Probability and Statistics
Mock Exam Paper B Reference Answer

1 Fill in the blanks (20 pt)

1. The number of calls received each minute by a call center has the Poisson distribution with parameter 2.5. Then the probability that this center receives 2 calls in a minute is $\underline{0.256515620699684}$, the average number of calls this center receives in a minute is $\underline{2.5}$, and the most probable number of calls this center receives in a minute is $\underline{2}$.

2. Suppose we are given observed values of a random sample from a distribution: 6, 10, 9, 3, 12. Then its second moment $a_2 = \underline{74}$, and its sample variance $s^2 = \underline{12.5}$.

3. Suppose that X_1, X_2, X_3 form a random sample from a particular distribution, $Y_1 = X_1 + X_2$, and $Y_2 = X_2 - X_3$, then $\frac{\mathrm{Var}(Y_1)}{\mathrm{Var}(Y_2)} = \underline{1}$ and the correlation coefficient $\rho(Y_1, Y_2) = \underline{\frac{1}{2}}$.

4. Suppose that random variables X_1, X_2 have bivariate normal distribution with covariance matrix $\begin{bmatrix} 1 & 2 \\ 2 & 9 \end{bmatrix}$, then the correlation coefficient of X_1 and X_2 is $\underline{\frac{2}{3}}$.

5. Suppose that X_1, X_2, X_3, X_4 form a random sample from $N(\mu, \sigma^2)$. Let $W = \sum_{i=1}^{4}(X_i - \overline{X})^2$ where $\overline{X} = \frac{X_1+X_2+X_3+X_4}{4}$. Then $E(W) = \underline{3\sigma^2}$, $\mathrm{Var}(W) = \underline{6\sigma^4}$.

Detailed explanation of 1.3: By assumption, X_1, X_2, X_3 are independent and have the same distribution. Let their variance be σ^2. Then

$$\mathrm{Var}(Y_1) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 2\sigma^2$$

$$\mathrm{Var}(Y_2) = \mathrm{Var}(X_2) + \mathrm{Var}(X_3) = 2\sigma^2$$

Therefore $\frac{\mathrm{Var}(Y_1)}{\mathrm{Var}(Y_2)} = 1$.

Now

$$\mathrm{Cov}(Y_1, Y_2) = \mathrm{Cov}(X_1 + X_2, X_2 - X_3) = \mathrm{Cov}(X_1, X_2) - \mathrm{Cov}(X_1, X_3) + \mathrm{Cov}(X_2, X_2) - \mathrm{Cov}(X_2, X_3)$$
$$= 0 - 0 + \mathrm{Var}(X_2) - 0 = \sigma^2.$$

Therefore

$$\rho(Y_1, Y_2) = \frac{\mathrm{Cov}(Y_1, Y_2)}{\sqrt{\mathrm{Var}(Y_1)}\sqrt{\mathrm{Var}(Y_2)}} = \frac{\sigma^2}{\sqrt{2\sigma^2}\sqrt{2\sigma^2}} = \frac{1}{2}.$$

Detailed explanation of 1.4: If X_1, X_2 have bivariate normal distribution with covariance matrix $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}$, then $\Sigma_{11} = \mathrm{Var}(X_1)$, $\Sigma_{22} = \mathrm{Var}(X_2)$, and $\Sigma_{12} = \mathrm{Cov}(X_1, X_2)$. From the given information we then have $\mathrm{Var}(X_1) = 1$, $\mathrm{Var}(X_2) = 9$, and $\mathrm{Cov}(X_1, X_2) = 2$, therefore

$$\rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}} = \frac{2}{3}.$$

2

(10 pt)

3

(10 pt) Three students: Alex, Bob, and Charlie, are going to take a test. The probabilities for them to pass the test are 0.8 for Alex, 0.7 for Bob, and 0.6 for Charlie. Furthermore, their test results are independent.

(a) Find the probability that exactly one student of the three will pass the test.

(b) Given that exactly one student of the three passes the test, what is the conditional probability that the student who passes is Alex?

Solution: Let A denote the event that Alex passes, B denote the event that Bob passes, and C denote the event that Charlie passes. Let U denote the event that exactly one student of the three will pass the test.

(a)

$$U = AB^cC^c \cup A^cBC^c \cup A^cB^cC$$

$$P(U) = P(AB^cC^c \cup A^cBC^c \cup A^cB^cC) = 0.8 \times 0.3 \times 0.4 + 0.2 \times 0.7 \times 0.4 + 0.2 \times 0.3 \times 0.6 = 0.188.$$

(b) The desired conditional probability is $P(A|U)$.

$$P(A|U) = \frac{P(AU)}{P(U)} = \frac{P(AB^cC^c)}{P(U)} = \frac{0.8 \times 0.3 \times 0.4}{0.188} = \frac{24}{47} = 0.51063830...$$

4

(12 pt) Let X_1 and X_2 be **independent** random variables. $Y_1 = \min(X_1, X_2)$ and $Y_2 = \max(X_1, X_2)$.

(a) Show that if $E(X_1), E(X_2), E(Y_1), E(Y_2)$ are all finite, then $E(Y_1) + E(Y_2) = E(X_1) + E(X_2)$.

(b) Suppose further that $X_1 \sim$ exponential(1) and $X_2 \sim$ exponential(2). Find $E(Y_1)$ and $E(Y_2)$.

Solution: (a) For every sample point s, either $Y_1(s) = X_1(s)$ and $Y_2(s) = X_2(s)$, or $Y_1(s) = X_2(s)$ and $Y_2(s) = X_1(s)$. In both cases, $Y_1(s) + Y_2(s) = X_1(s) + X_2(s)$. Thus $E(Y_1 + Y_2) = E(X_1 + X_2)$, and

$$E(Y_1) + E(Y_2) = E(X_1) + E(X_2).$$

(b) X_1 has c.d.f $F_{X_1}(x) = 1 - e^{-x}$. X_2 has c.d.f $F_{X_2}(x) = 1 - e^{-2x}$. Hence $Y_1 = \min(X_1, X_2)$ has c.d.f.

$$F_{Y_1}(x) = 1 - [1 - F_{X_1}(x)][1 - F_{X_2}(x)] = 1 - e^{-3x}$$

which means Y_1 has the exponential distribution with parameter 3. Therefore

$$E(Y_1) = \frac{1}{3}$$

and

$$E(Y_2) = E(X_1) + E(X_2) - E(Y_1) = 1 + \frac{1}{2} - \frac{1}{3} = \frac{7}{6}.$$

5

(8 pt) Let $F_\alpha(m, n)$ denote the upper α quantile of the $F(m, n)$ distribution. Show that

$$F_{1-\alpha}(n, m) = \frac{1}{F_\alpha(m, n)}.$$

Proof: Let $X \sim F(m, n)$. We know that $P(X > 0) = 1$.

$$P(X > F_\alpha(m,n)) = \alpha$$

$$P(\frac{1}{X} < \frac{1}{F_\alpha(m,n)}) = \alpha$$

$$P(\frac{1}{X} \geq \frac{1}{F_\alpha(m,n)}) = 1-\alpha$$

Now since all F distributions are continuous, we have $P(\frac{1}{X} = \frac{1}{F_\alpha(m,n)}) = 0$. Thus

$$P(\frac{1}{X} > \frac{1}{F_\alpha(m,n)}) = 1-\alpha$$

Now since $\frac{1}{X} \sim F(n,m)$, from the above equation we see that $F_{1-\alpha}(n,m) = \frac{1}{F_\alpha(m,n)}$.

6

(14 pt) Suppose that X and Y are random variables having the following joint p.d.f.:

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

(a) Determine the value of c.
(b) Compute $E(X)$ and $E(Y)$.
(c) Compute $\mathrm{Cov}(X,Y)$.

Solution: (a)

$$1 = \int_0^1 dx \int_0^{\sqrt{x}} cxydy = \frac{c}{2}\int_0^1 dx \cdot xy^2|_{y=0}^{y=\sqrt{x}} = \frac{c}{2}\int_0^1 x^2 dx = \frac{c}{6}.$$

Therefore $c = 6$.
(b)

$$E(X) = \int_0^1 dx \int_0^{\sqrt{x}} 6x^2ydy = \int_0^1 3x^3 dx = \frac{3}{4}$$

$$E(Y) = \int_0^1 dx \int_0^{\sqrt{x}} 6xy^2dy = \int_0^1 2x^{\frac{5}{2}} dx = \frac{4}{7}.$$

(c)

$$E(XY) = \int_0^1 dx \int_0^{\sqrt{x}} 6x^2y^2dy = \int_0^1 2x^{\frac{7}{2}} dx = \frac{4}{9}$$

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{4}{9} - \frac{3}{4} \times \frac{4}{7} = \frac{1}{63}.$$

7

(14 pt) Two players, A and B, are playing the following game: Player A tosses fair coins 1000 times, each time if a head is obtained, he receives 2 points, and if a tail is obtained, he receives 5 points. Player B rolls balanced dice 1000 times, each time a number in (1–6) appears on the die, he receive that number of points. Let X be the total points player A receives, and Y be the total points player B receives.

(a) Using the central limit theorem, determine the approximate distributions that X and Y have, respectively.

(b) Using correction for continuity, approximately evaluate the probabilities $P(X > 3600)$ and $P(Y > 3600)$.

Solution: (a) Let X_i denote the points that player A receives at his i-th toss, then $X = \sum_{i=1}^{1000} X_i$. Let Y_i denote the points that player A receives at his i-th roll, then $Y = \sum_{i=1}^{1000} Y_i$. We have

$$E(X_i) = \frac{2+5}{2} = 3.5, \quad \text{Var}(X_i) = \frac{(2-3.5)^2 + (5-3.5)^2}{2} = 2.25$$

and hence

$$E(X) = 3500, \quad \text{Var}(X) = 2250;$$

$$E(Y_i) = 3.5, \quad \text{Var}(Y_i) = \frac{\sum_{j=1}^{6}(j-3.5)^2}{6} = 2.9166667$$

and hence

$$E(Y) = 3500, \quad \text{Var}(Y) = 2916.6667.$$

Since the X_i's are independent of one another, and 1000 is a big number, by the central limit theorem, X approximately has the normal distribution with mean 3500 and variance 2250. Likewise, Y approximately has the normal distribution with mean 3500 and variance 2916.6667.

(b)

$$P(X > 3600) = P(X > 3600.5) = P\left(\frac{X - 3500}{\sqrt{2250}} > \frac{3600.5 - 3500}{\sqrt{2250}}\right) \approx 1 - \Phi(2.1187) = 0.017058$$

$$P(Y > 3600) = P(Y > 3600.5) = P\left(\frac{Y - 3500}{\sqrt{2916.6667}} > \frac{3600.5 - 3500}{\sqrt{2916.6667}}\right) \approx 1 - \Phi(1.8609) = 0.031379.$$

8

(10 pt) Suppose $X \sim N(\mu, 9)$ and $X_1, \dots, X_n$ form a random sample from X. If we want a confidence interval for μ with confidence level being 0.99, and the length of the interval no more than 2, then what is the smallest value of n?

Solution: The $1 - \alpha = 0.99$-level confidence interval for μ is $(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}})$. We have $\alpha = 0.01$. The length of the interval is $2\frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}$. So we are required

$$2\frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} \le 2$$

$$\frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} \le 1$$

$$n \ge \sigma^2 z_{\frac{\alpha}{2}}^2 = 9 z_{0.005}^2 = 9 \cdot 2.576^2 = 59.72.$$

Since n is integer, we have $n \ge 60$.

9

(12 pt) Suppose that the population distribution is a discrete distribution having the following p.f.:

x	1	2	3
$f(x)$	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

where θ is the parameter to be estimated. Suppose that we have obtained observed values of a random sample of size 5 from the distribution: **1,2,1,3,1**. Based on these data,

(a) Find the method-of-moments estimate of θ

(b) Find the maximum likelihood estimate of θ.

Solution: (a)

$$\mu_1 = \theta^2 \cdot 1 + 2\theta(1-\theta) \cdot 2 + (1-\theta)^2 \cdot 3 = 3 - 2\theta$$

$$\theta = 1.5 - 0.5\mu_1$$

$$\hat{\theta}_{MOM} = 1.5 - 0.5\frac{\sum_{i=1}^n x_i}{n} = 1.5 - 0.5 \times \frac{1+2+1+3+1}{5} = 0.7.$$

(b) For the particular observed values, the likelihood function

$$L(\theta) = (\theta^2)^3 \cdot 2\theta(1-\theta) \cdot (1-\theta)^2 = 2\theta^7(1-\theta)^3$$

$$\log L(\theta) = \log(2) + 7\log(\theta) + 3\log(1-\theta)$$

$$\frac{d\log L(\theta)}{d\theta} = \frac{7}{\theta} - \frac{3}{1-\theta}$$

The maximizer of $L(\theta)$ is one that makes $\frac{d\log L(\theta)}{d\theta} = 0$. Solving for this, we obtain $\theta = 0.7$. Therefore for these observed values,

$$\hat{\theta}_{MLE} = 0.7.$$