MTH 201: Probability and Statistics

Quiz 2

08/04/2025

Sanjit K. Kaul

No books, notes, or devices are allowed. Just a pen and eraser. Any exchange of information related to the quiz with a human or machine will be deemed as cheating. Institute rules will apply. Explain your answers. Show your steps. Approximate calculations are fine as long as the approximations are reasonable. Don't bother simplifying products and sums that are too time consuming. Leave your answer as products and sums, if you don't have the time. You have 70 minutes.

Question 1. 15 **marks** An experiment requires tossing a coin 50000 times. Outcomes of coin tosses are known to be independent and Bernoulli(0.5). Someone you know is interested in the probability that the experiment results in 10000 or more heads. An approximate value is sufficient. Derive an approximate value of the probability using the Central Limit Theorem approximation, which states that a sum of large enough number of independent and identical random variables is well approximated by a Gaussian distribution. Leave your probability in terms of the CDF $\Phi(x)$ of the standard Normal.

Question 2. 15 **marks** Suppose X_1 and X_2 are independent Bernoulli(0.5) random variables. Let $Y = X_1 + X_2$ and $Z = X_1 - X_2$. Calculate the covariance $E[YZ] - E[Y]E[Z]$ of Y and Z. Are Y and Z independent random variables?

Question 3. 40 **marks** Two professors begin their lectures. The lecture of one of the professors is X hours long, where X is exponential(1). The lecture of the other professor is Y hours long, where Y is exponential(2). The lengths of their lectures are independent of each other. Derive the following.

(a) The conditional PDF of X, given that $X > Y$.

(b) The conditional PDF of X, given that $X > Y$ and $Y > 0.5$.

Question 4. 30 **marks** Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed as random variable X. Consider the sample mean $M_n(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$. Answer the following questions.

(a) Derive the probability (or a bound on it, given that you may not have enough information to calculate it exactly) that the value $E[X]$ is in the interval $(M_n(X) - c, M_n(X) + c)$, $c > 0$. [Hint: See useful facts.]

(b) Derive the probability that the value $E[X]$ is in the interval $(M_n(X) - c, M_n(X) + c)$, $c > 0$, when X is Gaussian$(0, 1)$. Leave your result in terms of the CDF $\Phi(x)$ of the standard Normal. [Hint: The sum of independent Gaussian random variables is Gaussian. Scaling a Gaussian also results in a Gaussian.]

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. 15 **marks** An experiment requires tossing a coin 50000 times. Outcomes of coin tosses are known to be independent and Bernoulli(0.5). Someone you know is interested in the probability that the experiment results in 10000 or more heads. An approximate value is sufficient. Derive an approximate value of the probability using the Central Limit Theorem approximation, which states that a sum of large enough number of independent and identical random variables is well approximated by a Gaussian distribution. Leave your probability in terms of the CDF $\Phi(x)$ of the standard Normal.

Let N be the number of heads.

We are interested in an approximation of $P[N \geq 10000]$.

Note that N is a sum of 50000 Bernoulli(0.5) random variables. Specifically, let $X_1, X_2, \ldots, X_{50000}$ be the Bernoulli(0.5) rvs corresponding to the 50000 coin tosses.

$$N = \sum_{i=1}^{50000} X_i.$$

Recognizing the sum.

5/15

We can approximate the sum using a Gaussian $(E[N], Var[N])$.

$$E[N] = (50000)(0.5) = 25000.$$

$$Var[N] = Var\left[\sum_{i=1}^{50000} X_i\right]$$

$$= \sum_{i=1}^{50000} Var[X_i]$$

$$= (50000)(0.25)$$

$$= (500)(25)$$

$$= 12500.$$

Equivalent Gaussian

7/15

$$P[N \geq 10000]$$

$$= P[N - 25000 \geq -15000]$$

$$= P\left[\frac{N-25000}{\sqrt{12500}} \geq \frac{-15000}{\sqrt{12500}}\right]$$

3/15

Approximating N as Gaussian, observe that $\frac{N-25000}{\sqrt{12500}}$ is the standard normal Z.

$$\therefore P[N \geq 10000] \approx P\left[Z \geq \frac{-15000}{50\sqrt{5}}\right]$$

$$= 1 - \Phi\left(\frac{-300}{\sqrt{5}}\right)$$

$$\approx 1.$$

Question 2. 15 **marks** Suppose X_1 and X_2 are independent Bernoulli(0.5) random variables. Let $Y = X_1 + X_2$ and $Z = X_1 - X_2$. Calculate the covariance $E[YZ] - E[Y]E[Z]$ of Y and Z. Are Y and Z independent random variables?

$$E[YZ] = E[(X_1 + X_2)(X_1 - X_2)]$$
$$= E[X_1^2 - X_1X_2 + X_2X_1 - X_2^2]$$
$$= E[X_1^2] - E[X_2^2] = 0.$$

$$E[Y] = E[X_1 + X_2] = 2(0.5) = 1.$$
$$E[Z] = E[X_1 - X_2] = 0.$$

00
01
10
11

Covariance is $E[YZ] - E[Y]E[Z]$
$= 0.$

Random variables that have a covariance of zero may not be independent.

Calculating covariance

7.5/15

Consider the PMF of Z:

$$P_Z(z) = \begin{cases} 1/4 & z = -1 \\ 1/4 + 1/4 & z = 0 \\ 1/4 & z = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Consider $P[Z=0 \mid Y=1]$.

Since $Z = X_1 - X_2$ & $Y = X_1 + X_2$,

$$P[Z=0 \mid Y=1] = 0 \neq P[Z=0].$$

$\therefore$ Y & Z are not independent.

7.5/15

Any valid mathematical approach to showing the rvs are not independent.

Question 3. 40 **marks** Two professors begin their lectures. The lecture of one of the professors is X hours long, where X is exponential(1). The lecture of the other professor is Y hours long, where Y is exponential(2). The lengths of their lectures are independent of each other. Derive the following.

(a) The conditional PDF of X, given that $X > Y$.

(b) The conditional PDF of X, given that $X > Y$ and $Y > 0.5$.

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

First we calculate the conditional joint PDF

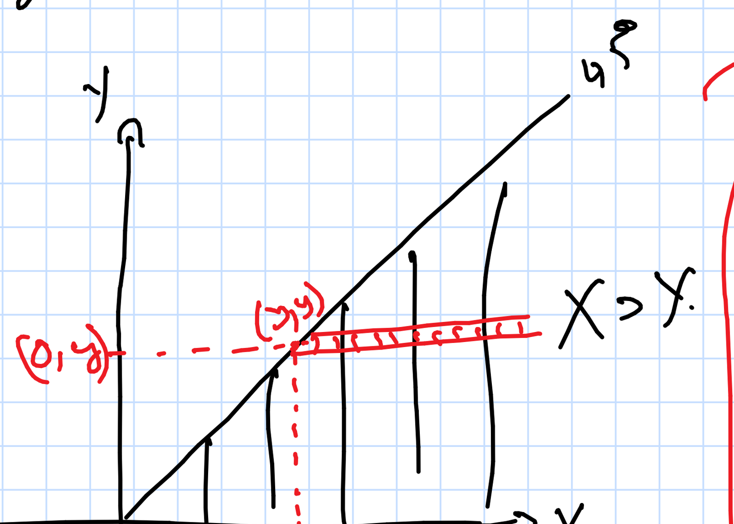
$$f_{X,Y|X>Y}(x,y).$$

$$f_{X,Y|X>Y}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{P(X>Y)} & x > y \\ 0 & \text{otherwise.} \end{cases}$$

$$f_{X,Y}(x,y) = e^{-x}\, 2e^{-2y}$$ Since X & Y are independent, the joint is product of the marginals.

5/20

$$P[X>Y] = \int_0^{\infty}\int_y^{\infty} f_{X,Y}(x,y)\,dx\,dy$$



$$= \int_0^{\infty}\left(\int_y^{\infty} e^{-x}\,dx\right) 2e^{-2y}\,dy$$

$$= \int_0^{\infty} e^{-y}\left(2e^{-2y}\right)dy$$

$$= \int_0^{\infty}\left(\frac{2}{3}\right) 3e^{-3y}\,dy$$

$$= \frac{2}{3}\int_0^{\infty} \underbrace{3e^{-3y}}_{\exp(3)}\,dy.$$

$$= 2/3.$$

5/20

Substituting, we get:

$$f_{X,Y|X>Y}(x,y) = \begin{cases} 3e^{-2y}e^{-x} & x > y \\ 0 & \text{otherwise.} \end{cases}$$

5/20

$$\therefore f_{X|X>Y}(x) = \int_0^{x} 3e^{-2y}e^{-x}\,dy$$

$$= \frac{3e^{-x}}{2}\int_0^{x} 2e^{-2y}\,dy$$

$$= \begin{cases} \frac{3}{2}e^{-x}\left(1-e^{-2x}\right) & x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

5/20

(b) $f_{X|X>Y,Y>0.5}(x)$?

First calculate

$$f_{X,Y|X>Y,Y>0.5}(x,y)$$

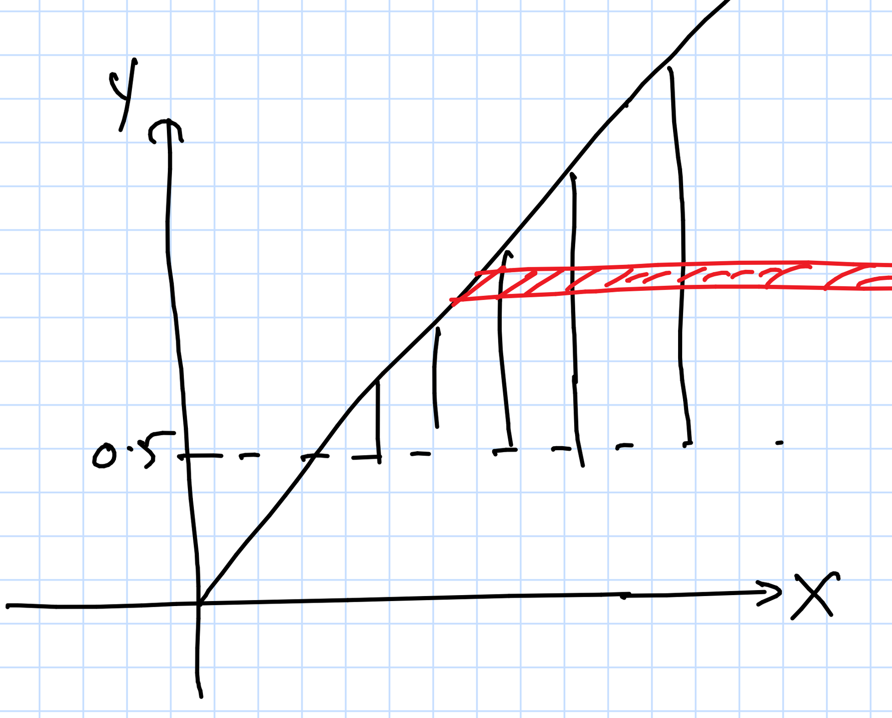
$$= \begin{cases} \dfrac{f_{X,Y}(x,y)}{P[X>Y, Y>0.5]} & 0.5 < y < x, \\ 0 & \text{otherwise} \end{cases}$$

5/20

$f_{X,Y}(x,y)$ is as in part (a).

$$P[X>Y, Y>0.5]$$

$$= \int_{0.5}^{\infty}\int_y^{\infty} f_{X,Y}(x,y)\,dx\,dy$$



$$= \int_{0.5}^{\infty}\left(e^{-y}\right)\left(2e^{-2y}\right)dy$$

$$= \frac{2}{3}\int_{0.5}^{\infty} 3e^{-3y}\,dy$$

$$= \left(\frac{2}{3}\right)e^{-3(0.5)}$$

$$= \frac{2}{3}e^{-3/2}.$$

5/20

Therefore:

$$f_{X,Y|X>Y,Y>0.5}(x,y) = \begin{cases} \dfrac{e^{-x}\left(2e^{-2y}\right)}{(2/3)\,e^{-3/2}} & 0.5 < y < x \\ 0 & \text{otherwise.} \end{cases}$$

5/20

The marginal

$$f_{X|X>Y,Y>0.5}(x)$$

$$= \frac{3e^{-x}}{2e^{-3/2}}\int_{0.5}^{x} 2e^{-2y}\,dy$$

$$= \begin{cases} \dfrac{3e^{-x}}{2e^{-3/2}}\left(e^{-1}-e^{-2x}\right) & x \geq 0.5, \\ 0 & \text{otherwise.} \end{cases}$$

5/20

Question 4. 30 **marks** Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed as random variable X. Consider the sample mean $M_n(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$. Answer the following questions.

(a) Derive the probability (or a bound on it, given that you may not have enough information to calculate it exactly) that the value $E[X]$ is in the interval $(M_n(X) - c, M_n(X) + c)$, $c > 0$. [Hint: See useful facts.]

(b) Derive the probability that the value $E[X]$ is in the interval $(M_n(X) - c, M_n(X) + c)$, $c > 0$, when X is Gaussian($\cancel{0}$ E[X], 1). Leave your result in terms of the CDF $\Phi(x)$ of the standard Normal. [Hint: The sum of independent Gaussian random variables is Gaussian. Scaling a Gaussian also results in a Gaussian.]

(a) $P\left[M_n(X) - c \le E[X] \le M_n(X) + c\right]$

$= P\left[-c \le M_n(X) - E(X) \le c\right]$

$= P\left[-c \le M_n(X) - E[M_n(X)] \le c\right]$ { Since $E[M_n(X)] = E[X]$.

$\ge 1 - \dfrac{Var[M_n(X)]}{c^2}$ { Chebyshev's inequality.

$= 1 - \dfrac{Var(X)}{nc^2}.$

Total 15

5/15

(b) Given the hint,

$M_n(X)$ is Gaussian with mean $E[X]$ and variance $\frac{1}{n}$.

5/15

In the question we say E[X]=1. The known mean makes the need to estimate redundant. In any case, given the typo, it is okay if students use 1 instead of E[X].

$P\left[M_n(X) - c \le E[X] \le M_n(X) + c\right]$

$= P\left[-c \le M_n(X) - E(X) \le c\right]$

$= P\left[-c \le M_n(X) - E[M_n(X)] \le c\right]$

$= P\left[\dfrac{-c}{1/\sqrt{n}} \le \dfrac{M_n(X) - E[M_n(X)]}{1/\sqrt{n}} \le \dfrac{c}{1/\sqrt{n}}\right]$

key steps.

10/15

$= \Phi(\sqrt{n}\,c) - \Phi(-\sqrt{n}\,c).$