# Probability and Statistics
# Mock Exam Paper A Reference Answer

## 1 Fill in the blanks (20 points)

1. The binomial distribution with parameters $n$ and $p$ has mean $\underline{np}$ and variance $\underline{np(1-p)}$.
2. The Poisson distribution with parameter $\lambda$ has mean $\underline{\lambda}$ and variance $\underline{\lambda}$.
3. The geometric distribution with parameter $p$ has mean $\underline{\frac{1}{p}}$ and variance $\underline{\frac{1-p}{p^2}}$.
4. The uniform distribution on interval $[a,b]$ has mean $\underline{\frac{a+b}{2}}$ and variance $\underline{\frac{(b-a)^2}{12}}$.
5. The exponential distribution with parameter $\beta$ has mean $\underline{\frac{1}{\beta}}$ and variance $\underline{\frac{1}{\beta^2}}$.

## 2

**(8 points)** A box contains 20 good products and 5 defective products. An inspector draws from the box three products one at a time, without replacement.
(a) What is the probability that the third product drawn is defective?
(b) What is the probability that the third product drawn is the second defective product drawn?

Solution: (a) $\frac{5}{20+5} = 0.2$.
(b) Let $D_i$ denote the event that the $i$-th drawn product is defective. Then the desired probability is $P(D_1 D_2^c D_3 \cup D_1^c D_2 D_3)$.

$$\begin{aligned} P(D_1 D_2^c D_3 \cup D_1^c D_2 D_3) &= P(D_1 D_2^c D_3) + P(D_1^c D_2 D_3) \\ &= P(D_1)P(D_2^c|D_1)P(D_3|D_1 D_2^c) + P(D_1^c)P(D_2|D_1^c)P(D_3|D_1^c D_2) \\ &= \frac{5}{25}\times\frac{20}{24}\times\frac{4}{23} + \frac{20}{25}\times\frac{5}{24}\times\frac{4}{23} = \frac{4}{69}. \end{aligned}$$

## 3

**(10 points)** Let $A$ and $B$ be two events. Define random variables $X$ and $Y$ as follows:

$$X(s) = \begin{cases} 1 & \text{for } s \in A \\ 0 & \text{otherwise.} \end{cases}, \quad Y(s) = \begin{cases} 1 & \text{for } s \in B \\ 0 & \text{otherwise.} \end{cases}.$$

Show that $\mathrm{Cov}(X,Y) = 0$ if and only if $A$ and $B$ are independent.

Solution:

$$E(X) = P(X=0)\cdot 0 + P(X=1)\cdot 1 = P(A)$$
$$E(Y) = P(Y=0)\cdot 0 + P(Y=1)\cdot 1 = P(B)$$
$$E(XY) = P(XY=0)\cdot 0 + P(XY=1)\cdot 1 = P(X=1,Y=1) = P(AB)$$

Hence

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = P(AB) - P(A)P(B)$$
$$\mathrm{Cov}(X,Y) = 0 \Leftrightarrow P(AB) = P(A)P(B)$$

## 4

**(12 points)** Suppose that random variables $X, Y$ are **independent** and $X \sim \text{normal}(0, \sigma_1^2)$ and $Y \sim \text{normal}(0, \sigma_2^2)$. Suppose further that $P(|5X + 12Y| < 13) < P(|4X + 3Y| < 5)$. Compare $\sigma_1$ and $\sigma_2$.

Solution: By assumption, $5X + 12Y \sim \text{normal}(0, 25\sigma_1^2 + 144\sigma_2^2)$ and $4X + 3Y \sim \text{normal}(0, 16\sigma_1^2 + 9\sigma_2^2)$. And

$$P(|5X + 12Y| < 13) = P(\frac{|5X + 12Y|}{\sqrt{25\sigma_1^2 + 144\sigma_2^2}} < \frac{13}{\sqrt{25\sigma_1^2 + 144\sigma_2^2}}) = 2\Phi(\frac{13}{\sqrt{25\sigma_1^2 + 144\sigma_2^2}}) - 1$$

$$P(|4X + 3Y| < 5) = P(\frac{|4X + 3Y|}{\sqrt{16\sigma_1^2 + 9\sigma_2^2}} < \frac{5}{\sqrt{16\sigma_1^2 + 9\sigma_2^2}}) = 2\Phi(\frac{5}{\sqrt{16\sigma_1^2 + 9\sigma_2^2}}) - 1$$

It then follows that

$$\begin{aligned}
\frac{13}{\sqrt{25\sigma_1^2 + 144\sigma_2^2}} &< \frac{5}{\sqrt{16\sigma_1^2 + 9\sigma_2^2}} \\
\frac{169}{25\sigma_1^2 + 144\sigma_2^2} &< \frac{25}{16\sigma_1^2 + 9\sigma_2^2} \\
2704\sigma_1^2 + 1521\sigma_2^2 &< 625\sigma_1^2 + 3600\sigma_2^2 \\
2079\sigma_1^2 &< 2079\sigma_2^2 \\
\sigma_1 &< \sigma_2.
\end{aligned}$$

## 5

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

$$f(x, y) = \begin{cases} 6xy & \text{for } 0 < y < \sqrt{x} < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute $P(X \leq Y)$.
(b) Find marginal density $f_Y(y)$.
(c) Find conditional density $f_{X|Y}(x|y)$.

Solution: (a)

$$P(X \leq Y) = \int_0^1 dx \int_x^{\sqrt{x}} 6xydy = \int_0^1 dx \cdot 3xy^2|_{y=x}^{y=\sqrt{x}} = \int_0^1 3(x^2 - x^3)dx = \frac{1}{4}.$$

(b) If $y \leq 0$ or $y \geq 1$, then $f_Y(y) = 0$. Otherwise,

$$f_Y(y) = \int_{y^2}^1 6xydx = 3x^2y|_{x=y^2}^{x=1} = 3y - 3y^5.$$

To sum up,

$$f_Y(y) = \begin{cases} 3y - 3y^5 & \text{for } 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(c) If $0 < y < 1$, then

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \begin{cases} \frac{2x}{1-y^4} & \text{for } y^2 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

## 6

**(12 points)** Let $X$ be the number of white cells in a ml of man's blood. It is known that $X$ has a mean of 7300, and a standard deviation of 700.

(a) Using Chebyshev's inequality, find an interval whose center is the mean so that the probability that $X$ is in that interval is at least $\frac{1}{2}$.

(b) If it is further supposed that $X$ has normal distribution, find the interval whose center is the mean so that the probability that $X$ is in that interval is $\frac{1}{2}$.

Solution: (a) By Chebyshev's inequality, $P(|X-\mu| \geq \kappa\sigma) \leq \frac{1}{\kappa^2}$, i.e. $P(|X-\mu| < \kappa\sigma) \geq 1 - \frac{1}{\kappa^2}$. Let $\kappa = \sqrt{2}$, then

$$P(|X-\mu| < \sqrt{2}\sigma) \geq \frac{1}{2}.$$

Therefore, that interval is $(\mu - \sqrt{2}\sigma, \mu + \sqrt{2}\sigma) = (6310.050506338834, 8289.949493661166)$.

(b) If $X \sim \text{normal}(\mu, \sigma^2)$, then $P(|X-\mu| < \kappa\sigma) = 2\Phi(\kappa) - 1$. To have $2\Phi(\kappa) - 1 = \frac{1}{2}$, $\kappa = \Phi^{-1}(\frac{3}{4}) = 0.6745$. Therefore, that interval is $(\mu - 0.6745\sigma, \mu + 0.6745\sigma) = (6827.85, 7772.15)$.

## 7

**(12 points)** When a computer adds up two real numbers and stores the result into an integer, the error has the uniform distribution on the interval $[-0.5, 0.5]$. Use the central limit theorem to approximately evaluate the following quantities:

(a) The probability that when 1501 real numbers are added, the total error is in the range $[-15, 15]$.

(b) How many real numbers can be added so that the probability for the total error to be in the range $[-10, 10]$ is at least 0.9?

Solution: (a) Let $E_i$ be the error of the $i$-th addition. Let $S_n = \sum_{i=1}^{n} E_i$. Then $E(X_i) = 0$ and $\text{Var}(X_i) = \frac{1}{12}$, and $E(S_n) = 0$ and $\text{Var}(S_n) = \frac{n}{12}$. The total error of addition of 1501 real numbers is $S_{1500}$ and

$$E(S_{1500}) = 0 \text{ and } \text{Var}(S_{1500}) = 125.$$

By the central limit theorem, $S_{1500}$ approximately $\sim$ normal$(0, 125)$. Thus

$$P|S_{1500}| \leq 15) = P(\frac{|S_{1500}|}{\sqrt{125}} \leq \frac{15}{\sqrt{125}}) \approx 2\Phi(\frac{15}{\sqrt{125}}) - 1 = 0.820287505121.$$

(b) By the central limit theorem, $S_n$ approximately $\sim$ normal$(0, \frac{n}{12})$. We want to have

$$0.9 \leq P(|S_n| \leq 10) = P(\frac{|S_n|}{\sqrt{\frac{n}{12}}} \leq \frac{10}{\sqrt{\frac{n}{12}}}) \approx 2\Phi(\sqrt{\frac{1200}{n}}) - 1.$$

$$\Phi(\sqrt{\frac{1200}{n}}) \geq 0.95$$

$$\sqrt{\frac{1200}{n}} \geq 1.645$$

$$n \leq 443.$$

Therefore at most 444 real numbers can be added to allow the desired relation to hold.

## 8

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

$$f(x;\theta) = \begin{cases} \frac{x}{\theta^2} e^{\frac{-x}{\theta}} & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

where $\theta > 0$ is the unknown parameter to be estimated.

(a) Derive the M.L.E. of $\theta$.

(b) Is the M.L.E. unbiased?

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

$$L(\theta) = \theta^{-2n} \prod_{i=1}^{n} x_i \cdot \exp(-\sum_{i=1}^{n} x_i/\theta).$$

$$\log L(\theta) = -2n \log(\theta) + \log(\prod_{i=1}^{n} x_i) - \sum_{i=1}^{n} x_i/\theta.$$

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

$$\frac{d \log L}{d\theta} = -\frac{2n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = 0,$$

$$\theta = \frac{\sum_{i=1}^{n} x_i}{2n}.$$

Thus the M.L.E. of $\theta$ is $\hat{\theta} = \frac{\sum_{i=1}^{n} X_i}{2n}$.

(b)

$$E(X_i) = \int_0^\infty \frac{x^2}{\theta^2} e^{\frac{-x}{\theta}} dx = \theta \int_0^\infty (\frac{x}{\theta})^2 e^{\frac{-x}{\theta}} d\frac{x}{\theta} = \theta\Gamma(3) = 2\theta.$$

Hence

$$E(\hat{\theta}) = E(\frac{\sum_{i=1}^{n} X_i}{2n}) = \frac{n \cdot 2\theta}{2n} = \theta,$$

which means $\hat{\theta}$ is an unbiased estimator of $\theta$.