CSE3504 Homework 2

Isaac Piegat

October 1, 2024

1 Problem 1:

Serial numbers of 1000 laptops were examined and the first digit of each number was noted. It was found that: (i) digit 1 occurs 305 times, (ii) digit 2 occurs 185 times, (iii) digit 3 occurs 124 times, (iv) digit 4 occurs 95 times, (v) digit 5 occurs 80 times, (vi) digit 6 occurs 64 times, (vii) digit 7 occurs 51 times, (viii) digit 8 occurs 49 times, and (ix) digit 9 occurs 47. Using R compute the probability mass function (PMF) and cumulative distribution function (CDF). Display the PMF and CDF in R. Represent the PMF and CDF in the form of a table. Plot the PMF and CDF. Please make sure you label the axes, and the plots with appropriate titles. Include your R code, with answers with your submission.

1.1 R Code:

```r
digits <- 1:9
frequencies <- c(305, 185, 124, 95, 80, 64, 51, 49, 47)
total <- sum(frequencies)

pmf <- frequencies / total

cdf <- cumsum(pmf)

pmf_cdf_table <- data.frame(Digit = digits, Frequency = frequencies
    , PMF = pmf, CDF = cdf)

print(pmf_cdf_table)

png("pmf_plot.png")
plot(digits, pmf, type = "b", col = "blue", pch = 16, xlab = "
    Digits", ylab = "PMF",
     main = "Probability Mass Function (PMF)", ylim = c(0, max(pmf)
        ))
grid()
dev.off()

png("cdf_plot.png")
plot(digits, cdf, type = "b", col = "green", pch = 16, xlab = "
    Digits", ylab = "CDF",
```

```
    main = "Cumulative Distribution Function (CDF)", ylim = c(0,
        1))
grid()
dev.off()
```

Listing 1: R Code for PMF and CDF Calculation

1.2 PMF and CDF Table

"Digit"	"Frequency"	"PMF"	"CDF"
1	305	0.305	0.305
2	185	0.185	0.49
3	124	0.124	0.614
4	95	0.095	0.709
5	80	0.08	0.789
6	64	0.064	0.853
7	51	0.051	0.904
8	49	0.049	0.953
9	47	0.047	1

Table 1: Probability Mass Function (PMF) and Cumulative Distribution Function (CDF) of Laptop Serial Numbers

1.3 PMF and CDF Plot

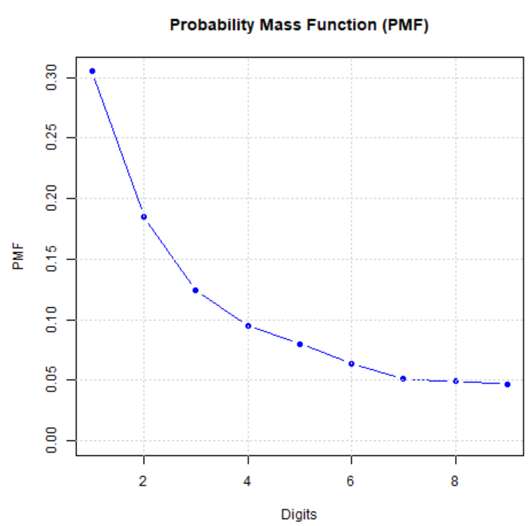


Figure 1: Probability Mass Function (PMF) of the first digit of laptop serial numbers.

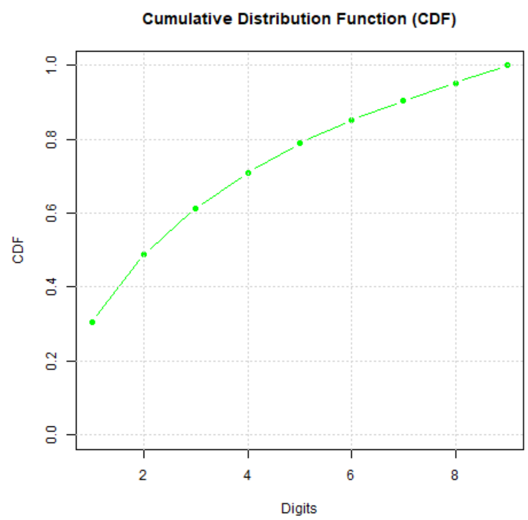


Figure 2: Cumulative Distribution Function (CDF) of the first digit of laptop serial numbers.

2 Problem 2:

Referring to Problem 1, we calculated the probabilities of different first digits of laptop serial numbers. Suppose we are sampling laptop serial numbers at random, and when we come across a serial number with a first digit as "9" we consider it a success. Write the expressions for the following probabilities, and calculate them using the pbinom and dbinom in R. Include your R code, with answers printed in the code. a: If we sample 10 serial numbers, what is the probability that the number "9" will be the first digit in at least 3 serial numbers? (9 points) b: If we sample 10 serial numbers, what is the probability that the number "9" will not be the first digit in any of the 10 numbers? (4 points) c: What is the minimum number of serial numbers that need to be sampled if we want to be at least .90 sure of finding at least one serial number with the first digit 9? In this problem, we need to calculate the probability of at least one success for different sample sizes. This problem involves experimentation with pbinom, and you need not write mathematical expressions for this problem. Write a loop to calculate the probability of at least one success for different sample sizes in R. Plot these probabilities as a function of sample size. Using abline, place a horizontal line on the graph at 0.90 to graphically identify the cutoff. Finally, find the minimum number of serial numbers. (22 points).

2.1 Expressions:

a: Probability that the number "9" will be the first digit in at least 3 serial numbers

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \sum_{k=0}^{2} \binom{n}{k} p^k (1-p)^{n-k}$$

where $n = 10$ and $p = 0.047$.

b: Probability that the number "9" will not be the first digit in any of the 10 numbers

$$P(X = 0) = \binom{n}{0} p^0 (1-p)^n = (1-p)^n$$

where $n = 10$ and $p = 0.047$.

c: Minimum number of serial numbers to be at least 90% sure of finding at least one serial number with the first digit 9

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (1-p)^n$$

We solve for n such that $1 - (1-p)^n \geq 0.90$, where $p = 0.047$.

2.2 R code:

```r
# Part (a)
n <- 10
p <- 0.047
prob_at_least_3 <- 1 - pbinom(2, size = n, prob = p)
print(prob_at_least_3) # answer is 0.009709584

# Part (b)
prob_zero_success <- dbinom(0, size = n, prob = p)
print(prob_zero_success) # answer is 0.6179154

# Part (c)
p <- 0.047

sample_sizes <- 1:100

# Vector to store probabilities of at least one success for each sample size
probs_at_least_one <- numeric(length(sample_sizes))

# Loop to calculate the probability of at least one success for each sample size
for (n in sample_sizes) {
  probs_at_least_one[n] <- 1 - pbinom(0, size = n, prob = p)
}

```

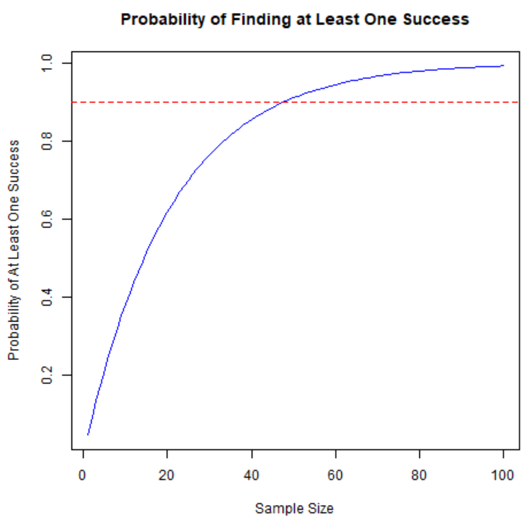
```r
# Plot the probabilities as a function of sample size
plot(sample_sizes, probs_at_least_one, type = "l", col = "blue",
     xlab = "Sample Size", ylab = "Probability of At Least One Success",
     main = "Probability of Finding at Least One Success")
abline(h = 0.90, col = "red", lty = 2)  # Add a horizontal line at 0.90

# Initialize variable for minimum sample size
min_sample_size <- NULL
prob_threshold <- 0.90  # 90% threshold

# Find the first sample size where the probability exceeds 90%
for (n in sample_sizes) {
  if (probs_at_least_one[n] >= prob_threshold) {
    min_sample_size <- n
    break  # Stop once we find the first sample size where the condition is met
  }
}

# Print the result
print(min_sample_size) # answer is 48
```

2.3 Part c plot:



3 Problem 3:

When Ava travels for professional commitments, the probability that she gains access to the Internet is 0.6 at each attempt. She makes each attempt sequentially. Write the expressions for the following probabilities, and compute them using the pgeom and qgeom functions in R. Include the R code/commands, and the answer in the R code. a: What is the probability that it takes no more than 4 attempts to get in? (6 points) b: What is the probability that more than 4 attempts will be required to gain access? (5 points) c: If Ava has already made two attempts without gaining access, what is the probability that she will gain access in the next four attempts? (7 points) d: What is the minimum number of attempts necessary to be at least .95 certain that Ava will gain access? (7 points) Write the expressions for the following probabilities, and compute them using the pgeom and qgeom functions in R. Include the R code/commands, and the answer in the R code.

3.1 Expressions:

textbfa: Probability that it takes no more than 4 attempts to get in

$$P(X \leq 4) = \sum_{k=1}^{4} (1-p)^{k-1} p = 1 - (1-p)^4$$

where $p = 0.6$.

b: Probability that more than 4 attempts will be required to gain access

$$P(X > 4) = 1 - P(X \leq 4) = (1-p)^4$$

where $p = 0.6$.

c: Probability that Ava will gain access in the next four attempts, given that she has already made two unsuccessful attempts

$$P(X \leq 4) = 1 - (1-p)^4$$

where $p = 0.6$. The previous two failed attempts do not affect this probability.

d: Minimum number of attempts necessary to be at least 95% certain that Ava will gain access

$$P(X \leq k) \geq 0.95$$

This can be written as:

$$k = \min\{k : 1 - (1-p)^k \geq 0.95\}$$

where $p = 0.6$.

3.2 R code:

```
# Part (a)
p_success <- 0.6

prob_no_more_than_4_attempts <- pgeom(3, prob = p_success)

print(prob_no_more_than_4_attempts) # answer is 0.9744

# Part (b)
p_success <- 0.6

# P(more than 4 attempts)= 1 - P(no more than 4 attempts)
prob_more_than_4_attempts <- 1 - pgeom(3, prob = p_success)

print(prob_more_than_4_attempts) # answer is 0.0256

# Part (c)
p_success <- 0.6

prob_more_than_4_attempts <- 1 - pgeom(3, prob = p_success)

print(prob_more_than_4_attempts) # answer is .9744 (same as
    part a)

# Part (d)
p_success <- 0.6

target_prob <- 0.95
min_attempts_for_95_percent <- qgeom(target_prob, prob = p_
    success)

min_attempts_for_95_percent <- min_attempts_for_95_percent + 1

print(min_attempts_for_95_percent) # answer is 4
```

4 Problem 4:

A new user of a smartphone receives an average of 3 messages per day. The arrival pattern of these messages is Poisson. Calculate the probabilities that: a: Exactly three messages will be received in any day? (6 points) b: A day will pass without receiving any message? (5 points) c: More than 3 messages will be received in any day? (9 points) Write expressions for the following probabilities, and compute them using R using dpois and ppois. Include your R code, along with your answers in the submission.

4.1 Expressions:

a: Exactly three messages will be received in any day

$$P(X = 3) = \frac{\lambda^3 e^{-\lambda}}{3!}$$

where $\lambda = 3$.

b: A day will pass without receiving any message

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

where $\lambda = 3$.

c: More than 3 messages will be received in any day

$$P(X > 3) = 1 - P(X \leq 3) = 1 - \sum_{k=0}^{3} \frac{\lambda^k e^{-\lambda}}{k!}$$

where $\lambda = 3$.

4.2 R code:

```
# Part (a)
lambda <- 3  # average messages per day

prob_exactly_3_messages <- dpois(3, lambda = lambda)

print(prob_exactly_3_messages) # answer is 0.2240418

# Part (b)
lambda <- 3  # average messages per day

prob_zero_messages <- dpois(0, lambda = lambda)

print(prob_zero_messages) # answer is 0.04978707

# Part (c)
lambda <- 3  # average messages per day

prob_more_than_3_messages <- 1 - ppois(3, lambda = lambda)

print(prob_more_than_3_messages) # answer is 0.3527681
```

5 Problem 5:

Suppose .01 of chips in a large batch on an assembly line are defective. What is the chance that from a sample 100 chips picked at random from this population, at least four will be defective? Write the expressions to compute this probability using a: Binomial distribution and b: Poisson approximation. Compute the final answers using pbinom, and ppois functions in R. Include the R commands and the answers in your submission.

5.1 Expressions:

Binomial distribution:

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - \sum_{k=0}^{3} \binom{n}{k} p^k (1-p)^{n-k}$$

where $n = 100$ and $p = 0.01$.

Poisson approximation:

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - \sum_{k=0}^{3} \frac{\lambda^k e^{-\lambda}}{k!}$$

where $\lambda = n \times p = 1$.

5.2 R code:

```
# Given data
n <- 100  # Sample size
p <- 0.01  # Probability of a defective chip
lambda <- n * p  # Poisson parameter (lambda)

# Binomial distribution
prob_binom <- 1 - pbinom(3, size = n, prob = p)
print(prob_binom)  # answer is 0.01837404

#  Poisson approximation
prob_poisson <- 1 - ppois(3, lambda = lambda)
print(prob_poisson)  # answer is 0.01837404
```

6 Problem 6:

Let X and Y be two random draws from a box containing three numbers 1, 2, 3. Calculate the distribution of S = X + Y, when X and Y are drawn a: without replacement, and b: with replacement.

6.1 Without replacement:

Combinations: (1,2),(1,3),(2,1),(2,3),(3,1),(3,2)
Sums: (3, 4, 3, 5, 4, 5)
Probabilities: S = 3, S = 4, and S = 5 all occur 2/6 times, thus each has a probability of $P = 1/3$.

6.2 With replacement:

Combinations: (1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)
Sums: (2, 3, 4, 3, 4, 5, 4, 5, 6)
Probabilities: $P(S = 2) = 1/9$, $P(S = 3) = 2/9$, $P(S = 4) = 3/9$, $P(S = 5) = 2/9$, $P(S = 6) = 1/9$.