Probability and Statistics – Lancaster/BJTU 2017 Final Exam Paper B with Answer

1 Fill in the blanks (20 points)

1. In a simple probability space that has 50 outcomes, the probability of an event containing 5 outcomes is $\underline{0.1}$.
2. If X has the uniform distribution on interval $[1, 5]$, then $\mathrm{Var}(X) = \underline{\frac{4}{3}}$. If X has the binomial distribution with parameter $n = 5$ and $p = 0.4$, then $\mathrm{Var}(X) = \underline{1.2}$.
3. Of the following six families of distributions: binomial, Poisson, geometric, uniform, normal, and exponential, the family of distributions for which the mean and the variance are always equal are the Poisson distributions; the family of distributions for which the variance is always the square of the mean are the exponential distributions.
4. Suppose that X and Y are two random variables and $\mathrm{Var}(X) > 0$. If $Y = -7X$, then their correlation coefficient $\rho(X, Y) = \underline{-1}$; if X and Y are independent, then $\rho(X, Y) = \underline{0}$.
5. A sampling distribution is the distribution of a statistic. For a sequence of estimators to be practically usable, it must be consistent, which means that it must converge in probability to the parameter being estimated.

2

(10 points) Suppose that A, B, and C are events such that A and B are independent, $P(A \cap B \cap C) = 0.04$, $P(C|A \cap B) = 0.25$, and $P(B) = 4P(A)$. Evaluate $P(A \cup B)$.

Solution: A and B are independent implies $P(A \cap B) = P(A)P(B)$, hence

$$0.25 = P(C|A \cap B) = \frac{P(A \cap B \cap C)}{P(A \cap B)} = \frac{P(A \cap B \cap C)}{P(A)P(B)} = \frac{0.04}{4P(A)^2},$$

and thus $P(A) = 0.2$ and $P(B) = 0.8$, $P(A \cap B) = P(A)P(B) = 0.16$. Finally

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.8 - 0.16 = 0.84.$$

3

(8 points) Two people are playing a game by rolling dice. The rule of the game is as follows:

- One of the people is designated as player A and the other as player B.
- Player A rolls the die first. If a 5 or a 6 appears on the die, then player A wins and the game is over, otherwise the game proceeds.
- Now player B rolls the die. If the number that appears on his die is equal to or greater than the number that appeared on player A's die, then player B wins, otherwise player A wins.

Determine the probability for player A to win the game and decide whether either player has any advantage.

Solution: Let X be the number that appears on player A's die, and let Y be the number that appears on player B's die.

$$P(\text{A wins}) = P(X=5) + P(X=6) + P(X=4, Y<4) + P(X=3, Y<3) + P(X=2, Y<2)$$
$$= P(X=5) + P(X=6) + P(X=4)P(Y<4) + P(X=3)P(Y<3) + P(X=2)P(Y<2)$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6}\frac{3}{6} + \frac{1}{6}\frac{2}{6} + \frac{1}{6}\frac{1}{6} = \frac{1}{2}.$$

Since there is no tie, it can be concluded that neither player has any advantage.

4

(14 points) Suppose that random variables X and Y have the following joint p.d.f.

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

Determine
(a) the marginal p.d.f. of X;
(b) $P(X < 1/2)$;
(c) the conditional p.d.f. of Y given that $X = x$.

Solution: (a)

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy.$$

When $0 < x < 1$, we have $f_X(x) = \int_x^1 2(x+y)dy = 1 + 2x - 3x^2$. Thus,

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

(b)

$$P(X < 1/2) = \int_0^{\frac{1}{2}} (1+2x-3x^2)dx = \frac{5}{8}.$$

(c)

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

5

(12 points) Suppose that $X_1, \dots, X_n$ form a random sample of size n from a continuous distribution with the following p.d.f.

$$f(x) = \begin{cases} 2x & \text{for } 0<x<1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y = \max(X_1, \dots, X_n)$. Evaluate $E(Y)$.

Solution: By integrating over f, we obtain the c.d.f of each of the random sample:

$$F(x) = \begin{cases} 0 & \text{for } x<0, \\ x^2 & \text{for } 0 \le x < 1, \\ 1 & \text{for } x \ge 1. \end{cases}$$

It then follows that

$$F_Y(x) = \begin{cases} 0 & \text{for } x < 0, \\ x^{2n} & \text{for } 0 \le x < 1, \\ 1 & \text{for } x \ge 1. \end{cases}$$

Hence the p.d.f of Y is

$$f_Y(x) = \begin{cases} 2nx^{2n-1} & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore

$$E(Y) = \int_0^1 2nx^{2n-1} \cdot x dx = \frac{2n}{2n+1}.$$

6

(9 points) In a small town there are 500 people and there is a shop. Each hour, each person in the town has 0.01 probability to go to the shop and they go independently of one another. Approximately evaluate the probability that in a given hour, less than 3 people go to the shop.

Solution: Let X be the number of shop customers in a hour. Then X has the binomial distribution with parameter $n = 500$ and $p = 0.01$. Then X approximately has the Poisson distribution with parameter $\lambda = np = 500 \times 0.01 = 5$. Hence

$$P(X < 3) \approx (1 + 5 + \frac{25}{2})e^{-5} = 0.124652019483081.$$

7

(11 points) In a building there are 400 lights, and they are on or off independently of one another. At night, each light has 0.7 probability to be on. Using correction for continuity, approximately evaluate the probability that in the building there are between 270 (inclusive) and 290 (inclusive) that are on.

Solution: Let X be the number of lights that are on in the building. X has the binomial distribution with parameter $n = 400$ and $p = 0.7$, and according to the central limit theorem, it approximately has the normal distribution with mean $\mu = 280$ and $\sigma^2 = 84$. Thus

$$P(270 \le X \le 290) = P(269.5 < X < 290.5) = P(\frac{269.5 - 280}{\sqrt{84}} < \frac{X - 280}{\sqrt{84}} < \frac{290.5 - 280}{\sqrt{84}})$$

$$\approx 2\Phi(1.1456) - 1 = 2 \times 0.874 - 1 = 0.748.$$

8

(16 points) Suppose that $X_1, \ldots, X_n$ form a random sample from a distribution for which the p.d.f. is as follows:

$$f(x; \theta) = \begin{cases} \theta x^{-(\theta+1)} & \text{for } x > 1, \\ 0 & \text{otherwise.} \end{cases}$$

where $\theta > 1$ is the parameter to be estimated.
(a) Derive the method-of-moments estimator of θ.
(b) Derive the maximum likelihood estimator of θ.

Solution: (5 points) (a)

$$\mu_1 = E(X) = \int_1^\infty x \cdot \theta x^{-(\theta+1)} dx = \int_1^\infty \theta x^{-\theta} dx = \frac{\theta}{\theta - 1}.$$

Thus $\theta = \frac{\mu_1}{\mu_1 - 1}$, and the method-of-moments estimator of θ is therefore

$$\hat{\theta}_{MOM} = \frac{A_1}{A_1 - 1} = 1 + \frac{1}{\overline{X}_n - 1}.$$

(11 points) (b) For observed values $x_1, \ldots, x_n$, the likelihood function is

$$L(\theta) = \theta^n (\prod_{i=1}^n x_i)^{-(\theta+1)}.$$

$$\log L(\theta) = n \log(\theta) - (\theta + 1) \log(\prod_{i=1}^n x_i).$$

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

$$\tfrac{d \log L}{d\theta} = \tfrac{n}{\theta} - \log(\textstyle\prod_{i=1}^n x_i) = 0 \ \ ,$$

$$\theta = \frac{n}{\log(\prod_{i=1}^n x_i)}.$$

Thus the maximum likelihood estimator of θ is $\hat{\theta}_{MLE} = \frac{n}{\log(\prod_{i=1}^n X_i)}$.