APMA 1650 Midterm 2 Solutions

November 15, 2018

BANNER ID:

NAME:

(LEAVE THIS BLANK)

Problem	Grade	Memo
1		
2		
3		
4		
5		
6		
7		
8		
Total		

This midterm is graded out of 100 points and has a 80 minute time limit. There are eight problems whose point values are indicated in parentheses. You are not allowed to use any outside references or calculators. Only solutions written on this paper will be graded.

You may leave answers in terms of basic arithmetic (+,-,$\times$,$\backslash$) and coefficients with known formulas (e.g. $\binom{10}{3}$ or 29!). Integrals should be solved completely.

Show all work and justify your answers.

1. (20 points) Let $f(x, y)$ be the joint density function of X and Y given by

$$f(x, y) = \begin{cases} 6(x^2 y) & 0 \le x \le y, x + y \le 2 \\ 0 & \text{otherwise} \end{cases}.$$

(a) (5 points) Find the marginal distribution $f_Y(y)$ of Y.

(b) (5 points) Find $f(x|y)$, the conditional pdf of X given Y.

(c) (5 points) Find $E[X|Y = y]$.

(d) (5 points) Find $P(X \le 1|Y = 1.5)$ and $P(X \ge 1|Y = 0.5)$.

Solutions. (a). First, we need to find the marginal distribution of Y. By definition,

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)dx \\ &= \begin{cases} \int_0^y 6x^2 y dx & \text{if } 0 \le y \le 1 \\ \int_0^{y-2} 6x^2 y dx & \text{if } 1 < y \le 2 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 2y^4 & \text{if } 0 \le y \le 1 \\ 2y(2-y)^3 & \text{if } 1 < y \le 2 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

(b). Therefore, the conditional pdf of X given Y is

$$f(x|y) = \begin{cases} \dfrac{3x^2}{y^3} & \text{if } 0 < y \le 1, 0 \le x \le y, x + y \le 2 \\ \dfrac{3x^2}{(2-y)^3} & \text{if } 1 < y < 2, 0 \le x \le y, x + y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

(c). By definition, if $0 < y \le 1$, we have

$$E[X|Y = y] = \int_0^y x \frac{3x^2}{y^3} dx = \frac{1}{y^3} \int_0^y 3x^3 dx = \frac{3}{4} y.$$

If $1 < y < 2$, we have

$$E[X|Y = y] = \int_0^{2-y} x \frac{3x^2}{(2-y)^3} dx = \frac{1}{(2-y)^3} \int_0^{2-y} 3x^3 dx = \frac{3}{4}(2 - y).$$

(d). Since $0 \le x \le y$ and $x + y \le 2$, if $y = 1.5$, x should be in $0 \le x \le 0.5$. Thus $P(X \le 1|Y = 1.5) = 1$. Or by definition,

$$P(X \le 1|Y = y) = \int_0^{2-y} \frac{3x^2}{(2-y)^3} dx = \frac{1}{(2-y)^3} \int_0^{2-y} 3x^2 dx = 1.$$

Since $f(x, y) = 0$ on $y = 0.5$ and $x \le 1$, $P(X \ge 1|Y = 0.5) = 0$. ■

2. (15 points) Let $X, Y \sim \mathcal{N}(0,1)$ be independent random variables. Show that

$$U_1 = \frac{X+Y}{\sqrt{2}}, \qquad U_2 = \frac{X-Y}{\sqrt{2}}$$

are independent.

Solution. Since X and Y are independent, the moment generating function of U_1 is

$$\begin{aligned}
m_{U_1}(t) &= E[e^{tU_1}] = E[e^{t/\sqrt{2}X + t/\sqrt{2}Y}] \\
&= E[e^{t/\sqrt{2}X}] \cdot E[e^{t/\sqrt{2}Y}] = m_X(t/\sqrt{2}) m_Y(t/\sqrt{2}) \\
&= e^{t^2/4 + t^2/4} = e^{t^2/2}.
\end{aligned}$$

Since the mgf of $\mathcal{N}(0,1)$ is $e^{t^2/2}$, we conclude that $U_1 \sim \mathcal{N}(0,1)$. Similarly, $U_2 \sim \mathcal{N}(0,1)$. Then the joint moment generating function of U_1 and U_2 is

$$\begin{aligned}
m_{\boldsymbol{U}}(t_1, t_2) &= E[e^{t_1 U_1 + t_2 U_2}] = E[e^{\frac{t_1+t_2}{\sqrt{2}}X + \frac{t_1-t_2}{\sqrt{2}}Y}] \\
&= E[e^{\frac{t_1+t_2}{\sqrt{2}}X}] E[e^{\frac{t_1-t_2}{\sqrt{2}}Y}] = m_X(\frac{t_1+t_2}{\sqrt{2}}) m_Y(\frac{t_1-t_2}{\sqrt{2}}) \\
&= \exp\left[\frac{\left(\frac{t_1+t_2}{\sqrt{2}}\right)^2}{2} + \frac{\left(\frac{t_1-t_2}{\sqrt{2}}\right)^2}{2}\right] \\
&= \exp\left[\frac{t_1^2 + t_2^2}{2}\right] = e^{t_1^2/2} e^{t_2^2/2} = m_{U_1}(t_1) m_{U_2}(t_2).
\end{aligned}$$

Therefore, U_1 and U_2 are independent.

■

3. (10 points) Ten hunters are waiting for ducks to fly by. When a flock of ducks flies overhead, the hunters fire at the same time, but each chooses his target at random, independently of the others. If each hunter independently hits his target with probability p, compute the expected number of ducks that escape unhurt when a flock of size 10 files overhead.

Solution. Let consider

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th duck escapes unhurt} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, \cdots, 10$. Let $X = \sum_{i=1}^{10} X_i$ which represents the number of ducks that escape unhurt. By the linearity of expectation,

$$E[X] = \sum_{i=1}^{10} E[X_i].$$

To compute $E[X_i] = P(X_i = 1)$, we note that each of the hunters will, independently, hit the i-th duck with probability $p/10$, so

$$P(X_i = 1) = \left(1 - \frac{p}{10}\right)^{10}.$$

Hence

$$E[X] = 10\left(1 - \frac{p}{10}\right)^{10}.$$

■

4. (15 points) Suppose that $\boldsymbol{X} = (X_1, X_2, X_3)$ is a random sample of size $n = 3$ from an exponential distribution with density function

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x \geq 0, \\ 0 & \text{otherwise} \end{cases}.$$

Consider the following estimators for θ:

$$\hat{\theta}_1 = X_1, \qquad \hat{\theta}_2 = aX_1 + (1-a)X_2 \qquad \hat{\theta}_3 = w_1 X_1 + w_2 X_2 + (1 - w_1 - w_2) X_3$$

where $a, w_1, w_2 > 0$.

(a) (10 points) Find the MSE of $\hat{\theta}_i$, $i = 1, 2, 3$.

(b) (2.5 points) Find a^* which minimizes the MSE of $\hat{\theta}_2$.

(c) (2.5 points) Find w_1^*, w_2^* which minimize the MSE of $\hat{\theta}_3$.

Solutions. (a). Note that

$$\text{MSE}(\hat{\theta}) = \text{Var}[\hat{\theta}] + \left(B(\hat{\theta})\right)^2$$

where $B(\hat{\theta}) = E[\hat{\theta} - \theta$. Also note that for $X \sim \text{Exp}(\theta^{-1})$, $E[X] = \theta$ and $\text{Var}(X) = \theta^2$. Then by linearity of expectation, one can check that

$$\begin{aligned} &E[\hat{\theta}_1] = E[X_1] = \theta \implies B(\hat{\theta}_1) = 0, \\ &E[\hat{\theta}_2] = aE[X_1] + (1-a)E[X_2] = \theta \implies B(\hat{\theta}_2) = 0, \\ &E[\hat{\theta}_3] = w_1 E[X_1] + w_2 E[X_2] + (1 - w_1 - w_2) E[X_3] = \theta \implies B(\hat{\theta}_3) = 0. \end{aligned}$$

Thus $\text{MSE}(\hat{\theta}_i) = \text{Var}(\hat{\theta}_i)$. For $\hat{\theta}_1$,

$$\text{MSE}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_1) = \text{Var}[X_1] = \theta^2.$$

For $\hat{\theta}_2$, since X_1 and X_2 are independent,

$$\text{MSE}(\hat{\theta}_2) = \text{Var}(\hat{\theta}_2) = a^2 \text{Var}[X_1] + (1-a)^2 \text{Var}[X_2] = \left(a^2 + (1-a)^2\right) \theta^2.$$

For $\hat{\theta}_3$, let $w_3 = 1 - w_1 - w_2$. Since X_1, X_2 and X_3 are independent,

$$\text{MSE}(\hat{\theta}_3) = \text{Var}(\hat{\theta}_3) = w_1^2 \text{Var}[X_1] + w_2^2 \text{Var}[X_2] + w_3^2 \text{Var}[X_3] = \left(w_1^2 + w_2^2 + w_3^2\right) \theta^2.$$

(b). It suffices to find the minimizer of $a^2 + (1-a)^2$. It readily follows that $a^* = \frac{1}{2}$.

(c). It follows from the inequality of arithmetic and geometric means that

$$\frac{x_1^2 + x_2^2 + x_3^2}{3} \geq \left(x_1^2 x_2^2 x_3^2\right)^{1/3}$$

and that equality holds if and only if $x_1 = x_2 = x_3$. Therefore, $w_1^2 + w_2^2 + w_3^2$ is minimized when $w_1 = w_2 = w_3$. Since $w_1 + w_2 + w_3 = 1$,

$$w_1^* = \frac{1}{3}, \quad w_2^* = \frac{1}{3}, \quad w_3^* = \frac{1}{3}.$$

Remark. In the sense that the MSE is minimized, the sample mean is optimal. ■

5. (10 points) An anthropologist wishes to estimate the average height of men for a certain race of people. The population standard deviation is assumed to be 3.6 inches and she randomly samples n men where $n \geq 30$. The anthropologist wants the difference between the sample mean and the population mean to be less than 0.6 inch, with probability 0.99. How many men should she sample to achieve this objective?

   Solution. Since

   $$P(-2.58 < Z < 2.58) = 0.99$$

   by CLT, we need

   $$\sqrt{n} \geq \frac{3.6}{0.6} \times 2.58 \implies n \geq (6 \times 2.58)^2 = 239.6304.$$

   ■

6. (10 points) Let $Y \sim \mathcal{U}(0,1)$ and $X|Y \sim B(n, Y)$. Find the marginal distribution of X.

Solution. Note that we have

$$f_Y(y) = \mathbb{1}_{(0,1)}(y), \qquad f(x|y) = \binom{n}{x} y^x (1-y)^{n-x}.$$

Therefore, the joint pdf is

$$f(x,y) = f(x|y) f_Y(y) = \binom{n}{x} y^x (1-y)^{n-x} \mathbb{1}_{(0,1)}(y).$$

By definition, the marginal pdf of X is

$$\begin{aligned} f_X(x) &= \int_0^1 f(x,y)dy = \int_0^1 \binom{n}{x} y^x (1-y)^{n-x} dy \\ &= \binom{n}{x} \int_0^1 y^x (1-y)^{n-x} dy = \binom{n}{x} B(x+1, n-x+1). \end{aligned}$$

where $B(\alpha, \beta)$ is the beta function. ■

7. (10 points) We want to estimate the proportion p of Democrats in the US population, by taking a random sample of size n. How large does our sample have to be to guarantee that our estimate (sample mean) will be within 5% (in relative terms) of the true value with probability at least 0.99?

Solution. It follows from the weak law of large number that

$$P(|\bar{X} - p| > \epsilon) \le \frac{pq}{n\epsilon^2}.$$

Since $pq \le \frac{1}{4}$, we above becomes

$$P(|\bar{X} - p| > \epsilon) \le \frac{pq}{n\epsilon^2} \le \frac{1}{4n\epsilon^2}.$$

By setting $\epsilon = 0.05$, we have

$$\frac{1}{4n\epsilon^2} = \frac{1}{4n(0.05)^2} = 0.01,$$

which gives

$$n = \frac{1}{4(0.05)^2(0.01)} = 10,000.$$

Therefore, we need at least $10,000$ samples. ■

8. (10 points) Suppose that buses are schedule to arrive at a stop at noon, but are always X minutes late, where X is an exponential random variable with parameter λ, i.e., $X \sim \text{Exp}(\lambda)$. Suppose you arrive at the bus stop precisely at noon.

   (a) (5 points) Find the probability that you have to wait for more than five minutes for a bus to arrive.

   (b) (5 points) Suppose that you have already been waiting for 10 minutes. Find the probability that you have to wait five or more additional minutes.

Solutions. Let $X \sim \text{Exp}(\lambda)$. Then its pdf is

$$f(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}_{[0,\infty)}(x).$$

(a).

$$P(X \geq 5) = \int_5^\infty \lambda e^{-\lambda x} dx = e^{-5\lambda}.$$

(b).

$$P(X \geq 15 | X \geq 10) = \frac{P(X \geq 15, X \geq 10)}{P(X \geq 10)} = \frac{P(X \geq 15)}{P(X \geq 10)} = e^{-5\lambda}.$$

Or by the memoryless property of exponential distribution, the answer of (b) should be the same as the answer of (a). ■