DATA 605: Computational Mathematics Homework 2

Gabriel Campos

Last edited November 09, 2024

Library

```r
library(expm)
```

```
## Loading required package: Matrix
##
## Attaching package: 'expm'
## The following object is masked from 'package:Matrix':
##
##     expm
```

Problem 1

1 (Bayesian)

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.

i. 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?

Following Bayesian Theorem and given the following

- **Prevalence** $P(Default)$= Probability borrower defaults is given as 0.02.
- **1-Prevalence** $P(\neg Default)$= Probability the borrower does not default which is $1 - P(D) = 0.98$
- **Sensitivity** $P(Positive|Default) = 0.90$ is the probability of a positive test given the borrower defaults, as per the question.
- **Specificity** $P(Negative|\neg Default) = 0.95$ is the probability that a borrower will not default is not flagged.
- **1-Specificity** $P(Positive|\neg Default) = 0.05$ is the false positive rate.

The question asks the probability a flagged borrower will default or *Default Given Positive*:

$P(Default|Positive) = \frac{P(Positive|Default) \cdot P(Default)}{P(Positive)}$

so we need to solve for

$P(Positive) = P(Positive|Default) \cdot P(Default) + P(Positive|\neg Default) \cdot P(\neg Default)$

and then apply to the formula above

```r
sensitivity <- 0.90                    # P(Positive | Default)
specificity <- 0.95                    # P(Negative | No Default)
prevalence <- 0.02                     # P(Default)
```

```r
false_positive_rate<- 1-specificity    # False positive rate
p_positive<- (sensitivity*prevalence)+
  (false_positive_rate*(1-prevalence)) # P(Positive)
p_default_given_positive<-(sensitivity * prevalence)/p_positive
p_default_given_positive
```

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

Answer: 26.86567%

ii. 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?

- $P(flagged) = (0.90 \times 0.02) + ((1 - 0.95) \times (1 - 0.02))$
- $Number\ flagged = 10,000 \times P(flagged)$
- $E[Defaults] \approx Number\ flagged \times P(Positive)$
- $Cost = E[Defaults] \times 200,000$

```r
n<-10000
cost_default<- 200000
cost_test<-500
p_flagged<- (0.90*0.02)+((1-0.95)*(1-0.02))
n_flagged<- n*p_flagged
e_default<-n_flagged*p_default_given_positive
cost<-(e_default*200000)+(cost_test*n)
format(cost, scientific = FALSE)
```

```
## [1] "41000000"
```

Answer:\$41,000,000

```r
rm(list = ls())
```

2 (Binomial)

i 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?

- $p = 0.7$
- $n = 8$
- $P(X = k) = \binom{n}{k} \cdot (p)^k \cdot (1 - p)^{n-k}$

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

```r
dbinom(6,8,0.7)
```

```
## [1] 0.2964755
```

Answer: 29.64755%

ii What is the probability that it pays dividends 6 or more times?

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

```r
sum(dbinom(6:8,8,0.7))
```

```
## [1] 0.5517738
```

Answer: 55.17738%

iii What is the probability that it pays dividends fewer than 6 times?

$P(X < 6) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$

```
sum(dbinom(0:5,8,0.7))
```

```
## [1] 0.4482262
```

Answer:44.82262%

iv What is the expected number of dividend payments over 8 quarters?

$Expected\ number = n \times p$

```
8*0.7
```

```
## [1] 5.6
```

Answer: 5 (rounded down)

v What is the standard deviation?

$Standard\ deviation = \sqrt{n \times p \times (1 - p)}$

$Standard\ deviation = \sqrt{8 \times 0.7 \times (1 - 0.7)}$

```
sqrt(8*0.7*(1-0.7))
```

```
## [1] 1.296148
```

Answer:1.296

3 (Poisson)

i 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?

Although *Poisson* is in the title, this problem is not suited for this method. Poisson uses large n and small p leading me to prefer the binomial approach. I will submit the binomial as my answer but will show Poisson along with the theme of the title.

If I were to use Poisson distribution

Poisson distribution according to the book is

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

where

- $lambda = 12$
- $k = 4$

we can just use R to solve

```
dpois(4, 12)
```

```
## [1] 0.005308599
```

Answer: 0.531%

ii What is the probability that more than 12 such days occur in a given month?

$P(X > 12) = 1 - P(X \leq 12)$

Using Poisson

```
1-ppois(12, 12)
```

```
## [1] 0.4240348
```

Answer: 42.40%

iii How many such days would you expect in a 6-month period?

$Expected\ days = n \times p = 20 \times 0.6 = 12$

Multiply by the number of months $12\ days \times 6\ months = 72\ days$

Answer: 72 days

iv What is the standard deviation of the number of such days?

Assuming the number of days is for a 6-month period

- $\sigma_{1\ month} = \sqrt{\lambda}$
- $\sigma_{6\ months} = 6 \times \sqrt{\lambda}$ or just $\sqrt{72}$ from our previous problem

```
sqrt(72)
```

```
## [1] 8.485281
```

Answer: 8.485281 days

v 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?

Assuming:

- $12\ trading\ days \times 12\ months = 144\ days$
- $\frac{Minimum\ Days\ Required}{Expected\ Days} \times 100 = \frac{70}{144} \times 100$

```
(70/144)*100
```

```
## [1] 48.61111
```

Answer: 48.6% of the profitable days is made up of the 70 days in question, so obviously it has a high probability of success (100%) since we're expected to achieve more profitable days. I would still say move forward with confidence but monitor the market for any sudden changes in trends.

4 (Hypergeometric)

i 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?

Hypergeometric probability formula:

$h(N, K, n, x) = \frac{\binom{k}{x} \cdot \binom{N-k}{n-x}}{\binom{N}{n}}$

Where we define the variables as the following

- $N = 25$ or Total number of stocks

- $k = 15$ or Number of high-risk stocks
- $n = 7$ or number of selected stocks
- $x = 5$ or number

If we plug-and-play we get $\frac{\binom{15}{5} \cdot \binom{25-15}{7-5}}{\binom{25}{7}}$

R has a helpful function dyper() that uses parameters m for k which will give us our answer

```
dhyper(x=5,m=15,n=10,k=7)
```

```
## [1] 0.2811213
```

Answer: 28.11%

ii How many high-risk and low-risk stocks would you expect to be selected?

page 193 of the probability textbook implies the formula for $E[high - risk] = n \cdot \frac{k}{N}$ whose result we can subtract from 7 to find the $low - risk$

```
# high-risk
7*(15/25)
```

```
## [1] 4.2
```

```
# low-risk
7-(7*(15/25))
```

```
## [1] 2.8
```

Answer: 4.2 high-risk and 2.8 low-risk

5 (Geometric)

i 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?

- $p = 0.005$
- $1 - p = 0.95$
- $n = 10$

$\therefore 1 - (1 - p)^{10}$

will give us our answer

```
p<-0.005
p_no_default<-1-p
1-p_no_default^10
```

```
## [1] 0.04888987
```

Answer: 4.89%

ii What is the probability that it will default in the next 15 years?

Similar to above but $n = 15$

```
1-p_no_default^15
```

```
## [1] 0.07243103
```

Answer: 7.24%

iii What is the expected number of years before the bond defaults?

Since were trying to see when the first default will occur we can use $E[Years\ to\ Default] = \frac{1}{p}$

```
paste(1/p,"years")
```

```
## [1] "200 years"
```

Answer: 200 years

iv If the bond has already survived 10 years, what is the probability that it will default in the next 2 years?

Since the problem does not explicitly state the bonds are not independently and identically distributed, and considering their are bonds, based on the scenario I will proceed assuming this.

```
1-(1-p)^2
```

```
## [1] 0.009975
```

```
rm(list = ls(pattern = "^p"))
```

Answer: 0.9975%

6 (Poisson)

i 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?

Poisson Distribution

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

We basically have to calculate for $1 - P(X \leq 2)$ Where $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$ and $k = X$ respectively

Our average number of failures is given as $\lambda = 1$ every 1500 hours.

We can use ppois() to solve this

```
1-ppois(q = 2, lambda = 1)
```

```
## [1] 0.0803014
```

**Answer: 8.03%*

ii What is the expected number of failures?

This was given as 1 failure every 1500 hours so $\lambda = 1$

Answer: 1 expected failure

7 (Uniform Distribution)

i 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?

Uniform Distribution function is

$m(\omega) = \frac{1}{n}\ for\ every\ \omega \in \Omega$

where (in this example)

$\Omega = \{(i, j) : 20 \leq i, j \leq 60$

and the interval between is 40 days.

We can view the probability as a ratio of the interval from 40 to 60 days, which we can represent as

$P(more\ than\ 40\ days) = \frac{60-40}{60-20} = \frac{20}{40}$

```
(60-40)/(60-20)
```

```
## [1] 0.5
```

Answer: 50%

ii 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?

at the 40th day the remaining days is $60 - 40 = 20\ days$

since we're looking for $P(40 \leq x \leq 50) = \frac{50-40}{60-20} = \frac{10}{40}$

```
10/40
```

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

Answer: 25%

iii What is the expected time for the stock to reach the target price?

$E[X] = \mu = \frac{20+60}{2} = 40\ days$

Answer: 40 days

8 (Exponential Distribution)

i 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?

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

$\therefore \lambda = \frac{1}{\mu}$

Given $\mu = 8,\ \lambda = \frac{1}{8}$

Either by reading carefully deciphering the significance of λ the expected time is 8 years.

Answer: 8 years

ii What is the standard deviation?

$\sigma = \mu$ so its 8 years as well.

Answer: 8 years

iii What is the probability that the start-up will go public or fail after 6 years?

page 72 of the probability text gives

$f(t) = \lambda e^{-\lambda t}$

in R we can us pexp() to solve

```
1-pexp(6, (1/8))
```

```
## [1] 0.4723666
```

Answer: 47.2%

iv 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?

$1 - f(t) \rightarrow 1 - \lambda e^{-\lambda t}$

```
pexp(2, (1/8))
```

```
## [1] 0.2211992
```

Answer: 22.2%

Problem 2

1 (Product Selection)

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?

We can find this out by adding the combinations for choosing 0 and 1 green pen:

The combinations for 0 green is $\binom{5}{0}$ for green and $\binom{7}{5}$ for red

while (without replacement) the combinations for 1 green is $\binom{5}{1}$ for green and $\binom{7}{4}$ for red.

So our answer is $\binom{5}{0} \cdot \binom{7}{5} + \binom{5}{1} \cdot \binom{7}{4}$

```
choose(5,0)*choose(7,5) + choose(5,1)*choose(7,4)
```

```
## [1] 196
```

ANSWER:196

2 (Team Formation for a Project)

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?

Case 1: Basically reads as choose 1 out of 14 and 4 out of 13. The combination for both implies multiplication. Case 2: All junior managers

```
(choose(13,4)*choose(14,1))+(choose(13,5)*choose(14,0))
```

```
## [1] 11297
```

Answer: 11297

3 (Marketing Campaign Outcomes)

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?

The question does not require any combinatrics, just multiplication.

$5 \times 2 \times 3 = 30$

Answer: 30

4 (Product Defect Probability)

A quality control team draws 3 products from a batch of size N 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?

- $n = N$
- $R = Defect\ Rate$
- $D = N \times R$
- $P(no\ defects) = \frac{\binom{N-D}{3}}{\binom{N}{3}} = \frac{\binom{N-N\times R}{3}}{\binom{N}{3}}$
- $P(At\ least\ 1\ defective) = 1 - \frac{\binom{N-N\times R}{3}}{\binom{N}{3}}$

Answer: $P(At\ least\ 1\ defective) = 1 - \frac{\binom{N-N\times R}{3}}{\binom{N}{3}}$ **OR** $P(at\ least\ 1\ defective) = 1 - (1-R)^3$ **Simplified**

5 (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?

$\binom{n}{r}$

where $n = 17 + 14 = 31$ $r = 5$

$\binom{31}{5}$

```r
choose(31,5)
```

```
## [1] 169911
```

Answer: 169911

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

Calculate the number of low risk combination and subtract from previous answer.

```r
choose(31,5)-choose(17,5)
```

```
## [1] 163723
```

Answer: 163723

6 (Event Scheduling)

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.

$Schedules\ for\ one\ combination = \binom{4}{x_{tech}} \times \binom{104}{x_{finance}} \times \binom{17}{x_{health}}$

```r
sessions <- 9
tech_num <- 4
finance_num <- 104
health_num <- 17

combinations_conference <- expand.grid(
```

```r
  tech = 1:min(tech_num, sessions-2),
  finance = 1:min(finance_num, sessions-2),
  healthcare = 1:min(health_num, sessions-2)
)

combinations_conference <- combinations_conference[rowSums(combinations_conference) == sessions,]

schedule_counts <- choose(tech_num, combinations_conference$tech) *
                   choose(finance_num, combinations_conference$finance) *
                   choose(health_num, combinations_conference$healthcare)

total_schedules <- sum(schedule_counts)

total_schedules_sci <- formatC(total_schedules, format = 'e', digits = 2)

total_schedules_sci
```

```
## [1] "2.83e+12"
```

Answer: 2.83e+12

7 (Book Selection for Corporate Training)

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.

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.

```r
# 0 Strategy Books
strat_0 <- choose(5, 0) * choose(6, 6) * choose(6, 6) * choose(7, 1)

# 1 Strategy Book
strat_1 <- choose(5, 1) * choose(6, 6) * choose(6, 6) * choose(7, 2)

# 2 Strategy Books
strat_2 <- choose(5, 2) * choose(6, 6) * choose(6, 6) * choose(7, 3)

# 3 Strategy Books
strat_3 <- choose(5, 3) * choose(6, 6) * choose(6, 6) * choose(7, 4)

# 4 Strategy Books
strat_4 <- choose(5, 4) * choose(6, 6) * choose(6, 6) * choose(7, 5)

# Total
format(strat_0 + strat_1 + strat_2 + strat_3 + strat_4, scientific = TRUE, digits = 3)
```

```
## [1] "9.17e+02"
```

Answer: 9.17×10^2

```r
rm(list = ls())
```

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.

```r
# Case 1: 0 Strategy Books
strat_0 <- choose(5, 0) * choose(6, 6) * choose(7, 1)

# Case 2: 1 Strategy Book
strat_1 <- choose(5, 1) * choose(6, 6) * choose(7, 1)

# Case 3: 2 Strategy Books
strat_2 <- choose(5, 2) * choose(6, 6) * choose(7, 1)

# Case 4: 3 Strategy Books
strat_3 <- choose(5, 3) * choose(6, 6) * choose(7, 1)

# Case 5: 4 Strategy Books
strat_4 <- choose(5, 4) * choose(6, 2) * choose(7, 1)

# Total Number of Different Reading Schedules
total_schedules <- strat_0 + strat_1 + strat_2 + strat_3 + strat_4

# Print the result
total_schedules
```

```
## [1] 707
```

```r
# Convert to scientific notation, rounded to 2 decimal places
format(total_schedules, scientific = TRUE, digits = 3)
```

```
## [1] "7.07e+02"
```

Answer: 7.07×10^2

```r
rm(list = ls())
```

8 (Product Arrangement)

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.

- The different ways we can arrange all electronics gadgets or accessories together is 5!
- This can be ordered with either group leading $\therefore 2 \times 5!5! = favorable\ arrangement$
- The random ways the arrangement of these groups can be ordered is 10!

$\therefore P(All\ gadgets\ \&\ accessories\ grouped) = \frac{2 \times 5!5!}{10!}$

```r
(2*factorial(5)*factorial(5))/factorial(10)
```

```
## [1] 0.007936508
```

Answer: 0.0079

9 (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 Given: - $P(X \leq Queen) = successful\ contract$ - $P(successful\ contract) = \frac{44}{52} = \frac{11}{13}$ this accounts for 2-10, Jack & Queens in 4 suits - $P(unsuccessful\ contract) = 1 - \frac{44}{52} = \frac{8}{52}$ this accounts for King and

Ace in 4 suits

Using the loss and gains with the probabilities we get $\$4 \times \frac{12}{13} + (-\$16) \times \frac{4}{52}$

```
4*(11/13)+(-16)*(8/52)
```

```
## [1] 0.9230769
```

Find the expected value of the deal. Round your answer to two decimal places. Losses must be expressed as negative values. **Answer: $0.92**

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.

$\$0.92 \times 833$

```
0.92*833
```

```
## [1] 766.36
```

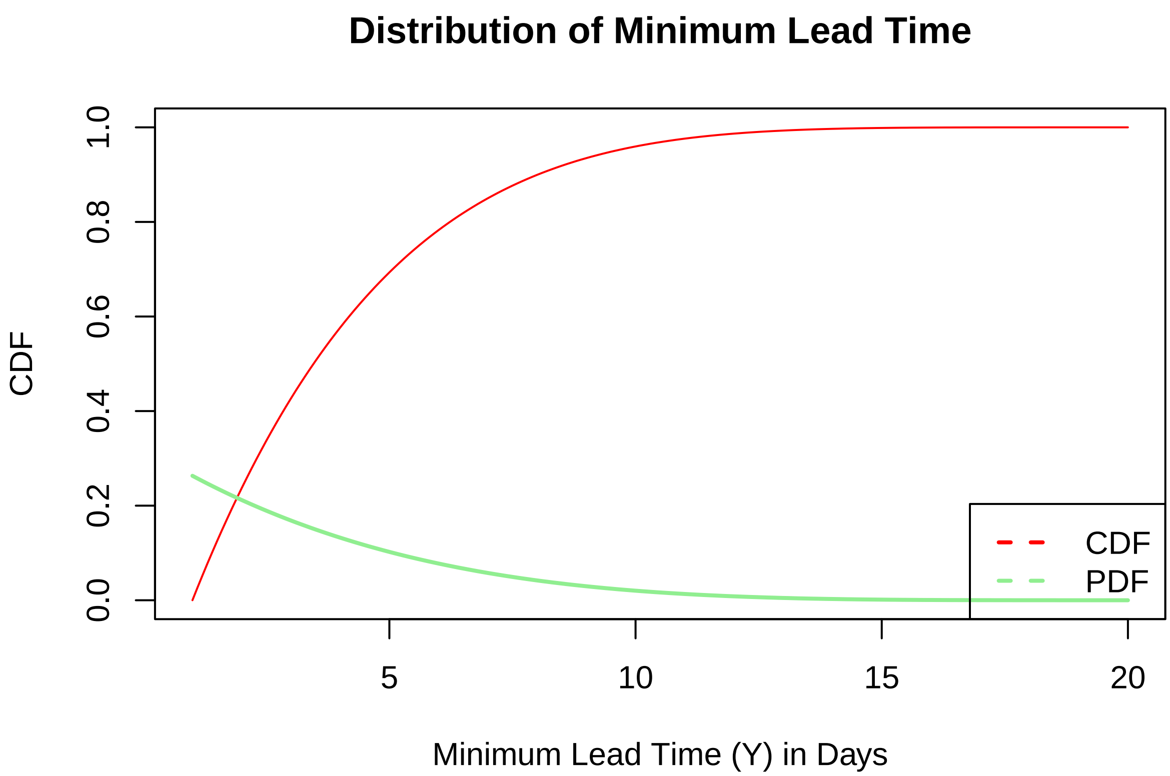
Answer: $766.36

Problem 3

1 (Supply Chain Risk Assessment)

Let $X_1, X_2, ..., X_n$ 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.

```
n <- 5
k <- 20
y_values <- seq(1, k, by = 0.1)
cdf <- 1 - (1 - (y_values - 1)/(k - 1))^n
pdf <- n * (1 - (y_values - 1)/(k - 1))^(n - 1) * (1/(k - 1))
plot(y_values,
     cdf, type = "l",
     col = "red",
     xlab = "Minimum Lead Time (Y) in Days",
     ylab = "CDF",
     main = "Distribution of Minimum Lead Time")
lines(y_values,
      pdf,
      type="l",
      col="lightgreen",
      lwd = 2)
legend("bottomright",
       legend = c("CDF", "PDF"),
   col = c("red",
           "lightgreen"),
   lty = 2,
   lwd  = 2)
```



Answer: To figure out a good lead time for production, we look at the CDF and PDF for a 20-day period. The CDF gives us the running total of the chances that supplies will arrive by each day, and the PDF shows the chance of arrival on a single day.

Although precise data points are not displayed, we approximate based on visual trends observed in the plot. For example, if a 90% confidence level in supply arrival is desired, it seems appropriate to select approximately 7 days as the lead time, as the CDF approaches this threshold by day 7.2. This approach offers a pragmatic balance between operational confidence and production planning needs.

```
rm(list = ls())
```

2 (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.

a. 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.

- $\lambda = 8$
- $E[T] = \frac{1}{\lambda}$
- $P(X > 6) = (1-p) \cdot (1-p) \cdot (1-p) \cdot (1-p) \cdot (1-p) \cdot (1-p) = (1-p)^6$
- $\sigma = \frac{\sqrt{1-p}}{p}$

```
n<-6
lambda <- 8
E_t <- 1/lambda
```

```r
p_x_6 <- (1 - E_t)^n
std_6 <- sqrt((1-E_t)/E_t^2)

print(paste("Probability of no failure within the first 6 years is:", p_x_6))
```

```
## [1] "Probability of no failure within the first 6 years is: 0.448795318603516"
```

```r
print(paste("The expected value is given: 8 years"))
```

```
## [1] "The expected value is given: 8 years"
```

```r
print(paste("Standard deviation is:", std_6))
```

```
## [1] "Standard deviation is: 7.48331477354788"
```

Answer:

- **Probability of no failure within the first 6 years is: 0.448795318603516**
- **The expected value is given: 8 years**
- **Standard deviation is: 7.48331477354788**

b. 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.

```r
pexp(n,E_t, lower.tail = FALSE)
```

```
## [1] 0.4723666
```

Answer:

- $P(X > 6)_{Exponential\ Model} = 47.24\%$
- $E[X] = 8\ years$ **or** $\frac{1}{lambda} = \frac{1}{8}$
- **Standar Deviation is 8 years or** $\frac{1}{lambda} = \frac{1}{8}$

c. 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.

```r
p_x_6_binom<-dbinom(0,n,1/lambda)
E_x_binom<-n*(1/lambda)
std__6_binom<-sqrt(n*(1/lambda)*(1-(1/lambda)))

print(paste("Probability of no failure within the first 6 years is:", p_x_6_binom))
```

```
## [1] "Probability of no failure within the first 6 years is: 0.448795318603516"
```

```r
print(paste("The expected value is ", E_x_binom))
```

```
## [1] "The expected value is  0.75"
```

```r
print(paste("Standard deviation is:", std__6_binom))
```

```
## [1] "Standard deviation is: 0.810092587300983"
```

Answer:

- **Probability of no failure within the first 6 years is: 0.448795318603516**
- **The expected value is 0.75**
- **Standard deviation is: 0.810092587300983**

d. 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.

```
lambda_pois<-n*(1/lambda)
std_6_pois<-sqrt(lambda_pois)
p_x_6_pois<-dpois(0,lambda_pois)

print(paste("Probability of no failure within the first 6 years is:", p_x_6_pois))
```

```
## [1] "Probability of no failure within the first 6 years is: 0.472366552741015"
```

```
print(paste("The expected value is ", lambda_pois))
```

```
## [1] "The expected value is  0.75"
```

```
print(paste("Standard deviation is:", std_6_pois))
```

```
## [1] "Standard deviation is: 0.866025403784439"
```

Answer:

- **Probability of no failure within the first 6 years is: 47.24%**
- **The expected value is 0.75**
- **Standard deviation is: 0.866025403784439**

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.

Question: 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.

Expectation

- $E[T_A] = \frac{1}{\lambda_a} = \frac{1}{0.5} = 2$
- $E[T_B] = \frac{1}{\lambda_a} = \frac{1}{0.3} \approx 3.333$
- $\therefore E[T] = 2 + 3.333 = 5.333\ hours$

Variance

- $\sigma^2 = Var(T) = Var(T_A) + Var(T_B)$
- $Var(T_A) = \frac{1}{(0.5)^2} = \frac{1}{0.25} = 4\ hours^2$
- $Var(T_B) = \frac{1}{(0.3)^2} = \frac{1}{0.09} \approx 11.11\ hours^2$
- $\sigma^2 = 4 + 11.11 = 15.11\ hours^2$

Moment Generating Function

- $M_T(s) = M_{T_A}(s) \cdot M_{T_B}(s)$ where $M_{T_x}(s) = \frac{\lambda x}{\lambda_x - s}\ for\ s < \lambda_x$
- Substituting to get the distribution: $M_t(s) = \frac{0.5}{0.5-s} \cdot \frac{0.3}{0.3-s}$

I am aware the question only asks for the distribution but I felt the parameters were needed to fully answer

Answer

- $E[T] = 5.333\ hours$
- $\sigma^2 = 15.11\ hour^2$

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

2 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: $X \sim Y(\mu = 5\%, \sigma_X^2 = 4\%)$
- Asset Y: $X \sim Y(\mu = 7\%, \sigma_Y^2 = 9\%)$

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

- $M_{(t)} = exp(t\mu + \frac{1}{2}t^2\sigma^2)$
- $Asset\ X = M_x(t) = exp(t \cdot 0.05 + \frac{1}{2}t^2 \cdot 0.04) = exp(0.05t + 0.02t^2)$
- $Asset\ Y = M_y(t) = exp(t \cdot 0.07 + \frac{1}{2}t^2 \cdot 0.09) = exp(0.07t + 0.045t^2)$
- $M_z(t) = exp(0.05t + 0.02t^2) + exp(0.07t + 0.045t^2)$
- $M_z(t) = exp(0.12t + 0.065t^2)$

Answer: $Z \sim N(12\%, 6.5\%)$

3 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: $X_A \sim Poisson(\lambda_A = 3)$
- Region B: $X_B \sim Poisson(\lambda_B = 5)$

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

- $M_x(t) = exp(\lambda(e^t - 1))$
- $M_{x_A}(t) = exp(\lambda_A(e^t - 1)) = exp(3(e^t - 1))$
- $M_{x_B}(t) = exp(\lambda_B(e^t - 1)) = exp(5(e^t - 1))$
- $M_Z(t) = exp(3(e^t - 1)) \cdot exp(5(e^t - 1))$
- $M_Z(t) = exp(8(e^t - 1))$
- $\lambda = 8$

Answer: MGF follows Poisson Distribution $\sigma^2 = 8,\ \lambda = 8 \therefore$ the distribution of the total number of calls is 8 calls per hour.

Problem 5

1 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.

```r
t_matrix <- matrix(c(0.80,0.15,0.05,
                              0.30,0.50,0.20,
                              0.00,0.00,1.00),
                            nrow = 3, byrow = TRUE)
rownames(t_matrix) <- c('Retention', 'At-Risk', 'Churn')
colnames(t_matrix) <- c('Retention', 'At-Risk', 'Churn')

t_matrix
```

```
##           Retention At-Risk Churn
## Retention       0.8    0.15  0.05
## At-Risk         0.3    0.50  0.20
## Churn           0.0    0.00  1.00
```

Answer:

$$\begin{pmatrix} 0.80 & 0.15 & 0.05 \\ 0.30 & 0.50 & 0.20 \\ 0.00 & 0.00 & 1.00 \end{pmatrix}$$

(b) If a customer starts as satisfied (State 1), what is the probability that they will eventually churn (move to State 3)?

- create sub-matrix for transient states (1&2, 3 doesn't have any) Q.
- sub-matrix from transient to absorption states R.
- Fundamental matrix $N = (I - Q)^{-1}$.
- Create the absorption probability matrix $B = N \times R$.

```r
Q <- t_matrix[1:2, 1:2]
R <- t_matrix[1:2, 3]
N <- solve(diag(2) - Q)
B <- N %*% R
B
```

```
##           [,1]
## Retention    1
## At-Risk      1
```

Answer: Probability of being absorbed from either is 1.0

(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?

We can keep multiplying the transition matrix by itself until it settles into a steady pattern. Here, we'll raise it to the 100th power and check out the final probabilities.

```r
round(t_matrix %^% 100)
```

```
##           Retention At-Risk Churn
## Retention         0       0     1
## At-Risk           0       0     1
## Churn             0       0     1
```

Answer: 100%

2 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.

```
t2_matrix <- matrix(c(0.70,0.25,0.05,
                              0.20,0.50,0.30,
                              0.10,0.40,0.50),
                            nrow = 3, byrow = TRUE)
rownames(t2_matrix) <- c('High', 'Medium', 'Low')
colnames(t2_matrix) <- c('High', 'Medium', 'Low')
t2_matrix
```

```
##        High Medium  Low
## High    0.7   0.25 0.05
## Medium  0.2   0.50 0.30
## Low     0.1   0.40 0.50
```

Answer:

$$\begin{pmatrix} 0.70 & 0.25 & 0.05 \\ 0.20 & 0.50 & 0.30 \\ 0.10 & 0.40 & 0.50 \end{pmatrix}$$

(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)?

```
p_limit <- t2_matrix %^% 100
p_limit
```

```
##             High    Medium       Low
## High   0.3466667 0.3866667 0.2666667
## Medium 0.3466667 0.3866667 0.2666667
## Low    0.3466667 0.3866667 0.2666667
```

Answer: $P(Low) = 26.6\%$

(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?

```
eig_decom <- eigen(t(t2_matrix))
stdy_sta <- eig_decom$vectors[,1]
stdy_sta <- stdy_sta/sum(stdy_sta)
stdy_sta
```

```
## [1] 0.3466667 0.3866667 0.2666667
```

Alternative method to the steady-state distribution, is the eigenvector corresponding to the dominant eigenvalue of the transposed transition matrix.

Answer: The normalized vectors of the dominant eigenvalue or 0.3466667, 0.3866667, 0.2666667