PS08-Midterm Review

Sejal Nimkar

2024-12-12

Toss a fair coin until we get 1 head and 2 tails. Random variables: X counts the number of tosses, U assigns 1 if the number of tosses is greater than 3, and 0 otherwise, $W = U1 + U2 + \cdots + U16$ where each Ui has the same distribution as U for i = 1, . . . , 16 and U_i is independent of U_j for $i = j$.

a. What is the sample space for this experiment? Explain and share at least 3 possible outcomes.

Ans:The sample space for this experiment includes all possible sequences of coin tosses that result in exactly 1 head and 2 tails. For example:

- HTT
- THT
- TTH

Each sequence represents a distinct outcome. The sample space is the set of all such sequences.

b. What is the range of X ? Is it countable? Clearly explain.

Ans: The random variable X, which counts the number of tosses required to achieve 1 head and 2 tails, has a range of values starting from 3 (the minimum number of tosses required) to infinity (if the process continues indefinitely). This range is countable since it consists of discrete integers.

c. Find $P(U = 1)$ (write this as a fraction), the median, and the IQR of U . Show your work.

Ans: To find $P(U = 1)$, note that $U = 1$ occurs when the number of tosses X is greater than 3. The probability of $X \leq 3$ can be computed by enumerating the sequences that result in exactly 1 head and 2 tails within 3 tosses: $P(X \leq 3) = P(\text{HTT}) + P(\text{THT}) + P(\text{TTH}) = 3 \cdot \left(\frac{1}{2}\right)^3 = \frac{3}{8}$. Thus, $P(U = 1) = 1 - P(X \leq 3) = 1 - \frac{3}{8} = \frac{5}{8}$.

The median of U is 0.5, and the interquartile range (IQR) is $1 - 0 = 1$.

d. Find $EW, Var(W)$, and present the R code to get $P(3 < W < 7)$.

Ans: To compute $\mathbb{E}[W]$ and $\text{Var}(W)$: $\mathbb{E}[U] = P(U = 1) = \frac{5}{8}, \quad \mathbb{E}[W] = 16 \cdot \mathbb{E}[U] = 16 \cdot \frac{5}{8} = 10$. $\text{Var}(U) = P(U = 1)(1 - P(U = 1)) = \frac{5}{8} \cdot \frac{3}{8} = \frac{15}{64}, \quad \text{Var}(W) = 16 \cdot \text{Var}(U) = 16 \cdot \frac{15}{64} = \frac{240}{64} = 3.75$.

The R code to calculate $P(3 < W < 7)$ is:

```
p_w <- pbinom(6, size = 16, prob = 5/8) - pbinom(3, size = 16, prob = 5/8)
p_w
```

```
## [1] 0.03698758
```

e. Let $V = \sum_{i=1}^{16} (-1)^i \cdot U_i$. Assume that you get 105 realizations of W and 105 realizations of V and the results are stored in R vectors vec.w and vec.v, respectively. Using the code below, do you expect res1 to be bigger than res2? Clearly explain.

res1 <- sum((vec.w - mean(vec.w))^2) res2 <- sum((vec.v - mean(vec.v))^2)

Ans: We can check the answer using simulations:

```
set.seed(123)
n <- 10^5
vec.w <- rbinom(n, size = 16, prob = 5/8)
vec.v <- cumsum((-1)^(1:16) * vec.w)
res1 <- sum((vec.w - mean(vec.w))^2)
res2 <- sum((vec.v - mean(vec.v))^2)
res1 > res2
```

```
## [1] FALSE
```

Imagine throwing darts at a dart board $B = \{(x, y) : 0 \le x \le y \le 2\}$. Suppose that the darts are thrown in such a way that they are certain to hit a point in B, and that each point in B is equally likely to be hit. Define the random variable X to be the value of the x-coordinate of the point that is hit, and define the random variable Y to be the value of the y-coordinate of the point that is hit.

a. Draw and write down the PDF of X

Ans:The probability density function (PDF) of X can be found from the uniform distribution over the region B. The length of the line $y = x$ to $y = 2$ varies with x, so the height of the distribution will be proportional to this length.

The PDF of X is given by:

$$f_X(x) = \begin{cases} 2 - x & \text{if } 0 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

b. Find $E[X]$, $P(X < \mu)$, q_2, and q_3

Ans:

1. $E[X]$: $E[X] = \int_0^2 x \cdot f_X(x)\,dx = \int_0^2 x(2-x)\,dx = \int_0^2 (2x - x^2)\,dx = \left[x^2 - \frac{x^3}{3}\right]_0^2 = \left(4 - \frac{8}{3}\right) = \frac{4}{3}$.
2. $P(X < \mu)$: Since $\mu = E[X] = \frac{4}{3}$, $P(X < \mu) = \int_0^{\frac{4}{3}} f_X(x)\,dx = \int_0^{\frac{4}{3}} (2-x)\,dx = \left[2x - \frac{x^2}{2}\right]_0^{\frac{4}{3}} = 2 \cdot \frac{4}{3} - \frac{\left(\frac{4}{3}\right)^2}{2} = \frac{8}{3} - \frac{8}{9} = \frac{16}{9}$.
3. q_2 (median): Solve for q_2 such that: $\int_0^{q_2} f_X(x)\,dx = 0.5$. $\int_0^{q_2} (2-x)\,dx = 0.5 \implies \left[2x - \frac{x^2}{2}\right]_0^{q_2} = 0.5$. $2q_2 - \frac{q_2^2}{2} = 0.5 \implies q_2^2 - 4q_2 + 1 = 0$. Solving this quadratic equation: $q_2 = 2 - \sqrt{3}$.
4. q_3 (75th percentile): Solve for q_3 such that: $\int_0^{q_3} f_X(x)\,dx = 0.75$. $2q_3 - \frac{q_3^2}{2} = 0.75 \implies q_3^2 - 4q_3 + 1.5 = 0$. Solving this quadratic equation: $q_3 = 2 - \sqrt{2.5}$.

c. Provide an example showing that X and Y are not independent.

Ans: If X and Y were independent, then $P(Y \le y | X = x) = P(Y \le y)$. However, given that $x \le y \le 2$, the conditional probability depends on x. For example: $P(Y \le 1 | X = 0.5) \ne P(Y \le 1)$.

Specifically: $P(Y \le 1 | X = 0.5) = \frac{\text{Length of } y=0.5 \text{ to } y=1}{\text{Length of } y=0.5 \text{ to } y=2} = \frac{1-0.5}{2-0.5} = \frac{1}{3}$.

d. Let the dart board be $B = \{(x, y) : 0 \le x \le y \le c\}$. Write down the PDF of X in terms of c.

Ans: The PDF of X for a general $c > 0$ is: $f_X(x) = \begin{cases} \frac{c-x}{\frac{c^2}{2}} & \text{if } 0 \leq x \leq c, \\ 0 & \text{otherwise.} \end{cases}$ The normalization factor is the area of the dartboard, which is $\frac{c^2}{2}$.

Buses pass by my stop at the top of the hour, 10 minutes past, and 30 minutes past the hour (e.g., at 8:00, 8:10, 8:30, 9:00, 9:10, 9:30, etc.). I arrive at the stop at a completely random time during the day. If the random variable Y represents the time I wait for the bus, we solve the following:

a. Draw and write down the PDF for Y.

Ans: The waiting time Y is distributed over the intervals defined by the bus schedule. Since buses arrive at 0, 10, and 30 minutes past the hour, the waiting time intervals are: - $0 \le Y \le 10$ with slope increasing from 0 to 1/10, - $10 \le Y \le 30$ with slope constant at 1/10, - $30 \le Y \le 60$ with slope decreasing from 1/10 to 0.

The PDF of Y is: $f_Y(y) = \begin{cases} \frac{1}{10}y & \text{if } 0 \le y \le 10, \\ \frac{1}{10} & \text{if } 10 < y \le 30, \\ \frac{6-y}{30} & \text{if } 30 < y \le 60, \\ 0 & \text{otherwise.} \end{cases}$

b. Find the following.

Ans:

1. The expected waiting time:

The expected value of Y is: $E[Y] = \int_0^{10} y \cdot \frac{1}{10} y\, dy + \int_{10}^{30} y \cdot \frac{1}{10}\, dy + \int_{30}^{60} y \cdot \frac{6-y}{30}\, dy.$

For $0 \le Y \le 10$: $\int_0^{10} y \cdot \frac{1}{10} y\, dy = \frac{1}{10} \int_0^{10} y^2\, dy = \frac{1}{10} \left[\frac{y^3}{3}\right]_0^{10} = \frac{1}{10} \cdot \frac{1000}{3} = \frac{100}{3}.$

For $10 \le Y \le 30$: $\int_{10}^{30} y \cdot \frac{1}{10}\, dy = \frac{1}{10} \int_{10}^{30} y\, dy = \frac{1}{10} \left[\frac{y^2}{2}\right]_{10}^{30} = \frac{1}{10}\left(\frac{900}{2} - \frac{100}{2}\right) = \frac{1}{10} \cdot 400 = 40.$

For $30 \le Y \le 60$: $\int_{30}^{60} y \cdot \frac{6-y}{30}\, dy = \frac{1}{30} \int_{30}^{60} (6y - y^2)\, dy = \frac{1}{30} \left[3y^2 - \frac{y^3}{3}\right]_{30}^{60}$. Substituting limits: $\frac{1}{30}\left(3(60^2) - \frac{60^3}{3} - 3(30^2) + \frac{30^3}{3}\right) = \frac{1}{30} \cdot 540 = 18.$

Adding these, $E[Y] = \frac{100}{3} + 40 + 18 = \frac{304}{3} \approx 101.33.$

2. The probability of waiting exactly 10 minutes:

Since Y is continuous, $P(Y = 10) = 0$.

3. The probability of waiting at most 5 minutes:

$P(Y \le 5) = \int_0^5 \frac{1}{10} y\, dy = \frac{1}{10} \int_0^5 y\, dy = \frac{1}{10} \cdot \frac{25}{2} = \frac{5}{4}.$

4. The IQR: Find the quartiles q_1 and q_3 by solving: $\int_0^{q_1} f_Y(y)\, dy = 0.25, \quad \int_0^{q_3} f_Y(y)\, dy = 0.75$. For q_1: $\int_0^{q_1} \frac{1}{10} y\, dy = 0.25 \implies \frac{1}{10} \cdot \frac{q_1^2}{2} = 0.25 \implies q_1 = \sqrt{5}.$

For q_3, note that it lies in the second interval $10 \leq q_3 \leq 30$: $\int_0^{10} \frac{1}{10} y\, dy + \int_{10}^{q_3} \frac{1}{10}\, dy = 0.75$. $\frac{1}{10} \cdot \frac{10^2}{2} + \frac{1}{10}(q_3 - 10) = 0.75 \implies 5 + \frac{q_3-10}{10} = 0.75$. Solving for q_3, $q_3 = 20$.

The IQR is: $q_3 - q_1 = 20 - \sqrt{5}$.

c. Bus schedule to produce $k = 5$ and generalization for $k \geq 1$.

Ans: For $k = 5$, divide the hour into 5 equal intervals: $0, 12, 24, 36, 48, 60$. The PDF will have 5 stair-steps.

To generalize for any $k \geq 1$, the bus schedule divides the hour into k equal intervals: $0, \frac{60}{k}, \frac{120}{k}, \ldots, 60$. The PDF for Y will have k stair-steps, with slopes proportional to the interval length.