

Homework 2: Data 605

Umer Farooq

2024-11-10

Problem 1:

Bayesian:

Problem Statement:

A new credit scoring system has been developed to predict the likelihood of loan defaults. The system has a 90% sensitivity, meaning that it correctly identifies 90% of those who will default on their loans. It also has a 95% specificity, meaning that it correctly identifies 95% of those who will not default. The default rate among borrowers is 2%. Given these prevalence, sensitivity, and specificity estimates, what is the probability that a borrower flagged by the system as likely to default will actually default? If the average loss per defaulted loan is \$200,000 and the cost to run the credit scoring test on each borrower is \$500, what is the total first-year cost for evaluating 10,000 borrowers?

Solution:

Here is the solution using R Language:

```
# Given values
sensitivity <- 0.90
specificity <- 0.95
prevalence <- 0.02

# Complement of prevalence (probability of no default)
no_default_prevalence <- 1 - prevalence

# Probability of a positive test result given no default (False Positive Rate)
false_positive_rate <- 1 - specificity

# Probability of a positive result (P(Positive))
p_positive <- (sensitivity * prevalence) + (false_positive_rate * no_default_prevalence)

# Positive Predictive Value (PPV): P(Default | Positive)
ppv <- (sensitivity * prevalence) / p_positive
ppv
```

```
## [1] 0.2686567
```

```
# Cost calculations
cost_per_test <- 500 # Cost in dollars
num_borrowers <- 10000 # Number of borrowers evaluated

# Total cost for evaluating all borrowers
total_testing_cost <- cost_per_test * num_borrowers

# Expected number of borrowers flagged as likely to default
```

```

expected_flagged_as_default <- p_positive * num_borrowers

# Expected number of actual defaults among flagged borrowers
expected_actual_defaults <- ppv * expected_flagged_as_default

# Average loss per defaulted loan
loss_per_default <- 200000 # Loss in dollars

# Total expected loss due to defaults
total_loss_due_to_defaults <- expected_actual_defaults * loss_per_default

# Total first-year cost (testing cost + expected loss due to defaults)
total_first_year_cost <- total_testing_cost + total_loss_due_to_defaults
total_first_year_cost

## [1] 4.1e+07

```

Binomial:

Problem Statement:

The probability that a stock will pay a dividend in any given quarter is 0.7. What is the probability that the stock pays dividends exactly 6 times in 8 quarters? What is the probability that it pays dividends 6 or more times? What is the probability that it pays dividends fewer than 6 times? What is the expected number of dividend payments over 8 quarters? What is the standard deviation?

Solution:

The probability that a stock will pay a dividend in any given quarter is 0.7. Over 8 quarters, we need to calculate:

1. The probability that the stock pays dividends exactly 6 times.
2. The probability that it pays dividends 6 or more times.
3. The probability that it pays dividends fewer than 6 times.
4. The expected number of dividend payments over 8 quarters.
5. The standard deviation of dividend payments.

The dividend payments follow a **Binomial distribution** where:

$$n = 8 \quad \text{and} \quad p = 0.7$$

Solution

1. Probability of Exactly 6 Dividends in 8 Quarters

Using the Binomial probability formula:

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

Given $n = 8$, $k = 6$, and $p = 0.7$, we find $P(X = 6)$:

$$P(X = 6) = \binom{8}{6} (0.7)^6 (0.3)^2$$

$$= \frac{8!}{6!(8-6)!} \times (0.7)^6 \times (0.3)^2$$

$$\approx 0.2965$$

Thus, the probability that the stock pays dividends exactly 6 times in 8 quarters is approximately 0.2965.

2. Probability of 6 or More Dividends in 8 Quarters

To find $P(X \geq 6)$, we sum the probabilities for 6, 7, and 8 dividends:

$$P(X \geq 6) = P(X = 6) + P(X = 7) + P(X = 8)$$

Using a binomial cumulative distribution:

$$P(X \geq 6) \approx 0.5518$$

So, the probability that it pays dividends 6 or more times is approximately 0.5518.

3. Probability of Fewer Than 6 Dividends

To find $P(X < 6)$, we sum the probabilities for 0 through 5 dividends:

$$P(X < 6) = P(X = 0) + P(X = 1) + \dots + P(X = 5)$$

Using the cumulative distribution function:

$$P(X < 6) \approx 0.4482$$

Thus, the probability that it pays dividends fewer than 6 times is approximately 0.4482.

4. Expected Number of Dividend Payments

The expected number of successes $E(X)$ in a Binomial distribution is given by:

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

Substitute $n = 8$ and $p = 0.7$:

$$E(X) = 8 \times 0.7 = 5.6$$

So, the expected number of dividend payments over 8 quarters is 5.6.

5. Standard Deviation of Dividend Payments

The standard deviation σ for a Binomial distribution is given by:

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)}$$

Substitute $n = 8$ and $p = 0.7$:

$$\sigma = \sqrt{8 \times 0.7 \times 0.3} \approx 1.2961$$

Thus, the standard deviation of dividend payments over 8 quarters is approximately 1.2961.

Here is the same solution using R language

```

# Given parameters
n <- 8          # Number of trials (quarters)
p <- 0.7        # Probability of a dividend payment per quarter

# 1. Probability of exactly 6 dividends in 8 quarters
k <- 6
prob_6_dividends <- dbinom(k, n, p)
cat("Probability of exactly 6 dividends in 8 quarters:", round(prob_6_dividends, 4), "\n")

## Probability of exactly 6 dividends in 8 quarters: 0.2965

# 2. Probability of 6 or more dividends in 8 quarters
#  $P(X \geq 6) = P(X = 6) + P(X = 7) + P(X = 8)$ 
prob_6_or_more_dividends <- pbinom(5, n, p, lower.tail = FALSE)
cat("Probability of 6 or more dividends in 8 quarters:", round(prob_6_or_more_dividends, 4), "\n")

## Probability of 6 or more dividends in 8 quarters: 0.5518

# 3. Probability of fewer than 6 dividends in 8 quarters
#  $P(X < 6) = P(X \leq 5)$ 
prob_fewer_than_6_dividends <- pbinom(5, n, p)
cat("Probability of fewer than 6 dividends in 8 quarters:", round(prob_fewer_than_6_dividends, 4), "\n")

## Probability of fewer than 6 dividends in 8 quarters: 0.4482

# 4. Expected number of dividend payments
expected_dividends <- n * p
cat("Expected number of dividend payments over 8 quarters:", expected_dividends, "\n")

## Expected number of dividend payments over 8 quarters: 5.6

# 5. Standard deviation of dividend payments
std_deviation_dividends <- sqrt(n * p * (1 - p))
cat("Standard deviation of dividend payments:", round(std_deviation_dividends, 4), "\n")

## Standard deviation of dividend payments: 1.2961

```

Poisson:

Problem Statement:

A financial analyst notices that there are an average of 12 trading days each month when a certain stock's price increases by more than 2%. What is the probability that exactly 4 such days occur in a given month? What is the probability that more than 12 such days occur in a given month? How many such days would you expect in a 6-month period? What is the standard deviation of the number of such days? If an investment strategy requires at least 70 days of such price increases in a year for profitability, what is the percent utilization and what are your recommendations?

Solution:

A financial analyst observes that there are, on average, 12 trading days each month when a certain stock's price increases by more than 2%. We need to calculate:

1. The probability that exactly 4 such days occur in a given month.
2. The probability that more than 12 such days occur in a given month.
3. The expected number of such days in a 6-month period.

4. The standard deviation of the number of such days.
5. If an investment strategy requires at least 70 days of such price increases in a year for profitability, calculate the percent utilization and provide recommendations.

We assume that the number of days follows a **Poisson distribution** with a mean $\lambda = 12$ (days per month).

Here is the solution using R-Code:

```
# Given parameters
lambda <- 12 # Average number of days per month with price increase > 2%

# 1. Probability of exactly 4 days in a month
k <- 4
prob_4_days <- dpois(k, lambda)
cat("Probability of exactly 4 days in a month:", round(prob_4_days, 4), "\n")

## Probability of exactly 4 days in a month: 0.0053

# 2. Probability of more than 12 days in a month
prob_more_than_12 <- ppois(12, lambda, lower.tail = FALSE)
cat("Probability of more than 12 days in a month:", round(prob_more_than_12, 4), "\n")

## Probability of more than 12 days in a month: 0.424

# 3. Expected number of days in a 6-month period
expected_6_months <- 6 * lambda
cat("Expected number of days in a 6-month period:", expected_6_months, "\n")

## Expected number of days in a 6-month period: 72

# 4. Standard deviation of the number of days in a 6-month period
std_dev_6_months <- sqrt(lambda)
cat("Standard deviation of days over 6 months:", round(std_dev_6_months, 4), "\n")

## Standard deviation of days over 6 months: 3.4641

# 5. Percent utilization and recommendation
# Expected number of days in a year
expected_year <- 12 * lambda
target_days <- 70
percent_utilization <- (target_days / expected_year) * 100
cat("Percent utilization for 70 target days:", round(percent_utilization, 2), "%\n")

## Percent utilization for 70 target days: 48.61 %
```

Recommendation: The investment strategy is likely to be profitable based on the expected number of days.

Hypergeometric:

Problem Statement:

A hedge fund has a portfolio of 25 stocks, with 15 categorized as high-risk and 10 as low-risk. The fund manager randomly selects 7 stocks to closely monitor. If the manager selected 5 high-risk stocks and 2 low-risk stocks, what is the probability of selecting exactly 5 high-risk stocks if the selection was random? How many high-risk and low-risk stocks would you expect to be selected?

Solution:

A hedge fund has a portfolio of 25 stocks, with 15 categorized as high-risk and 10 as low-risk. The fund manager randomly selects 7 stocks to monitor. We are asked to determine:

1. The probability that exactly 5 of the selected stocks are high-risk.
2. The expected number of high-risk and low-risk stocks in the sample.

This scenario follows a **Hypergeometric distribution**, where: - $N = 25$ (total number of stocks), - $K = 15$ (number of high-risk stocks), - $n = 7$ (number of stocks selected), - $k = 5$ (number of high-risk stocks chosen in the sample).

Solution

1. Probability of Selecting Exactly 5 High-Risk Stocks

The probability of selecting exactly 5 high-risk stocks is given by the Hypergeometric formula:

$$P(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

Substitute $N = 25$, $K = 15$, $n = 7$, and $k = 5$:

$$P(X = 5) = \frac{\binom{15}{5} \cdot \binom{10}{2}}{\binom{25}{7}}$$

Calculating each component:

$$\binom{15}{5} = \frac{15!}{5!(15-5)!} = 3003, \quad \binom{10}{2} = 45, \quad \binom{25}{7} = 480700$$

Substitute these values:

$$P(X = 5) = \frac{3003 \times 45}{480700} \approx 0.2812$$

Thus, the probability of selecting exactly 5 high-risk stocks is approximately 0.2812.

2. Expected Number of High-Risk and Low-Risk Stocks

For a Hypergeometric distribution, the expected number of high-risk stocks in the sample is:

$$E(\text{high-risk}) = n \cdot \frac{K}{N} = 7 \cdot \frac{15}{25} = 4.2$$

Similarly, the expected number of low-risk stocks is:

$$E(\text{low-risk}) = n \cdot \frac{N-K}{N} = 7 \cdot \frac{10}{25} = 2.8$$

Therefore, we expect to select approximately 4.2 high-risk and 2.8 low-risk stocks.

Here is the same solution using R-Code:

```
# Given parameters
N <- 25      # Total number of stocks
K <- 15      # Number of high-risk stocks
n <- 7       # Number of stocks selected
k <- 5       # Number of high-risk stocks chosen
```

```

# 1. Probability of selecting exactly 5 high-risk stocks
prob_5_high_risk <- dhyper(k, K, N - K, n)
cat("Probability of exactly 5 high-risk stocks:", round(prob_5_high_risk, 4), "\n")

## Probability of exactly 5 high-risk stocks: 0.2811

# 2. Expected number of high-risk and low-risk stocks
expected_high_risk <- n * (K / N)
expected_low_risk <- n * ((N - K) / N)
cat("Expected number of high-risk stocks:", round(expected_high_risk, 2), "\n")

## Expected number of high-risk stocks: 4.2

cat("Expected number of low-risk stocks:", round(expected_low_risk, 2), "\n")

## Expected number of low-risk stocks: 2.8

```

Geometric:

Problem Statement:

The probability that a bond defaults in any given year is 0.5%. A portfolio manager holds this bond for 10 years. What is the probability that the bond will default during this period? What is the probability that it will default in the next 15 years? What is the expected number of years before the bond defaults? If the bond has already survived 10 years, what is the probability that it will default in the next 2 years?

Solution:

The probability that a bond defaults in any given year is 0.5%. A portfolio manager holds this bond for a specified period. We are asked to determine:

1. The probability that the bond will default within 10 years.
2. The probability that the bond will default within 15 years.
3. The expected number of years before the bond defaults.
4. The probability that the bond defaults within 2 years, given that it has already survived for 10 years.

This scenario follows a **Geometric distribution**, where: - $p = 0.005$ (probability of default each year).

Solution

1. Probability of Default within 10 Years

For a Geometric distribution, the probability that the bond defaults within n years is given by the cumulative distribution function:

$$P(X \leq n) = 1 - (1 - p)^n$$

For $p = 0.005$ and $n = 10$:

$$\begin{aligned}
 P(X \leq 10) &= 1 - (1 - 0.005)^{10} \\
 &= 1 - (0.995)^{10} \approx 0.0488
 \end{aligned}$$

Thus, the probability that the bond will default within 10 years is approximately 4.88%.

2. Probability of Default within 15 Years

Similarly, for $n = 15$:

$$\begin{aligned} P(X \leq 15) &= 1 - (1 - 0.005)^{15} \\ &= 1 - (0.995)^{15} \approx 0.0729 \end{aligned}$$

The probability of default within 15 years is approximately 7.29%.

3. Expected Number of Years Before Default

The expected number of years before default for a Geometric distribution is given by:

$$E(X) = \frac{1}{p}$$

Substituting $p = 0.005$:

$$E(X) = \frac{1}{0.005} = 200$$

Thus, the expected number of years before default is 200 years.

4. Probability of Default within 2 Years Given Survival for 10 Years

Given that the bond has already survived for 10 years, the probability that it defaults within the next 2 years is:

$$P(X \leq 2) = 1 - (1 - p)^2$$

For $p = 0.005$:

$$P(X \leq 2) = 1 - (0.995)^2 \approx 0.009975$$

Therefore, the probability that the bond defaults in the next 2 years, given it has survived 10 years, is approximately 0.9975%.

Here is the same solution using R-Code:

```
# Given parameters
p <- 0.005           # Probability of default in a given year

# 1. Probability of default within 10 years
n1 <- 10
prob_within_10_years <- 1 - (1 - p)^n1
cat("Probability of default within 10 years:", round(prob_within_10_years, 4), "\n")

## Probability of default within 10 years: 0.0489

# 2. Probability of default within 15 years
n2 <- 15
prob_within_15_years <- 1 - (1 - p)^n2
cat("Probability of default within 15 years:", round(prob_within_15_years, 4), "\n")

## Probability of default within 15 years: 0.0724

# 3. Expected number of years before default
expected_years_before_default <- 1 / p
cat("Expected number of years before default:", expected_years_before_default, "\n")

## Expected number of years before default: 200
```



```
# 4. Probability of default within 2 years given survival for 10 years
n3 <- 2
prob_within_2_years_given_10_years <- 1 - (1 - p)^n3
cat("Probability of default within 2 years after surviving 10 years:", round(prob_within_2_years_given_

## Probability of default within 2 years after surviving 10 years: 0.01
```

Poisson:

Problem Statement:

A high-frequency trading algorithm experiences a system failure about once every 1500 trading hours. What is the probability that the algorithm will experience more than two failures in 1500 hours? What is the expected number of failures?

Solution:

A high-frequency trading algorithm experiences a system failure once every 1500 trading hours on average. We need to calculate:

1. The probability that the algorithm will experience more than two failures in 1500 hours.
2. The expected number of failures over 1500 hours.

Since this is a rare event over a fixed interval, we can model the number of failures with a **Poisson distribution**. The average rate (λ) for 1500 hours is 1 failure.

Solution

1. Expected Number of Failures

The expected number of failures in 1500 hours, given that failures occur on average once every 1500 hours, is:

$$E(X) = \lambda = 1$$

So, we expect 1 failure over 1500 hours.

2. Probability of More Than 2 Failures in 1500 Hours

For a Poisson distribution, the probability of observing more than k events is calculated as:

$$P(X > k) = 1 - P(X \leq k)$$

In this case, we need the probability of more than 2 failures ($k = 2$).

First, we calculate $P(X \leq 2)$, the cumulative probability for 0, 1, and 2 failures:

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

Each term in the Poisson probability mass function is given by:

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

For $\lambda = 1$:

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

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

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

Now, summing these probabilities:

$$P(X \leq 2) = 0.3679 + 0.3679 + 0.1839 = 0.9197$$

Then,

$$P(X > 2) = 1 - P(X \leq 2) = 1 - 0.9197 = 0.0803$$

Thus, the probability of experiencing more than two failures in 1500 hours is approximately 8.03%.

\end{document}

R-Code:

```
# Given parameter
lambda <- 1 # Mean number of failures in 1500 hours

# 1. Expected number of failures
expected_failures <- lambda
cat("Expected number of failures in 1500 hours:", expected_failures, "\n")

## Expected number of failures in 1500 hours: 1

# 2. Probability of more than 2 failures
prob_more_than_2 <- 1 - ppois(2, lambda)
cat("Probability of more than 2 failures in 1500 hours:", round(prob_more_than_2, 4), "\n")

## Probability of more than 2 failures in 1500 hours: 0.0803
```

Uniform Distribution:

Problem Statement:

An investor is trying to time the market and is monitoring a stock that they believe has an equal chance of reaching a target price between 20 and 60 days. What is the probability that the stock will reach the target price in more than 40 days? If it hasn't reached the target price by day 40, what is the probability that it will reach it in the next 10 days? What is the expected time for the stock to reach the target price?

Solution:

An investor monitors a stock expected to reach a target price between 20 and 60 days with equal probability for any day in this range. We are asked to calculate:

1. The probability that the stock will reach the target price in more than 40 days.
2. Given it hasn't reached the target price by day 40, the probability it will reach it in the next 10 days.
3. The expected time for the stock to reach the target price.

Since the stock reaching the target price is uniformly distributed, we can use the properties of the **Uniform distribution** on the interval $[a, b]$.

Solution

1. Probability of Reaching the Target Price in More Than 40 Days

For a continuous uniform distribution over $[a, b]$, the probability that the time X is greater than a certain day c is:

$$P(X > c) = \frac{b - c}{b - a}$$

Substituting $c = 40$, $a = 20$, and $b = 60$:

$$P(X > 40) = \frac{60 - 40}{60 - 20} = \frac{20}{40} = 0.5$$

So, the probability that the stock will reach the target price in more than 40 days is 0.5 (or 50%).

2. Probability of Reaching the Target Price Between Days 40 and 50, Given It Hasn't Reached by Day 40

Given that the target hasn't been reached by day 40, we are interested in the conditional probability that it reaches the target in the next 10 days (between days 40 and 50). This is calculated as:

$$P(40 < X \leq 50 \mid X > 40) = \frac{P(40 < X \leq 50)}{P(X > 40)}$$

We already know $P(X > 40) = 0.5$. Now we calculate $P(40 < X \leq 50)$ as:

$$P(40 < X \leq 50) = \frac{50 - 40}{60 - 20} = \frac{10}{40} = 0.25$$

Thus,

$$P(40 < X \leq 50 \mid X > 40) = \frac{0.25}{0.5} = 0.5$$

So, given it hasn't reached by day 40, the probability of reaching the target in the next 10 days is also 0.5 (or 50%).

3. Expected Time to Reach the Target Price

The expected value $E(X)$ for a uniform distribution on $[a, b]$ is given by:

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

Substituting $a = 20$ and $b = 60$:

$$E(X) = \frac{20 + 60}{2} = \frac{80}{2} = 40$$

So, the expected time for the stock to reach the target price is 40 days.

R-Code:

```
# Given parameters
a <- 20 # Minimum time in days
b <- 60 # Maximum time in days
c <- 40 # Given day for probability calculations

# 1. Probability that the stock reaches the target price in more than 40 days
prob_more_than_40 <- (b - c) / (b - a)
cat("Probability of reaching the target price in more than 40 days:", prob_more_than_40, "\n")
```

```
## Probability of reaching the target price in more than 40 days: 0.5
# 2. Probability that the stock reaches the target in the next 10 days given it hasn't by day 40
prob_40_to_50 <- (50 - 40) / (b - a) # Probability of reaching between day 40 and 50
cond_prob_40_to_50 <- prob_40_to_50 / prob_more_than_40
cat("Conditional probability of reaching target between days 40 and 50:", cond_prob_40_to_50, "\n")

## Conditional probability of reaching target between days 40 and 50: 0.5
# 3. Expected time to reach the target price
expected_time <- (a + b) / 2
cat("Expected time to reach the target price:", expected_time, "\n")

## Expected time to reach the target price: 40
```

Exponential Distribution:

Problem Statement:

A financial model estimates that the lifetime of a successful start-up before it either goes public or fails follows an exponential distribution with an expected value of 8 years. What is the expected time until the start-up either goes public or fails? What is the standard deviation? What is the probability that the start-up will go public or fail after 6 years? Given that the start-up has survived for 6 years, what is the probability that it will go public or fail in the next 2 years?

Solution:

A financial model estimates that the lifetime of a successful start-up before it either goes public or fails follows an Exponential distribution with an expected lifetime of 8 years. We need to calculate:

1. The expected time until the start-up either goes public or fails.
2. The standard deviation of the lifetime.
3. The probability that the start-up will go public or fail after 6 years.
4. Given that the start-up has survived for 6 years, the probability it will go public or fail in the next 2 years.

Solution

1. Expected Time Until the Start-up Goes Public or Fails

For an Exponential distribution, the expected value $E(X)$ is given as:

$$E(X) = \frac{1}{\lambda}$$

Since the expected lifetime is already provided as 8 years, we have:

$$E(X) = 8$$

2. Standard Deviation of the Lifetime

The standard deviation of an Exponential distribution is also $\frac{1}{\lambda}$, which is equal to the expected value. Therefore:

$$\text{Standard Deviation} = 8$$

3. Probability that the Start-up Will Go Public or Fail After 6 Years

For an Exponential distribution, the probability that the lifetime X exceeds a specific time t is:

$$P(X > t) = e^{-\lambda t}$$

Substituting $\lambda = 0.125$ and $t = 6$:

$$P(X > 6) = e^{-0.125 \times 6} = e^{-0.75} \approx 0.4724$$

So, the probability that the start-up will go public or fail after 6 years is approximately 0.4724.

4. Probability of Going Public or Failing in the Next 2 Years Given Survival for 6 Years

This is a conditional probability calculation. For an Exponential distribution, the memoryless property allows us to calculate it as:

$$P(X > 6 \text{ and } X \leq 8 \mid X > 6) = P(X \leq 2) = 1 - e^{-\lambda \times 2}$$

Substituting $\lambda = 0.125$:

$$P(X \leq 2) = 1 - e^{-0.125 \times 2} = 1 - e^{-0.25} \approx 0.2212$$

Thus, given that the start-up has already survived 6 years, there is approximately a 22.12% probability that it will go public or fail in the next 2 years.

R-Code:

```
# Given parameters
expected_lifetime <- 8    # Expected lifetime in years
lambda <- 1 / expected_lifetime # Rate parameter

# 1. Expected time until the start-up goes public or fails
expected_time <- expected_lifetime
cat("Expected time until going public or failure:", expected_time, "years\n")

## Expected time until going public or failure: 8 years

# 2. Standard deviation of the lifetime
std_dev <- 1 / lambda
cat("Standard deviation of the lifetime:", std_dev, "years\n")

## Standard deviation of the lifetime: 8 years

# 3. Probability that the start-up will go public or fail after 6 years
t1 <- 6
prob_after_6_years <- exp(-lambda * t1)
cat("Probability of going public or failing after 6 years:", round(prob_after_6_years, 4), "\n")

## Probability of going public or failing after 6 years: 0.4724

# 4. Probability of going public or failing in the next 2 years given survival for 6 years
t2 <- 2
prob_next_2_years_given_6 <- 1 - exp(-lambda * t2)
cat("Probability of going public or failing in the next 2 years given survival for 6 years:", round(prob_next_2_years_given_6, 4), "\n")

## Probability of going public or failing in the next 2 years given survival for 6 years: 0.2212
```

Problem 2:

Product Selection:

Problem Statement:

A company produces 5 different types of green pens and 7 different types of red pens. The marketing team needs to create a new promotional package that includes 5 pens. How many different ways can the package be created if it contains fewer than 2 green pens?

Solution:

```
# Number of ways to select 5 red pens from 7
case1 <- choose(7, 5)

# Number of ways to select 1 green pen from 5 and 4 red pens from 7
case2 <- choose(5, 1) * choose(7, 4)

# Total number of ways
total_ways <- case1 + case2
total_ways

## [1] 196
```

Team Formation for a Project:

Problem Statement:

A project committee is being formed within a company that includes 14 senior managers and 13 junior managers. How many ways can a project team of 5 members be formed if at least 4 of the members must be junior managers?

Solution:

```
# Case 1: 4 junior managers from 13 and 1 senior manager from 14
case1 <- choose(13, 4) * choose(14, 1)

# Case 2: 5 junior managers from 13 and 0 senior managers
case2 <- choose(13, 5)

# Total number of ways
total_ways <- case1 + case2
total_ways

## [1] 11297
```

Marketing Campaign Outcomes:

Problem Statement:

A marketing campaign involves three stages: first, a customer is sent 5 email offers; second, the customer is targeted with 2 different online ads; and third, the customer is presented with 3 personalized product recommendations. If the email offers, online ads, and product recommendations are selected randomly, how many different possible outcomes are there for the entire campaign?

Solution:

```

# Number of email offers
email_offers <- 5

# Number of online ads
online_ads <- 2

# Number of product recommendations
product_recommendations <- 3

# Total possible outcomes
total_outcomes <- email_offers * online_ads * product_recommendations
total_outcomes

## [1] 30

```

Product Defect Probability:

Problem Statement:

A quality control team draws 3 products from a batch without replacement. What is the probability that at least one of the products drawn is defective if the defect rate is known to be consistent? Express your answer as a fraction or a decimal number rounded to four decimal places.

Solution:

```

# Defect rate
p <- 0.1

# Probability that at least one product is defective
prob_at_least_one_defective <- 1 - (1 - p)^3
round(prob_at_least_one_defective, 4)

## [1] 0.271

```

Business Strategy Choices:

A business strategist is choosing potential projects to invest in, focusing on 17 high-risk, high-reward projects and 14 low-risk, steady-return projects.

Step 1: How many different combinations of 5 projects can the strategist select?

Step 2: How many different combinations of 5 projects can the strategist select if they want at least one low-risk project?

```

# Total projects
total_projects <- 31

# High-risk and low-risk projects
high_risk <- 17
low_risk <- 14

# Step 1: Total combinations of 5 projects from 31
total_combinations <- choose(total_projects, 5)

# Step 2: Combinations of 5 projects with only high-risk projects

```

```

high_risk_only <- choose(high_risk, 5)
# Combinations with at least one low-risk project
at_least_one_low_risk <- total_combinations - high_risk_only

# Output results
total_combinations

## [1] 169911
at_least_one_low_risk

## [1] 163723

```

Event Scheduling:

Problem Statement:

A business conference needs to schedule 9 different keynote sessions from three different industries: technology, finance, and healthcare. There are 4 potential technology sessions, 104 finance sessions, and 17 healthcare sessions to choose from. How many different schedules can be made? Express your answer in scientific notation rounding to the hundredths place.

Solution:

```

# Calculate the combinations for each category
technology_choices <- choose(4, 1)
finance_choices <- choose(104, 4)
healthcare_choices <- choose(17, 4)

# Calculate the total number of schedules
total_schedules <- technology_choices * finance_choices * healthcare_choices

# Format in scientific notation, rounded to hundredths
total_schedules_scientific <- format(total_schedules, scientific = TRUE, digits = 3)
cat("The total number of different schedules is approximately:", total_schedules_scientific, "\n")

## The total number of different schedules is approximately: 4.38e+10

```

Book Selection for Corporate Training

Problem Statement:

An HR manager needs to create a reading list for a corporate leadership training program, which includes 13 books in total. The books are categorized into 6 novels, 6 business case studies, 7 leadership theory books, and 5 strategy books.

Solution:

Step 1: If the manager wants to include no more than 4 strategy books, how many different reading schedules are possible? Express your answer in scientific notation rounding to the hundredths place.

```

# Number of books in each category
novels <- 6
business_case_studies <- 6
leadership_theory <- 7
strategy <- 5

```



```

# Initialize total schedules
total_schedules <- 0

# Loop over each possible count of strategy books from 0 to 4
for (s in 0:4) {
  # Calculate the number of ways to select `s` strategy books from 5
  strategy_choices <- choose(strategy, s)

  # Remaining books needed from other categories to reach 13 books
  remaining_books_needed <- 13 - s

  # Calculate the number of ways to select the remaining books from novels, business case studies, and
  remaining_choices <- choose(novels + business_case_studies + leadership_theory, remaining_books_needed)

  # Multiply the choices for strategy books with the choices for remaining books
  total_schedules <- total_schedules + (strategy_choices * remaining_choices)
}

# Display result in scientific notation, rounded to two decimal places
paste0("Answer: ", format(total_schedules, scientific = TRUE, digits = 3))

## [1] "Answer: 2.42e+06"

```

Step 2: If the manager wants to include all 6 business case studies, how many different reading schedules are possible? Express your answer in scientific notation rounding to the hundredths place.

```

# Fixed choice of all 6 business case studies
business_case_choices <- 1

# Calculate the number of ways to select 7 more books from novels, leadership theory, and strategy books
remaining_choices_step2 <- choose(novels + leadership_theory + strategy, 7)

# Total schedules when all 6 business case studies are included
total_schedules_step2 <- business_case_choices * remaining_choices_step2

# Display result in scientific notation, rounded to two decimal places
paste0("Answer: ", format(total_schedules_step2, scientific = TRUE, digits = 3))

## [1] "Answer: 3.18e+04"

```

Product Arrangement

Problem Statement:

A retailer is arranging 10 products on a display shelf. There are 5 different electronic gadgets and 5 different accessories. What is the probability that all the gadgets are placed together and all the accessories are placed together on the shelf? Express your answer as a fraction or a decimal number rounded to four decimal places.

Solution:

```

# Factorials of the necessary numbers
total_arrangements <- factorial(10)
favorable_arrangements <- factorial(2) * factorial(5) * factorial(5)

# Probability calculation

```

```
probability <- favorable_arrangements / total_arrangements
round(probability, 4)
```

```
## [1] 0.0079
```

Expected Value of a Business Deal:

A company is evaluating a deal where they either gain \$4 for every successful contract or lose \$16 for every unsuccessful contract. A “successful” contract is defined as drawing a queen or lower from a standard deck of cards. (Aces are considered the highest card in the deck.)

Step 1: Find the expected value of the deal. Round your answer to two decimal places. Losses must be expressed as negative values.

```
# Probabilities of success and failure
p_success <- 48 / 52
p_failure <- 4 / 52

# Gains and losses
gain_success <- 4
loss_failure <- -16

# Expected value calculation
expected_value <- (p_success * gain_success) + (p_failure * loss_failure)
round(expected_value, 2)
```

```
## [1] 2.46
```

Step 2: If the company enters into this deal 833 times, how much would they expect to win or lose? Round your answer to two decimal places. Losses must be expressed as negative values.

```
# Number of deals
num_deals <- 833

# Total expected outcome
total_expected_outcome <- expected_value * num_deals
round(total_expected_outcome, 2)
```

```
## [1] 2050.46
```

Problem 3:

Supply Chain Risk Assessment:

Let $X_1, X_2, X_3, \dots, X_n$ represent the lead times (in days) for the delivery of key components from $n = 5$ different suppliers. Each lead time is uniformly distributed across a range of 1 to $k = 20$ days, reflecting the uncertainty in delivery times. Let Y denote the minimum delivery time among all suppliers. Understanding the distribution of Y is crucial for assessing the earliest possible time you can begin production. Determine the distribution of Y to better manage your supply chain and minimize downtime.

Step-by-Step Calculation for the Distribution of Y

Step 1: Distribution of Each Supplier’s Lead Time Each supplier’s lead time X_i follows a uniform distribution $U(1, k)$, where $k = 20$.

- **PDF** of X_i :

$$f_{X_i}(x) = \frac{1}{k-1}, \quad \text{for } 1 \leq x \leq k.$$

- **CDF** of X_i :

$$F_{X_i}(x) = \frac{x-1}{k-1}, \quad \text{for } 1 \leq x \leq k.$$

Step 2: Definition of Y

The minimum Y of independent variables X_1, X_2, \dots, X_n has a cumulative distribution function:

$$F_Y(y) = P(Y \leq y).$$

The event $Y \leq y$ means that at least one of the X_i is less than or equal to y . Using the independence of X_1, X_2, \dots, X_n , the complementary probability is:

$$P(Y > y) = P(X_1 > y, X_2 > y, \dots, X_n > y).$$

Step 3: Calculate $P(Y > y)$ For any single X_i :

$$P(X_i > y) = 1 - P(X_i \leq y) = 1 - F_{X_i}(y).$$

Substituting $F_{X_i}(y) = \frac{y-1}{k-1}$:

$$P(X_i > y) = 1 - \frac{y-1}{k-1} = \frac{k-y}{k-1}.$$

Since the X_i 's are independent, the probability that all n variables are greater than y is:

$$P(Y > y) = P(X_1 > y) \cdot P(X_2 > y) \cdot \dots \cdot P(X_n > y).$$

$$P(Y > y) = \left(\frac{k-y}{k-1} \right)^n.$$

Step 4: Calculate $F_Y(y)$ Using the relationship $F_Y(y) = 1 - P(Y > y)$:

$$F_Y(y) = 1 - \left(\frac{k-y}{k-1} \right)^n, \quad \text{for } 1 \leq y \leq k.$$

Step 5: Calculate $f_Y(y)$ The PDF of Y is the derivative of the CDF $F_Y(y)$:

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

Differentiate $F_Y(y) = 1 - \left(\frac{k-y}{k-1} \right)^n$:

$$f_Y(y) = n \cdot \frac{1}{k-1} \cdot \left(\frac{k-y}{k-1} \right)^{n-1}.$$

Step 6: Final Formulas - **CDF** of Y :

$$F_Y(y) = 1 - \left(\frac{k-y}{k-1} \right)^n, \quad \text{for } 1 \leq y \leq k.$$

- **PDF** of Y :

$$f_Y(y) = n \cdot \frac{1}{k-1} \cdot \left(\frac{k-y}{k-1} \right)^{n-1}, \quad \text{for } 1 \leq y \leq k.$$

Step 7: Example Calculation (with $n = 5, k = 20$): 1. **CDF** at $y = 5$:

$$F_Y(5) = 1 - \left(\frac{20-5}{20-1} \right)^5 = 1 - \left(\frac{15}{19} \right)^5.$$

- 2. **PDF** at $y = 5$:

$$f_Y(5) = 5 \cdot \frac{1}{19} \cdot \left(\frac{15}{19} \right)^4.$$

Step 8: Visualization in R

```
# Parameters
n <- 5 # Number of suppliers
k <- 20 # Maximum delivery time

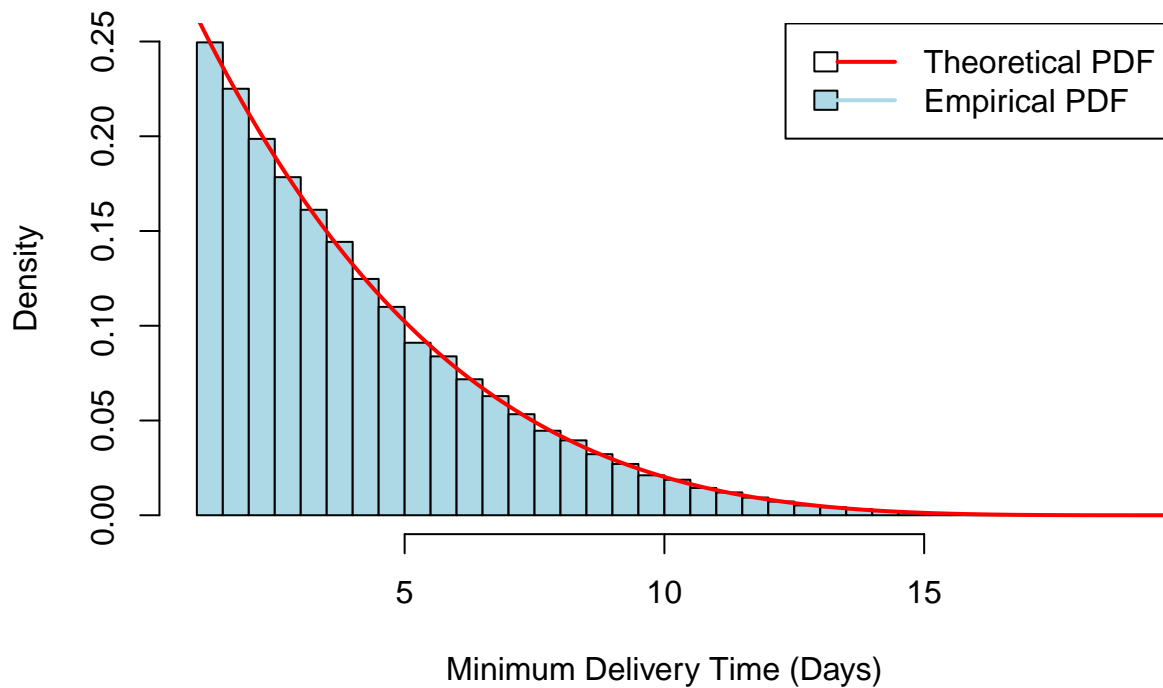
# Theoretical PDF of Y
f_Y <- function(y, n, k) {
  n * (1 / (k - 1)) * (1 - (y - 1) / (k - 1))^(n - 1)
}

# Simulating minimum delivery times
set.seed(123)
num_simulations <- 100000
lead_times <- matrix(runif(num_simulations * n, 1, k), ncol = n)
min_delivery_times <- apply(lead_times, 1, min)

# Plot theoretical vs empirical PDF
y_vals <- seq(1, k, length.out = 1000)
theoretical_pdf <- f_Y(y_vals, n, k)

hist(min_delivery_times, breaks = 30, probability = TRUE, col = "lightblue",
     main = "PDF of Minimum Delivery Time (Simulated vs Theoretical)",
     xlab = "Minimum Delivery Time (Days)")
lines(y_vals, theoretical_pdf, col = "red", lwd = 2)
legend("topright", legend = c("Theoretical PDF", "Empirical PDF"),
     col = c("red", "lightblue"), lwd = 2, fill = c(NA, "lightblue"))
```

PDF of Minimum Delivery Time (Simulated vs Theoretical)



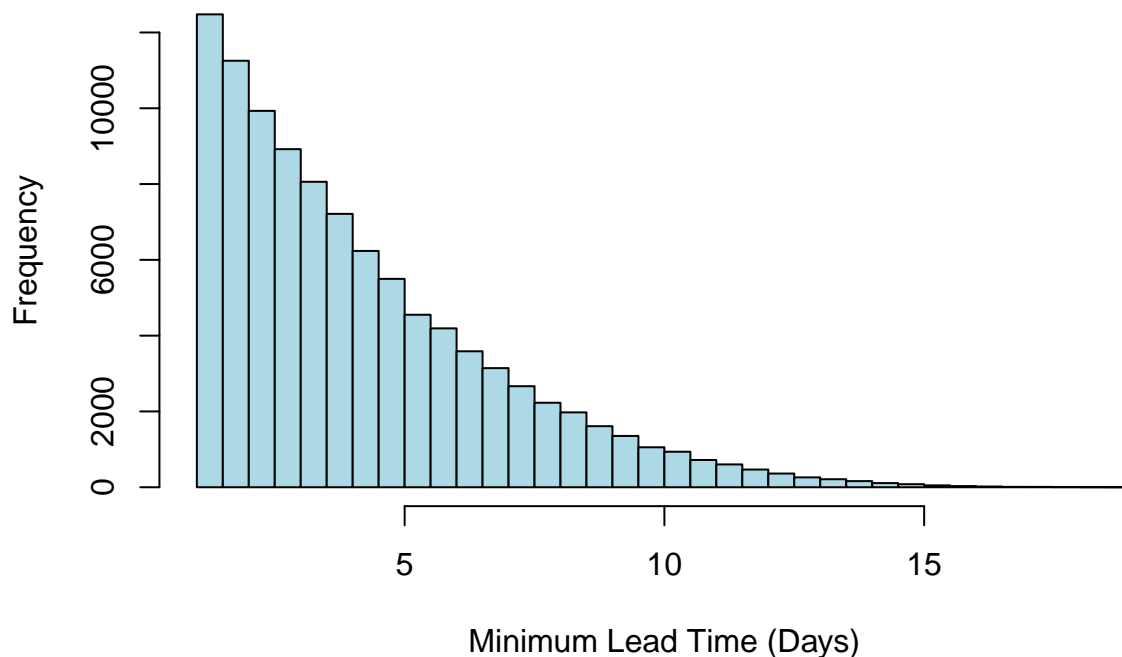
```
# Set parameters
n <- 5          # Number of suppliers
k <- 20         # Maximum lead time (days)
num_samples <- 100000 # Number of simulations

# Generate random samples for the lead times for n suppliers
set.seed(123) # For reproducibility
lead_times <- matrix(runif(num_samples * n, min = 1, max = k), ncol = n)

# Find the minimum lead time for each simulation
min_lead_times <- apply(lead_times, 1, min)

# Plot the distribution of Y (minimum lead time)
hist(min_lead_times, breaks = 50, main = "Distribution of Minimum Lead Time (Y)",
     xlab = "Minimum Lead Time (Days)", ylab = "Frequency", col = "lightblue",
     border = "black")
```

Distribution of Minimum Lead Time (Y)



```
# Display summary statistics
summary(min_lead_times)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.000   2.059   3.456   4.161   5.600   18.646
```

Maintenance Planning for Critical Equipment:

Your organization owns a critical piece of equipment, such as a high-capacity photocopier (for a law firm) or an MRI machine (for a healthcare provider). The manufacturer estimates the expected lifetime of this equipment to be 8 years, meaning that, on average, you expect one failure every 8 years. It's essential to understand the likelihood of failure over time to plan for maintenance and replacements.

Geometric Model: Calculate the probability that the machine will not fail for the first 6 years. Also, provide the expected value and standard deviation. This model assumes each year the machine either fails or does not, independently of previous years.

```
# Parameters
p <- 1 / 8 # Failure probability

# Probability of no failure for the first 6 years
geom_prob <- (1 - p)^6

# Expected value and standard deviation
geom_exp <- 1 / p
geom_sd <- sqrt((1 - p) / p^2)
```

```

# Results
cat("Geometric Model:\n")

## Geometric Model:
cat("P(X > 6):", round(geom_prob, 4), "\n")

## P(X > 6): 0.4488
cat("Expected Value (E[X]):", round(geom_exp, 2), "\n")

## Expected Value (E[X]): 8
cat("Standard Deviation (SD):", round(geom_sd, 2), "\n")

## Standard Deviation (SD): 7.48

```

Exponential Model: Calculate the probability that the machine will not fail for the first 6 years. Provide the expected value and standard deviation, modeling the time to failure as a continuous process.

```

# Parameters
lambda <- 1 / 8 # Failure rate

# Probability of no failure for the first 6 years
exp_prob <- exp(-lambda * 6)

# Expected value and standard deviation
exp_exp <- 1 / lambda
exp_sd <- 1 / lambda

# Results
cat("\nExponential Model:\n")

##
## Exponential Model:
cat("P(T > 6):", round(exp_prob, 4), "\n")

## P(T > 6): 0.4724
cat("Expected Value (E[T]):", round(exp_exp, 2), "\n")

## Expected Value (E[T]): 8
cat("Standard Deviation (SD):", round(exp_sd, 2), "\n")

## Standard Deviation (SD): 8

```

Binomial Model: Calculate the probability that the machine will not fail during the first 6 years, given that it is expected to fail once every 8 years. Provide the expected value and standard deviation, assuming a fixed number of trials (years) with a constant failure probability each year.

```

# Parameters
n <- 6
k <- 0

# Probability of no failure in the first 6 years
binom_prob <- dbinom(k, n, p)

```

```

# Expected value and standard deviation
binom_exp <- n * p
binom_sd <- sqrt(n * p * (1 - p))

# Results
cat("\nBinomial Model:\n")

##
## Binomial Model:
cat("P(X = 0):", round(binom_prob, 4), "\n")

## P(X = 0): 0.4488
cat("Expected Value (E[X]):", round(binom_exp, 2), "\n")

## Expected Value (E[X]): 0.75
cat("Standard Deviation (SD):", round(binom_sd, 2), "\n")

## Standard Deviation (SD): 0.81

```

Poisson Model: Calculate the probability that the machine will not fail during the first 6 years, modeling the failure events as a Poisson process. Provide the expected value and standard deviation.

```

# Parameters
mu <- lambda * 6 # Mean failure count

# Probability of no failures in the first 6 years
poisson_prob <- dpois(0, mu)

# Expected value and standard deviation
poisson_exp <- mu
poisson_sd <- sqrt(mu)

# Results
cat("\nPoisson Model:\n")

##
## Poisson Model:
cat("P(X = 0):", round(poisson_prob, 4), "\n")

## P(X = 0): 0.4724
cat("Expected Value (E[X]):", round(poisson_exp, 2), "\n")

## Expected Value (E[X]): 0.75
cat("Standard Deviation (SD):", round(poisson_sd, 2), "\n")

## Standard Deviation (SD): 0.87

```

Problem 4:

1. Scenario:

You are managing two independent servers in a data center. The time until the next failure for each server follows an exponential distribution with different rates:

Server A has a failure rate of $\lambda_A = 0.5$ failures per hour. Server B has a failure rate of $\lambda_B = 0.3$ failures per hour.

What is the distribution of the total time until both servers have failed at least once? Use the moment generating function (MGF) to find the distribution of the sum of the times to failure.

Solution:

Let T_A and T_B denote the time to failure for Server A and Server B, respectively. $T_A \sim \text{Exponential}(\lambda_A)$ and $T_B \sim \text{Exponential}(\lambda_B)$, where $\lambda_A = 0.5$ and $\lambda_B = 0.3$. The total time to failure is $T = T_A + T_B$.

The moment-generating function (MGF) of an exponential random variable $T \sim \text{Exponential}(\lambda)$ is:

$$M_T(s) = \frac{\lambda}{\lambda - s}, \quad s < \lambda.$$

Since T_A and T_B are independent, the MGF of $T = T_A + T_B$ is the product of their MGFs:

$$M_T(s) = M_{T_A}(s) \cdot M_{T_B}(s).$$

Substituting the MGFs for T_A and T_B :

$$M_T(s) = \frac{\lambda_A}{\lambda_A - s} \cdot \frac{\lambda_B}{\lambda_B - s}.$$

Substituting $\lambda_A = 0.5$ and $\lambda_B = 0.3$:

$$M_T(s) = \frac{0.5}{0.5 - s} \cdot \frac{0.3}{0.3 - s}.$$

To find the probability density function (PDF) of T , we perform partial fraction decomposition:

$$M_T(s) = \frac{A}{0.5 - s} + \frac{B}{0.3 - s}.$$

Equating terms:

$$\frac{0.15}{(0.5 - s)(0.3 - s)} = \frac{A}{0.5 - s} + \frac{B}{0.3 - s}.$$

Multiply through by $(0.5 - s)(0.3 - s)$:

$$0.15 = A(0.3 - s) + B(0.5 - s).$$

Expanding:

$$0.15 = 0.3A - As + 0.5B - Bs.$$

Equating coefficients of s and the constant terms:

$$-A - B = 0, \quad 0.3A + 0.5B = 0.15.$$

From $-A - B = 0$, we get $B = -A$. Substitute into $0.3A + 0.5B = 0.15$:

$$0.3A + 0.5(-A) = 0.15,$$

$$-0.2A = 0.15, \quad A = -0.75.$$

Thus, $B = 0.75$.

Substituting back:

$$M_T(s) = \frac{-0.75}{0.5 - s} + \frac{0.75}{0.3 - s}.$$

The probability density function (PDF) of T is obtained by taking the inverse Laplace transform:

$$f_T(t) = -0.75 \cdot 0.5e^{-0.5t} + 0.75 \cdot 0.3e^{-0.3t}.$$

Simplify:

$$f_T(t) = 0.375e^{-0.3t} - 0.375e^{-0.5t}, \quad t \geq 0.$$

Therefore, the total time $T = T_A + T_B$ has the PDF:

$$f_T(t) = 0.375(e^{-0.3t} - e^{-0.5t}), \quad t \geq 0.$$

Sum of Independent Normally Distributed Random Variables:

Scenario: An investment firm is analyzing the returns of two independent assets, Asset X and Asset Y. The returns on these assets are normally distributed:

Asset X: Asset Y:

Question: Find the distribution of the combined return of the portfolio consisting of these two assets using the moment generating function (MGF).

Given that X and Y are independent normally distributed random variables, we have:

$$X \sim N(\mu_X = 5\%, \sigma_X^2 = 4\%) \quad \text{and} \quad Y \sim N(\mu_Y = 7\%, \sigma_Y^2 = 9\%).$$

The sum of two independent normal random variables, $Z = X + Y$, is also normally distributed with the mean and variance given by:

$$\mu_Z = \mu_X + \mu_Y, \quad \sigma_Z^2 = \sigma_X^2 + \sigma_Y^2.$$

Step 1: Calculate the mean of Z

$$\mu_Z = \mu_X + \mu_Y = 5\% + 7\% = 12\%.$$

Step 2: Calculate the variance of Z

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 = 4\% + 9\% = 13\%.$$

Step 3: Standard deviation of Z

$$\sigma_Z = \sqrt{\sigma_Z^2} = \sqrt{13\%} \approx 3.61\%.$$

Thus, the distribution of the combined return Z is given by:

$$Z \sim N(\mu_Z = 12\%, \sigma_Z^2 = 13\%).$$

Scenario:

A call center receives calls independently from two different regions. The number of calls received from Region A and Region B in an hour follows a Poisson distribution:

Region A: Region B:

Question: Determine the distribution of the total number of calls received in an hour from both regions using the moment generating function (MGF).

Given that the number of calls received from Region A and Region B in an hour follows a Poisson distribution:

$$X_A \sim \text{Poisson}(\lambda_A = 3) \quad \text{and} \quad X_B \sim \text{Poisson}(\lambda_B = 5).$$

The total number of calls received in an hour from both regions is:

$$X = X_A + X_B.$$

Since the sum of two independent Poisson random variables also follows a Poisson distribution with the rate parameter equal to the sum of the individual rates, we have:

$$X \sim \text{Poisson}(\lambda_A + \lambda_B).$$

Step 1: Calculate the combined rate parameter

$$\lambda_X = \lambda_A + \lambda_B = 3 + 5 = 8.$$

Step 2: Moment-Generating Function (MGF) of the Poisson Distribution

The MGF of a Poisson random variable X with parameter λ is given by:

$$M_X(t) = \exp(\lambda(e^t - 1)).$$

For the total number of calls $X = X_A + X_B$, the MGF is:

$$M_X(t) = \exp((\lambda_A + \lambda_B)(e^t - 1)) = \exp(8(e^t - 1)).$$

Thus, the total number of calls received in an hour from both regions follows a Poisson distribution with parameter:

$$X \sim \text{Poisson}(\lambda = 8).$$

Problem 5:**Customer Retention and Churn Analysis**

Scenario: A telecommunications company wants to model the behavior of its customers regarding their likelihood to stay with the company (retention) or leave for a competitor (churn). The company segments its customers into three states:

State 1: Active customers who are satisfied and likely to stay (Retention state).

State 2: Customers who are considering leaving (At-risk state).

State 3: Customers who have left (Churn state).

The company has historical data showing the following monthly transition probabilities:

From State 1 (Retention): 80% stay in State 1, 15% move to State 2, and 5% move to State 3.

From State 2 (At-risk): 30% return to State 1, 50% stay in State 2, and 20% move to State 3.

From State 3 (Churn): 100% stay in State 3.

The company wants to analyze the long-term behavior of its customer base.

Question: (a) Construct the transition matrix for this Markov Chain. (b) If a customer starts as satisfied (State 1), what is the probability that they will eventually churn (move to State 3)? (c) Determine the steady-state distribution of this Markov Chain. What percentage of customers can the company expect to be in each state in the long run?

Solution:

```
# Transition matrix
P <- matrix(c(0.80, 0.15, 0.05,
              0.30, 0.50, 0.20,
              0.00, 0.00, 1.00),
            nrow = 3, byrow = TRUE)

# (b) Probability that a customer starting in State 1 eventually moves to State 3 (Churn)
# Create the fundamental matrix for the transient states
Q <- P[1:2, 1:2] # Submatrix for transient states
I <- diag(2)      # Identity matrix for transient states
N <- solve(I - Q) # Fundamental matrix

# Calculate the absorbing probabilities
R <- P[1:2, 3, drop = FALSE] # Submatrix for transitions to the absorbing state (State 3)
B <- N %*% R                  # Absorbing probabilities for starting in States 1 and 2

prob_churn_from_state_1 <- B[1]

# (c) Steady-state distribution
# Solving for the steady-state distribution: pi %*% P = pi
library(expm)

## Loading required package: Matrix
##
## Attaching package: 'Matrix'
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
## Attaching package: 'expm'
## The following object is masked from 'package:Matrix':
##
##     expm

steady_state <- eigen(t(P))$vectors[,1]
steady_state <- steady_state / sum(steady_state) # Normalize to sum to 1

# Results
list(
  Transition_Matrix = P,
  Prob_Churn_From_State_1 = prob_churn_from_state_1,
```

```

Steady_State_Distribution = steady_state
)

## $Transition_Matrix
##      [,1] [,2] [,3]
## [1,]  0.8 0.15 0.05
## [2,]  0.3 0.50 0.20
## [3,]  0.0 0.00 1.00
##
## $Prob_Churn_From_State_1
## [1] 1
##
## $Steady_State_Distribution
## [1] 0 0 1

```

Inventory Management in a Warehouse:

Scenario: A warehouse tracks the inventory levels of a particular product using a Markov Chain model. The inventory levels are categorized into three states:

State 1: High inventory (More than 100 units in stock).

State 2: Medium inventory (Between 50 and 100 units in stock).

State 3: Low inventory (Less than 50 units in stock).

The warehouse has the following transition probabilities for inventory levels from one month to the next:

From State 1 (High): 70% stay in State 1, 25% move to State 2, and 5% move to State 3.

From State 2 (Medium): 20% move to State 1, 50% stay in State 2, and 30% move to State 3.

From State 3 (Low): 10% move to State 1, 40% move to State 2, and 50% stay in State 3.

The warehouse management wants to optimize its restocking strategy by understanding the long-term distribution of inventory levels.

Question: (a) Construct the transition matrix for this Markov Chain. (b) If the warehouse starts with a high inventory level (State 1), what is the probability that it will eventually end up in a low inventory level (State 3)? (c) Determine the steady-state distribution of this Markov Chain. What is the long-term expected proportion of time that the warehouse will spend in each inventory state?

Solution:

```

# Transition matrix
P <- matrix(c(0.70, 0.25, 0.05,
              0.20, 0.50, 0.30,
              0.10, 0.40, 0.50),
            nrow = 3, byrow = TRUE)

# (b) Probability that a warehouse starting in State 1 (High) eventually ends up in State 3 (Low)
# Since this is not an absorbing Markov Chain, we will look at powers of the transition matrix
# to analyze long-term behavior.

# (c) Steady-state distribution
# Solving for steady-state distribution: pi %*% P = pi
steady_state <- eigen(t(P))$vectors[,1]
steady_state <- steady_state / sum(steady_state) # Normalize to sum to 1

```

```
# Results
list(
  Transition_Matrix = P,
  Steady_State_Distribution = steady_state
)
```

```
## $Transition_Matrix
##      [,1] [,2] [,3]
## [1,]  0.7 0.25 0.05
## [2,]  0.2 0.50 0.30
## [3,]  0.1 0.40 0.50
##
## $Steady_State_Distribution
## [1] 0.3466667 0.3866667 0.2666667
```

To calculate the probability of eventually reaching a low inventory state from an initial high state, we would need to examine the evolution of the system over multiple time steps using powers of the transition matrix P^n as n increases. This would show the probability distribution across states after several transitions.
