Probability and Statistics
Model Answer of 2021 Final Exam Paper A
BJTU Lancaster University College

Yiping Cheng

November 10, 2022

1. (20pt) Fill in the blanks:

(1) Suppose $P(A - B) = P(A) - P(B)$. Then $P(B - A) = \underline{\mathbf{0}}$.

(2) A bag contains 3 white balls and 7 black balls. Take 2 balls, without replacement, from the bag. The probability that at least one of the two balls taken is white is $\underline{\mathbf{1 - \frac{C_7^2}{C_{10}^2} = \frac{8}{15}}}$.

(3) Suppose the PDF of random variable X is $f(x)$. Let $Y = 1 - 2X$. Then the PDF of Y is $g(y) = \underline{\mathbf{\frac{1}{2} f(\frac{1-y}{2})}}$.

(4) Assume that $X \sim B(40, 0.1)$, then the distribution of X can be approximated by the Poisson distribution with $\lambda = \underline{\mathbf{4}}$.

(5) The continuous random variable X has memoryless property, then X follows the **exponential** distribution..

(6) Let $X \sim U(1, 7)$, $Y \sim N(0, 4)$, and X and Y are independent. Let $Z = 3X - 2Y$. Then $E(Z) = \underline{\mathbf{12}}$, $\mathrm{Var}(Z) = \underline{\mathbf{43}}$.

(7) Given sample data $(8, 4, 0, 3, 5)$. The sample mean is $\underline{\mathbf{4}}$, and the sample variance is $\underline{\mathbf{8.5}}$.

(8) There are three criteria for assessing estimators: consistency, **bias** or **unbiasedness**, and efficiency.

2. (8pt) Prove that for every events A and B, the following inequality holds:

$$\max(0, P(A) + P(B) - 1) \le P(AB) \le \min(P(A), P(B)).$$

Proof: $AB \subseteq A$, hence $P(AB) \le P(A)$. Likewise, $P(AB) \le P(B)$. Therefore $P(AB) \le \min(P(A), P(B))$.

Obviously, $0 \le P(AB)$. On the other hand,

$$1 \ge P(A \cup B) = P(A) + P(B) - P(AB),$$

$$1 - P(A) - P(B) \ge -P(AB),$$

$$P(A) + P(B) - 1 \le P(AB).$$

Therefore $\max(0, P(A) + P(B) - 1) \leq P(AB)$.

3. (10pt) Customers are used to evaluate preliminary product designs. In the past, 95% of highly successful products received good reviews, 60% of moderately successful products received good reviews, and 10% of poor products received good reviews. In addition, 40% of products have been highly successful, 35% have been moderately successful, and 25% have been poor products.

(a) (5pt) What is the probability that a product attains a good review?

(b) (5pt) If a new design attains a good review, what is the probability that it will be a highly successful product?

Solution: (a)

$$\begin{aligned}
P(\text{“good review”}) &= P(\text{“good review”}|\text{“highly successful”})P(\text{“highly successful”}) \\
&\quad + P(\text{“good review”}|\text{“moderately successful”})P(\text{“moderately successful”}) \\
&\quad + P(\text{“good review”}|\text{“poor”})P(\text{“poor”}) \\
&= 0.95 \times 0.4 + 0.6 \times 0.35 + 0.1 \times 0.25 \\
&= 0.615.
\end{aligned}$$

(b)

$$P(\text{“highly successful”}|\text{“good review”}) = \frac{P(\text{“highly successful and good review”})}{P(\text{“good review”})}$$

$$\begin{aligned}
&= \frac{P(\text{“good review”}|\text{“highly successful”})P(\text{“highly successful”})}{P(\text{“good review”})} \\
&= \frac{0.95 \times 0.4}{0.615} \\
&= 0.617886179.
\end{aligned}$$

4. (10pt) Random variables X and Y are independent, $X \sim N(1, 4)$, $Y \sim N(0, 1)$. Determine $P(2X - Y > 10)$.

Solution: Let $Z = 2X - Y$. Then $Z \sim N(\mu, \sigma^2)$ where

$$\mu = 2 \times 1 - 0 = 2,$$

$$\sigma^2 = 4 \times 4 + 1 \times 1 = 17.$$

$$\begin{aligned}
P(Z > 10) &= P(\frac{Z - \mu}{\sigma} > \frac{10 - 2}{\sqrt{17}}) \\
&= 1 - \Phi(1.940285) \\
&= 1 - 0.973827 = 0.026173.
\end{aligned}$$

5. (14pt) Suppose that X and Y have the following joint PDF:

$$f(x, y) = \begin{cases} c(x + y), & \textbf{if } 0 < x < y < 1 \\ 0, & \textbf{otherwise.} \end{cases}$$

(a) (3pt) Find the constant c;

(b) (3pt) Find the marginal PDF $f_X(x)$;

(c) (3pt) Determine $P(X \leq 1/2)$;

(d) (3pt) Find the conditional PDF of Y given that $X = x$.

(e) (2pt) Are X and Y independent? Why?

Solution: (a)

$$\begin{aligned} 1 &= \int_0^1 \int_0^y c(x+y)dxdy \\ &= \int_0^1 c\frac{3}{2}y^2 dy \\ &= \frac{c}{2} \end{aligned}$$

Thus $c = 2$.
(b) If $x \leq 0$ or $x \geq 1$, then $f_X(x) = 0$. If $0 < x < 1$, then

$$\begin{aligned} f_X(x) &= \int_x^1 2(x+y)dy \\ &= 2x + 1 - 3x^2 \end{aligned}$$

To sum up,

$$f_X(x) = \begin{cases} 2x + 1 - 3x^2, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(c)

$$\begin{aligned} P(X \leq 1/2) &= \int_0^{\frac{1}{2}} (2x + 1 - 3x^2)dx \\ &= \frac{5}{8} \end{aligned}$$

(d) If $0 < x < 1$, then

$$f_{Y|X}(y|x) = \begin{cases} \frac{2(x+y)}{2x+1-3x^2}, & \text{if } 0 < x < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(e) X and Y are not independent, because the support of the joint PDF is not rectangular.

6. (16pt) The joint PDF of X and Y is given by

$$f(x,y) = \begin{cases} \frac{x+y}{8}, & \textbf{if } 0 < x < 2, 0 < y < 2 \\ 0, & \textbf{otherwise.} \end{cases}$$

Determine

(a) (4pt) $E(X)$ and $E(Y)$;

(b) (5pt) $\mathrm{Var}(X)$ **and** $\mathrm{Var}(Y)$**;**

(c) (7pt) $\mathrm{Cov}(X,Y)$ **and** $\rho_{X,Y}$**.**

Solution: (a)

$$
\begin{aligned}
E(X) &= \int_0^2 \int_0^2 x \cdot \frac{1}{8}(x+y)dydx \\
&= \int_0^2 \frac{x}{8}[xy + \frac{y^2}{2}]_{y=0}^{y=2} dx \\
&= \int_0^2 \frac{x^2+x}{4} dx \\
&= \frac{7}{6}
\end{aligned}
$$

Likewise, $E(Y) = \frac{7}{6}$.
(b)

$$
\begin{aligned}
E(X^2) &= \int_0^2 \int_0^2 x^2 \cdot \frac{1}{8}(x+y)dydx \\
&= \int_0^2 \frac{x^2}{8}[xy + \frac{y^2}{2}]_{y=0}^{y=2} dx \\
&= \int_0^2 \frac{x^3+x^2}{4} dx \\
&= \frac{5}{3}
\end{aligned}
$$

$$
\mathrm{Var}(X) = E(X^2) - E^2(X) = \frac{11}{36}.
$$

Likewise, $\mathrm{Var}(Y) = \frac{11}{36}$.
(c)

$$
\begin{aligned}
E(XY) &= \int_0^2 \int_0^2 xy \cdot \frac{1}{8}(x+y)dydx \\
&= \int_0^2 \frac{x}{8}[\frac{xy^2}{2} + \frac{y^3}{3}]_{y=0}^{y=2} dx \\
&= \int_0^2 (\frac{x^2}{4} + \frac{x}{3})dx \\
&= \frac{4}{3}
\end{aligned}
$$

$$
\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{4}{3} - \frac{7}{6} \times \frac{7}{6} = -\frac{1}{36}.
$$

$$
\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}Y}} = -\frac{1}{11}.
$$

7. (12pt) Let $X_1, X_2, \cdots, X_{30}$ be independent random variables each having a discrete distribution with PF

$$f(x) = \begin{cases} \frac{1}{4}, & \textbf{if } x = 0, \textbf{ or } 2 \\ \frac{1}{2}, & \textbf{if } x = 1, \\ 0, & \textbf{otherwise.} \end{cases}$$

Use the central limit theorem and the correction for continuity to approximate the probability that $X_1 + X_2 + \cdots + X_{30}$ is at most 33.

Solution: For each i, we have

$$E(X_i) = 1$$
$$E(X_i^2) = \frac{3}{2}$$
$$\text{Var}(X_i) = E(X_i^2) - E^2(X_i) = \frac{1}{2}.$$

Thus,

$$E(X) = 30$$
$$Var(X) = 15$$

By the central limit theorem, X approximately follows $N(30, 15)$.

$$\begin{aligned} P(X \leq 33) &= P(X < 33.5) \\ &= P(\frac{X - 30}{\sqrt{15}} < \frac{33.5 - 30}{\sqrt{15}}) \\ &\approx \Phi(\frac{33.5 - 30}{\sqrt{15}}) \\ &= \Phi(0.903696114) \\ &= 0.81692172 \end{aligned}$$

8. (10pt) Let $X_1, X_2, \cdots, X_n$ be a random sample from population $X \sim Exp(\lambda)$. Derive the maximum likelihood estimator of the parameter λ.

Solution: For observed values $x_1, \ldots, x_n$, the likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i).$$

$$\log L(\lambda) = n \log(\lambda) - \lambda \sum_{i=1}^{n} x_i.$$

The maximizer of $\log L(\lambda)$ is the solution of the following equation:

$$0 = \frac{d \log L}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i.$$

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i}.$$

Thus the maximum likelihood estimator of λ is $\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^{n} X_i}$.