PS03

Sejal Nimkar

2024-10-05

1. Use the urn {1,1,1,1,2,2,2,3,3,4}. The experiment is to draw two tickets, with replacement, and let the random variable X be the sum of these two tickets.

a. Find the 1st, 2nd, and 3rd quartiles of X.

Ans:

The total number of tickets in the urn are 10. Thus total number of combinations are $10 \times 10 = 100$

The range of random variable X where X assigns the sum of two tickets is:

$X(S) = \{2, 3, 4, 5, 6, 7, 8\}$.

The probability for $X = 2$ can be calculated as:

Possible combinations: (1,1)

Frequency: $4(\text{for } 1) \times 4(\text{for } 1) = 16$

$\therefore$ Probability $P(X = 2) = \dfrac{16}{100} = 0.16$

Similarly, probability for $X = 3$

Possible combinations: (1,2), (2,1)

Frequency (2,1) : $4(\text{for } 1) \times 3(\text{for } 2) = 12$

Frequency (1,2): $3(\text{for } 2) \times 4(\text{for } 1) = 12$

$\therefore$ Total frequency = 12 + 12 = 24

Probability: $P(X) = \dfrac{24}{100} = 0.24$

Calculating for the rest, we get:

$$P(x) = \begin{cases} 0.16 & \text{if } X = 2 \\ 0.24 & \text{if } X = 3 \\ 0.25 & \text{if } X = 4 \\ 0.20 & \text{if } X = 5 \\ 0.10 & \text{if } X = 6 \\ 0.04 & \text{if } X = 7 \\ 0.01 & \text{if } X = 8 \end{cases}$$

Therefore the cumulative distribution function is:

$$F(x) = \begin{cases} 0.16 & \text{if } X = 2 \\ 0.40 & \text{if } X = 3 \\ 0.65 & \text{if } X = 4 \\ 0.85 & \text{if } X = 5 \\ 0.95 & \text{if } X = 6 \\ 0.99 & \text{if } X = 7 \\ 1 & \text{if } X = 8 \end{cases}$$

For any quantile q: $P(X < q) \leq \alpha$ and $P(X > q) \leq 1 - \alpha$

1st quartile $= q_1 = 25^{th}$ percentile of the distribution can be given as:

To calculate q_1 for $\alpha = 0.25$:

$F(2) = 0.16$ and $F(3) = 0.40$

Thus, α will lie in between $X = 2$ and $X = 3$

$\therefore$ First quartile $= q_1 = 3$

Similarly, for q_2 i.e. $2^{nd} quartile$, the value of $\alpha = 0.50$ will lie in between $F(3) = 0.40$ and $F(4) = 0.65$

$\therefore$ Second quartile $= q_2 = 4$

For q_3 i.e. $3^{rd} quartile$, the value of $\alpha = 0.75$ will lie in between $F(4) = 0.65$ and $F(5) = 0.85$

$\therefore$ Third quartile $= q_3 = 5$

```
# Values of X and their corresponding probabilities
values <- c(2, 3, 4, 5, 6, 7, 8)
probabilities <- c(0.16, 0.24, 0.25, 0.20, 0.10, 0.04, 0.01)

# Repeating each value by its probability (scaled by a large factor)
data <- rep(values, times = probabilities * 1000)

# Calculate the quartiles (Q1, Q2, Q3)
quartiles <- quantile(data, probs = c(0.25, 0.50, 0.75))

# Print the quartiles
quartiles
```

```
## 25% 50% 75%
##   3   4   5
```

b. Find the IQR of X.

Ans: Interquartile range is given by IQR $= q_3 - q_1$

$\therefore IQR = 5 - 3 = 2$

```
IQR(data)
```

```
## [1] 2
```

2. Consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by:

$$f(x) = \begin{cases} 0, & x < 0 \\ cx, & 0 < x < 1.5 \\ c(3-x), & 1.5 < x < 3 \\ 0, & x > 3 \end{cases}$$

where c is an undetermined constant.

(a) For what value of c is f a probability density function?

Ans: For f(x) to be a probability density function, two conditions should be satisfied:

a. $f(x) \geq 0$ for all $x \in R$

b. The integral of $f(x)$ over all $x \in R$ must equal 1.

$\therefore \int_{-\infty}^{\infty} f(x)\,dx = 1$

$\therefore \int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{1.5} cx\,dx + \int_{1.5}^{3} c(3-x)\,dx = 1$

$\therefore \int_{-\infty}^{\infty} f(x)\,dx = c\int_{0}^{1.5} x\,dx + 3c\int_{1.5}^{3} dx \,-c\int_{1.5}^{3} x\,dx$

$\therefore \int_{-\infty}^{\infty} f(x)\,dx = 1.125c + 3c \times 1.5 - 3.375c$

$\therefore \int_{-\infty}^{\infty} f(x)\,dx = 1.125c + 4.5c - 3.375c = 2.625c$

$\therefore \int_{-\infty}^{\infty} f(x)\,dx = 2.25c = 1$

$\therefore c = \frac{1}{2.25} = \frac{4}{9} = 0.44$

(b) Suppose that a continuous random variable X has probability density function f . Compute EX . (Hint: Draw a picture of the pdf.)

Ans:

The Triangle here is an isosceles triangle. Thus it is symmetric.

Area of triangle is given by : $\frac{1}{2} \times base \times height$

Height can be determined by:

$f(1.5) = \frac{4}{9} \times 1.5 = \frac{2}{3}$

Therefore area of half triangle is :

Area of half triangle $= \frac{1}{2} \times 1.5 \times \frac{4}{9} = 0.5$

Figure 1: Graph

Therefore $E(X)$ is given by:

$E(X) = \int_{-\infty}^{\infty} xf(x)dx$

$\therefore E(X) = [1.5 - 0] \times 0.5 + [3 - 1.5] \times 0.5 = 1.5$

(c) Compute P (X > 2).

Ans: $P(X > 2) = Areaunder\{2 < X < 3\}$

$\therefore P(X > 2) = Areaoftrianglefrom\{2 < X < 3\} = \frac{1}{2} \times$ base $\times$ height

base $= 3 - 2 = 1$

height $= f(2) = c(3 - x) = \frac{4}{9}(3 - 2) = \frac{4}{9}$

$\therefore P(X > 2) = \frac{1}{2} \times 1 \times \frac{4}{9} = \frac{4}{18} = 0.22$

(d) Suppose that $Y \sim Uniform(0, 3)$. Which random variable has the larger variance, X or Y ? (Hint: Draw a picture of the two pdfs.)

Ans:

The purple coloured graph represents the distribution of Y while the green coloured graph represents the distribution of random variable X.

Intuitively, by checking the graphs, the uniform distribution Y seems more spread out towards the ends than the graph of the random variable X.

Thus it is possible that the variance of Y is greater than the variance of X.

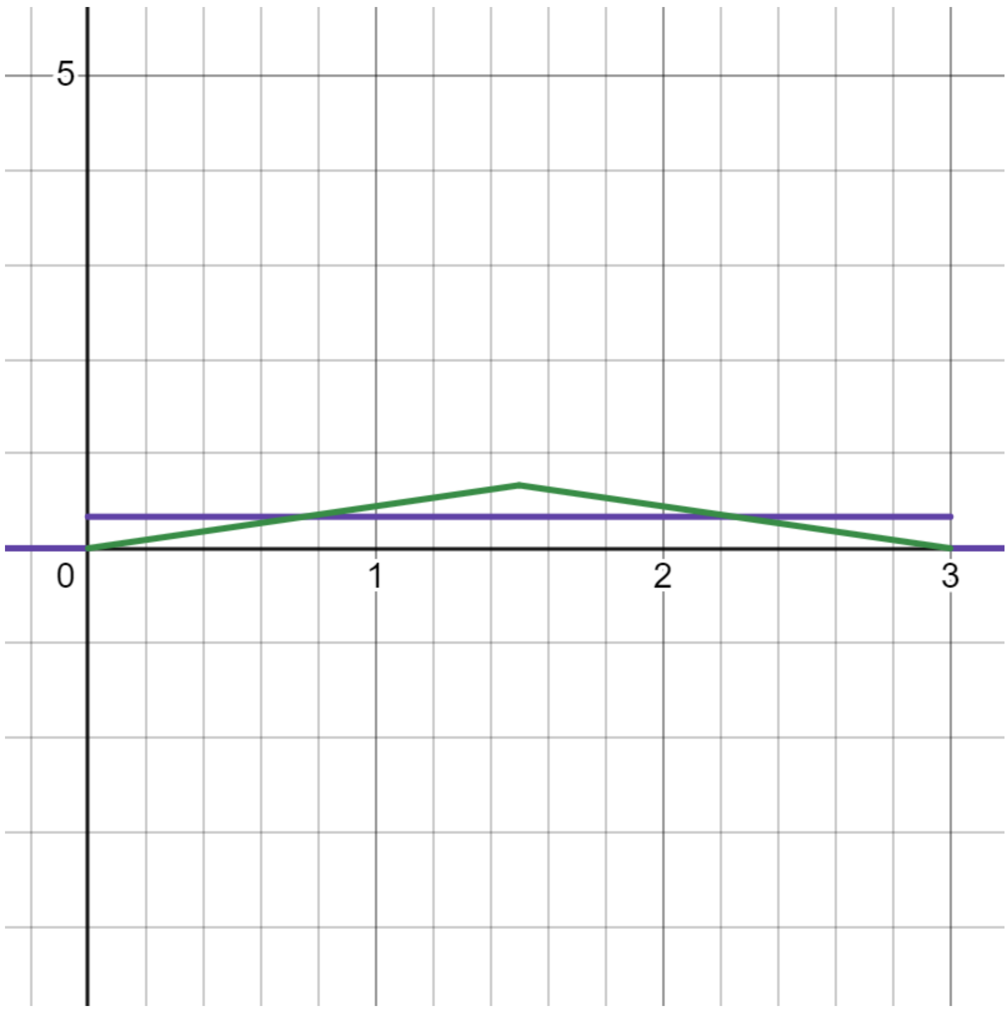

Figure 2: Graph of random variables X and Y

(e) What is the median of X?

Ans: For a symmetric continuous graph of random variable X, the median coincides with the mean $E(X)$. For random variable X, the graph has equal areas on the either side of x = 1.5 therefore the median of X is 1.5.

(f) What is the IQR of Y?

Ans: IQR of Y can be given by calculating the 25th percentile and the 75th percentile of Y :

Let's calculate this using R:

```
q1= qunif(p=0.25,min=0,max=3)
cat("First quartile q1 =", q1)
```

```
## First quartile q1 = 0.75
```

```
q3= qunif(p=0.75,min=0,max=3)
cat("First quartile q3 =", q3)
```

```
## First quartile q3 = 2.25
```

```
IQR= q3 - q1
cat("IQR = ", IQR)
```

```
## IQR =  1.5
```

3. Buses pass my stop precisely 20 minutes apart throughout the day. I arrive at the stop at a completely random time during the day. If the random variable T_2 represents the time I wait for the bus, determine the following:

a. What is the expected waiting time?

Ans: The random variable T_2 represents waiting for a bus, then the distribution of random variable T_2 is uniform.

For a uniform distribution $X \sim \text{Uniform}(a, b)$, the expected value (mean) is:

$E(X) = \dfrac{a+b}{2}$

$\therefore E(T_2) = \frac{0+20}{2} = 10$ minutes

b. What is the probability that I must wait at least 15 minutes for the bus?

Ans: Let's calculate this using R:

```
p_wait_15_or_more <- 1 - punif(15, min = 0, max = 20)
p_wait_15_or_more
```

```
## [1] 0.25
```

c. What is the 87th percentile?

Ans: Let's calculate this using R:

```
percentile_87 <- qunif(0.87, min = 0, max = 20)
print(percentile_87)
```

```
## [1] 17.4
```

The 87th percentile is 17.4 minutes.

4. Buses pass my stop precisely 20 minutes and 40 minutes past the hour (e.g., ~8:20, 8:40, 9:20, 9:40, etc.). I arrive at the stop at a completely random time during the day. If the random variable T_3 represents the time I wait for the bus, determine the following:

a. What is the expected waiting time?

Ans: This question is similar to the previous question, where the waiting time between each bus is 20 minutes, although the interval defined here is different from the previous one.

Thus distribution of T_3 is also uniform.

For a uniform distribution $X \sim \text{Uniform}(a, b)$, the expected value (mean) is:

$E(X) = \dfrac{a+b}{2}$

$\therefore E(T_3) = \frac{0+20}{2} = 10$ minutes

b. What is the probability that I must wait at least 15 minutes for the bus?

Ans: Let's calculate this using R:

```
p_wait_15_or_more <- 1 - punif(15, min = 0, max = 20)
p_wait_15_or_more
```

```
## [1] 0.25
```

c. What is the 87th percentile?

Ans: Let's calculate this using R:

```
percentile_87 <- qunif(0.87, min = 0, max = 20)
print(percentile_87)
```

```
## [1] 17.4
```

The 87th percentile is 17.4 minutes.

5. We let X_1 and X_2 represent the heights of women's and men's heights, respectively, and we assume that

$X_1 \sim \mathcal{N}(63.7, 2.7^2)$ and $X_2 \sim \mathcal{N}(69.6, 2.8^2)$.

a) What is the probability that the man is over six feet?

Ans: Here,

For X_1:

Standard deviation $= \sqrt{\sigma^2} = 2.7$ inches

Mean $= \mu = 63.7$ inches

For X_2:

Standard deviation $= \sqrt{\sigma^2} = 2.8$ inches

Mean $= \mu = 69.6$ inches

To determine the probability that a man is over 6 feet i.e. 72 inches, we'll use R.

The probability that a man is 6 feet is $P(X_2 \leq 72)$:

```
x2_sd = 2.8
x2_mean= 69.6

prob_equal_to_72 = pnorm(q= 72, sd= x2_sd, mean = x2_mean)
print(prob_equal_to_72)
```

```
## [1] 0.804317
```

$P(X_2 \leq 72) = 0.804317$

Therefore the probability than a man is over 6 feet is $P(X > 72)$:

```
prob_over_72=1-prob_equal_to_72
print(prob_over_72)
```

```
## [1] 0.195683
```

$P(X_2 > 72) = 0.195683$

b) Let $S = X_1 + X_2$. Use S to determine the probability that the sum of the man's and woman's height is over 11 feet.

Ans: SinceX_1 and X_2 are independent, S can be given as:

$S \sim \mathcal{N}(63.7 + 69.6, 2.7^2 + 2.8^2)$

The mean and the standard deviation for the distribution of S can be given as:

mean $= \mu_s = \mu_x 1 + \mu_x 2 = 63.7 + 69.6$

standard deviation $= \sqrt{\sigma_s^2} = 2.7^2 + 2.8^2$

Thus the probability that the sum of heights of man's and woman's is over 11 feet i.e. 132 inches is given by $P(S > 11) =$

```r
mean_S <- 63.7 + 69.6
sd_S <- sqrt(2.7^2 + 2.8^2)

# Calculate the probability
prob_over_11_feet <- 1 - pnorm(132, mean = mean_S, sd = sd_S)
prob_over_11_feet
```

```
## [1] 0.6308907
```

$\therefore P(S > 11) = 0.6308907$

c) What is the IQR of S?

Ans: IQR of S can be calculated using R:

mean $= 63.7 + 69.6 = 133.3$

Standard deviation = ()

```r
mean_S <- 63.7 + 69.6
sd_S <- sqrt(2.7^2 + 2.8^2)

q_1 = qnorm(p = 0.25, mean = mean_S,sd = sd_S )
cat("First Quartile: ",q_1,"\n")
```

```
## First Quartile:  130.6764
```

```r
q_3 = qnorm(p= 0.75, mean = mean_S, sd = sd_S)
cat("Third quartile:", q_3, "\n")
```

```
## Third quartile: 135.9236
```

```r
IQR= q_3 - q_1
cat(" IQR :", IQR)
```

```
##  IQR : 5.247166
```

d) Let $D = X_1 - X_2$. Use D to determine the probability that the random man is shorter than the random woman.

Ans: X_1 is the height of a random woman while X_2 is the height of a random man.

To determine the probability that the height of a random man is less than random woman, the value of D should be positive i.e. greater than zero.

Therefore we have to find the probability of $P(D > 0)$.

$S \sim \mathcal{N}(63.7 - 69.6, 2.7^2 - 2.8^2)$

The mean and the standard deviation for the distribution of S can be given as:

mean $= \mu_s = \mu_x 1 + \mu_x 2 = 63.7 - 69.6$

standard deviation $= \sqrt{\sigma_s^2} = 2.7^2 - 2.8^2$

```
mean_D <- 63.7 - 69.6
sd_D <- sqrt(2.7^2 + 2.8^2)

# Probability P(D > 0)
prob_d_greater_than_zero <- 1 - pnorm(0, mean = mean_D, sd = sd_D)
cat("Probability :", prob_d_greater_than_zero, "\n")
```

```
## Probability : 0.06465673
```

Therefore the probability that the height of a random man is less than that of a woman is:

$P(D > 0) = 0.06465673$

e) What is the 95th percentile of D?

Ans:

```
mean_D <- 63.7 - 69.6
sd_D <- sqrt(2.7^2 + 2.8^2)
D_95_percentile= qnorm(p=0.95,mean = mean_D ,sd = sd_D)
cat("95th percentile of D is :",D_95_percentile)
```

```
## 95th percentile of D is : 0.4980366
```