}{\sqrt{80}}\right)$$

with confidence 80%.

Correct conf interval 3/10

Question 3. 20 marks An experiment trial results in a pair of values drawn from the joint distribution of the pair of random variables X and Y. We perform such trials repeatedly till we observe $X < Y$ for the 10^{th} time. Assume X and Y are independent. X is exponential(1) and Y is uniform(0, 1). Answer the following. Show your steps.

1) (10 marks) Derive the distribution of the total number of performed trials.
2) (10 marks) Derive the conditional distribution of X, given that $X < Y$.

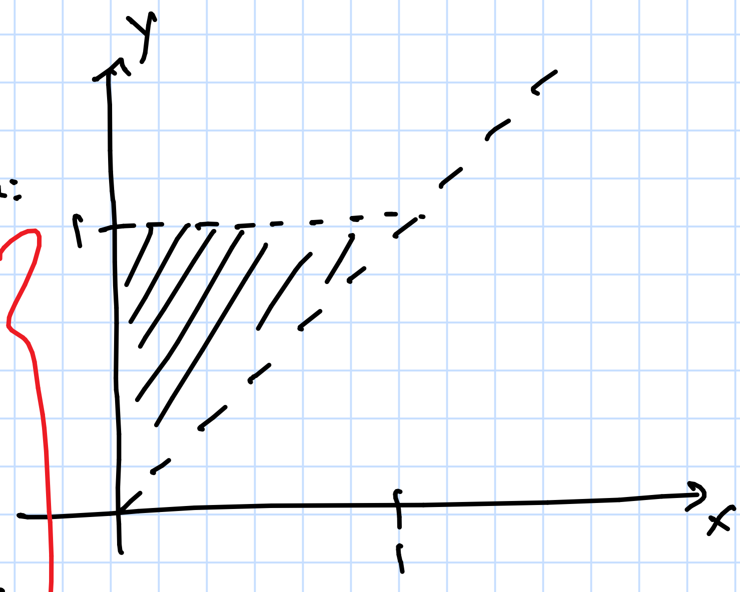
1) Let event $A = \{X < Y\}$

We perform trials till we see A for the 10^{th} time. Let N be the total number of trials.

$$P[N=n] = \begin{cases} {}^{n-1}C_9 \left(P[A]\right)^9 \left(1-P[A]\right)^{n-10} P[A] & n = 10, 11, \dots \\ 0 & \text{otherwise} \end{cases}$$

Correct PMF (range of n is important) Up to 3/10.

$P[A] = P[X < Y]$



X & Y are independent. Therefore:

$$f_{X,Y}(x,y) = \begin{cases} f_X(x) f_Y(y) & 0 < x, \ 0 < y < 1. \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} e^{-x} & 0 < x, \ 0 < y < 1. \\ 0 & \text{otherwise.} \end{cases}$$

Joint PDF Up to 3/10

$$P[X < Y] = \int_0^1 \int_0^y e^{-x} \, dx \, dy$$

$$= \int_0^1 (1 - e^{-y}) \, dy$$

$$= 1 - (1 - e^{-1})$$

$$= 1/e.$$

Up to 4/10

2) $f_{X|X<Y}(x) = ?$

$$f_{X,Y|X<Y}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{P[X<Y]} & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \dfrac{e^{-x}}{e^{-1}} & 0 < x < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Up to 5/10

$$f_{X|X<Y}(x) = \int_x^1 e \, e^{-x} \, dy \qquad 0 < x < 1$$

$$= \begin{cases} e \, e^{-x} (1 - x) & 0 < x < 1. \\ 0 & \text{otherwise.} \end{cases}$$

Up to 5/10

Question 4. **20 marks** A bin consists of 100 balls of which 20 are colored red, 30 are black, 10 are blue, 25 are green and 15 are purple. You draw balls randomly and with replacement and stop once you draw a green ball for the first time. Answer the following questions.

1) (5 marks) Derive the distribution of the number of draws from first principles.
2) (15 marks) Derive the distribution of the number of red balls drawn from the bin.

1) Let X be the no. of draws.

The X^{th} draw is the first draw in which a green ball is observed.

Let p be the probability that a green ball is obtained in any draw.

$$p = \frac{25}{100} = 1/4.$$

$$P[X=x] = \begin{cases} (1-p)^{x-1} p & x = 1, 2, \dots, \\ & \text{otherwise.} \end{cases}$$

Up to 5 marks.

2) Let N be the no. of red balls.

Consider the conditional PMF of N, given $X=x$.

$$P[N=n \mid X=x] = \begin{cases} {}^{x-1}C_n \left(\frac{20}{75}\right)^n \left(\frac{55}{75}\right)^{x-1-n} & n = 0, 1, \dots, x-1 \\ 0 & \text{otherwise.} \end{cases}$$

Up to 5/15

Note that conditioned on $X=x$, a green can't be drawn for the first $x-1$ trials. Therefore, probability that a red ball is drawn is $\frac{20}{75}$ (20 red balls in the total of 75 balls that are not green in color).

$$P[\text{Red is chosen} \mid \text{Green isn't}] = \frac{P[\text{Red is chosen}]}{P[\text{Green isn't}]} = \frac{20/100}{75/100} = 20/75.$$

Up to 5/15

We require $P(N=n)$.

$$P[N=n] = \sum_{x=n+1}^{\infty} P[N=n, X=x]$$

$$= \sum_{x=n+1}^{\infty} P[N=n \mid X=x] \, P[X=x]$$

$$= \sum_{x=n+1}^{\infty} \left({}^{x-1}C_n \left(\frac{20}{75}\right)^n \left(\frac{55}{75}\right)^{x-1-n} \right) \left(\frac{3}{4}\right)^{x-1} \left(\frac{1}{4}\right).$$

Up to 5/15

The range over which x is being summed is important.

Question 5. 30 **marks** An experiment trial involves tossing two coins A and B. For each trial, before the toss, coin A chooses a probability of heads uniformly and randomly from $(0, 1)$. Similarly, coin B chooses a probability of heads uniformly and randomly from $(0, 0.5)$, independently of the choice made by coin A. We perform two such trails. Choices of probabilities made by the coins for the two trails are independent of each other. Coin A is declared winner in case it gives its first heads in an earlier trial than coin B. For example, Coin A gives its first heads in trial 2 and Coin B doesn't given any heads. In case both coins give their first heads in the same trial or if no heads are observed, we declare a draw. (10 **marks) Derive the probability that coin A wins**.

Now consider a slightly modified setting. As before, choices in different trials are independent of each other. However, for either trial, while coin A chooses a probability of heads uniformly and randomly from $(0, 1)$, coin B chooses a probability of heads such that for a choice of probability p by coin A, the conditional distribution from which coin b chooses its probability of heads is $1/p$ over the interval $(0, p)$ and is 0 otherwise. (20 **marks) As before, derive the probability that coin A wins**.

(1) Suppose p_i and q_i are the probabilities of heads chosen by coins A & B respectively in trial i, $i=1,2$.

For a given choice of p_1, q_1, p_2, q_2,

$$P[\text{Coin A wins}] = p_1(1-q_1) + (1-p_1)(1-q_1)p_2(1-q_2).$$

If chosen for a given set of values: Up to 3/10

As given in the question p_i, q_i are chosen from distributions. Let P_i and Q_i be the corresponding r.v.s.

$P_1 \sim \text{uniform}(0,1)$
$P_2 \sim \text{uniform}(0,1)$
$Q_1 \sim \text{uniform}(0,0.5)$
$Q_2 \sim \text{uniform}(0,0.5)$

Since P_1, Q_1, P_2, Q_2 are independent

$$P[\text{Coin A wins}] = \int_0^{0.5}\int_0^1 p(1-q)\, f_{P_1}(p) f_{Q_1}(q)\, dp\, dq$$

Up to 5/10 with correct substitution for the marginal PDFs.

$$+ \left(\int_0^{0.5}\int_0^1 (1-p)(1-q) f_{P_1}(p) f_{Q_1}(q)\, dp\, dq\right)\left(\int_0^{0.5}\int_0^1 p(1-q) f_{P_2}(p) f_{Q_2}(q)\, dp\, dq\right)$$

Product, since the two trials are independent.

Up to 5/10 with correct substitution for the marginal PDFs.

$$= (0.5)(0.75) + (0.5)(0.75)(0.5)(0.75)$$
$$= (0.5)(0.75)\left[1 + (0.5)(0.75)\right]$$
$$= \frac{1}{2}\left(\frac{3}{4}\right)\left[1 + \frac{3}{8}\right]$$
$$= \frac{3}{8}\left[\frac{11}{8}\right] = \frac{33}{64}$$

(2) In part 2, P_i, Q_i are dependent r.v.s for a given i.

$$f_{P_i,Q_i}(p,q) = f_{Q_i|P_i}(q|p)\, f_{P_i}(p),$$

where $f_{Q_i|P_i}(q|p) = \begin{cases} \frac{1}{p} & q \in (0,p) \\ 0 & \text{otherwise.} \end{cases}$

$$\therefore f_{P_i,Q_i}(p,q) = \frac{1}{p} \qquad 0<q<p<1, \text{ for } i=1,2.$$

Correctly stating the conditional and/or joint. Up to 5/20

$$P[A \text{ wins}] = \int_0^1\int_q^1 p(1-q)\, f_{P_1,Q_1}(p,q)\, dp\, dq$$

Up to 5/20

$$+ \left(\int_0^1\int_q^1 (1-p)(1-q) f_{P_1,Q_1}(p,q)\, dp\, dq\right)\left(\int_0^1\int_q^1 p(1-q) f_{P_2,Q_2}(p,q)\, dp\, dq\right)$$

Up to 8/20

Note that f_{P_1,Q_1} & f_{P_2,Q_2} are identical distributions (as given in the question)

$$P[A \text{ wins}] = \int_0^1\int_q^1 p(1-q)\frac{1}{p}\, dp\, dq + \left(\int_0^1\int_q^1 (1-p)(1-q)\frac{1}{p}\, dp\, dq\right)\left(\int_0^1\int_q^1 p(1-q)\frac{1}{p}\, dp\, dq\right).$$

Final answer Up to 2/20

$$\int_0^1\int_q^1 p(1-q)\frac{1}{p}\, dp\, dq$$
$$= \int_0^1 (1-q)\int_q^1 dp\, dq$$
$$= \int_0^1 \left(1 + \frac{q^2}{2} - 2q\right) dq$$
$$= 1 + \frac{1}{2}\left(\frac{1}{3}\right) - \frac{2}{2} = \frac{1}{6}.$$

$$= \underbrace{\int_0^1\int_q^1 \frac{1}{p}(1-q)\, dp\, dq}_{\downarrow} - \underbrace{\int_0^1\int_q^1 p(1-q)\frac{1}{p}\, dp\, dq}_{1/6}$$

$$\int_0^1 (1-q)\int_q^1 \frac{dp}{p}\, dq$$
$$= \int_0^1 (1-q)\left(\ln p\right)_q^1\, dq$$
$$= \int_0^1 (1-q)(-\ln q)\, dq$$
$$= \int_0^1 (-\ln q + q\ln q)\, dq$$
$$= -\int_0^1 \ln q\, dq + \int_0^1 q\ln q\, dq$$
$$= -\left[(q\ln q)_0^1 - \int_0^1 q\frac{1}{q}\, dq\right] + \frac{1}{2}\left((q^2\ln q)_0^1 - \int_0^1 \frac{q^2}{q}\, dq\right)$$
$$= 1 + \frac{1}{2}(-1)\frac{1}{2} = 1 - \frac{1}{4} = \frac{3}{4}.$$

$$\therefore P[A \text{ wins}] = \frac{3}{4} + \frac{1}{6} = \frac{22}{24} = \frac{11}{12}$$