19/20 Solution

Tuesday, November 9, 2021 1:48 AM

NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER I EXAMINATION 2019-2020

MH2500 – Probability and Introduction to Statistics

December 2019 TIME ALLOWED: 2 HOURS

INSTRUCTIONS TO CANDIDATES

1. This examination paper contains **FIVE (5)** questions and comprises **FIVE (5)** printed pages including appendix.
2. Answer **ALL** questions. The marks for each question are indicated at the beginning of each question.
3. Answer each question beginning on a **FRESH** page of the answer book.
4. This is a **RESTRICTED OPEN BOOK** exam. Candidates are allowed **ONE** double-sided A4 size help sheet.
5. Candidates may use calculators. However, they should write down systematically the steps in the workings.
6. The table with the values of the standard normal distribution is included at the end of this examination paper.
7. All questions are the property of Nanyang Technological University.

QUESTION 1 **(15 marks)**

An urn contains 4 blue balls and 3 red balls. A fair die is rolled. If the number shown on the die is i ($i = 1, 2, \cdots, 6$), then i balls are drawn from the urn uniformly at random without replacement.

(a) (5 marks) When $i = 3$, what is the probability that all balls drawn are red?

(b) (10 marks) Given that all balls drawn are red, what is the probability that the remaining balls in the urn are all blue?

(a) $P(\text{all balls are red}) = \dfrac{\binom{4}{0}\binom{3}{3}}{\binom{7}{3}} = \dfrac{1}{35}$ //

(b) Let R be the event that all balls drawn are Red.

Let B be the event that remaining balls in the urn are all Blue.

$$P(B|R) = \frac{P(B \cap R)}{P(R)}$$

All 4 Blue left + All balls are red.

Bayes' Rule.

$$= \frac{P(R_3)}{P(R_1) + P(R_2) + P(R_3)}$$

$$= \frac{P(R_3 \mid i=3)\, P(i=3)}{P(R_1 \mid i=1)\, P(i=1) + P(R_2 \mid i=2) P(i=2) + P(R_3 \mid i=3) P(i=3)}$$

Note that $P(i=1) = P(i=2) = P(i=3) = \frac{1}{6}$

$$= \frac{\frac{1}{35}}{\frac{\binom{4}{0}\binom{3}{1}}{\binom{7}{1}} + \frac{\binom{4}{0}\binom{3}{2}}{\binom{7}{2}} + \frac{1}{35}} = \frac{\frac{1}{35}}{\frac{3}{7} + \frac{1}{7} + \frac{1}{35}} = \frac{1}{21}$$ //

(a) Note that $D_\theta = A_0 \cap A_\theta$

A_0 : $\{f(x,y): |x\cos 0 - y\sin 0| < 1\} = \{f(x,y): |x| < 1\}$

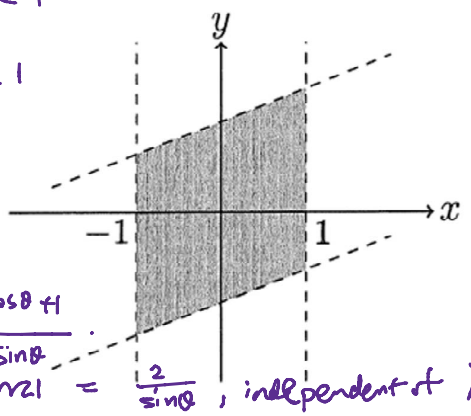
$= -1 < x < 1$.

A_θ

$|x\cos\theta - y\sin\theta| < 1$

$\Leftrightarrow -1 < x\cos\theta - y\sin\theta < 1$

$\Leftrightarrow \frac{x\cos\theta - 1}{\sin\theta} < y < \frac{x\cos\theta + 1}{\sin\theta}$



$\therefore$ Region of D_θ

$\Rightarrow -1 < x < 1$ and $\frac{x\cos\theta - 1}{\sin\theta} < y < \frac{x\cos\theta + 1}{\sin\theta}$.

X is uniformly distributed, length of Y's interval $= \frac{2}{\sin\theta}$, independent of X.

$f_X(x) = \frac{1}{1-(-1)} = \frac{1}{2}$ if $-1 < x < 1$.

MH2500

QUESTION 2 (30 marks)

For each $\theta \in [0, \pi/2)$, we define a planar region A_θ as

$$A_\theta = \{f(x,y) : |x\cos\theta - y\sin\theta| < 1\}, \quad 0 \le \theta < \frac{\pi}{2}.$$

For each $\theta \in (0, \pi/2)$, we define a planar region D_θ as

$$D_\theta = A_0 \cap A_\theta, \quad 0 < \theta < \frac{\pi}{2}.$$

The shaded region on the right is a sketch of $D_{6/5}$.

Let (X, Y) be a point uniformly distributed in D_θ for some $\theta \in (0, \pi/2)$.

(a) (5 marks) Find the marginal distribution of X.

(b) (5 marks) Find the conditional density function of Y given $X = x$.

(c) (5 marks) Find the conditional expectation $\mathbb{E}(Y|X)$.

(d) (5 marks) Find $\mathbb{E}(Y)$.

(e) (5 marks) Find $\mathrm{Var}(Y)$.

(f) (5 marks) Are X and Y correlated? If they are correlated, are they positively or negatively correlated? If they are not correlated, are they independent? State your reason.

Extra:

$f_{XY}(x,y) = \frac{1}{\text{Area of Region}}$

$= \frac{1}{2 \times \frac{2}{\sin\theta}}$

$= \frac{\sin\theta}{4}$

if $-1 < x < 1$,

$\frac{x\cos\theta - 1}{\sin\theta} < y < \frac{x\cos\theta + 1}{\sin\theta}$.

(b) $f_{Y|X=x}(y|x)$ (Also uniform.) $= \begin{cases} \frac{1}{\frac{2}{\sin\theta}} = \frac{\sin\theta}{2} & \text{if } y \in \left(\frac{x\cos\theta - 1}{\sin\theta}, \frac{x\cos\theta + 1}{\sin\theta}\right) \\ 0 & \text{otherwise.} \end{cases}$

(c) $E(Y|X) = \dfrac{\frac{X\cos\theta - 1}{\sin\theta} + \frac{X\cos\theta + 1}{\sin\theta}}{2}$ (uniform distribution)

$= \dfrac{X\cos\theta}{\sin\theta}$ //

(d) $E(Y) = E[E(Y|X)]$

$= E\left[\dfrac{X\cos\theta}{\sin\theta}\right]$

$= \dfrac{\cos\theta}{\sin\theta} E(X)$

$= \dfrac{\cos\theta}{\sin\theta}\left[\dfrac{-1+1}{2}\right]$ → Marginal density of X is uniform.

$= 0$ //

(e) $\mathrm{Var}(Y) = \mathrm{Var}[E(Y|X)] + E[\mathrm{Var}(Y|X)]$

$= \mathrm{Var}\left[\frac{X\cos\theta}{\sin\theta}\right] + \frac{\left(\frac{X\cos\theta+1}{\sin\theta} - \frac{X\cos\theta-1}{\sin\theta}\right)^2}{12}$

$= \frac{\cos^2\theta}{\sin^2\theta}\mathrm{Var}(X) + \frac{1}{3\sin^2\theta}$

$= \frac{\cos^2\theta}{\sin^2\theta}\left[E(X^2) - \{E(X)\}^2\right] + \frac{1}{3\sin^2\theta}$

$= \frac{\cos^2\theta}{\sin^2\theta}\left[\int_{-1}^{1} x^2\left(\frac{1}{2}\right)dx - 0^2\right] + \frac{1}{3\sin^2\theta}$

$= \frac{\cos^2\theta}{\sin^2\theta}\left[\frac{1}{2}\left[\frac{x^3}{3}\right]_{-1}^{1}\right] + \frac{1}{3\sin^2\theta}$

$= \frac{\cos^2\theta + 1}{3\sin^2\theta}$ //

(f) $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$ ($E(X)$ is a wavy line, $= 0$)

$= E(XY)$

$= E(XE(Y|X))$

$= E\left(X\cdot\left(\frac{X\cos\theta}{\sin\theta}\right)\right)$

$= \frac{\cos\theta}{\sin\theta}E(X^2)$

$= \frac{\cos\theta}{3\sin\theta} > 0$

positively correlated.

QUESTION 3 **(20 marks)**

Suppose that a fair die is rolled n times independently. For each $i = 1, \cdots, 6$, let X_i denote the number of times i appears.

(a) (2 marks) Find the probability mass function of X_2.

(b) (2 marks) Find the joint probability mass function of (X_2, X_4).

(c) (2 marks) Find the conditional probability mass function of X_2 given that $X_4 = l$ $(0 \le l \le n)$.

(d) (4 marks) Find the probability mass function of $X_2 + X_4$.

(e) (10 marks) Find $\mathrm{Cov}(X_2, X_4)$.

(a) $X_2 \sim \mathrm{Bin}(n, \frac{1}{6})$.

pmf of X_2: $\binom{n}{x_2}\left(\frac{1}{6}\right)^{x_2}\left(\frac{5}{6}\right)^{n-x_2}$ $P(X_2 = x_2)$

(b) Joint pmf of X_2, X_4:

$$P(X_2 = x_2, X_4 = x_4) = \binom{n}{x_2, x_4}\left(\frac{1}{6}\right)^{x_2}\left(\frac{1}{6}\right)^{x_4}\left(\frac{4}{6}\right)^{n-x_2-x_4}$$

(c) $$P(X_2 \mid X_4 = l) = \binom{n-l}{x_2}\left(\frac{1}{5}\right)^{x_2}\left(\frac{4}{5}\right)^{n-l-x_2}$$

$X_2 \mid X_4 = l \sim \mathrm{Bin}(n-l, \frac{1}{5})$

(a) Let $Z = X_2 + X_4$, $X_2 + X_4 \sim \mathrm{Bin}(n, \frac{1}{3})$

pmf of Z: $\binom{n}{z}\left(\frac{1}{3}\right)^{n-z}\left(\frac{2}{3}\right)^{z}$

(e) Preferred way:

$$Var(X_2+X_4) = Var(X_2) + Var(X_4) + Cov(X_2,X_4) + Cov(X_4,X_2)$$

$$Var(X_2+X_4) = Var(X_2) + Var(X_4) + 2Cov(X_2,X_4)$$

$X_2+X_4 \sim Bin(n, \frac{1}{3})$, $X_2 \sim Bin(n, \frac{1}{6})$, $X_4 \sim Bin(n, \frac{1}{6})$

$Var(X_2+X_4) = n(\frac{1}{3})(1-\frac{1}{3})$ $\qquad$ $Var(X_2) = Var(X_4) = n(\frac{1}{6})(\frac{5}{6})$

$= \frac{2}{9}n$ $\qquad$ $= \frac{5}{36}n$.

$$\therefore \quad \frac{2}{9}n = \frac{5}{36}n + \frac{5}{36}n + 2Cov(X_2,X_4)$$

$$\therefore Cov(X_2,Y_4) = -\frac{n}{36} //$$

QUESTION 4 **(20 marks)**

For each integer $m \geq 1$, we define $2m$ numbers $b_{m,1}, \cdots, b_{m,2m}$ as follows:

$$b_{m,i} = \begin{cases} 1, & 1 \leq i \leq m; \\ -1, & m+1 \leq i \leq 2m. \end{cases}$$

} Constants!

Let $Y_1, Y_2, \cdots$ be i.i.d. random variables such that $\Pr\{Y_i = 0\} = \Pr\{Y_i = 1\} = 1/2$.

(a) (10 marks) Show that there exists a constant $c_1 > 0$ (which does not depend on m) such that for all $m \geq 1$,

$$\Pr\left\{\left|\sum_{i=1}^{2m} b_{m,i} Y_i\right| > c_1\sqrt{m}\right\} \leq \frac{1}{10}.$$

(b) (10 marks) Show that there exists a constant $c_2 > 0$ (which does not depend on m) such that for all $m \geq 1$,

$$\lim_{m\to\infty} \Pr\left\{\left|\sum_{i=1}^{2m} b_{m,i} Y_i\right| > c_2\sqrt{m}\right\} > \frac{9}{10}.$$

(a) $E\left(\sum_{i=1}^{2m} b_{m,i} Y_i\right) = E\left[\sum_{i=1}^{m} b_{m,i} Y_i\right] + E\left[\sum_{i=m+1}^{2m} b_{m,i} Y_i\right]$

$E[Y_i] = \frac{1}{2}(0) + \frac{1}{2}(1) = \frac{1}{2}$

$\text{Var}[Y_i] = E[Y_i^2] - \{E[Y_i]\}^2$
$= \frac{1}{2} - \left(\frac{1}{2}\right)^2$
$= \frac{1}{4}$

$= E\left[\sum_{i=1}^{m} Y_i\right] + E\left[-\sum_{i=m+1}^{2m} Y_i\right]$

$= \frac{1}{2}m - \frac{1}{2}m = 0$

Note: Question 5 (Hypothesis Testing) is not tested for the Final Exam.

END OF PAPER

$$\mathrm{Var}\left(\sum_{i=1}^{2m} b_{m,i} Y_i\right) = \mathrm{Var}\left(\sum_{i=1}^{m} Y_i\right) + \mathrm{Var}\left(\sum_{i=m+1}^{2m} Y_i\right)$$

$$= 2m\,\mathrm{Var}(Y_i)$$

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

$$= \frac{m}{2}$$

By Chebyshev's Inequality,

$$P\left(\left|\sum_{i=1}^{2m} b_{m,i} Y_i - 0\right| > c_1\sqrt{m}\right) \le \frac{\mathrm{Var}\left(\sum_{i=1}^{2m} b_{m,i} Y_i\right)}{(c_1\sqrt{m})^2}$$

$$P\left(\left|\sum_{i=1}^{2m} b_{m,i} Y_i\right| > c_1\sqrt{m}\right) \le \frac{\frac{m}{2}}{c_1^2 m}$$

$$= \frac{1}{2c_1^2}$$

Take any $c_1 \ge \sqrt{5}$, $\frac{1}{2c_1^2} \le \frac{1}{10}$

$\therefore$ c_1 exists.

(b) $$\sum_{i=1}^{2m} b_{m,i} Y_i = \sum_{i=1}^{m} b_{m,i} Y_i + \sum_{i=m+1}^{2m} b_{m,i} Y_i$$

$$= \sum_{i=1}^{m} Y_i + \sum_{i=m+1}^{2m} -Y_i$$

$$= \sum_{i=1}^{m} (Y_i - Y_{m+i})$$

we see that $Y_i - Y_{m+i}$ for $i = 1, \ldots, m$

is i.i.d., $E(Y_i - Y_{m+i}) = 0$ $\quad Var(Y_i - Y_{m+i})$

$= 2 Var(Y_i)$

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

$= \frac{1}{2}$

$$P\left(\left| \sum_{i=1}^{2m} b_{m,i} Y_i \right| > c_2 \sqrt{m} \right)$$

$$= 1 - P\left(\left| \sum_{i=1}^{2m} b_{m,i} Y_i \right| \leq c_2 \sqrt{m} \right)$$

$$= 1 - P\left(-c_2 \sqrt{m} \leq \sum_{i=1}^{2m} b_{m,i} Y_i \leq c_2 \sqrt{m} \right)$$

$$= 1 - P\left(-C_2\sqrt{m} \le \sum_{i=1}^{m} (Y_i - Y_{m+i}) \le C_2\sqrt{m}\right)$$

Applying CLT,

$$= 1 - P\left(\frac{-C_2\sqrt{m} - 0}{\sqrt{\frac{1}{2}m}} \le Z \le \frac{C_2\sqrt{m} - 0}{\sqrt{\frac{1}{2}m}}\right)$$

$$= 1 - P\left(\sqrt{2}C_2 \le Z \le \sqrt{2}C_2\right)$$

$$= 1 - 2\left[\phi(\sqrt{2}C_2) - 0.5\right]$$

$$= 2 - 2\phi(\sqrt{2}C_2) > 0.9$$

$$-2\phi(\sqrt{2}C_2) > 1.1$$

$$\phi(\sqrt{2}C_2) < 0.55$$

Since $C_2 > 0$, $\sqrt{2}C_2 < \phi^{-1}(0.55)$

$$0 < C_2 < \frac{1}{\sqrt{2}}\phi^{-1}(0.55)$$

$$0 < C_2 < \frac{0.12}{\sqrt{2}}$$

$\therefore C_2$ exists.