Applied Stochastic Assignment 2

Manyara Baraka - mbaraka

October 1, 2024

Question 1: Discrete Random Variables

Part 1

- Details given

$$p = 0.8n = 5$$

a. Total Probability = 1

- The total summation of all probabilities should be equal to 1.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(X = k) = \binom{5}{k} (0.8)^k (0.2)^{5-k}, \quad \text{for } k = 0, 1, 2, 3, 4, 5$$

$$P(X = 0) = \binom{5}{0} (0.8)^0 (0.2)^5 = 1 \times 1 \times 0.2^5 = 0.00032$$

$$P(X = 1) = \binom{5}{1} (0.8)^1 (0.2)^4 = 5 \times 0.8 \times 0.2^4 = 0.0064$$

$$P(X = 2) = \binom{5}{2} (0.8)^2 (0.2)^3 = 10 \times 0.64 \times 0.2^3 = 0.0512$$

$$P(X = 3) = \binom{5}{3} (0.8)^3 (0.2)^2 = 10 \times 0.512 \times 0.2^2 = 0.2048$$

$$P(X = 4) = \binom{5}{4} (0.8)^4 (0.2)^1 = 5 \times 0.4096 \times 0.2 = 0.4096$$

$$P(X = 5) = \binom{5}{5} (0.8)^5 (0.2)^0 = 1 \times 0.32768 \times 1 = 0.32768$$

$$\sum_{k=0}^{5} P(X = k) = 0.00032 + 0.0064 + 0.0512 + 0.2048 + 0.4096 + 0.32768 = 1$$

b. Expected Value and Variance

Expected Value of X:

For a binomial distribution, the expected value $E(X)$:

$$E(X) = n \cdot p$$

substitute n and p:

$$E(X) = 5 \cdot 0.8 = 4$$

Variance of X:

The variance $\text{Var}(X)$ for a binomial distribution is given by the formula:

$$\text{Var}(X) = n \cdot p \cdot (1 - p)$$

$$\text{Var}(X) = 5 \cdot 0.8 \cdot 0.2 = 0.8$$

Interpretation

- The expected value of 4 means that, on average, the factory can expect 4 out of 5 components to pass the quality check daily.
- The variance of 0.8 indicates the level of variability around the mean.

Using This Information:

- Estimating Daily Production Quality: The factory can use the expected value of 4 to predict that, on average, 80% of the components tested daily will pass the quality check.
- Monitoring Production: Incase the number of components deviates from the expected value will indicate a problem in the production.

c. Probability of only 3 out of 5 will pass

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Where:

$$\begin{aligned} n &= 5 \quad \text{(number of components tested)} \\ k &= 3 \quad \text{(number of components that pass)} \\ p &= 0.8 \quad \text{(probability of a component passing)} \\ 1 - p &= 0.2 \quad \text{(probability of failure)} \end{aligned}$$

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5 \times 4}{2 \times 1} = 10$$

Substituting:

$$P(X=3) = \binom{5}{3}(0.8)^3(0.2)^2$$

$$P(X=3) = 10 \times (0.8)^3 \times (0.2)^2$$

$$P(X=3) = 10 \times 0.512 \times 0.04$$

$$P(X=3) = 10 \times 0.02048 = 0.2048$$

Therefore: $P(X=3) = 0.2048$.
How Rare or Frequent is this Scenario?

- The probability of 0.2048, indicates that this scenario will occur about 20% of the time.

Implications for Managing Production Quality:

- This will be below the average mean of 4 out of 5, this will be a lower rate of the average production, and it will indicate a problem in production or an improvement in the process.
- This could indicate some natural variability in the production process and hence this will help in managing variability.

Part 2

- Being a Poison solution we have:

$$\lambda = 4$$

a. PMF for X of all values in range = 1

Poisson's PMF for random variable X is given by:

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{for } k = 0, 1, 2, 3, \ldots$$

Using $\lambda = 4$, it becomes:

$$P(X=k) = \frac{4^k e^{-4}}{k!}, \quad \text{for } k = 0, 1, 2, 3, \ldots$$

To show that the PMF is valid over all ranges:

$$\sum_{k=0}^{\infty} P(X=k) = \sum_{k=0}^{\infty} \frac{4^k e^{-4}}{k!} = 1$$

Re-write and take constants out:

$$e^{-4}\sum_{k=0}^{\infty}\frac{4^k}{k!}$$

Using Taylor's series expansion: e^x:

$$e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}$$

our expression is at $x = 4$ substituting it:

$$\sum_{k=0}^{\infty}\frac{4^k}{k!} = e^4$$

Re-writing the equation:

$$e^{-4} \cdot e^4 = 1$$

Since the total sum of probabilities equals 1, the PMF is valid. Therefore, the Poisson PMF with parameter $\lambda = 4$ correctly describes a probability distribution.

b. Expected value and Varience

Expected Value

Poison's expected value $E(X)$ is given by:

$$E(X) = \lambda$$

Given that $\lambda = 4$, therefore:

$$E(X) = 4$$

Variance

The variance $\mathrm{Var}(X)$ of a Poisson-distributed random variable X is also equal to λ:

$$\mathrm{Var}(X) = \lambda = 4$$

Significance:

- Expected value $E(X) = 4$: On average, 4 defective products are returned per period. This can be used to set the target goal of production if the number goes lower this will indicate an error in the production.
- Variance $\mathrm{Var}(X) = 4$: The variation in the number of returns will have a variance of 4. This will help in identifying trends if any number is recorded this will indicate an unusual trend.

c. Probability of receiving fewer than 3

Calculate $P(X < 3)$

Since $\lambda = 4$, we are to calculate $P(X = 0)$, $P(X = 1)$, and $P(X = 2)$.

1. For $k = 0$:

$$P(X = 0) = \frac{4^0 e^{-4}}{0!} = e^{-4} \approx 0.0183$$

2. For $k = 1$:

$$P(X = 1) = \frac{4^1 e^{-4}}{1!} = 4e^{-4} \approx 0.0732$$

3. For $k = 2$:

$$P(X = 2) = \frac{4^2 e^{-4}}{2!} = 8e^{-4} \approx 0.1465$$

Summing the Probabilities

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X < 3) \approx 0.0183 + 0.0732 + 0.1465 = 0.2380$$

Real life implication:

- **Planning inventory:** 23.8% indicates a minimal likelihood of having a defect, if the warehouse knows this they will optimize the stock levels, which will potentially reduce the excess inventory costs.

- **Customer service response:** having a reasonable chance of receiving fewer than 3 defective products the warehouse may enhance things like customer messaging when having fewer defects to ensure the product quality is promoted and also streamline the return process.

Question 2: Joint Probability and Covariance

Details given:

- X is the number of defective components in a batch of $n = 100$ components

Part1: Proof of Concept

a. Show that as $n \to \infty$ and $p \to 0$, such that $\lambda = np$ remains constant

The binomial PMF is:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Substitute $p = \frac{\lambda}{n}$, where $\lambda = np$ remains constant:

$$P(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

As $n \to \infty$, we use the approximation $\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}$.

The binomial coefficient can be approximated as $\binom{n}{k} \approx \frac{n^k}{k!}$.
Substitute:

$$P(X = k) \approx \frac{n^k}{k!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot e^{-\lambda}$$

Simplify the powers of n:

$$P(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

The PMF of a Poisson distribution with λ. Hence, the binomial distribution approaches a Poisson distribution as $n \to \infty$ and $p \to 0$.

b. Show that

$$\lim_{n\to\infty, p\to 0} P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where $\lambda = np$.

The binomial PMF is:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Substitute $p = \frac{\lambda}{n}$:

$$P(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

From part (a) we have:

$$P(X=k) \approx \frac{n^k}{k!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot e^{-\lambda}$$

Simplifying:

$$P(X=k) \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

c. Why poison is a good approximation for binomial distribution

- Binomila indicates the number of successes in n independent trials with a success probability of p while poison indicates the number of events happening in a fixed number of intervals with an average rate of λ. When n is large and p is small, it shows that the number of trials is many/ large while the number of successes in each trial is low. In this scenario, binomial distribution starts behaving like poison.

Part 2: Poison Approximation:

a. Probability for $P(X=3)$

The Poisson PMF:

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where $\lambda = np$,

$$p = 0.01$$

therefore:

$$\lambda = 100 \times 0.01 = 1$$

Substitute $k=3$ and $\lambda = 1$ into the formula:

$$P(X=3) = \frac{1^3 e^{-1}}{3!} = \frac{1 \times e^{-1}}{6} \approx 0.0613$$

b. Probability for $P(X \leq 5)$

$$P(X \leq 5) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$$

Poisson PMF for each value of k, with $\lambda = 1$:

$$P(X=k) = \frac{1^k e^{-1}}{k!}$$

$$P(X=0) = \frac{1^0 e^{-1}}{0!} = e^{-1} \approx 0.3679$$

$$P(X=1) = \frac{1^1 e^{-1}}{1!} = e^{-1} \approx 0.3679$$

$$P(X=2) = \frac{1^2 e^{-1}}{2!} = \frac{e^{-1}}{2} \approx 0.1839$$

$$P(X=3) = \frac{1^3 e^{-1}}{3!} = \frac{e^{-1}}{6} \approx 0.0613$$
$$P(X=4) = \frac{1^4 e^{-1}}{4!} = \frac{e^{-1}}{24} \approx 0.0153$$
$$P(X=5) = \frac{1^5 e^{-1}}{5!} = \frac{e^{-1}}{120} \approx 0.0031$$

Sum:

$$P(X \leq 5) \approx 0.3679 + 0.3679 + 0.1839 + 0.0613 + 0.0153 + 0.0031 = 0.9994$$

$$\text{Thus,} P(X \leq 5) \approx 0.9994$$

c. Implication for quality control:

- The probability of finding exactly 3 defective components in a batch of 100 is relatively small ($\approx 6.13\%$), indicating that it is not a frequent event. However, the probability of finding at most 5 defective components is almost certain ($\approx 99.94\%$), which implies that the factory is very likely to encounter between 0 and 5 defective components in any given batch of 100.

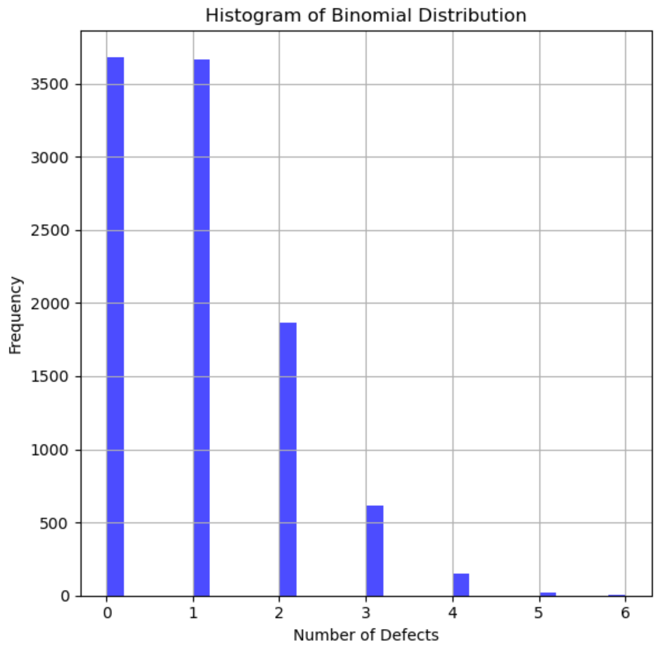
Part 3: Simulation

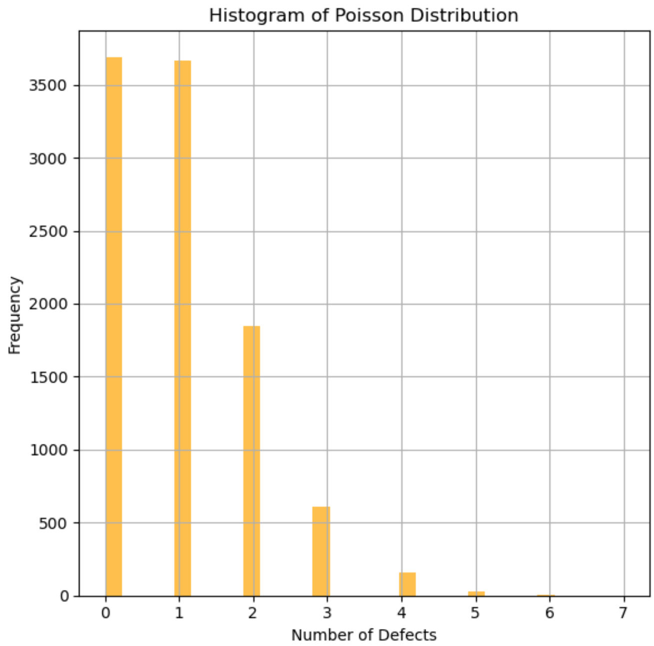
```
#Import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Parameters
n_batches = 10000
n_components = 100
prob_defect = 0.01
lambda_value = 1

# Simulate binomial distribution B(100, 0.01)
binomial_samples = np.random.binomial(n=n_components, p=prob_defect, size=n_batches)

# Simulate Poisson distribution with  = 1
poisson_samples = np.random.poisson(lam=lambda_value, size=n_batches)

# Calculate means and variances
binomial_mean = np.mean(binomial_samples)
binomial_var = np.var(binomial_samples)
poisson_mean = np.mean(poisson_samples)
poisson_var = np.var(poisson_samples)

print(f"Binomial Mean: {binomial_mean:.2f}, Variance: {binomial_var:.2f}")
print(f"Poisson Mean: {poisson_mean:.2f}, Variance: {poisson_var:.2f}")
```

```
# Plot histograms comparing the two distributions
plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 1)
plt.hist(binomial_samples, bins=30, color='blue', alpha=0.7)
plt.title('Histogram of Binomial Distribution')
plt.xlabel('Number of Defects')
plt.ylabel('Frequency')
plt.grid()

plt.subplot(1, 2, 2)
plt.hist(poisson_samples, bins=30, color='orange', alpha=0.7)
plt.title('Histogram of Poisson Distribution')
plt.xlabel('Number of Defects')
plt.ylabel('Frequency')
plt.grid()

plt.tight_layout()
plt.show()
```

Results:

- **Binomial**:
  - Mean: 0.99
  - Variance: 0.99
- **Poisson**:
  - Mean: 1.01
  - Variance: 1.02





Analysis

- Poisson distribution approximates the binomial distribution well with the values almost within the same range, the defects are small in the probability with the large numbers.

Question 3: Poisson Distribution

Details given:

- $\lambda = 5$ (the average number of calls per minute before the reduction),
- $\lambda = 7$ (the average number of calls per minute after the reduction),
- 7 : 3 Ratio of intra-network to inter-network calls.

  Before the Tariff Reduction

  - Intra-network tariff: 70 RwF per minute
  - Inter-network tariff: 90 RwF per minute

  After the Tariff Reduction

  - Intra-network tariff: 60 RwF per minute
  - Inter-network tariff: 80 RwF per minute

Part 1

a. Find the probability of P(X = 7)

Use Poisson distribution PMF :

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$k = 7$ and $\lambda = 5$
Substitute the values:

$$P(X = 7) = \frac{5^7 e^{-5}}{7!}$$

$$P(X = 7) = \frac{78125 \times 0.006738}{5040}$$

$$P(X = 7) \approx \frac{526.466875}{5040} \approx 0.1045$$

Therefore, P(X = 7):

$$P(X = 7) \approx 0.1045$$

b. The probability of $P(X < 3)$

$\lambda = 7$:
Therefore:

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

The Poisson PMF is given by:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$P(X = 0)$

$$P(X = 0) = \frac{7^0 e^{-7}}{0!} = e^{-7}$$

$$P(X = 0) \approx 0.00091188$$

$P(X = 1)$

$$P(X = 1) = \frac{7^1 e^{-7}}{1!} = 7e^{-7}$$

$$P(X = 1) \approx 0.00638316$$

$P(X = 2)$

$$P(X = 2) = \frac{7^2 e^{-7}}{2!} = \frac{49e^{-7}}{2}$$

$$P(X = 2) \approx 0.02229156$$

Summing:

$$P(X < 3) \approx 0.00091188 + 0.00638316 + 0.02229156 \approx 0.0295866$$

The probability that fewer than 3 calls are made in a given minute after the tariff reduction is approximately 0.0296, or 2.96%.

c. Expected Number of Intra and Inter-network calls before:

Expected Number of Intra-Network Calls

$$= \frac{7}{10} \times 5 = 3.5 \text{ calls per minute}$$

Expected Number of Inter-Network Calls

$$= \frac{3}{10} \times 5 = 1.5 \text{ calls per minute}$$

d. Expected number of Intra and Inter-network calls after

Expected Number of Intra-Network Calls

$$= \frac{7}{10} \times 7 = 4.9 \text{ calls per minute}$$

Expected Number of Inter-Network Calls

$$= \frac{3}{10} \times 7 = 2.1 \text{ calls per minute}$$

e. Expected revenue per minute before and after the tariff reduction

Revenue from Intra-Network Calls Before Reduction

$$= 3.5 \times 70 = 245 \text{ RwF per minute}$$

Revenue from Inter-Network Calls Before Reduction

$$= 1.5 \times 90 = 135 \text{ RwF per minute}$$

Total Revenue Before Tariff Reduction

$$= 245 + 135 = 380 \text{ RwF per minute}$$

Revenue from Intra-Network Calls After Reduction

$$= 4.9 \times 60 = 294 \text{ RwF per minute}$$

Revenue from Inter-Network Calls After Reduction

$$= 2.1 \times 80 = 168 \text{ RwF per minute}$$

Total Revenue After Tariff Reduction

$$= 294 + 168 = 462 \text{ RwF per minute}$$

- The expected revenue **before the tariff reduction** was 380 RwF per minute.
- The expected revenue **after the tariff reduction** is 462 RwF per minute.

f. Simulation

```
import numpy as np

# Given values
lambda_before = 5  # Poisson parameter before tariff reduction
lambda_after = 7   # Poisson parameter after tariff reduction

intra_tariff_before = 70  # RwF per minute (intra-network) before reduction
inter_tariff_before = 90  # RwF per minute (inter-network) before reduction

intra_tariff_after = 60   # RwF per minute (intra-network) after reduction
inter_tariff_after = 80   # RwF per minute (inter-network) after reduction

intra_ratio = 7 / 10      # Ratio for intra-network calls
inter_ratio = 3 / 10      # Ratio for inter-network calls

repetitions = 100000      # Number of repetitions for simulation

# Simulate Poisson-distributed number of calls for 100,000 repetitions
calls_before = np.random.poisson(lambda_before, repetitions)
calls_after = np.random.poisson(lambda_after, repetitions)
```

```
# Simulate revenue before tariff reduction
intra_calls_before = calls_before * intra_ratio
inter_calls_before = calls_before * inter_ratio
revenue_before = (intra_calls_before * intra_tariff_before) + (inter_calls_before * i

# Simulate revenue after tariff reduction
intra_calls_after = calls_after * intra_ratio
inter_calls_after = calls_after * inter_ratio
revenue_after = (intra_calls_after * intra_tariff_after) + (inter_calls_after * inter

# Calculate the probability that revenue exceeds 500 RwF
prob_before_exceeds_500 = np.mean(revenue_before > 500)
prob_after_exceeds_500 = np.mean(revenue_after > 500)

print(f'The probability before reduction {prob_before_exceeds_500:.2f}')
print(f'The Probability after reduction {prob_after_exceeds_500:.2f}')
```

Results:

- The probability **before reduction: 0.24 = 24%**
- The Probability **after reduction: 0.40 = 40%**

Part 2: Simulation

```
#Import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Parameters
lambda_before = 5  # avg calls per minute before reduction
lambda_after = 7  # avg calls per minute after reduction

# Simulate calls for 1440 minutes (1 day)
calls_before = np.random.poisson(lambda_before, 1440)
calls_after = np.random.poisson(lambda_after, 1440)

# Cap the number of calls to a maximum of 10
calls_before[calls_before > 10] = 10
calls_after[calls_after > 10] = 10

# Tariff information
intra_network_before = 70  # RwF per minute
inter_network_before = 90  # RwF per minute
intra_network_after = 60  # RwF per minute
inter_network_after = 80  # RwF per minute
```

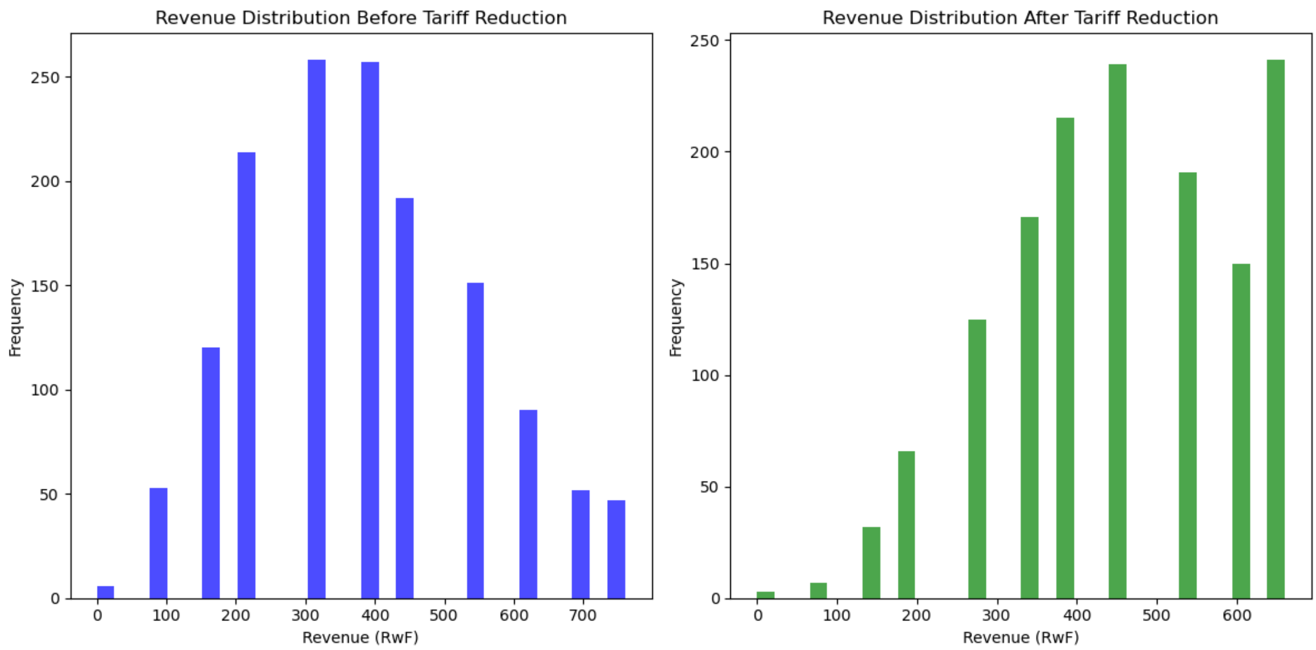
```
# Ratio of calls
ratio_intra = 7 / 10
ratio_inter = 3 / 10

# Create a DataFrame to hold the results
call_data = pd.DataFrame({
    'Calls_Before_Reduction': calls_before,
    'Calls_After_Reduction': calls_after
})

# Calculate revenues
call_data['Intra_Network_Before'] = call_data['Calls_Before_Reduction'] * ratio_intra
call_data['Inter_Network_Before'] = call_data['Calls_Before_Reduction'] * ratio_inter
call_data['Revenue_Before'] = (call_data['Intra_Network_Before'] * intra_network_befo
                               call_data['Inter_Network_Before'] * inter_network_befo

call_data['Intra_Network_After'] = call_data['Calls_After_Reduction'] * ratio_intra
call_data['Inter_Network_After'] = call_data['Calls_After_Reduction'] * ratio_inter
call_data['Revenue_After'] = (call_data['Intra_Network_After'] * intra_network_after
                              call_data['Inter_Network_After'] * inter_network_after)

# Plot revenue distributions using histograms
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(call_data['Revenue_Before'], bins=30, color='blue', alpha=0.7)
plt.title('Revenue Distribution Before Tariff Reduction')
plt.xlabel('Revenue (RwF)')
plt.ylabel('Frequency')

plt.subplot(1, 2, 2)
plt.hist(call_data['Revenue_After'], bins=30, color='green', alpha=0.7)
plt.title('Revenue Distribution After Tariff Reduction')
plt.xlabel('Revenue (RwF)')
plt.ylabel('Frequency')

plt.tight_layout()
plt.show()

# Use line plots to visualize the trend in revenue over time
plt.figure(figsize=(12, 6))
plt.plot(call_data['Revenue_Before'], label='Revenue Before Reduction', color='blue')
plt.plot(call_data['Revenue_After'], label='Revenue After Reduction', color='green')
plt.title('Revenue Trends Over Time (1 Day)')
plt.xlabel('Minutes')
plt.ylabel('Revenue (RwF)')
plt.legend()
plt.grid()
```

```
plt.show()

# Calculate total revenue for both before and after the tariff reduction
total_revenue_before = call_data['Revenue_Before'].sum()
total_revenue_after = call_data['Revenue_After'].sum()

# Compute the percentage increase in revenue after the tariff reduction
percentage_increase = ((total_revenue_after - total_revenue_before) / total_revenue_b

total_revenue_before, total_revenue_after, percentage_increase
```

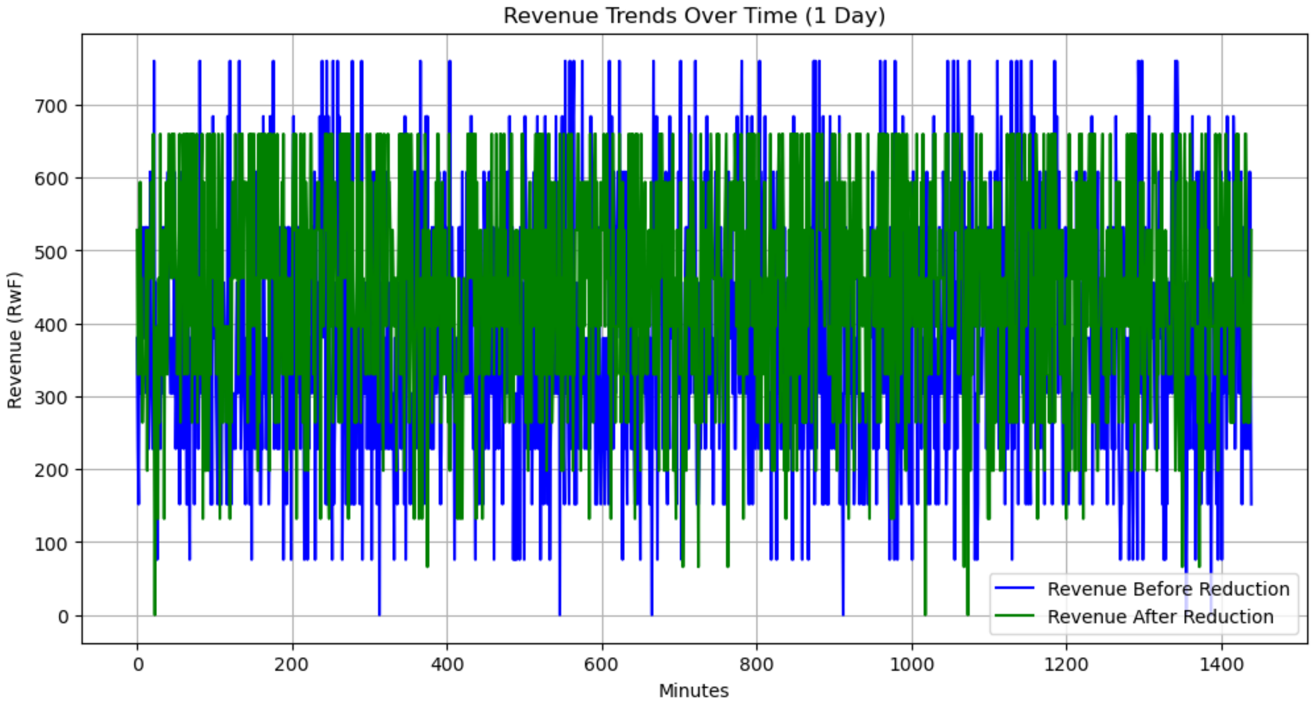
Results:

- Total revenue Before Reduction 541044.00
- Total revenue After Reduction 651750.00
- Percentage increase 20.46

Revenue Distribution hist Graph



Revenue Trends line graph



Insights:

- There is a 20% increase in the revenue collection in the tariff collections.
- The histograms show that the revenue collection that averaged around 300 and 400 RwF shifted to 400 and 500 RwF which is a high revenue increase.
- The line plot displays the high revenue after the reduction of the tariff call prices.

Question 4: Expectation, Variance, and Moments of Discrete Random Variables

Part 1: Intern Earning

Details given:

- Probability of solving 0 tasks: p_0
- Probability of solving 1 task: $p_1 = 2p_0$
- Probability of solving 2 tasks: $p_2 = 4p_0$
- Probability of solving 3 tasks: $p_3 = 3p_0$

The sum of these probabilities must equal 1:

$$p_0 + 2p_0 + 4p_0 + 3p_0 = 1$$

$$10p_0 = 1$$

$$p_0 = \frac{1}{10}$$

Thus, the probabilities are:

- $p_0 = \frac{1}{10}$
- $p_1 = \frac{2}{10} = \frac{1}{5}$
- $p_2 = \frac{4}{10} = \frac{2}{5}$
- $p_3 = \frac{3}{10}$

(a) Expected Money

The expected value $E(X)$:

$$E(X) = (0) \cdot \frac{1}{10} + (1) \cdot \frac{1}{5} + (2) \cdot \frac{2}{5} + (3) \cdot \frac{3}{10}$$

$$E(X) = 0 + \frac{1}{5} + \frac{4}{5} + \frac{9}{10}$$

$$E(X) = \frac{2}{10} + \frac{8}{10} + \frac{9}{10}$$

$$E(X) = \frac{19}{10} = \$1.9$$

(b) Variance

The variance $Var(X)$:

$$Var(X) = E(X^2) - [E(X)]^2$$

$E(X^2)$:

$$\begin{aligned} E(X^2) &= (0^2) \cdot \frac{1}{10} + (1^2) \cdot \frac{1}{5} + (2^2) \cdot \frac{2}{5} + (3^2) \cdot \frac{3}{10} \\ &= 0 + \frac{1}{5} + 4 \cdot \frac{2}{5} + 9 \cdot \frac{3}{10} \\ &= 0.2 + 1.6 + 2.7 \\ &= 4.5 \end{aligned}$$

Substitute:

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ &= 4.5 - (1.9)^2 \\ &= 4.5 - 3.61 \\ &= \$0.89 \end{aligned}$$

Part 2:

a. Probability of the First Success on the 4th Trial $P(X_1 = 4)$

- Details given:

- $p = 0.3$ probability of success.
- X_1 number of trials until the first success
- $k = 4$

The PMF of a geometric random variable:

$$P(X_1 = k) = (1 - p)^{k-1} \times p$$

Substituting:

$$P(X_1 = 4) = (1 - 0.3)^{4-1} \times 0.3 = (0.7)^3 \times 0.3 = 0.1029$$

b. Geometric and negative binomial random Variable

(i): Probability of Securing the First Catch on the x_1-th Cast

Geometric PMF:

$$P(X_1 = x_1) = (1 - p)^{x_1-1} \times p$$

Substituting:

$$P(X_1 = 1) = (1 - 0.1)^{1-1} \times 0.1 = (0.9)^0 \times 0.1 = 0.1$$

(ii): Expected Number of Casts to Catch the First Fish, $E[X_1]$

The expected value of a geometric random variable:

$$E[X_1] = \frac{1}{p}$$

$$E[X_1] = \frac{1}{0.1} = 10$$

(iii): Probability of Exactly x_4 Casts to Catch 4 Fish
Negative binomial distribution PMF:

$$P(X_r = x_r) = \binom{x_r - 1}{r - 1} p^r (1-p)^{x_r - r}$$

Substituting:

$$P(X_4 = 4) = \binom{4-1}{4-1}(0.1)^4(0.9)^{4-4} = \binom{3}{3} \times (0.1)^4 \times (0.9)^0 = 0.0001$$

(iv): Expected Number of Casts to Catch 6 Fish, $E[X_6]$
The expected value of a negative binomial random variable:

$$E[X_r] = \frac{r}{p}$$

Substitute $r = 6$ and $p = 0.1$:

$$E[X_6] = \frac{6}{0.1} = 60$$

Question 5: Expectation, Variance, and Moments of Continuous Random Variables

Part 1. Exponential Distribution and Reliability

1. PDF of the Exponential Distribution and Expectation $E[X]$

The PDF for exponential distribution is:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Given $\mu = 10$ hours,:

$$\lambda = \frac{1}{\mu} = \frac{1}{10} = 0.1$$

PDF is:

$$f_X(x) = 0.1e^{-0.1x}, \quad x \geq 0$$

$E[X]$ is:

$$E[X] = \frac{1}{\lambda} = \mu = 10 \text{ hours}$$

2. Variance of X

The variance of exponentially distributed random variables is:

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

Substituting $\lambda = 0.1$:

$$\text{Var}(X) = \frac{1}{(0.1)^2} = 100 \text{ hours}$$

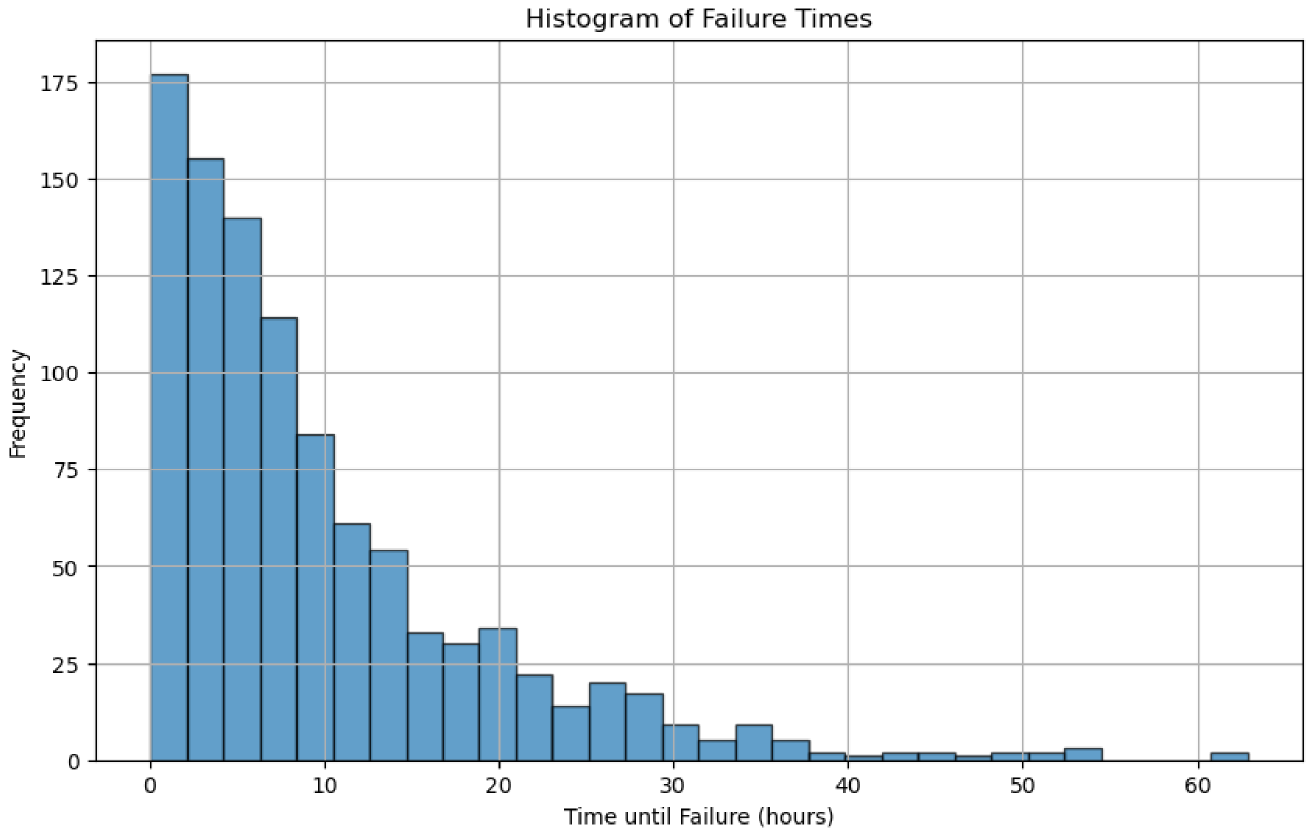
Interpretation: The variance indicates a large variability between failures, this means that while the mean failure time is 10 hours, individual failures can occur both earlier or later by a considerable amount.

3. Simulation

```
#Import Libraries
import numpy as np
import matplotlib.pyplot as plt

# Set the parameters
mean_time = 10  # mean time between failures
num_failures = 1000  # number of failures

# Simulate the failure times
failure_times = np.random.exponential(scale=mean_time, size=num_failures)

# Calculate total downtime
```

```
total_downtime = np.sum(failure_times)

# Plot the histogram of failure times
plt.hist(failure_times, bins=50, density=True, alpha=0.75, color='blue')
plt.title('Histogram of Failure Times')
plt.xlabel('Failure Time (hours)')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()

# Print total downtime
print(f"Total downtime over 1000 failures: {total_downtime:.2f} hours")
```

Total downtime over 1000 failures: 9552.41 hours



Insights: The histogram of failure shows that most failures occur around the mean of 10 hours, while some failures take a bit longer. This will aid the company in having strategic maintenance to be adopted to minimize unexpected downtimes.

Part 2: Normal distribution and Production Quality

1. PDF of the Normal Distribution, Expectation, and Variance

PDF for a normal distribution:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

Where:

$$\mu = 500\text{grams is the mean (expectation),}$$
$$\sigma = 10\text{grams is the standard deviation.}$$

PDF:

$$f_X(x) = \frac{1}{10\sqrt{2\pi}} e^{-\frac{(x-500)^2}{200}}, \quad x \in \mathbb{R}$$

Expectation E[X]:

$$E[X] = \mu = 500 \text{ grams}$$

Varience Var(X):

$$\text{Var}(X) = \sigma^2 = 10^2 = 100 \text{ grams}^2$$

2. Probability That a Product Weighs Between 490 and 510 Grams
Standardize the values using the standard normal distribution:

$$Z = \frac{X - \mu}{\sigma}$$

For $X = 490$:

$$Z_a = \frac{490 - 500}{10} = -1 = 0.1587$$

For $X = 510$:

$$Z_b = \frac{510 - 500}{10} = 1 = 0.8413$$

Find the probability by subtracting:

$$Z_b - Z_a = 0.8413 - 0.1587 = 0.6826$$

Therefore;

$$P(490 \leq X \leq 510) = P(-1 \leq Z \leq 1) = 0.6826$$

Implicataion: This indicates that on average 68.26% of the products weigh between 490 and 510 grams. This guides the company to know the range accepted or in case of any downfall or failure on the products will call upon adjusting the manufacturing process.

3. Adjusting the Standard Deviation for 95% of Products to Weigh Between 495 and 505 Grams
For a 95% confidence interval, the z-values are $Z = -1.96$ and $Z = 1.96$:

$$495 = 500 - 1.96\sigma$$

$$505 = 500 + 1.96\sigma$$

Solve for σ:

$$500 - 495 = 1.96\sigma \quad \Rightarrow \quad \sigma = \frac{5}{1.96} \approx 2.55 \text{ grams}$$

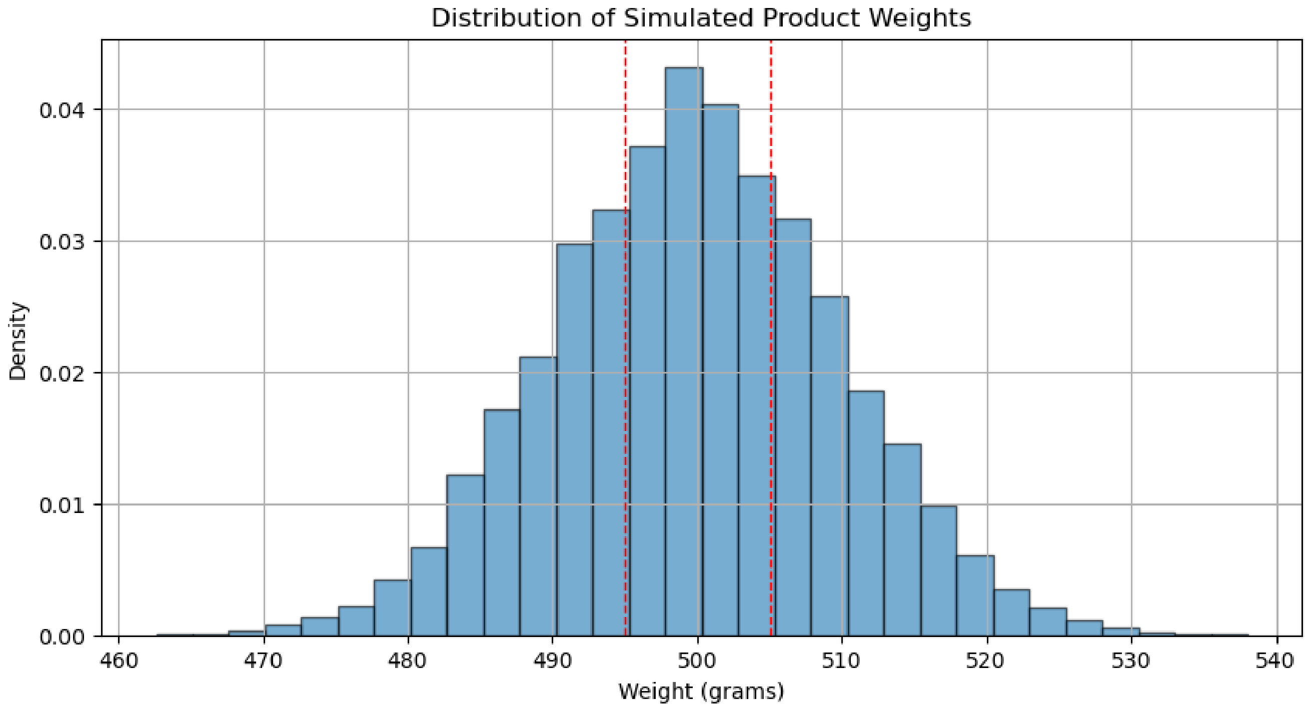
Implication: To ensure that 95% of products weigh between 495 and 505 grams, the standard deviation must be reduced to approximately 2.55 grams. This indicates that the manufacturing process must be improved to reduce the variability in product weight, tightening the distribution significantly.

4. Simulating 10,000 Product Weights

```
# Import libraries
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Values given
mu = 500  # mean weight in grams
sigma = 10  # standard deviation in grams

# Simulate 10,000 product weights
np.random.seed(0)  # for reproducibility
weights = np.random.normal(mu, sigma, 10000)

# Plot the distribution
plt.figure(figsize=(10, 5))
plt.hist(weights, bins=30, density=True, alpha=0.6, edgecolor='black')
plt.title('Distribution of Simulated Product Weights')
plt.xlabel('Weight (grams)')
plt.ylabel('Density')
plt.grid(True)
plt.axvline(495, color='red', linestyle='dashed', linewidth=1)
plt.axvline(505, color='red', linestyle='dashed', linewidth=1)
plt.show()

# Percentage of products that fall within the range of 495 to 505 grams
range_count = ((weights >= 495) & (weights <= 505)).sum()
percentage_in_range = (range_count / len(weights)) * 100

# Calculation using CDF of normal distribution
theory_percentage = (norm.cdf(505, mu, sigma) - norm.cdf(495, mu, sigma)) * 100

print(f'The percentage range is {percentage_in_range:.2f}')
print(f'The theorithical percentage {theory_percentage:.2f}')
```

Results:

- The percentage range is **39.08%**
- The theoretical percentage **38.29%**



Comparison Insights:

- The simulated results are almost similar to the calculated percentage, the difference is due to the randomness of the simulation, this gives the confidence of the simulation getting the value near to the calculated percentage.

Part 3: Lognormal Distribution and Stock Returns

1. PDF, Expectation, and Variance of the Lognormal Distribution

Lognormal distribution PDF:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, \quad x > 0$$

Where: - $\mu = 0.001$ - $\sigma = 0.02$

Expectation:

$$E[X] = e^{\mu+\frac{\sigma^2}{2}} = e^{0.001+\frac{(0.02)^2}{2}} = e^{0.0012} \approx 1.0012$$

Variance:

$$\text{Var}(X) = (e^{\sigma^2}-1)e^{2\mu+\sigma^2} = (e^{0.0004}-1)e^{0.002+0.0004} \approx 0.00040096$$

2. Probability of Returns Greater Than 5%

$$P(X > 1.05) = 1 - P(X \leq 1.05) = 1 - P\left(Z \leq \frac{\ln(1.05)-\mu}{\sigma}\right)$$

Z follows a standard normal distribution.

$$\ln(1.05) \approx 0.04879$$

$$P\left(Z \leq \frac{0.04879 - 0.001}{0.02}\right) = P(Z \leq 2.3895) \approx 0.9913$$

Thus:

$$P(X > 1.05) \approx 1 - 0.9913 = 0.0087$$

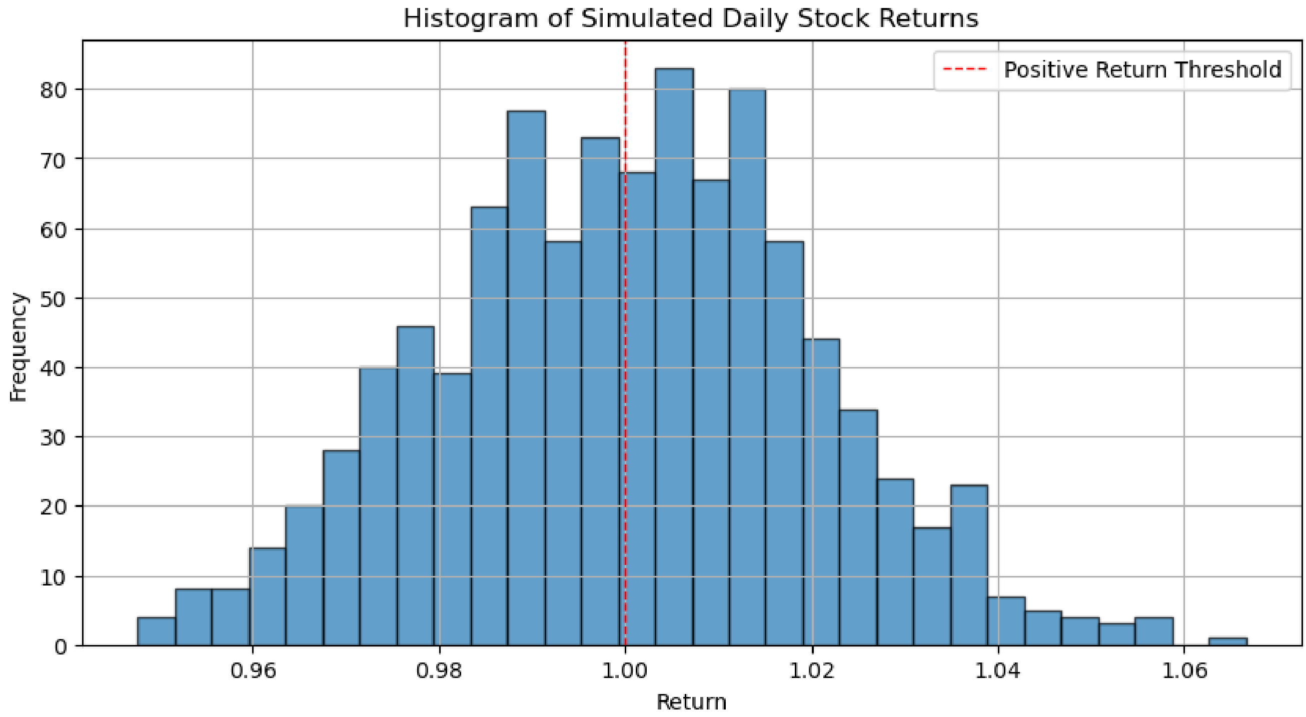
Implications: This indicates a 0.87% chance of the stock yielding more than 5% returns in a single day, suggesting it is a high-risk, high-reward asset.

3. Simulating Stock Returns

```
# Import libraries
import numpy as np
import matplotlib.pyplot as plt

# Parameters for the lognormal distribution
mu = 0.0005   # mean of daily return
sigma = 0.02   # standard deviation of daily return
size = 1000    # number of days

# Simulate daily stock returns using lognormal distribution
returns = np.random.lognormal(mean=mu, sigma=sigma, size=size)

# Plot histogram of daily returns
plt.figure(figsize=(10, 5))
plt.hist(returns, bins=30, alpha=0.7, edgecolor='black')
plt.title('Histogram of Simulated Daily Stock Returns')
plt.xlabel('Return')
plt.ylabel('Frequency')
plt.axvline(1, color='red', linestyle='dashed', linewidth=1)
plt.legend(['Positive Return Threshold'])
plt.grid()
plt.show()

# Calculate the proportion of days with positive returns
proportion_positive = np.sum(returns > 1) / size
proportion_positive
```

Proportion positive = 0.514

Insights: For the stock returns the proportion of days with returns greater than 1 is approximately 50%, which suggests that the stock has an equal likelihood of having positive or negative returns on any given day. In the simulation, we got 51.4%, this aligns perfectly with the tendency of markets to have more positive days over time, which reflects a long-term growth trend.



4. 95th Percentile of Daily Returns

$$X_{0.95} = e^{\mu + \sigma \cdot Z_{0.95}} = e^{0.001 + 0.02 \cdot 1.645} \approx e^{0.0339} \approx 1.0345$$

Significance: This indicates that 95% of the time, the stock's daily return will be less than approximately 3.45%, which is crucial for understanding the upper bounds of potential returns in risk management.

Question 6: Joint Probability Distributions and Covariance

Part 1:

(a) Find $P(X \leq 2$ and $Y \leq 2)$

$$
\begin{aligned}
P(X \leq 2 \text{ and } Y \leq 2) &= P(X=1, Y=1) + P(X=1, Y=2) + P(X=2, Y=1) + P(X=2, Y=2) \\
&= 0.1 + 0 + 0.3 + 0 \\
&= 0.4
\end{aligned}
$$

Insights: The total probability of 0.4 indicates that in most households, the probability of owning at most two cars and two sets of televisions is that 40% of the households have at least 2 or two televisions and cars.

(b) Marginal Distribution for the Number of Cars X

$$
\begin{aligned}
P(X=1) &= P(X=1, Y=1) + P(X=1, Y=2) + P(X=1, Y=3) + P(X=1, Y=4) \\
&= 0.1 + 0 + 0.1 + 0 \\
&= 0.2 \\
P(X=2) &= P(X=2, Y=1) + P(X=2, Y=2) + P(X=2, Y=3) + P(X=2, Y=4) \\
&= 0.3 + 0 + 0.1 + 0.2 \\
&= 0.6 \\
P(X=3) &= P(X=3, Y=1) + P(X=3, Y=2) + P(X=3, Y=3) + P(X=3, Y=4) \\
&= 0 + 0.2 + 0 + 0 \\
&= 0.2
\end{aligned}
$$

Therefore;

$$
\begin{aligned}
P(X=1) &= 0.2 \\
P(X=2) &= 0.6 \\
P(X=3) &= 0.2
\end{aligned}
$$

Insight: The distribution indicates that most households (60%) own two cars, and having one and three cars share (40%) each, this indicates a common car ownership pattern.

(c) Marginal Distribution for the Number of Television Sets Y

$$
\begin{aligned}
P(Y = 1) &= P(X = 1, Y = 1) + P(X = 2, Y = 1) + P(X = 3, Y = 1) \\
&= 0.1 + 0.3 + 0 \\
&= 0.4 \\
P(Y = 2) &= P(X = 1, Y = 2) + P(X = 2, Y = 2) + P(X = 3, Y = 2) \\
&= 0 + 0 + 0.2 \\
&= 0.2 \\
P(Y = 3) &= P(X = 1, Y = 3) + P(X = 2, Y = 3) + P(X = 3, Y = 3) \\
&= 0.1 + 0.1 + 0 \\
&= 0.2 \\
P(Y = 4) &= P(X = 1, Y = 4) + P(X = 2, Y = 4) + P(X = 3, Y = 4) \\
&= 0 + 0.2 + 0 \\
&= 0.2
\end{aligned}
$$

Summary of Marginal Distribution:

$$
\begin{aligned}
P(Y = 1) &= 0.4 \\
P(Y = 2) &= 0.2 \\
P(Y = 3) &= 0.2 \\
P(Y = 4) &= 0.2
\end{aligned}
$$

Interpretation: The marginal distribution reveals that a significant portion (40%) of households own one television set, suggesting that this is the most common ownership level in the suburb.

(d) Independence of X and Y

X and Y are independent if:

$$P(X = x, Y = y) = P(X = x)P(Y = y) \quad \text{for all } x, y$$

Let us use when $X = 1, Y = 1$:

$$
\begin{aligned}
P(X = 1, Y = 1) &= 0.1 \\
P(X = 1) &= 0.2 \\
P(Y = 1) &= 0.4 \\
P(X = 1)P(Y = 1) &= 0.2 \times 0.4 = 0.08 \\
\Rightarrow \quad & 0.1 \neq 0.08
\end{aligned}
$$

Therefore, X and Y are **not independent**.

(e) Conditional Probability $P(Y = 2|X = 2)$

Conditional probability:

$$P(Y = 2|X = 2) = \frac{P(X = 2, Y = 2)}{P(X = 2)}$$

$$\text{substituting}$$

$$P(X = 2, Y = 2) = 0$$

$$P(X = 2) = 0.6$$

$$P(Y = 2|X = 2) = \frac{0}{0.6} = 0$$

Explanation: The result indicates that if a household owns exactly two cars, they can't own exactly two television sets.

(f) Expected Values $E[X]$ and $E[Y]$

Expected value:

$$\begin{aligned}
E[X] &= \sum_x x \cdot P(X = x) \\
&= 1 \cdot 0.2 + 2 \cdot 0.6 + 3 \cdot 0.2 \\
&= 0.2 + 1.2 + 0.6 \\
&= 2.0 \\
E[Y] &= \sum_y y \cdot P(Y = y) \\
&= 1 \cdot 0.4 + 2 \cdot 0.2 + 3 \cdot 0.2 + 4 \cdot 0.2 \\
&= 0.4 + 0.4 + 0.6 + 0.8 \\
&= 2.2
\end{aligned}$$

Interpretation: On average, a household owns approximately 2 cars and approximately $2.2 \approx 3$ television sets.

Part 2: Daily Returns of Two Correlated Stocks

Given details:

- $\mu_X = 0.001$
- $\mu_Y = 0.002$
- $\sigma_X = 0.02$
- $\sigma_Y = 0.03$
- $\rho_{X,Y} = 0.8$

(a) Joint PDF of the Lognormal Distribution of X and Y

The joint PDF for X and Y:

$$\begin{aligned} f_{X,Y}(x,y) = &\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \\ &\cdot \exp\left(-\frac{1}{2(1-\rho_{X,Y}^2)}\left(\frac{(\ln x-\mu_X)^2}{\sigma_X^2}+\frac{(\ln y-\mu_Y)^2}{\sigma_Y^2}\right.\right. \\ &\left.\left.-2\rho_{X,Y}\frac{(\ln x-\mu_X)(\ln y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right) \end{aligned} \tag{1}$$

(b) Marginal Distributions of X and Y

The marginal distributions of X and Y for lognormal distributions are given by:

- For X:

  $$f_X(x) = \frac{1}{x\sigma_X\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu_X)^2}{2\sigma_X^2}\right)$$

  substituting

  $$f_X(x) = \frac{1}{0.02x\sqrt{2\pi}}\exp\left(-\frac{(\ln x-0.001)^2}{2\times 0.02^2}\right)$$

- For Y:

  $$f_Y(y) = \frac{1}{y\sigma_Y\sqrt{2\pi}}\exp\left(-\frac{(\ln y-\mu_Y)^2}{2\sigma_Y^2}\right)$$

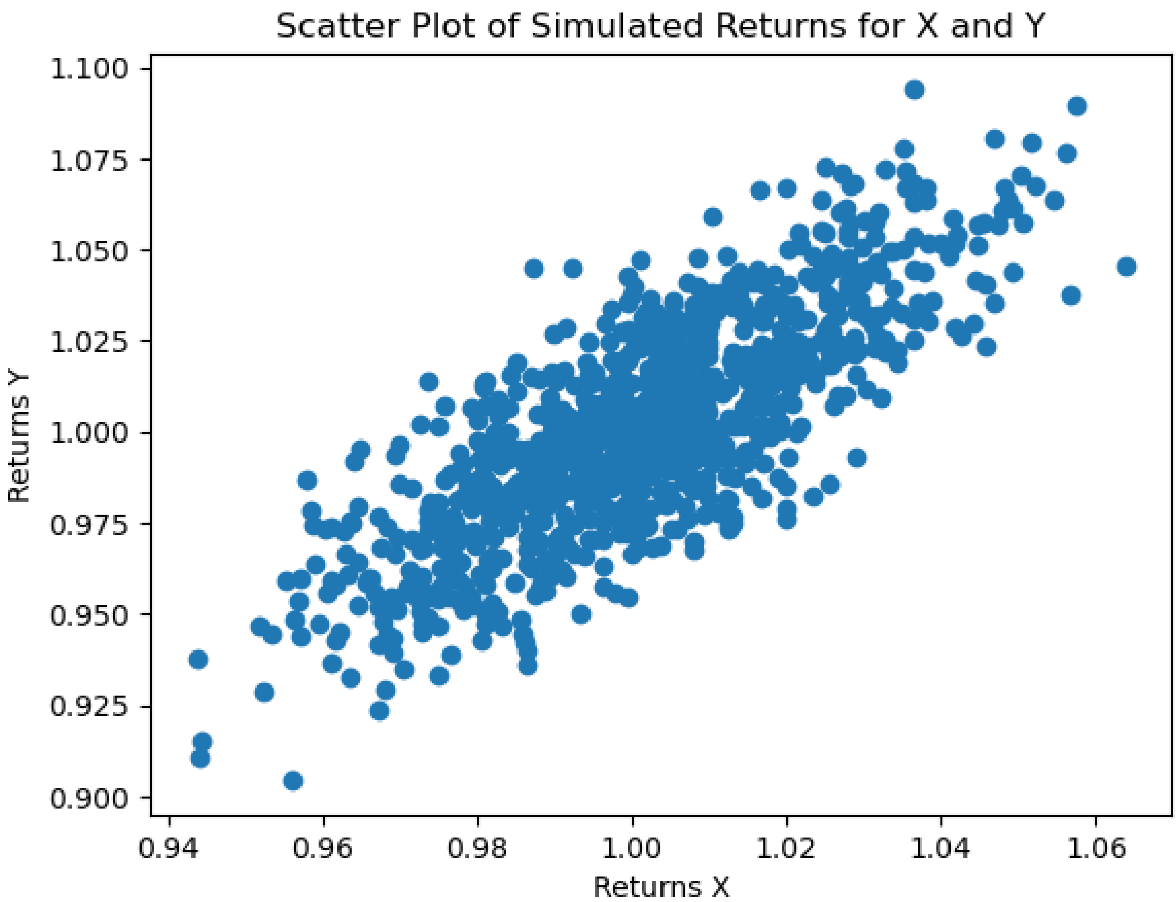
  substituting values:

  $$f_Y(y) = \frac{1}{0.03y\sqrt{2\pi}}\exp\left(-\frac{(\ln y-0.002)^2}{2\times 0.03^2}\right)$$

(c) Simulating 1,000 Days of Returns for Both Stocks

```
#Importing Libraries
import numpy as np
import matplotlib.pyplot as plt

# Parameters
mu_X = 0.001
mu_Y = 0.002
sigma_X = 0.02
sigma_Y = 0.03
rho_XY = 0.8
num_days = 1000

# Covariance matrix
cov_matrix = [[sigma_X**2, rho_XY * sigma_X * sigma_Y],
              [rho_XY * sigma_X * sigma_Y, sigma_Y**2]]
```

```
# Simulate the bivariate normal distribution
mean_vector = [mu_X, mu_Y]
log_returns = np.random.multivariate_normal(mean_vector, cov_matrix, num_days)

# Converting the log-returns to returns by exponentiating
returns_X = np.exp(log_returns[:, 0])
returns_Y = np.exp(log_returns[:, 1])

# Plot the scatter plot of returns
plt.scatter(returns_X, returns_Y)
plt.title("Scatter Plot of Simulated Returns for X and Y")
plt.xlabel("Returns X")
plt.ylabel("Returns Y")
plt.show()

# Calculate the empirical correlation
empirical_corr = np.corrcoef(returns_X, returns_Y)[0, 1]
print(f"Empirical Correlation: {empirical_corr}")
```



Empirical Correlation: 0.7951206043358056

(d) Percentage of Days with Positive Returns

```
# Calculate the percentage of days where both stocks have positive returns
both_positive = np.sum((returns_X > 1) & (returns_Y > 1)) / num_days * 100
print(f"Percentage of having Positive Returns: {both_positive:.2f}%")
```

Percentage of having Positive Returns: 40.90%
Comparison of $\rho_{X,Y} = 0.8$ with the 0.409 - this deviation could be because of the randomness of the simulation.

Part 3: Covariance of Combined Claims

Details Given:

- $E(X) = 5$
- $E(X^2) = 27.4$
- $E(Y) = 7$
- $E(Y^2) = 51.4$
- $\text{Var}(X + Y) = 8$

- Variance of X:

$$\text{Var}(X) = E(X^2) - (E(X))^2 = 27.4 - 5^2 = 27.4 - 25 = 2.4$$

- Variance of Y:

$$\text{Var}(Y) = E(Y^2) - (E(Y))^2 = 51.4 - 7^2 = 51.4 - 49 = 2.4$$

Covariance of X and Y:

$$\begin{gathered}\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \\ 8 = 2.4 + 2.4 + 2\text{Cov}(X, Y) \\ 8 = 4.8 + 2\text{Cov}(X, Y) \\ 3.2 = 2\text{Cov}(X, Y) \\ \text{Cov}(X, Y) = 1.6\end{gathered}$$

Finding C_1 and C_2:

- $C_1 = X + Y$
- $C_2 = X + 1.2Y$

$$\begin{gathered}\text{Cov}(C_1, C_2) = \text{Cov}(X + Y, X + 1.2Y) \\ = \text{Cov}(X, X) + \text{Cov}(X, 1.2Y) + \text{Cov}(Y, X) + \text{Cov}(Y, 1.2Y) \\ = \text{Var}(X) + 1.2\text{Cov}(X, Y) + \text{Cov}(Y, X) + 1.2\text{Var}(Y) \\ = \text{Var}(X) + (1.2 + 1)\text{Cov}(X, Y) + 1.2\text{Var}(Y) \\ = 2.4 + (2.2)(1.6) + 1.2(2.4) \\ = 2.4 + 3.52 + 2.88 \\ = 8.8\end{gathered}$$

Covariance of $\text{Cov}(C_1, C_2)$ is **8.8**