Model Answers of Quiz IV (Total points:15)

1. (2 points) Let $\{X_n\}_{n\geq 1}$ be a sequence of independent exponential random variables with $E(X_n) = 10 - \sum_{i=1}^{n} \frac{5}{2^{i-1}}$ for all $n \geq 1$. Then which of the following statements is/are true?

   (a) $X_n \to 0$ in distribution
   (b) $X_n \to 0$ in probability
   (c) $X_n \to X$ in distribution, where $X \sim \text{Poi}(10)$
   (d) $X_n \to X$ in probability, where $X \sim \text{Poi}(10)$

   Solution: Note that

$$\begin{aligned} E(X_n) &= 10 - \sum_{i=1}^{n} \frac{5}{2^{i-1}} \\ &= 10 - 5\left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{n-1}}\right) \\ &= 10 - 5\left(\frac{1 - \frac{1}{2^n}}{1 - \frac{1}{2}}\right) \\ \implies E(X_n) &= 10\left(\frac{1}{2}\right)^n \end{aligned}$$

   Thus, X_n follows exponential distribution with mean $10(\frac{1}{2})^n$. So, $X_n \sim \text{Exp}(\lambda_n = \frac{2^n}{10})$.

   Now, to check X_n converges in probability to 0, we need to show that $\lim_{n\to\infty} P(|X_n - 0| \leq \epsilon) = 1$ for all $\epsilon > 0$; which also implies $\lim_{n\to\infty} P(|X_n - 0| > \epsilon) = 0$ for all $\epsilon > 0$. Now,

$$\begin{aligned} \lim_{n\to\infty} P(|X_n - 0| > \epsilon) &= \lim_{n\to\infty} P(|X_n| > \epsilon) \\ &= \lim_{n\to\infty} P(X_n > \epsilon) \\ &= \lim_{n\to\infty} (1 - P(X_n \leq \epsilon)) \\ &= \lim_{n\to\infty} e^{-\frac{2^n}{10}\epsilon} \\ \implies \lim_{n\to\infty} P(|X_n - 0| > \epsilon) &= 0. \end{aligned}$$

   Thus, $X_n \to 0$ in probability. So, option (b) is correct and option (d) is incorrect.

   Again, we know that Convergence in probability implies convergence in distribution. Thus, $X_n \to 0$ in distribution. So, option (a) is correct and option (c) is incorrect.

2. (2 points) Let $\{X_n : n \geq 1\}$ be a sequence of independent and identically distributed random variables with common probability mass function

$$f(m) = \begin{cases} \frac{(\ln 3)^m}{3(m!)} & \text{if m} = 0, 1, 2, 3, ... \\ 0 & \text{otherwise.} \end{cases}$$

If $S_n = X_1 + X_2 + ... + X_n$ for $n \geq 1$, then which of the following statements is/are true?

(a) $\dfrac{S_n - n\ln(3)}{n} \to 0$ in probability.

(b) $\dfrac{S_n - n\ln(3)}{\sqrt{n\ln(3)}} \to \text{Z}$ in distribution, where $\text{Z} \sim N(0,1)$.

(c) $\dfrac{S_n - n\ln(3)}{\sqrt{n\ln(3)}} \to 0$ in probability.

(d) $\dfrac{S_n}{n} - 2\ln(3) \to 0$ in 2nd mean.

Solution: Note that for $m = 0, 1, 2, 3, ...$, the pmf of X_n can be modified as

$$f(m) = \frac{(\ln 3)^m}{3(m!)} = \frac{1}{3}\frac{(\ln 3)^m}{(m)!} = e^{-\ln(3)}\frac{(\ln 3)^m}{(m)!}.$$

Thus, $X_n \sim \text{Poi}(\lambda = \ln(3))$. Therefore, $\text{E}(X_n) = \ln(3)$ and $\text{Var}(X_n) = \ln(3)$. Using SLLN,

$$\frac{S_n}{n} \to \ln 3$$

with probability 1. Therefore, option (a) is correct.

Again, $E(S_n) = n\ln 3$ and $Var(S_n) = n\ln 3$. By Central Limit Theorem, $\dfrac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$ converges in distribution to Z, where $Z \sim N(0,1)$. Hence, option (b) is correct and option (c) is incorrect.

Note that

$$\begin{aligned} E\left[\left(\frac{S_n}{n} - 2\ln 3\right)^2\right] &= E\left[\left(\frac{S_n}{n} - \ln 3 - \ln 3\right)^2\right] \\ &= Var\left(\frac{S_n}{n}\right) + (\ln 3)^2 \\ &= \frac{\ln 3}{n} + (\ln 3)^2 \not\to 0 \text{ as n} \to \infty. \end{aligned}$$

So, option (d) is not correct.

3. (3 points) Let $\{X_n : n \geq 1\}$ be a sequence of independent and identically distributed random variables with mean 0, variance 1. Let

$$Z_n = \sqrt{n}\frac{X_1X_2 + X_3X_4 + ... + X_{2n-1}X_{2n}}{X_1^2 + X_2^2 + ... + X_{2n}^2} \text{ for all } n \geq 1.$$

Suppose that $Z_n \to X$ in distribution. Then $\text{E}(X^2) + \text{E}(X^3)$ equals

(a) 0 (b) 0.25 (c) 0.50 (d) 0.75

Solution: For $i = 1, 2, ...,$ let us define

$$U_i = X_{2i-1}X_{2i} \qquad \text{and} \qquad V_j = X_j^2.$$

Then, $Z_n = \sqrt{n}\left(\frac{\sum_{i=1}^{n} U_i}{\sum_{j=1}^{2n} V_j}\right)$. Now, Y_i's are i.i.d. with

$$E(U_i) = E(X_{2i-1})E(X_{2i}) = 0,$$

and

$$Var(U_i) = E(X_{2i-1}^2)E(X_{2i}^2) = Var(X_{2i-1})Var(X_{2i}) = 1.$$

Thus, using CLT, we have $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i$ converges in distribution to $N(0,1)$.

Now, V_i's are i.i.d. with $E(V_i) = 1$. Therefore, using SLLN, $\frac{1}{2n}\sum_{i=1}^{2n} V_i \to 1$ with probability 1. By Slutsky's Theorem, $2Z_n \xrightarrow{\mathcal{D}} N(0,1) \implies Z_n \xrightarrow{\mathcal{D}} N(0,\frac{1}{4})$. Thus, X is normally distributed with mean 0 and variance $\frac{1}{4}$. Therefore,

$$E(X^2) + E(X^3) = Var(X) + (E(X))^2 + 0 = \frac{1}{4} = 0.25.$$

Hence, option (b) is correct option.

4. (2 points) Let X_1 and X_2 be independent and identically distributed $N(0,1)$ random variables. If $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$, then which of the following statements is/are true?

   (a) Y_1 and Y_2 are independent random variables

   (b) $\frac{Y_1}{|Y_2|}$ and $\frac{2Y_2}{3|Y_1|}$ are identically distributed

   (c) $Y_1^2 + Y_2^2 \sim \chi_2^2$

   (d) $\frac{2Y_2^2}{3Y_1^2} \sim F_{2,3}$

Solution: Note that the joint distribution of X_1 and X_2 is bi-variate normal with mean vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and variance-covariance matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Now $l_1Y_1 + l_2Y_2 = l_1(X_1 + X_2) + l_2(X_1 - X_2) = (l_1 + l_2)\,X_1 + (l_1 - l_2)\,X_2$. If $(l_1,\, l_2) \neq (0,0)$, then $(l_1 + l_2, l_1 - l_2) \neq (0,0)$, and hence, $l_1Y_1 + l_2Y_2$ has a univariate normal distribution. Thus, the joint distribution of Y_1 and Y_2 is bi-variate normal. As

$$Cov(Y_1, Y_2) = Cov(X_1 + X_2, X_1 - X_2) = Var(X_1) - Var(X_2) = 0,$$

Y_1 and Y_2 are independent random variables. Thus, option (a) is correct.

Now,

$$\frac{Y_1}{|Y_2|} = \frac{\frac{Y_1}{\sqrt{2}}}{\sqrt{\frac{Y_2^2}{2}}} \sim t_1.$$

Similarly, $\frac{Y_2}{|Y_1|} \sim t_1 \implies \frac{2}{3}\frac{Y_2}{|Y_1|}$ has a different distribution than t_1. Hence, $\frac{Y_1}{|Y_2|}$ and $\frac{2}{3}\frac{Y_2}{|Y_1|}$ are not identically distributed. Thus, option (b) is not correct.

Again,

$$\begin{aligned}
&\frac{Y_1}{\sqrt{2}} \sim N(0,1); \frac{Y_2}{\sqrt{2}} \sim N(0,1)\\
\implies &\frac{Y_1^2 + Y_2^2}{2} \sim \chi_2^2\\
\implies &Y_1^2 + Y_2^2 \sim 2\chi_2^2.
\end{aligned}$$

Thus, option (c) is not correct.
Also,

$$\frac{Y_2^2}{Y_1^2} = \frac{Y_2^2/2}{Y_1^2/2} \sim F_{1,1}.$$

Thus, expectation of $\frac{2Y_2^2}{3Y_1^2}$ does not exist. However, the expectation of $F_{2,3}$ distribution exists. Thus, option (d) is not correct.

5. (2 points) Let $X_1, X_2, \ldots, X_9$ be a random sample of size $n = 9$ from a population having $N(0,1)$ distribution. Let $\bar{X} = \frac{1}{9}\sum_{i=1}^{9} X_i$ and $S^2 = \frac{1}{8}\sum_{i=1}^{9}(X_i - \bar{X})^2$. Then, find the value of $\frac{1}{2}E(\bar{X}S) + \frac{1}{2}\operatorname{Var}(\bar{X}S)$.

(a) $\frac{1}{9}$ (b) $\frac{1}{8}$ (c) 1 (d) 0

Solution: Since $X_i \sim N(0,1)$ independently, we have

$$\bar{X} \sim N\left(0, \frac{1}{9}\right).$$

Also, it is known that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Here, $\sigma^2 = 1$ and $n = 9$, so

$$8S^2 \sim \chi_8^2.$$

Therefore,

$$E[S^2] = \frac{8}{8} = 1.$$

For a normal sample, $\bar{X}$ and S^2 are independent random variables. Now,

$$E(\bar{X}S) = E[\bar{X}]E[S] = 0. \quad \text{[Since, } \bar{X} \text{ and } S \text{ are independent]}$$

and

$$\begin{aligned}
\text{Var}(\bar{X}S) &= E[\bar{X}^2 S^2] - [E(\bar{X}S)]^2 \\
&= E[\bar{X}^2]\, E[S^2] - 0 \quad \text{[Since, } \bar{X} \text{ and } S^2 \text{ are independent]} \\
&= [\text{Var}(\bar{X}) + E^2[\bar{X}]]\, E[S^2] \\
&= [\frac{1}{9} + 0] \times 1 \\
&= \frac{1}{9} \qquad (1)
\end{aligned}$$

Hence,

$$\frac{1}{2}E(\bar{X}S) + \frac{1}{2}\text{Var}(\bar{X}S) = \frac{1}{2}(0) + \frac{1}{2}\left(\frac{1}{9}\right) = \frac{1}{18}.$$

Remark: The value is not given in any option. Therefore, marks for this question are given to all.

6. (2 points) Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n\,(\geq 2)$ from a population having Poisson distribution with mean λ, where $\lambda > 0$ is an unknown parameter. Define

$$T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \text{and} \quad T_2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - T_1)^2.$$

Consider the following statements:

P: T_1 is an unbiased estimator of λ.

Q: T_2 is an unbiased estimator of λ.

Which of the following statements is true?

(a) P only (b) Q only (c) Both P and Q (d) Neither P nor Q.

Solution: Note that

$$E[T_1] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\cdot n\lambda = \lambda.$$

Hence, T_1 is an unbiased estimator of λ.

$$E[T_2] = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right].$$

Expanding the square:

$$E[T_2] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right].$$

Simplify:

$$E[T_2] = \frac{1}{n-1}\left(E\left[\sum_{i=1}^{n} X_i^2\right] - nE[\bar{X}^2]\right). \tag{1}$$

Since $X_i \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$, we have

$$E[X_i] = \lambda, \quad \text{Var}(X_i) = \lambda.$$

Hence,

$$E[X_i^2] = \text{Var}(X_i) + [E(X_i)]^2 = \lambda + \lambda^2.$$

Also,

$$E[\bar{X}^2] = \text{Var}(\bar{X}) + [E(\bar{X})]^2 = \frac{\lambda}{n} + \lambda^2.$$

Substituting these into equation (1):

$$E[T_2] = \frac{1}{n-1}\left[n(\lambda + \lambda^2) - n\left(\frac{\lambda}{n} + \lambda^2\right)\right].$$

Simplify:

$$E[T_2] = \frac{1}{n-1}\left[n\lambda + n\lambda^2 - \lambda - n\lambda^2\right] = \frac{1}{n-1}(n\lambda - \lambda) = \lambda.$$

Hence, both statements P and Q are true.

7. (2 points) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population having $U(0, \theta)$ distribution, where $\theta > 0$ is unknown. Let $\hat{M}$ be the maximum likelihood estimator (MLE) of the median of X_1.

   Consider the following statements:

   P: $\hat{M} = \frac{1}{2}\max(X_1, X_2, \ldots, X_n)$.

   Q: $\text{Bias}(\hat{M}) = -\frac{\theta}{(n+1)}$.

   Which of the following statements is true?

   (a) P only (b) Q only (c) Both P and Q (d) Neither P nor Q.

Solution: The probability density function (PDF) of X_i is:

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

$$L(\theta \mid X_1, \ldots, X_n) = \prod_{i=1}^{n} f(X_i \mid \theta) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} I(0 < X_i < \theta),$$

where $I(\cdot)$ is the indicator function.

The likelihood $L(\theta)$ is a decreasing function of θ for $\theta > \max(X_1, \ldots, X_n)$. Hence, $L(\theta)$ attains its maximum at

$$\widehat{\theta}_{\text{MLE}} = \max(X_1, X_2, \ldots, X_n) = X_{(n)}.$$

For $X \sim U(0, \theta)$, the median M satisfies

$$\int_0^M f_X(x)\, dx = \frac{1}{2}.$$

That is,

$$\int_0^M \frac{1}{\theta}\, dx = \frac{1}{2} \quad \Rightarrow \quad M = \frac{\theta}{2}.$$

Using the invariance property of MLEs:

$$\widehat{M} = \frac{\widehat{\theta}}{2} = \frac{X_{(n)}}{2}.$$

Hence, statement P is true. We know that if $X_i \overset{\text{iid}}{\sim} U(0, \theta)$, then

$$E[X_{(n)}] = \frac{n\theta}{n+1}.$$

Now,

$$E[\widehat{M}] = E\left[\frac{X_{(n)}}{2}\right] = \frac{1}{2} E[X_{(n)}] = \frac{n\theta}{2(n+1)}.$$

Hence, the bias is

$$\text{Bias}(\widehat{M}) = E[\widehat{M}] - M = \frac{n\theta}{2(n+1)} - \frac{\theta}{2} = -\frac{\theta}{2(n+1)}.$$

Therefore, statement Q is false.