

YZV231E - PROB. STAT. FOR DATA SCIENCE

Nazrin Abdinli, 150220925

26/12/2023

Question 1

#Q1 - PART1 SOLUTION

```
def calculate_skewness(data):
    """
    Args:
        data (numpy.ndarray): A skewed normal distribution.
    Returns:
        int: Skewness value of the input distribution.
    """
    ##### CODE HERE #####
    N = len(data)          #determining the number of data points
    mean = np.mean(data)    #calculating the mean
    m2 = np.sum((data - mean) ** 2) / N    #calculating the second central moment
    m3 = np.sum((data - mean) ** 3) / N    #calculating the third central moment

    #calculating the numerators and denominators for the skewness formula
    numerator_1 = np.sqrt(N * (N - 1))
    denominator_1 = (N - 2)
    numerator_2 = m3
    denominator_2 = np.power(m2, 3/2)

    #calculating the skewness value using Fisher-Pearson coefficient formula
    skewness = (numerator_1/denominator_1) * (numerator_2/denominator_2)
    return skewness

    ##### CODE HERE #####

# Your calculation
est_skewness = calculate_skewness(data_skewed)

# Library call
true_skewness = stats.skew(data_skewed, bias=False)
print(est_skewness, true_skewness)
```

#Q1 - PART2 SOLUTION

```
##### CODE HERE #####
#given data
np.random.seed(42)
sample_size = 1000
param_skewness = 10
skewed_data = stats.skewnorm.rvs(a=param_skewness, size=sample_size)

#calculating mean and median
mean_skewed = np.mean(data_skewed)
median_skewed = np.median(data_skewed)

#plotting the histogram of the skewed data
plt.hist(skewed_data, bins=20, density=True, color='skyblue', edgecolor='black', alpha=0.7)

#plotting the theoretical distribution
x_min, x_max = plt.xlim()
x_line = np.linspace(x_min, x_max, 100)
pdf = stats.skewnorm.pdf(x_line, param_skewness)
plt.plot(x_line, pdf, 'k', linewidth=2)

#designing plot
plt.axvline(mean_skewed, color='red', linestyle='dashed', linewidth=1, label='Mean')
plt.axvline(median_skewed, color='green', linestyle='dashed', linewidth=1, label='Median')

plt.title("Skewed Normal Distribution with Mean and Median")
plt.xlabel("Value")
plt.ylabel("Density")
plt.legend()
plt.show()

#displaying the calculated mean and median values
print(f"Mean value: {mean_skewed}")
print(f"Median value: {median_skewed}")

##### CODE HERE #####
```

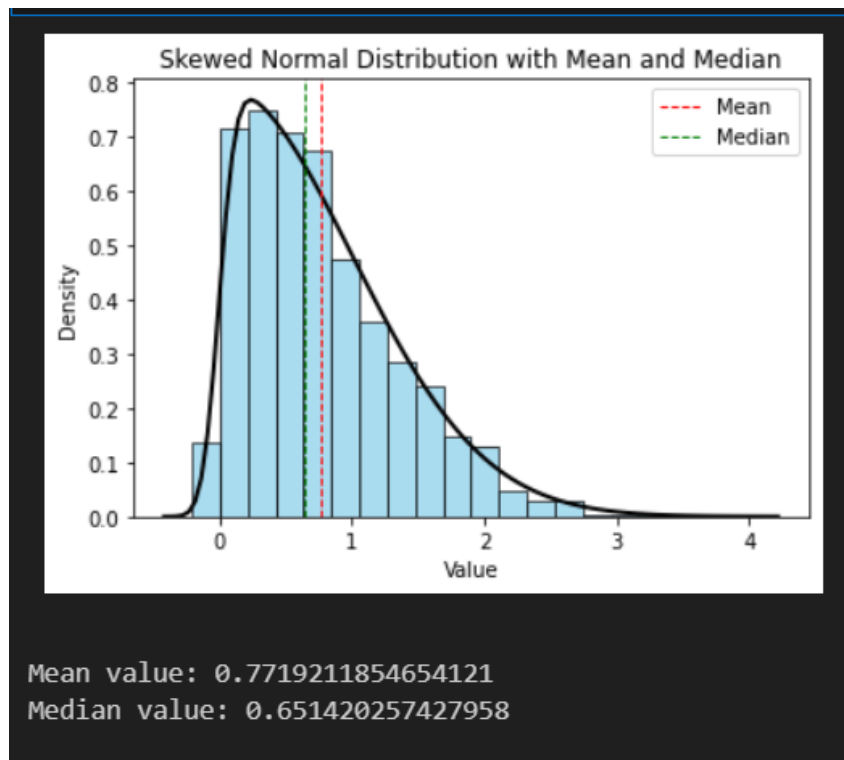


Figure 1: Mean and Median Values of the Distribution

#Q1 - PART3 SOLUTION

Positive skewness: $\text{mean} > \text{median}$ (right-skewed)

Negative skewness: $\text{mean} < \text{median}$ (left-skewed)

Skewness close to zero: $\text{mean} \approx \text{median}$ (approximately symmetric distribution)

When a set of numbers has a positive skewness, it means that most of the numbers are clustered on the left side, and there are a few very large values pulling the mean higher than the median.

Conversely, if there's a negative skewness, it indicates that most numbers are on the right side, with a few very small values dragging the mean lower than the median.

When the skewness is close to zero, it means that the numbers are fairly spread out and balanced, so the mean and the median are roughly around the same point. This suggests a more symmetric distribution without a noticeable leaning towards either larger or smaller values.

Question 2

#Q2 - PART1 SOLUTION

Number of dice rolls for each experiment

num_rolls = [50, 100, 200, 400]

Number of experiments

num_experiments = 1000

Create a 2x2 grid of subplots

fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(10, 8), sharey=True)

Flatten the axes array to simplify indexing in the loop

axes = axes.flatten()

for i, ax in enumerate(axes):

np.random.seed(42)

CODE HERE

sum_dice_rolls = []

for _ in range(num_experiments):

dice_rolls = np.random.randint(1, 7, size=num_rolls[i]) #simulating dice rolls

sum_dice_rolls.append(np.sum(dice_rolls)) #calculating the sum

#plotting histogram

ax.hist(sum_dice_rolls, bins=30, density=True, color='skyblue', edgecolor='black',
 ↪ alpha=0.7)

ax.set_title(f'number of rolls = {num_rolls[i]}')

ax.set_ylabel('Frequency')

CODE HERE

Adjust layout to prevent subplot titles from overlapping

plt.tight_layout()

Show the plot

plt.show()

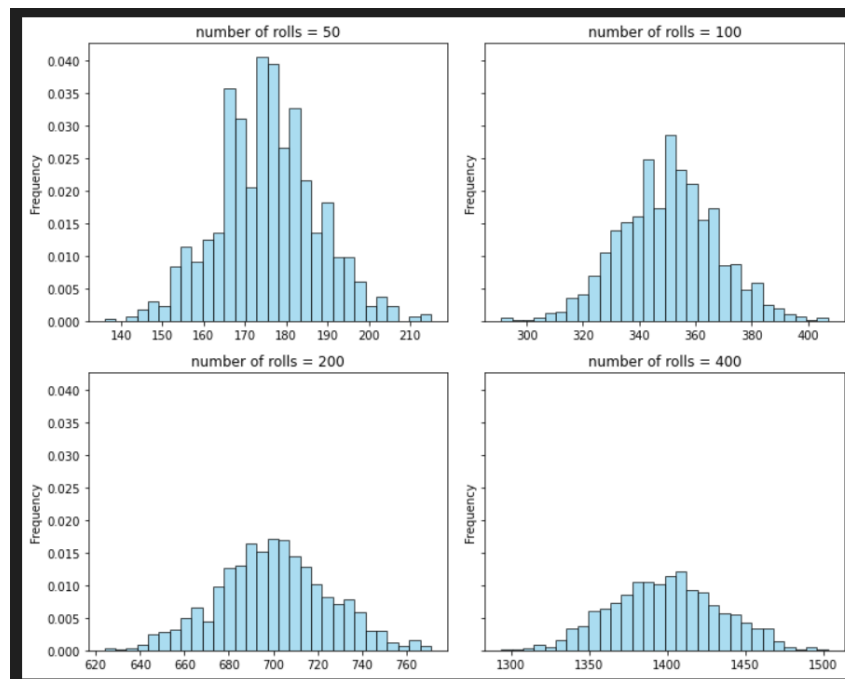


Figure 2: Distribution of Sum of the Dice Rolled for Each Experiment

#Q2 - PART2 SOLUTION

Number of dice rolls for each experiment

```
num_rolls = [50, 100, 200, 400]
```

Number of experiments

```
num_experiments = 1000
```

Create a 2x2 grid of subplots

```
fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(10, 8), sharex=True, sharey=True)
```

Flatten the axes array to simplify indexing in the loop

```
axes = axes.flatten()
```

```
for i, ax in enumerate(axes):
```

```
    ##### CODE HERE #####
```

```
    np.random.seed(42)
```

```
    average_dice_rolls = []
```

```
    for _ in range(num_experiments):
```

```
        dice_rolls = np.random.randint(1, 7, size=num_rolls[i]) #simulating dice rolls
```

```
        average_dice_rolls.append(np.mean(dice_rolls)) #calculating the average
```

```
    #plotting histogram for the current number of rolls
```

```
ax.hist(average_dice_rolls, bins=30, density=True, color='skyblue', edgecolor='black',
        alpha=0.7)
ax.set_title(f'number of rolls = {num_rolls[i]}')
ax.set_ylabel('Frequency')
```

```
##### CODE HERE #####
```

```
# Adjust layout to prevent subplot titles from overlapping
plt.tight_layout()
```

```
# Show the plot
plt.show()
```

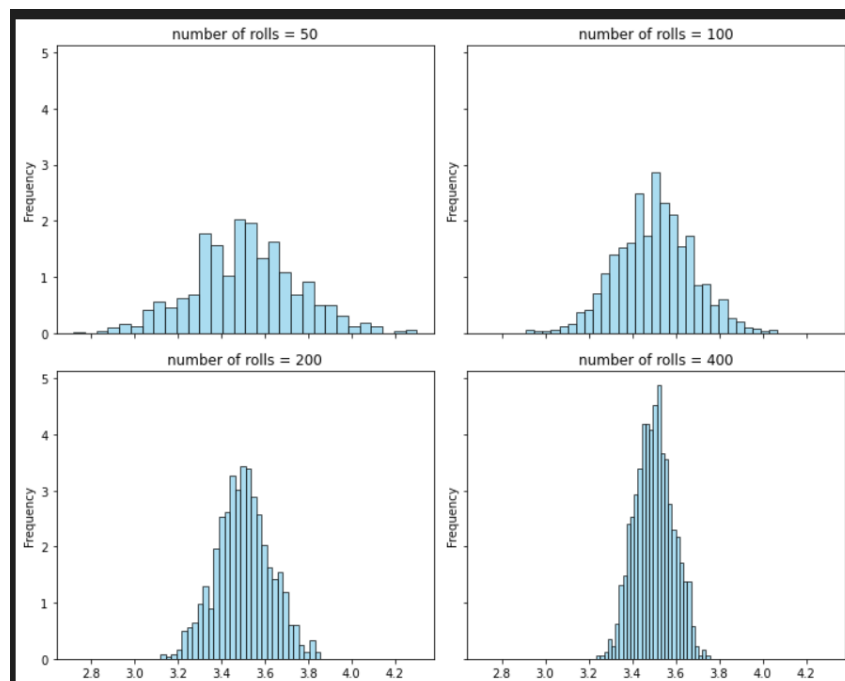


Figure 3: Distribution of Average of the Dice Rolled for Each Experiment

#Q3 - PART3 SOLUTION

When you increase the number of the dice rolls, you're effectively increasing the range of the sum of possible outcomes. For example, when you use a single dice roll, you can get the sum outcomes from 1 to 6. As you incline the number of rolls in an experiment, the possible range of the sum grows accordingly.

When you average the outcomes of the dice rolls, the averaging process tends to reduce the variability. The reduction in variability happens because when you take the average of several values, the influence of extreme values decreases, particularly as the sample size gets larger. As a result, the range of average values becomes narrower and more closely centered around the anticipated mean.

1 Problem 1 Handwriting

1.1 Subsection A

$$\begin{aligned}f_X(x) &= \int_0^1 f_{X,Y}(x,y) dy \rightarrow \\&\rightarrow f_X(x) = \int_0^1 \frac{12}{49} (2 + x + xy + 4y^2) dy \rightarrow \\&\rightarrow f_X(x) = \frac{12}{49} \left[2y + xy + \frac{xy^2}{2} + \frac{4y^3}{3} \right]_0^1 \rightarrow \\&\rightarrow f_X(x) = \frac{12}{49} \left[2 + x + \frac{x}{2} + \frac{4}{3} \right] \rightarrow \\&\rightarrow f_X(x) = \frac{12}{49} \left[\frac{10}{3} + \frac{3x}{2} \right] \rightarrow \\&\rightarrow f_X(x) = \frac{40}{49} + \frac{18x}{49} = \frac{18x + 40}{49}\end{aligned}$$

1.2 Subsection B

$$\begin{aligned}f_Y(y) &= \int_0^1 f_{X,Y}(x,y) dx \rightarrow \\&\rightarrow f_Y(y) = \int_0^1 \frac{12}{49} (2 + x + xy + 4y^2) dx \rightarrow \\&\rightarrow f_Y(y) = \frac{12}{49} \left[2x + \frac{x^2}{2} + \frac{x^2 y}{2} + 4y^2 x \right]_0^1 \rightarrow \\&\rightarrow f_Y(y) = \frac{12}{49} \left[2 + \frac{1}{2} + \frac{1}{2}y + 4y^2 \right] \rightarrow \\&\rightarrow f_Y(y) = \frac{12}{49} \left[\frac{5}{2} + \frac{1}{2}y + 4y^2 \right] \rightarrow \\&\rightarrow f_Y(y) = \frac{30}{49} + \frac{6}{49}y + \frac{48}{49}y^2 = \frac{48y^2 + 6y + 30}{49}\end{aligned}$$

1.3 Subsection C

Two random variables, X and Y , are independent if and only if their joint probability function $f_{X,Y}(x,y)$ can be expressed as the product of their individual probability functions, $f_X(x)$ and $f_Y(y)$. In other words, for independent variables, the joint density function equals the product of the marginal density functions.

for independence:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

we are given $f_{X,Y}(x,y) = \frac{12}{49} (2 + x + xy + 4y^2)$ we found $f_X(x) = \frac{18x+40}{49}$ and $f_Y(y) = \frac{48y^2+6y+30}{49}$
calculation:

$$f_X(x) \cdot f_Y(y) = \frac{(18x + 40)(48y^2 + 6y + 30)}{49^2} \neq f_{X,Y}(x,y)$$

Since $f_{X,Y}(x,y) \neq f_X(x) \cdot f_Y(y)$ for all x and y within their ranges, X and Y are not independent variables

2 Problem 2 Handwriting

2.1 Subsection A

the possible values of $X : x = 1, 2, 3, 4, 5, 6$

the possible values of Y : depend on the value of X

$$f_{X,Y}(x, y) = P(X = x \& Y = y)$$

The number of heads obtained when flipping x coins follows a binomial distribution with parameters x (number of trials) and $p = 1/2$ (probability of success in each trial since it's a fair coin). In this case, Y follows a binomial distribution.

The joint PMF can be represented as:

$$f_{X,Y}(x, y) = \begin{cases} \left(\frac{1}{6}\right) \cdot \binom{x}{y} \cdot \left(\frac{1}{2}\right)^x, & \text{if } 0 \leq y \leq x \& 1 \leq x \leq 6 \\ 0, & \text{otherwise} \end{cases}$$

$\binom{x}{y} \rightarrow$ represents the number of ways to choose y successes (heads) out of x trials (coin flips) $\left(\frac{1}{2}\right)^x \rightarrow$ the probability of getting y heads in x flips, assuming a fair coin.

2.2 Subsection B

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$E(XY) = \sum_{X=1}^6 \sum_{y=0}^x xy \cdot f_{X,Y}(x, y)$$

we found $f_{XY}(x, y) = \frac{1}{6} \cdot \binom{x}{y} \cdot \left(\frac{1}{2}\right)^x$ for $1 \leq x \leq 6$ and $0 \leq y \leq x$, we can compute $E(XY)$:

$$E(XY) = \sum_{x=1}^6 \sum_{y=0}^x xy \cdot \frac{1}{6} \cdot \binom{x}{y} \cdot \left(\frac{1}{2}\right)^x$$

$$E(XY) = \frac{1}{6} \sum_{x=1}^6 x \cdot \underbrace{\sum_{y=0}^x y \cdot \binom{x}{y} \cdot \left(\frac{1}{2}\right)^x}_{x \cdot 2^{x-1}}$$

$$E(XY) = \frac{1}{6} \sum_{x=1}^6 x \cdot x \cdot 2^{x-1} \cdot \left(\frac{1}{2}\right)^x$$

$$E(XY) = \frac{1}{6} \sum_{x=1}^6 x^2 \cdot \frac{1}{2}$$

$$E(XY) = \frac{1}{12} \sum_{x=1}^6 x^2$$

$$E(XY) = \frac{1}{12} (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2)$$

$$E(XY) = \frac{91}{12}$$

$$E(X) = \sum_{x=1}^6 x \cdot P(X = x)$$

$$E(X) = \sum_{x=1}^6 x \cdot \frac{1}{6}$$

$$E(X) = \frac{1}{6} \cdot \frac{6 \cdot 7}{2} \rightarrow E(X) = \frac{7}{2}$$

$$E(Y) = E(x \cdot p) = E(x) \cdot P = \frac{7}{2} \cdot 0.5 = \frac{7}{4} \rightarrow E(Y) = \frac{7}{4}$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$\text{Cov}(X, Y) = \frac{91}{12} - \frac{7}{2} \cdot \frac{7}{4} = \frac{91}{12} - \frac{49}{8} = \frac{35}{24}$$

$$\text{Cov}(X, Y) = \frac{35}{24}$$

2.3 Subsection C

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}$$

$$\text{Var}(Y) = \sum_{x=1}^6 x \cdot p \cdot (1 - p) = 21 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{21}{4}$$

we found $\text{Cov}(X, Y) = \frac{35}{24}$

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \sum_{x=1}^6 x^2 \cdot P(X = x) = \sum_{x=1}^6 x^2 \cdot \frac{1}{6} = \frac{91}{6}$$

$$\text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{182 - 147}{12} = \frac{35}{12}$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{35/24}{\sqrt{\frac{35}{12} \cdot \frac{21}{4}}} = \frac{\frac{35}{24}}{\frac{7\sqrt{5}}{2}} =$$

$$= \frac{35}{24} \cdot \frac{2}{7\sqrt{5}} = \frac{5}{12\sqrt{5}} = \frac{5\sqrt{5}}{12 \cdot 5} = \frac{\sqrt{5}}{12}$$

$$\text{Corr}(X, Y) = \frac{\sqrt{5}}{12}$$

2.4 Subsection D

- The positive value indicates a tendency for both variables X and Y to move in the same direction. As one variable increases, the other tends to increase and vice versa
- However, the strength of this relation is not particularly strong. There is a discernible trend that the variables move together, but the relationship is not extremely strong
- Correlation does not mean causation. Even though these variables might show a relationship, it does not mean that changes in one variable cause changes in the other
- in summary, $\text{Corr}(X, Y) = \frac{\sqrt{5}}{12}$ indicates a moderate positive linear relation between X and Y , suggesting that X and Y tend to move together in the same direction but not very strongly

3 Problem 3 Handwriting

3.1 Subsection A

Finding all the values of X and $Y = 3$ according to the given joint probability:

$$f_{X,Y}(x | y) = \frac{f_{X,Y}(x, y)}{P(Y = y)}$$
$$P(Y = 3) = f_{X,Y}(2, 3) + f_{X,Y}(3, 3) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

Computing the conditional probability distribution for X given that $Y = 3$:

$$f_{X|Y=3}(x | 3) = \frac{f_{X,Y}(x, 3)}{P(Y = 3)}$$
$$f_{X|Y=3}(2 | 3) = \frac{f_{X,Y}(2, 3)}{P(Y = 3)} = \frac{1/5}{2/5} = \frac{1}{2}$$
$$f_{X|Y=3}(3 | 3) = \frac{f_{X,Y}(3, 3)}{P(Y = 3)} = \frac{1/5}{2/5} = \frac{1}{2}$$

Calculating the conditional expectation of X given $Y = 3$:

$$E[X | Y = 3] = \sum_x x \cdot f_{X|Y=3}(x | 3)$$
$$E[X | Y = 3] = 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = 1 + 1.5 = 2.5$$
$$E[X | Y = 3] = 2.5$$

3.2 Subsection B

Finding all the values of Y and $X = 3$ according to the given joint probability:

$$f_{Y,X}(y | x) = \frac{f_{X,Y}(x, y)}{P(X = x)}$$
$$P(X = 3) = f_{X,Y}(3, 2) + f_{X,Y}(3, 3) + f_{X,Y}(3, 17) =$$
$$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5}$$

Computing the conditional probability distribution for Y given that $X = 3$:

$$f_{Y|X=3}(y | 3) = \frac{f_{X,Y}(3, y)}{P(X = 3)}$$
$$f_{Y|X=3}(2 | 3) = \frac{f_{X,Y}(3, 2)}{P(X = 3)} = \frac{1/5}{3/5} = \frac{1}{3}$$
$$f_{Y|X=3}(3 | 3) = \frac{f_{X,Y}(3, 3)}{P(X = 3)} = \frac{1/5}{3/5} = \frac{1}{3}$$
$$f_{Y|X=3}(17 | 3) = \frac{f_{X,Y}(3, 17)}{P(X = 3)} = \frac{1/5}{3/5} = \frac{1}{3}$$

Calculating the conditional expectation of X given $Y = 3$:

$$E[Y | X = 3] = \sum_y y \cdot f_{Y|X=3}(y | 3)$$
$$E[Y | X = 3] = 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} + 17 \cdot \frac{1}{3} = \frac{2 + 3 + 17}{3} = \frac{22}{3}$$
$$E[Y | X = 3] = \frac{22}{3}$$

3.3 Subsection C

$$P(Y = 2) = f_{X|Y}(3, 2) + f_{X|Y}(2, 2) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

$$P(Y = 3) = f_{X|Y}(2, 3) + f_{X|Y}(3, 3) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

$$P(Y = 17) = f_{X|Y}(3, 17) = \frac{1}{5}$$

$$E[X | Y = 2] = \sum_X x \cdot f_{X|Y=2}(x|2)$$

$$E[X | Y = 2] = 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = 1 + 1.5 = 2.5$$

$$E[X | Y = 2] = 2.5$$

$$E[X | Y = 3] = 2.5$$

$$E[X | Y = 17] = \sum_x x \cdot f_{X|Y=17}(x | 17)$$

$$E[X | Y = 17] = 1 \cdot 3 = 3$$

$$E[X | Y = 17] = 3$$

$$E[X | Y] = 2.5 \cdot \frac{2}{5} + 2.5 \cdot \frac{2}{5} + 3 \cdot \frac{1}{5} = \frac{13}{5} = 2.6$$

$$E[X | Y] = 2.6$$

3.4 Subsection D

$$P(X = 2) = f_{X|Y}(2, 3) + f_{X|Y}(2, 2) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

$$P(X = 3) = f_{X|Y}(3, 2) + f_{X|Y}(3, 3) + f_{X|Y}(3, 17) = \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5}$$

$$E[Y | X = 2] = \sum_y y \cdot f_{Y|X=2}(y | 2)$$

$$E[Y | X = 2] = 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = \frac{5}{2} = 2.5$$

$$E[Y | X = 3] = \frac{22}{3}$$

$$E[Y | X] = \frac{5}{2} + \frac{22}{3} = \frac{15 + 44}{6} = \frac{59}{6}$$

4 Problem 4 Handwriting

$X_i \rightarrow$ the value shown on the i^{th} die when rolled

$$Z_n = X_1^2 + X_2^2 + \dots + X_n^2$$

Each X_i follows a discrete uniform distribution over $\{1, 2, 3, 4, 5, 6\}$, because we are rolling a fair 6-sided die

The mean of X_i^2 is:

$$\begin{aligned} E[X_i^2] &= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} \\ E[X_i^2] &= \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6} \end{aligned}$$

$Y_i = X_i^2$ for each i .

According to the strong law of large numbers, if Y_1, Y_2, \dots, Y_n are independent and identically distributed random variables with finite mean $\mu = \frac{91}{6}$, then:

$$\frac{1}{n} \sum_{i=1}^n Y_i \rightarrow \mu$$

In this case, $\frac{1}{n} Z_n = \frac{1}{n} (X_1^2 + X_2^2 + \dots + X_n^2) = \frac{1}{n} \sum_{i=1}^n Y_i$ converges almost surely to $\frac{91}{6}$ by the strong law of large numbers

So we have $m = \frac{91}{6}$ as the limit of $\frac{1}{n} Z_n$ as n approaches infinity.

5 Problem 5 Handwriting

5.1 Subsection A

$\lambda = \frac{1}{2}$ is given

$$\mu = \frac{1}{\lambda} \rightarrow \mu = \frac{1}{\frac{1}{2}} = 2 \text{ minutes}$$

$$\sigma = \frac{1}{\lambda} \rightarrow \sigma = \frac{1}{\frac{1}{2}} = 2$$

$$\mu_{\bar{x}} = \mu = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{16}} = \frac{2}{4} = 0.5 \text{ minutes} \quad x = 2.5 \text{ minutes}$$

$$\mu_{\bar{x}} = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = 0.5 \text{ minutes}$$

$$Z = \frac{x - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \rightarrow Z = \frac{2.5 - 2}{0.5} = \frac{0.5}{0.5} = 1$$

$P(Z < 1) \rightarrow$ From the standard normal distribution table, $P(Z < 1) = 0.8413$

5.2 Subsection B

$$\lambda = \frac{1}{2} \text{ is given}$$

$$\mu = \frac{1}{\lambda} \rightarrow \mu = \frac{1}{\frac{1}{2}} = 2 \text{ minutes}$$

$$\sigma = \frac{1}{\lambda} \rightarrow \sigma = \frac{1}{\frac{1}{2}} = 2$$

$$\mu_{\bar{x}} = \mu = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{36}} = \frac{2}{6} = \frac{1}{3} \text{ minutes}$$

$$x = 2.5 \text{ minutes}$$

$$\mu_{\bar{x}} = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = \frac{1}{3} \text{ minutes}$$

$$Z = \frac{x - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \rightarrow Z = \frac{2.5 - 2}{\frac{1}{3}} = \frac{0.5}{\frac{1}{3}} = 1.5$$

$P(Z < 1.5) \rightarrow$ From the standard normal distribution table, $P(Z < 1.5) = 0.9332$

5.3 Subsection C

$$\lambda = \frac{1}{2} \text{ is given}$$

$$\mu = \frac{1}{\lambda} \rightarrow \mu = \frac{1}{\frac{1}{2}} = 2 \text{ minutes}$$

$$\sigma = \frac{1}{\lambda} \rightarrow \sigma = \frac{1}{\frac{1}{2}} = 2$$

$$\mu_{\bar{x}} = \mu = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{100}} = \frac{2}{10} = \frac{1}{5} \text{ minutes}$$

$$x = 2.5 \text{ minutes}$$

$$\mu_{\bar{x}} = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = \frac{1}{5} \text{ minutes}$$

$$Z = \frac{x - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \rightarrow Z = \frac{2.5 - 2}{\frac{1}{5}} = \frac{0.5}{\frac{1}{5}} = 2.5$$

$P(Z < 2.5) \rightarrow$ From the standard normal distribution table, $P(Z < 2.5) = 0.9938$

5.4 Subsection D

The CLT suggests that as n increases:

1. The probability that the average service time of the first n customers is less than 2.5 is going to increase
2. The distribution of sample means tends to become more symmetric normal distribution. it occurs irrespective of the initial distribution of the population, demonstrating a more standardized and even shape over time
3. The variability of the sample mean distribution decreases. This decrease in standard error causes a narrower range around the true population mean.
4. Larger sample sizes enhance the accuracy of statistical conclusions. They lead to more precise estimations, narrower confidence intervals, and more reliable hypothesis testing, improving the overall accuracy of inferences drawn from the data