Probability and Statistics
Model Answer of 2022 Midterm Exam
BJTU Lancaster University College

Yiping Cheng

November 12, 2022

1. (20pt) Fill in the blanks:

(1) Events A and B are disjoint and independent, $P(A) = 0.5$, then $P(A) + P(B) = \underline{\mathbf{0.5}}$.

(2) Out of the students in a class, 50% are geniuses, 70% love chocolate. Let p be the conditional probability that a randomly selected student is a chocolate lover given that he/she is a genius. Then the theoretically lowest value of p is $\underline{\mathbf{0.4}}$.

(3) $2n$ students are assigned randomly to 2 classes each having n students. A couple of boyfriend and girlfriend are in the $2n$ students. The probability that this couple are in the same class is $\underline{\frac{\mathbf{n-1}}{\mathbf{2n-1}}}$. (or $\underline{\frac{\mathbf{2C_{2n-2}^{n}}}{\mathbf{C_{2n}^{n}}}}$)

(4) When it rains, there is a 40% chance that students play football on that day. On the other hand, when it does not rain there is a 90% chance that they play. The probability that it will rain tomorrow is 0.2. The probability that students play football tomorrow is $\underline{\mathbf{0.8}}$.

(5) Let $X \sim B(20, 0.8)$, then the most likely value of X is $\underline{\mathbf{16}}$.

(6) The random variable X has PDF $f(x) = Ce^{-3|x|}$. Then $C = \underline{\frac{\mathbf{3}}{\mathbf{2}}}$ and $P(|X| < 1) = \underline{\mathbf{1 - e^{-3}}}$.

(7) Let $X \sim U(0, 1)$, $Y \sim U(0, 1)$, and X and Y be independent. Then $P(X = Y) = \underline{\mathbf{0}}$ and $P((X - 1/2)^2 + (Y - 1/2)^2 < 1/4) = \underline{\frac{\pi}{\mathbf{4}}}$.

(8) The joint CDF of X and Y is $F(x, y) = \begin{cases} \Phi(x)(1 - e^{-y}), & \text{for } y > 0 \\ 0, & \text{otherwise.} \end{cases}$ Then the joint PDF of X and Y is $f(x, y) = \underline{\begin{cases} \varphi(x)e^{-y}, & \text{for } y > 0 \\ 0, & \text{otherwise.} \end{cases}}$

2. (10pt) Approximately 0.04% of human have liver cancer. A person with liver cancer has a 95% chance of a positive test, while a person without liver cancer has a 2% chance of a false positive result. What is the probability a person has liver cancer given that the person just had a positive test?

Solution: Let A denote the event that he is tested positive. Let D denote the event that he has liver cancer. Then

$$P(D) = 0.0004, P(A|D) = 0.95, P(A|D^C) = 0.02. \quad \textbf{(3pt)}$$

$$P(D^C) = 1 - P(D) = 0.9996. \quad \textbf{(1pt)}$$

$$\begin{aligned} P(D|A) &= \frac{P(AD)}{P(A)}) && \textbf{(1pt)} \\ &= \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D^C)P(D^C)} && \textbf{(2pt)} \\ &= \frac{0.95 * 0.0004}{0.95 * 0.0004 + 0.02 * 0.9996} && \textbf{(1pt)} \\ &= 0.018653053 && \textbf{(2pt)} \end{aligned}$$

3. (10pt) Suppose that random variable X and X itself are independent. Show that there must exist a constant C such that $P(X = C) = 1$. (Hint: consider its CDF and use the properties of CDFs)

Proof: For any $x \in \mathbb{R}$, since X and X are independent,

$$P(X \le x)P(X \le x) = P(X \le x \wedge X \le x) = P(X \le x). \quad \textbf{(3pt)}$$

Hence $P(X \le x) = 0$ or 1. **(2pt)**

Let C be the infimum of the set $\{x \in \mathbb{R} | P(X \le x) = 1\}$. For any $a < C$, $P(X \le a) = 0$, hence $P(X < C) = 0$. **(2pt)** For any $b > C$, $P(X \le b) = 1$. By right-continuity of the CDF, we have $P(X \le C) = 1$. **(2pt)**

Therefore the CDF of X has a jump of magnitude 1 at C, that is,

$$P(X = C) = P(X \le C) - P(X < C) = 1 - 0 = 1. \quad \textbf{(2pt)}$$

4. (10pt) A fair coin is tossed two times independently. Let

$$H_i = \begin{cases} 1, & \text{if head is obtained at the } i\text{th toss,} \\ 0, & \text{otherwise.} \end{cases}$$

Let $X = H_1 + H_2$, $Y = |H_1 - H_2|$. Give the joint PMF of X and Y in table form.

Solution: X can take 0,1,2 and Y can take 0,1.**(2pt)**

$$P(X = 0, Y = 0) = P(H_1 = 0, H_2 = 0) = \frac{1}{4} \quad \textbf{(1pt)}$$

$$P(X = 0, Y = 1) = 0 \quad \textbf{(1pt)}$$

$$P(X = 1, Y = 0) = 0 \quad \textbf{(1pt)}$$

$$P(X = 1, Y = 1) = P(H_1 = 0, H_2 = 1) + P(H_1 = 1, H_2 = 0) = \frac{1}{2} \quad \textbf{(2pt)}$$

$$P(X = 2, Y = 0) = P(H_1 = 1, H_2 = 1) = \frac{1}{4} \quad \textbf{(1pt)}$$

$$P(X = 2, Y = 1) = 0. \quad \textbf{(1pt)}$$

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	$\frac{1}{4}$	0	$\frac{1}{4}$
$Y = 1$	0	$\frac{1}{2}$	0

(1pt)

5. (10pt) A book has 300 pages. For each page, whether it contains typing error has Bernoulli distribution $B(1, 0.02)$. The events for each page to contain typing error are independent. What is (approximately) the probability that this book contains at most 3 pages with typing error?

Solution: Let X be the number of pages with typing error. Then $X \sim B(300, 0.02)$. **(2pt)**

Hand computing probabilities for binomial distributions is troublesome when n is large. So we use approximation. It is known that X approximately has $Poisson(\lambda = 6)$. **(2pt)**

Hence

$$\begin{aligned} P(X \leq 3) &\approx \sum_{k=0}^{3} \frac{6^k}{k!} e^{-6} = e^{-6}(1 + 6 + 18 + 36) && \textbf{(4pt)} \\ &= 0.151204. && \textbf{(2pt)} \end{aligned}$$

6. (12pt) Scores on a certain standardized test, IQ scores, are approximately normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 15$. Here we are referring to the distribution of scores over a very large population, and we approximate that discrete cumulative distribution function by a normal continuous cumulative distribution function.

(a) An individual is selected at random. What is the probability that his score X satisfies $120 < X < 130$?

(b) Two individuals are selected at random, independently. What is the probability that their scores X and Y satisfy $\max(X, Y) > 120$?

Solution: (a)

$$\begin{aligned} P(120 < X < 130) &= P\left(\frac{120 - 100}{15} < \frac{X - \mu}{\sigma} < \frac{130 - 100}{15}\right) && \textbf{(2pt)} \\ &= \Phi(2) - \Phi\left(\frac{4}{3}\right) && \textbf{(2pt)} \\ &= 0.068461088. && \textbf{(1pt)} \end{aligned}$$

(b)

$$\begin{aligned} P(\max(X, Y) > 120) &= 1 - P(\max(X, Y) \leq 120) && \textbf{(2pt)} \\ &= 1 - P(X \leq 120)P(Y \leq 120) && \textbf{(2pt)} \\ &= 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{120 - 100}{15}\right)P\left(\frac{Y - \mu}{\sigma} \leq \frac{120 - 100}{15}\right) && \textbf{(1pt)} \\ &= 1 - \Phi^2\left(\frac{4}{3}\right) = 0.174102953. && \textbf{(2pt)} \end{aligned}$$

7. (14pt) Suppose that the joint distribution of X and Y is uniform over the region in the xy-plane bounded by the four lines $x = -1$, $x = 1$, $y = x + 1$, and $y = x - 1$. Determine

(a) $P(XY > 0)$

(b) the joint PDF of X and Y

(c) the PDF of X

(d) the conditional PDF of Y given that $X = x$

(e) $P(Y > 0 | X = 1/2)$**.**

Solution: (a) The area of the region bounded by the four lines is 4, and the area of the region $xy > 0$ is 3, so $P(XY > 0) = \frac{3}{4}$. **(2pt)**

(b)

$$f(x,y) = \begin{cases} \frac{1}{4} & \text{for } -1 < x < 1, x-1 < y < x+1 \\ 0 & \text{otherwise.} \end{cases} \quad \textbf{(2pt)}$$

(c)

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{for } -1 < x < 1 \\ 0 & \text{otherwise.} \end{cases} \quad \textbf{(3pt)}$$

(d) If $-1 < x < 1$, **(1pt)**

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{2} & \text{for } -1 < x < 1, x-1 < y < x+1 \\ 0 & \text{otherwise.} \end{cases} \quad \textbf{(2pt)}$$

(e)

$$\begin{aligned} P(Y > 0|X = 1/2) &= \int_0^\infty f_{Y|X}(y|\frac{1}{2})dy \quad \textbf{(2pt)} \\ &= \int_0^{\frac{3}{2}} \frac{1}{2}dy = \frac{3}{4} \quad \textbf{(2pt)} \end{aligned}$$

8. (14pt) Let X and Y be random variables for which the joint PDF is as follows:

$$f(x,y) = \begin{cases} 8xy & \textbf{for } 0 < x < y < 1, \\ 0 & \textbf{otherwise.} \end{cases}$$

Find the PDF of $Z = (X-Y)^2$.

Solution: Let $f_Z(\cdot)$ be the PDF of Z. Let $W = X - Y$ and $f_W(\cdot)$ be the PDF of W. Then

$$f_W(w) = \int_{-\infty}^{\infty} f(x, x-w)dx. \quad \textbf{(2pt)}$$

If $w \le -1$ or $w \ge 0$, then $f_W(w) = 0$. **(1pt)**
If $-1 < w < 0$, then

$$\begin{aligned} f_W(w) &= \int_0^{1+w} 8x(x-w)dx = \int_0^{1+w} (8x^2 - 8xw)dx \quad \textbf{(2pt)} \\ &= (\frac{8}{3}x^3 - 4x^2w)|_{x=0}^{x=1+w} = \frac{4}{3}(1+w)^2(2-w). \quad \textbf{(2pt)} \end{aligned}$$

If $z \le 0$ or $z \ge 1$, then $f_Z(z) = 0$. **(1pt)**
If $0 < z < 1$, then

$$\begin{aligned} f_Z(z) &= f_W(-\sqrt{z})|\frac{d(-\sqrt{z})}{dz}| \quad \textbf{(2pt)} \\ &= \frac{4}{3}(1-\sqrt{z})^2(2+\sqrt{z})\frac{1}{2\sqrt{z}} = \frac{2}{3}(z + \frac{2}{\sqrt{z}} - 3). \quad \textbf{(2pt)} \end{aligned}$$

To sum up,

$$f_Z(z) = \begin{cases} \frac{2}{3}(z + \frac{2}{\sqrt{z}} - 3) & \text{for } 0 < z < 1 \\ 0 & \text{otherwise.} \end{cases} \quad \textbf{(2pt)}$$