PS04

Sejal Nimkar

2024-10-14

1. A small liberal arts college receives applications for admission from 1000 high school seniors. The college has dormitory space for a freshman class of 95 students and will have to arrange for off-campus housing for any additional freshmen. In previous years, an average of 64 percent of the students that the college has accepted have elected to attend another school. Clearly the college should accept more than 95 students, but its administration does not want to take too big a chance that it will have to accommodate more than 95 students. After some delibration, the administrators decide to accept 225 students. Answer the following questions as well as you can with the information provided.

a. How many freshmen do you expect that the college will have to accommodate?

Ans: In the above question, we are given the following:

n = total number of students accepted: 225

q = probability of students that choose to attend another school = 64

p = probability of students that will attend the college = $1 - p = 1 - 0.64 = 0.36$

Thus, the following distribution is binomial:

$X \sim \text{Binomial}(n = 225, p = 0.36)$

Hence expected value for binomial distribution is : $np = 225 \times 0.36 = 81$

b. What is the probability that the college will have to arrange for some freshmen to live off-campus?

Ans: Let X be a random variable which takes values of the number of students that choose to attend the college.

Therefore we have to find the probability that the number of students who choose to attend college is greater than 95 such that the college will have prepare accomodations off-campus.

Thus we want to calculate $P(X > 95)$.

$\therefore P(X > 95) = 1 - P(X \leq 95)$

It is tedious to calculate $P(X \leq 95)$ manually hence we will use R.

By Central Limit Theorem, the binomial distribution becomes more symmetric and resembles a normal distribution.

Thus we can use the normal distribution to calculate $P(X > 95)$.

```r
# Define parameters
mu <- 225 * 0.36
sigma <- sqrt(225 * 0.36 * 0.64)

# Calculate the probability P(X > 95) using the normal distribution
probability <- 1 - pnorm(95, mean = mu, sd = sigma)
probability
```

```
## [1] 0.02592094
```

2. Imagine throwing darts at a circular dart board, B . Let us measure the dart board in units for which the radius of B is 1 , so that the area of B is π . 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. Thus, if $A \subset B$, then the probability of hitting a poin in A is:

$$P(A) = \frac{area(A)}{area(B)} = \frac{area(A)}{\pi}$$

Define the random variable X to be the distance from the center of B to the point that is hit.

a. What are the possible values of X ?

Ans: The random variable X is a continuous random variable

The possible values of X range from 0 (center of the dartboard) to 1 (radius of the dartboard).

b. Compute $P(X \leq 0.5)$.

Ans: As given in the question: $P(A) = \dfrac{area(A)}{\pi}$

$$\therefore P(X \leq 0.5) = \frac{area(0 < X < 0.5)}{\pi}$$

Area of circle with radius 0.5 is $\pi \times 0.5^2$

$$\therefore P(X \leq 0.5) = \frac{\pi \times 0.5^2}{\pi} = 0.5^2 = 0.25$$

c. Compute $P(0.5 < X \leq 0.7)$.

Ans: Here, to find the probability of $P(0.5 < X \leq 0.7)$, we will determine the difference between the probabilities of $P(X < 0.5)$ and $P(X < 0.7)$.

We found $P(X < 0.5) = 0.25$ in the earlier question

$$P(X \leq 0.7) = \frac{area(0 < X < 0.7)}{\pi} = \frac{\pi \times 0.7^2}{\pi} = 0.7^2 = 0.49$$

$$\therefore P(0.5 < X \leq 0.7) = 0.49 - 0.25 = 0.24$$

d. Determine and graph the cumulative distribution function of X .

Ans: The cumulative distribution function (CDF) of X, $F(x)$, gives the probability that X is less than or equal to a value x.

For $X \in [0, 1]$:

$$F(x) = P(X \leq x) = \frac{\text{Area of a circle with radius } x}{\pi} = \frac{\pi x^2}{\pi} = x^2$$

So, the CDF is:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

The graph for the above is shown below:

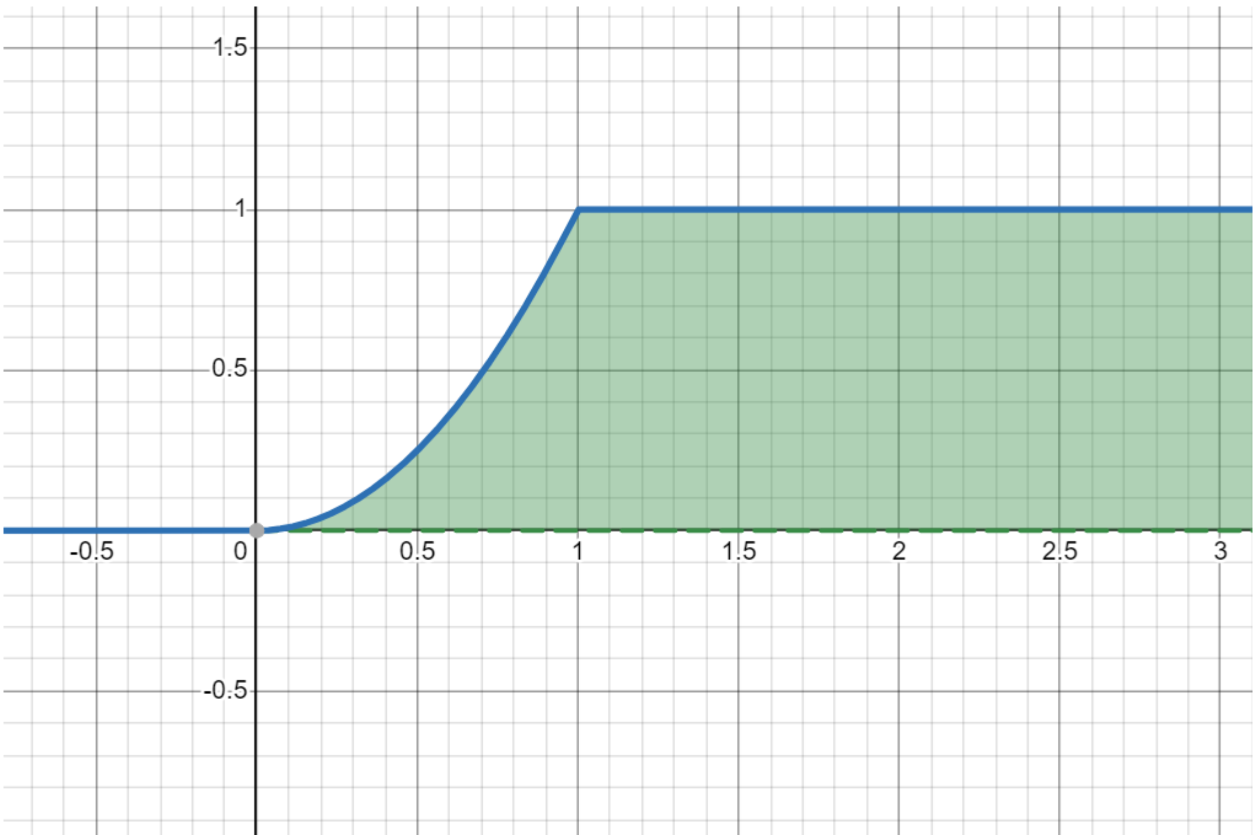

Figure 1: CDF

3. Suppose that X is a continuous random variable with probability density function

$$f(x) = \begin{cases} 0 & x < 0 \\ x & x \in (0,1) \\ \dfrac{3-x}{4} & x \in (1,3) \\ 0 & x > 3 \end{cases}$$

a. Compute $q_2(X)$, the population median.

Ans: 2nd quartile is the value of X at which the probability is 0.5. The below graph represents the pdf of f(x).

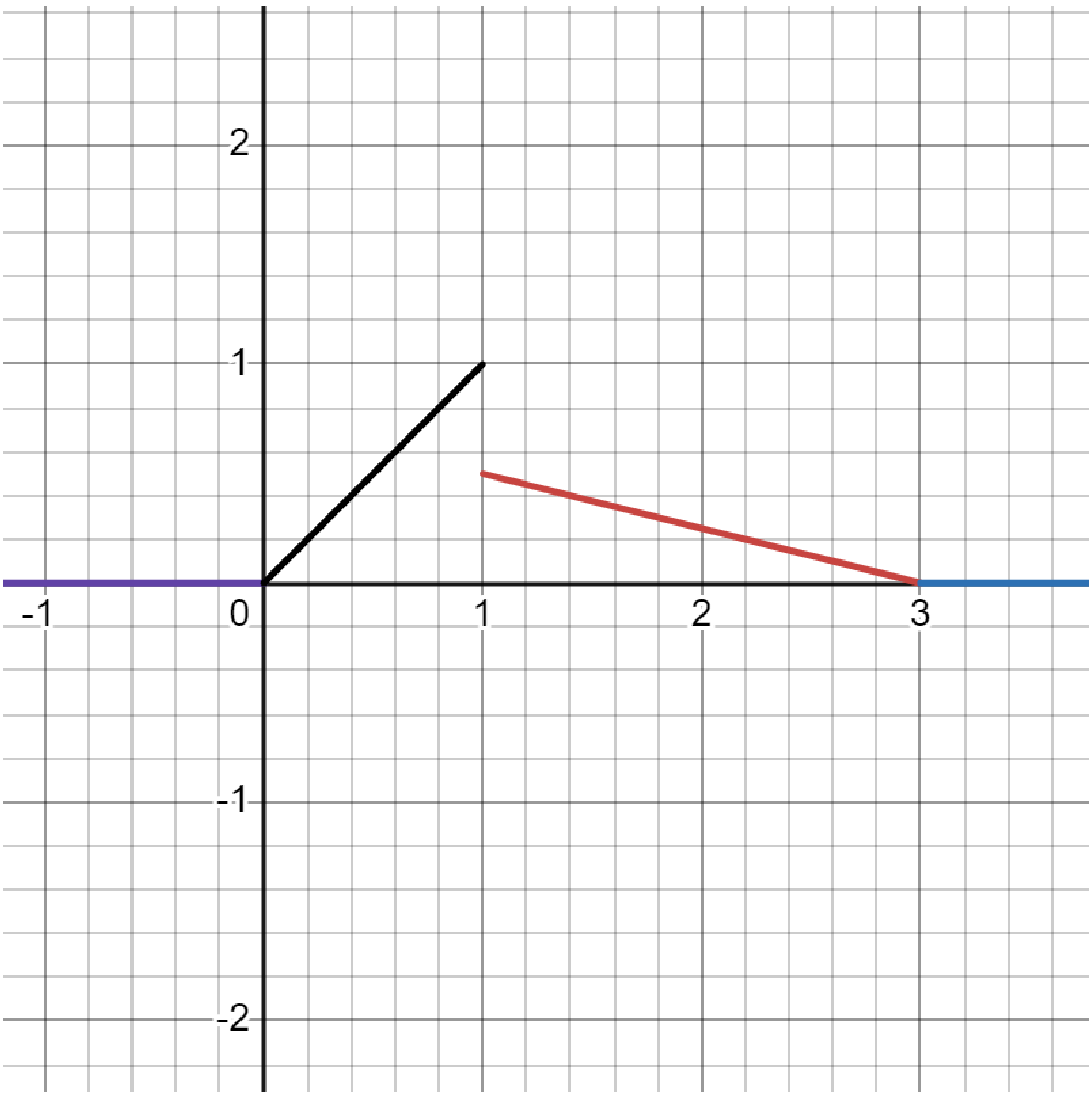

Figure 2: Graph for Q3

Let's determine the probability when $0 < X < 1$.

$\therefore P(0 < X < 1) =$ area under triangle when X ranges from 0 to 1

$\therefore P(0 < X < 1) = \frac{1}{2} \times$ base $\times$ height

Here, base $= 1$ unit

height $= 1$ unit

$\therefore P(0 < X < 1) = \frac{1}{2} \times \; 1 \times 1 = \frac{1}{2} = 0.5$

$q_2 =$ second quartile $= 1$

b. Which is greater, $q_2(X)$ or $E[X]$? Explain your reasoning.

Ans: From the graph, we can see that the data is skewed towards the right side than the left. henec the mena i.e. the expected value will shift towards the right side, which will make the expected value greater than the 2nd quartile.

$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$

$\therefore E(X) = 0.5 \times 1 + 0.5 \times (3 - 1) = 0.5 + 1 = 1.5$

From above we get $q_2 = 1$

$\therefore E(X) > q_2$

c. Compute $P(0.5 < X < 1.5)$.

Ans: We will subtract $P(0 < X < 0.5)$ i.e the area of graph under $0 < X < 0.5$ from $P(0 < X < 1.5)$ i.e. the area of graph under($0 <$ X < 1.5)

base of triangle under $(0 < X < 0.5) = 0.5$

height of triangle under $(0 < X < 0.5) = x = 0.5$

$P(0 < X < 0.5) = \frac{1}{2} \times 0.5 \times 0.5 = 0.125$

$P(0 < X < 1.5)$ i.e area under $(0 < X < 1.5)$ is equal to area under $(0 < X < 1)+$ area under $(1 < X < 1.5$

Area under $(1 < X < 1.5 =$ area of trapezium

Height of trapezium is $1.5 - 1 = 0.5$

Length of parallel sides is $\frac{3-1}{4}$ and $\frac{3-1.5}{4}$ i.e 0.5 and 0.375

Therefore Area under $(1 < X < 1.5 = \frac{0.5 + 0.375}{2} \times 0.5 = 0.21875$

$P(0 < X < 1.5) = 0.5 + 0.21875 = 0.71875$

$\therefore P(0.5 < X < 1.5) = P(0 < X < 1.5) - P(0 < X < 0.5) = 0.71875 - 0.125 = 0.59375$

d. Compute iqr(X), the population interquartile range.

Ans: q_1 i.e. the first quartile is the probability where $P(X) = 0.25$

We can use the area of triangle to determine q_1:

$\frac{1}{2} \times q_1 \times q_1 = 0.25$

$\therefore q_1^2 = 0.25 \times 2 = 0.5$

$\therefore q_1 = \sqrt{0.5} = 0.7071$

The area of a trapezoid is given by:

$A = \frac{1}{2}(b_1 + b_2)h$

where: - $b_1 = f(1) = \frac{1}{2}$ - $b_2 = f(q_3) = \frac{3 - q_3}{4}$ - $h = q_3 - 1$ (the length of the interval)

We want this area to equal 0.25. So, the equation becomes:

$\frac{1}{2}\left(\frac{1}{2} + \frac{3 - q_3}{4}\right)(q_3 - 1) = 0.25$

Simplifying the terms inside the parentheses:

$\frac{1}{2} + \frac{3 - q_3}{4} = \frac{5 - q_3}{4}$

Now the equation is:

$\frac{1}{2} \times \frac{5 - q_3}{4} \times (q_3 - 1) = 0.25$

$\frac{5 - q_3}{4} \times (q_3 - 1) = 0.5$

$(5 - q_3)(q_3 - 1) = 2$

$5q_3 - q_3^2 - 5 + q_3 = 2$

Simplifying:

$-q_3^2 + 6q_3 - 5 = 2$

$q_3^2 - 6q_3 + 7 = 0$

Solving the quadratic equation using the quadratic formula:

$q_3 = \frac{-(-6) \pm \sqrt{(-6)^2 - 4(1)(7)}}{2(1)} = \frac{6 \pm \sqrt{36 - 28}}{2} = \frac{6 \pm \sqrt{8}}{2} = \frac{6 \pm 2\sqrt{2}}{2}$

$q_3 = 3 \pm \sqrt{2}$

Since q_3 must lie between 1 and 3, we take the smaller root:

$q_3 = 3 - \sqrt{2} \approx 1.586$

$\therefore iqr = q_3 - q_1 = 1.586 - 0.707 = 0.879$

4. Imagine throwing darts at a triangular dart board, $B = (x, y) : 0 \leq y \leq x \leq 1$. 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. Consider the dart-throwing experiment described and compute the following quantities:

a. $q_2(X)$

Ans:
Median of X, $q_2(X)\}$ To find the median, we solve: $F_X(q_2) = 0.5$. Since $F_X(x) = x^2$, as F_X is calculated using the area of the graph, we set: $q_2^2 = 0.5 \implies q_2 = \sqrt{0.5} \approx 0.707$.

b. $q_2(Y)$

Ans:
Median of Y, $q_2(Y)\}$ The conditional median of Y, given X, is $\frac{X}{2}$. To find the overall median: $q_2(Y) = \int_0^1 \frac{x}{2} \cdot 2x \, dx = \int_0^1 x^2 \, dx = \frac{1}{3}$.

c. iqr(X)

Ans:
Interquartile Range of X, IQR$(X)\}$ To compute IQR(X), find the first and third quartiles.

- First quartile q_1 satisfies $F_X(q_1) = 0.25$: $q_1^2 = 0.25 \implies q_1 = 0.5$.
- Third quartile q_3 satisfies $F_X(q_3) = 0.75$: $q_3^2 = 0.75 \implies q_3 = \sqrt{0.75} \approx 0.866$.
- Therefore, IQR$(X) = q_3 - q_1 = 0.866 - 0.5 = 0.366$.

d. iqr(Y)

Ans:
Interquartile Range of Y, IQR$(Y)\}$

To find the interquartile range of Y, we need to consider the conditional distribution of Y given X.

1. **Conditional distribution of** $Y \mid X = x$: Y is uniformly distributed over $[0, x]$, so its cumulative distribution function given $X = x$ is: $F_{Y|X}(y \mid X = x) = \frac{y}{x}$.
2. **First quartile** $q_1(Y \mid X = x)$: Solve $F_{Y|X}(q_1 \mid X = x) = 0.25$: $\frac{q_1}{x} = 0.25 \implies q_1 = 0.25x$.
3. **Third quartile** $q_3(Y \mid X = x)$: Solve $F_{Y|X}(q_3 \mid X = x) = 0.75$: $\frac{q_3}{x} = 0.75 \implies q_3 = 0.75x$.
4. **Interquartile range of** Y **given** $X = x$: IQR$(Y \mid X = x) = q_3 - q_1 = 0.75x - 0.25x = 0.5x$.
5. **Compute the overall interquartile range of** Y: IQR$(Y) = \int_0^1 0.5x \cdot 2x \, dx = \int_0^1 x^2 \, dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$.

Thus, the interquartile range of Y is $\frac{1}{3}$.