Probability & Statistics for EECS: Final Exam

Due on June 25, 2024 at 23:59

Name:
Student ID:

June 25, 2024

Problem 1

(20 points) Given two i.i.d. random variables X and Y satisfying distribution $\mathrm{Unif}(0, 5)$.

(a) (10 points) Find $P(\max(X, Y) \leq 1)$.

(b) (10 points) Find $P(\min(X, Y) \leq 1)$.

Solution

(a) By definition, we have

$$\begin{aligned} P(\max(X, Y) \leq 1) &= P(X \leq 1, Y \leq 1) \\ &= P(X \leq 1)P(Y \leq 1) \\ &= \frac{1-0}{5-0} \times \frac{1-0}{5-0} = \frac{1}{25}. \end{aligned}$$

(b) By definition, we have

$$\begin{aligned} P(\min(X, Y) \geq 1) &= P(X \geq 1, Y \geq 1) \\ &= P(X \geq 1)P(Y \geq 1) \\ &= \left(1 - \frac{1-0}{5-0}\right) \times \left(1 - \frac{1-0}{5-0}\right) = \frac{16}{25}. \end{aligned}$$

Therefore, we have

$$P(\min(X, Y) \leq 1) = 1 - P(\min(X, Y) \geq 1) = \frac{9}{25}.$$

Problem 2

(20 points) Let X and Y be two continuous random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} c & \text{if } 0 < x < 1, 0 < y < 2x \\ 0 & \text{otherwise} \end{cases}$$

(a) (10 points) Find the value of constant c.

(b) (5 points) Find the marginal distribution of X.

(c) (5 points) Find the marginal distribution of Y.

Solution

(a) The constant c is determined by the requirement that the total probability over the entire range of X and Y must be 1, we thus have

$$\int_0^1 \int_0^{2x} c dy dx = 1,$$

and

$$\int_0^1 \int_0^{2x} c dy dx = \int_0^1 [cy]_0^{2x} dx = \int_0^1 2cx dx = [cx^2]_0^1 = c$$

Setting this equal to 1 and solving for c gives us $c = 1$.

(b) The marginal distribution of X is obtained by integrating the joint PDF over all values of Y that X can take. Therefore, we have

$$f_X(x) = \int_0^{2x} f_{X,Y}(x,y) dy = 2x, 0 < x < 1.$$

(c) Similarly, we have

$$f_Y(y) = \int_{y/2}^1 f_{X,Y}(x,y) dx = \frac{2-y}{2}, 0 < y < 2.$$

Problem 3

(20 points) Given a random variable $X \sim \text{Expo}(\lambda)$, where $\lambda > 0$.

(a) (10 points) Find $E(X \mid X > 2030)$.

(b) (10 points) Find $E(X \mid X < 2030)$.

Solution

(a)

$$E(X \mid X \geq 2030) = E(X - 2030 \mid X \geq 2030) + 2030 = E(X) + 2030 = \frac{1}{\lambda} + 2030$$

(b) By LOTE and the memoryless property of exponential random variables, we have

$$\begin{aligned}
E(X) &= E(X \mid X < 2030)P(X < 2030) + E(X \mid X \geq 2030)P(X \geq 2030) \\
&= E(X \mid X < 2030)P(X < 2030) + E(X - 2030 \mid X \geq 2030)P(X \geq 2030) + 2030 \times P(X \geq 2030) \\
&= E(X \mid X < 2030)P(X < 2030) + E(X)P(X \geq 2030) + 2030 \times P(X \geq 2030).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
E(X \mid X < 2024) &= \frac{E(X)P(X < 2030) - 2030 \times P(X \geq 2030)}{P(X < 2030)} \\
&= \frac{E(X)P(X < 2030) - 2030 + 2030 \times P(X < 2030)}{P(X < 2030)} \\
&= E(X) + 2030 - \frac{2030}{P(X < 2030)}.
\end{aligned}$$

Since the CDF of $X \sim \text{Expo}(\lambda)$ is $F(x) = 1 - e^{-\lambda x}$, we have

$$E(X \mid X < 2030) = \frac{1}{\lambda} + 2030 - \frac{2030}{1 - e^{-2030\lambda}}.$$

Problem 4

(5 points) Let $X \sim \text{Pois}(\lambda)$, where $\lambda > 0$. Find $E\left[\binom{X}{k}\right]$, where k is a nonnegative integer.

Solution

The expectation can be written as:

$$\begin{aligned}
E\left[\binom{X}{k}\right] &= \sum_{x=k}^{\infty} \frac{x!}{k!(x-k)!} \cdot \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \frac{1}{k!} \sum_{x=k}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-k)!} \\
&= \frac{1}{k!} \sum_{y=0}^{\infty} \frac{\lambda^{y+k} e^{-\lambda}}{y!} \\
&= \frac{\lambda^k e^{-\lambda}}{k!} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \\
&= \frac{\lambda^k e^{-\lambda}}{k!} \cdot e^{\lambda} = \frac{\lambda^k}{k!}
\end{aligned}$$

Problem 5

(5 points) U, V, W are three independent random variables satisfying the uniform distribution $\text{Unif}(0,1)$. Let $S = U + V + W$. Find the value of $E(\text{Var}(U \mid S))$.

Solution

By the law of total variance, we have:

$$\text{Var}(U) = E(\text{Var}(U \mid S)) + \text{Var}(E(U \mid S))$$

Since $U \sim \text{Unif}(0,1)$, we have:

$$\text{Var}(U) = \frac{1}{12}$$

The variance of the sum S of three independent $\text{Unif}(0,1)$ variables is:

$$\text{Var}(S) = \text{Var}(U + V + W) = 3 \cdot \text{Var}(U) = 3 \cdot \frac{1}{12} = \frac{1}{4}$$

By symmetry and the properties of the independent uniform distribution:

$$E(U \mid S) = \frac{S}{3}$$

Therefore, we have

$$\text{Var}(E(U \mid S)) = \text{Var}\left(\frac{S}{3}\right) = \frac{1}{9}\,\text{Var}(S) = \frac{1}{9} \cdot \frac{1}{4} = \frac{1}{36}$$

Substitute the known variances into the expression of $\text{Var}(U)$:

$$\frac{1}{12} = E(\text{Var}(U \mid S)) + \frac{1}{36}$$

Therefore, we have

$$E(\text{Var}(U \mid S)) = \frac{1}{12} - \frac{1}{36} = \frac{1}{18}$$

Problem 6

(5 points) Let $X \sim \mathrm{Unif}(0,1)$, and express X as a continued fraction in the following format:

$$X = \frac{1}{Y_1 + \frac{1}{Y_2 + \frac{1}{Y_3 + \cdots}}}.$$

Find the joint distribution of Y_1 and Y_2.

Solution

We have that

$$\begin{aligned}
\mathbb{P}\left(Y_1 = u, Y_2 = v\right) =& \mathbb{P}\left(X \leq \frac{1}{u + \frac{1}{v+1}}\right) - \mathbb{P}\left(X \leq \frac{1}{u + \frac{1}{v}}\right) \\
=& \frac{v+1}{u(v+1)+1} - \frac{v}{uv+1} \\
=& \frac{1}{(uv+1)(uv+u+1)}, \quad u, v = 1, 2, \ldots
\end{aligned}$$

Problem 7

(5 points) Given a random variable X satisfying exponential distribution, i.e., $X \sim \text{Expo}(\lambda)$, where $\lambda > 0$. Let $Y = \lfloor X \rfloor$ be the integer part of X, and $Z = X - Y$ be the fractional part of X. Find the joint distribution of Y and Z.

Solution

By definition, we have

$$\begin{aligned}
P(Y = y, Z \leq z) &= P(y \leq X \leq y + z) \\
&= \int_{y}^{y+z} \lambda e^{-\lambda x} dx \\
&= \left(1 - e^{-\lambda z}\right) e^{-\lambda y}, y \in \mathbb{N}, z \in (0, 1).
\end{aligned}$$

Or, we may have

$$\begin{aligned}
P(Y \leq y, Z \leq z) &= \sum_{m=0}^{y} P(m \leq X \leq m + z) \\
&= \sum_{m=0}^{y} \int_{m}^{m+z} \lambda e^{-\lambda x} dx \\
&= \left(1 - e^{-\lambda z}\right) \sum_{m=0}^{y} e^{-\lambda m} \\
&= \left(1 - e^{-\lambda z}\right) \frac{1 - e^{-\lambda y}}{1 - e^{-\lambda}}, y \in \mathbb{N}, z \in (0, 1).
\end{aligned}$$

Problem 8

(10 points) Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with mean μ and variance σ^2, and $n \geq 2$. A bootstrap sample of $X_1, \ldots, X_n$ is a sample of n r.v.s $X_1^*, \ldots, X_n^*$ formed from the $X_j, \forall j \in \{1, \ldots, n\}$ by sampling with replacement with equal probabilities. Let $\bar{X}^*$ denote the sample mean of the bootstrap sample:

$$\bar{X}^* = \frac{1}{n}\left(X_1^* + \cdots + X_n^*\right)$$

(a) (5 points) Calculate $E\left(X_j^*\right)$ and $\operatorname{Var}\left(X_j^*\right)$ for each $j \in \{1, \ldots, n\}$.

(b) (5 points) Calculate $E\left(\bar{X}^*\right)$ and $\operatorname{Var}\left(\bar{X}^*\right)$. Then explain intuitively why $\operatorname{Var}(\bar{X}) < \operatorname{Var}\left(\bar{X}^*\right)$.

Solution

(a) Define random variable I_j that marks what index $i \in \{1, \ldots, n\}$ is the actual value of j, i.e.,

$$X_j^* \mid (I_j = i) = X_i.$$

Conditioning on I_j, we get that

$$E\left(X_j^*\right) = E\left(E\left(X_j^* \mid I_j\right)\right) = E\left(E\left(X_{I_j}\right)\right) = E(\mu) = \mu$$

and

$$\begin{aligned}
\operatorname{Var}\left(X_j^*\right) &= \operatorname{Var}\left(E\left(X_j^* \mid I_j\right)\right) + E\left(\operatorname{Var}\left(X_j^* \mid I_j\right)\right) \\
&= \operatorname{Var}\left(E\left(X_{I_j}\right)\right) + E\left(\operatorname{Var}\left(X_{I_j}\right)\right) \\
&= \operatorname{Var}(\mu) + E\left(\sigma^2\right) = \sigma^2.
\end{aligned}$$

(b) Use Adam's law (conditioning on $X_1, \ldots, X_n$) to obtain the required mean

$$E\left(\bar{X}^*\right) = E\left(E\left(\bar{X}^* \mid X_1, \ldots, X_n\right)\right) = E\left(\bar{X}_n\right) = \mu,$$

and Eve's law to obtain the variance

$$\begin{aligned}
\operatorname{Var}\left(\bar{X}^*\right) &= \operatorname{Var}\left(E\left(\bar{X}^* \mid X_1, \ldots, X_n\right)\right) + E\left(\operatorname{Var}\left(\bar{X}^* \mid X_1, \ldots, X_n\right)\right) \\
&= \operatorname{Var}\left(\bar{X}_n\right) + E\left(\frac{1}{n^2}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\bar{X}_n^2\right) \\
&= \frac{\sigma^2}{n} + \frac{1}{n^2}\sum_{i=1}^{n} E\left(X_i^2\right) - \frac{1}{n}E\left(\bar{X}_n^2\right) \\
&= \frac{\sigma^2}{n} + \frac{1}{n^2}\sum_{i=1}^{n} E\left[\operatorname{Var}\left(X_i\right) + \left[E\left(X_i\right)\right]^2\right] - \frac{1}{n}\left[\operatorname{Var}\left(\bar{X}_n\right) + \left[E\left(\bar{X}_n\right)\right]^2\right] \\
&= \frac{\sigma^2}{n} + \frac{1}{n^2}\sum_{i=1}^{n}\left(\sigma^2 + \mu^2\right) - \frac{1}{n}\left(\frac{\sigma^2}{n} + \mu^2\right) \\
&= \frac{(2n-1)\,\sigma^2}{n^2}.
\end{aligned}$$

Observe that we have that

$$\operatorname{Var}(\bar{X}) = \frac{(2n-1)\,\sigma^2}{n^2} < \frac{2\sigma^2}{n} = \operatorname{Var}\left(\bar{X}^*\right).$$

This is intuitively because of the fact that bootstrap sampling gives another touch of uncertainty to the starting basic uncertainty of variables $X_1, \ldots, X_n$.

Problem 9

(10 points) There are two urns with a total of $2N$ distinguishable balls. Initially, the first urn has N white balls and the second urn has N black balls. At each stage, we pick a ball at random from each urn and interchange them. Let X_n be the number of black balls in the first urn at time n. This is a Markov chain on the state space $\{0, 1, \ldots, N\}$.

(a) (5 points) Find the transition probabilities of the chain.

(b) (5 points) Find the stationary distribution of the chain.

Solution

(a) We first note that

$$P\left(X_{n+1} = 1 \mid X_n = 0\right) = 1, P\left(X_{n+1} = N-1 \mid X_n = N\right) = 1.$$

If $X_n = i, i \in \{1, \ldots, N-1\}$, we have $X_{n+1} \in \{i-1, i, i+1\}$.

- Observe that $X_{n+1} = i-1$ if and only if we have chosen black ball from the first urn and white ball from the second:
$$P\left(X_{n+1} = i-1 \mid X_n = i\right) = \frac{i}{N} \cdot \frac{i}{N} = \frac{i^2}{N^2}.$$
- Similarly we have that $X_{n+1} = i+1$ if and only if we have chosen white ball from the first urn and black ball from the second:
$$P\left(X_{n+1} = i+1 \mid X_n = i\right) = \frac{N-i}{N} \cdot \frac{N-i}{N} = \frac{(N-i)^2}{N^2}.$$
- Number of black balls in the first urn remains the same if and only if we have picked different colors:
$$P\left(X_{n+1} = i \mid X_n = i\right) = \frac{i}{N} \cdot \frac{N-i}{N} + \frac{N-i}{N} \cdot \frac{i}{N} = \frac{2i(N-i)}{N^2}.$$

(b) Note two important observations:

- The Markov chain is irreducible.
- The Markov chain is a step-by-step analogy to the story of Hypergeometric.

This two observations lead to the guess of stationary distribution $\mathbf{s} = [s_0, \ldots, s_i, \ldots, s_N]$ with the PMF of Hypergeometric distribution, i.e.,
$$s_i = \frac{\binom{N}{i}\binom{N}{N-i}}{\binom{2N}{N}}.$$
Due to irreducibility, we justify the proposed distribution by checking the detailed balance equation,

$$s_i q_{ij} = s_j q_{ji}, \forall i, j \in \{0, 1, \ldots, N\}.$$

For state $i = 0$, the only non-trivial case that we should check is state $j = 1$ since there is no direct transition to other states. Therefore, we have that

$$\frac{\binom{N}{0}\binom{N}{N}}{\binom{2N}{N}} \cdot 1 = \frac{\binom{N}{1}\binom{N}{N-1}}{\binom{2N}{N}} \cdot \frac{1^2}{N^2},$$

which is true since $\binom{N}{1}\binom{N}{N-1} = N^2$. Similarly, for $i = N$, The only non-trivial case is for $j = N-1$ which is true using the similar calculations.

For $i = 1, \ldots, N-1$, non-trivial cases happen for $j = i-1$ and $j = i+1$. We are going to show that the equation holds for $1 < i \leq N-1$ and for $j = i-1$ (all other calculations are similar or we have already showed). We have that

$$
\begin{aligned}
& s_i q_{i,i-1} = s_{i-1} q_{i-1,i} \\
\Leftrightarrow\ & \frac{\binom{N}{i}\binom{N}{N-i}}{\binom{2N}{N}} \cdot \frac{i^2}{N^2} = \frac{\binom{N}{i-1}\binom{N}{N-i+1}}{\binom{2N}{N}} \cdot \frac{(N-i+1)^2}{N^2} \\
\Leftrightarrow\ & \binom{N}{i}\binom{N}{N-i} i^2 = \binom{N}{i-1}\binom{N}{N-i+1}(N-i+1)^2 \\
\Leftrightarrow\ & \frac{N!}{(i-1)!(N-i)!} = \frac{N!}{(i-1)!(N-i)!}.
\end{aligned}
$$

Therefore, we have showed that the chain is reversible, hence, $\mathbf{s}$ is the stationary distribution.