Statistical data analysis, Assignment 4

Name: 徐竣霆
ID: 110060012

Problem 1

- code

```
### Problem 1
lambda <- 35.304
cases_observed <- 18

probability <- ppois(cases_observed, lambda)

cat("The probability of observing 18 or fewer cases is", probability, "\n")
```

- result

```
> source("/home/xavier/course/statistics/HW4/hw4.R", encoding = "UTF-8")
The probability of observing 18 or fewer cases is 0.001024734
```

The probability is extremely low, this suggests that the national incidence rate might not accurately reflect the situation in the county, as the observed number of cases is relatively lower than expected based on the national rate.

Problem 2

- code

```r
### Problem 2
# Set up
mu <- 18.5
sigma <- 5
n <- 50
p <- 0.3
size <- 5
prob <- 0.3

# Normal distribution
normal_lt_16 <- pnorm(16, mean = mu, sd = sigma)
normal_leq_16 <- pnorm(16, mean = mu, sd = sigma, lower.tail = TRUE)
normal_eq_25 <- dnorm(25, mean = mu, sd = sigma)
normal_gt_20 <- 1 - pnorm(20, mean = mu, sd = sigma)

# Binomial distribution
binom_lt_16 <- pbinom(15, size = n, prob = p)
binom_leq_16 <- pbinom(16, size = n, prob = p)
binom_eq_25 <- dbinom(25, size = n, prob = p)
binom_gt_20 <- 1 - pbinom(20, size = n, prob = p)

# Negative binomial distribution
neg_binom_lt_16 <- pnbinom(15, size = size, prob = prob)
neg_binom_leq_16 <- pnbinom(16, size = size, prob = prob)
neg_binom_eq_25 <- dnbinom(25, size = size, prob = prob)
neg_binom_gt_20 <- 1 - pnbinom(20, size = size, prob = prob)

# Print the table
cat("Distribution\tP(X < 16)\tP(X ≤ 16)\tP(X = 25)\tP(X > 20)\n")
cat("Normal\t\t", normal_lt_16, "\t", normal_leq_16, "\t", normal_eq_25, "\t", normal_gt_20, "\n")
cat("Binomial\t", binom_lt_16, "\t", binom_leq_16, "\t", binom_eq_25, "\t", binom_gt_20, "\n")
cat("Negative Binomial", neg_binom_lt_16, "\t", neg_binom_leq_16, "\t", neg_binom_eq_25, "\t", neg_binom_gt_20, "\n")
```

- result

```
Distribution      P(X < 16)      P(X ≤ 16)  P(X = 25)      P(X > 20)
Normal             0.3085375      0.3085375  0.03427372      0.3820886
Binomial           0.5691784      0.6838786  0.001436369     0.04776384
Negative Binomial 0.7624922      0.8016185  0.007739968     0.09047192
```

Problem 3

- code

```r
### Problem 3
mu <- 18.5
sigma <- 5
n <- 50
p <- 0.3
size <- 5
prob <- 0.3

# Normal distribution
normal_quantile_25 <- round(qnorm(0.25, mean = mu, sd = sigma))
normal_quantile_75 <- round(qnorm(0.75, mean = mu, sd = sigma))
normal_quantile_85 <- round(qnorm(0.85, mean = mu, sd = sigma))

# Binomial distribution
binom_quantile_25 <- qbinom(0.25, size = n, prob = p)
binom_quantile_75 <- qbinom(0.75, size = n, prob = p)
binom_quantile_85 <- qbinom(0.85, size = n, prob = p)

# Negative binomial distribution
neg_binom_quantile_25 <- qnbinom(0.25, size = size, prob = prob)
neg_binom_quantile_75 <- qnbinom(0.75, size = size, prob = prob)
neg_binom_quantile_85 <- qnbinom(0.85, size = size, prob = prob)

# Print the table
cat("Distribution\t(P(X < x) = 0.25)\t(P(X > x) = 0.25)\t(P(X ≥ x) = 0.15)\n")
cat("Normal\t\t", normal_quantile_25, "\t\t\t", normal_quantile_75, "\t\t\t", normal_quantile_85, "\n")
cat("Binomial\t", binom_quantile_25, "\t\t\t", binom_quantile_75, "\t\t\t", binom_quantile_85, "\n")
cat("Neg Binomial\t", neg_binom_quantile_25, "\t\t\t", neg_binom_quantile_75, "\t\t\t", neg_binom_quantile_85, "\n")
```

- result

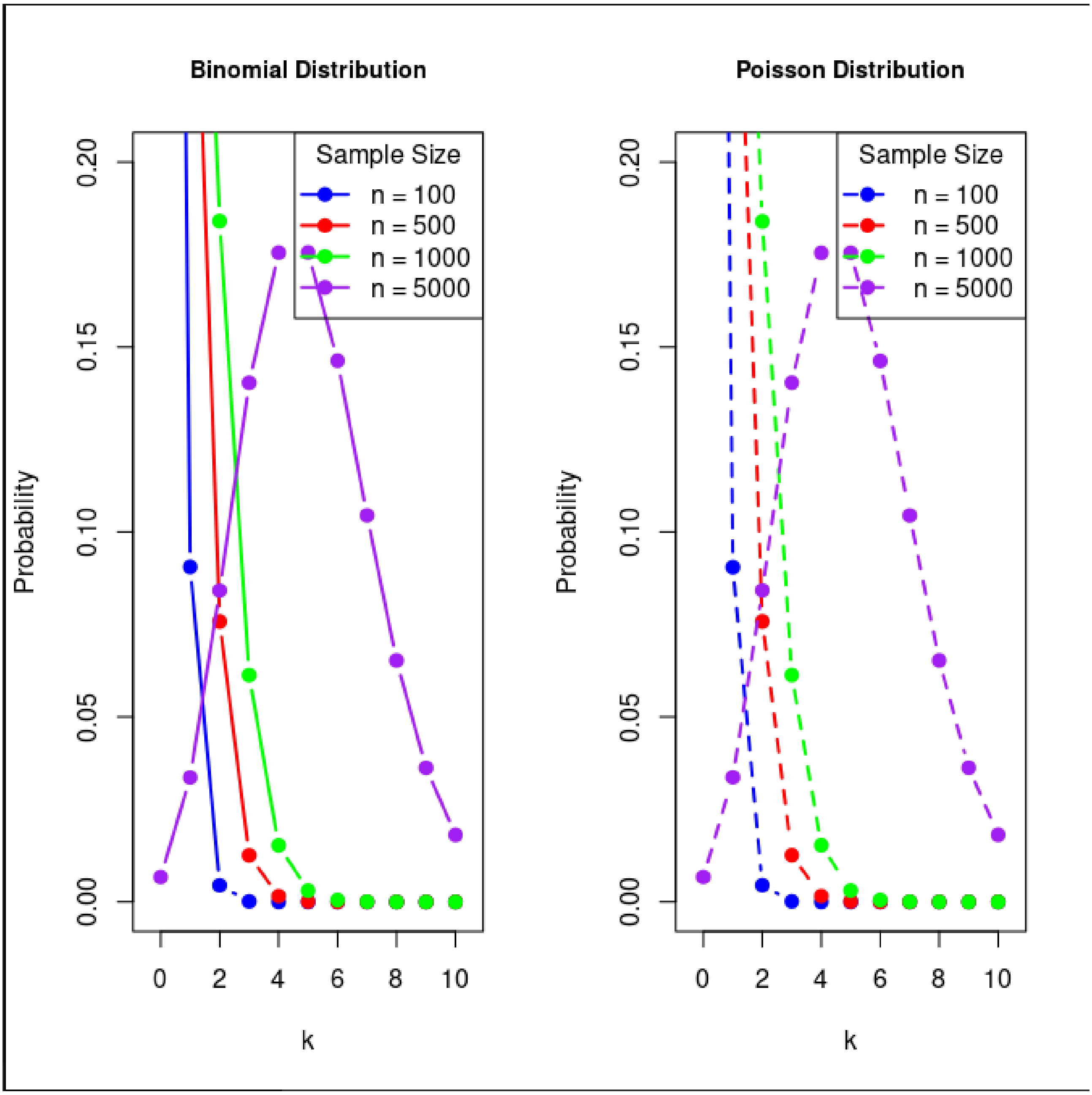
```
Distribution      (P(X < x) = 0.25)       (P(X > x) = 0.25)       (P(X ≥ x) = 0.15)
Normal             15                      22                      24
Binomial           13                      17                      18
Neg Binomial       7                       15                      18
```

Problem 4

- code

```r
### Problem 4
# Define parameters
p <- 0.001
n_values <- c(100, 500, 1000, 5000)
k_values <- 0:10

# Initialize empty lists to store probabilities for each distribution
binomial_probs <- list()
poisson_probs <- list()

# Calculate probabilities for each n value
for (n in n_values) {
  # Binomial distribution
  binomial_probs[[as.character(n)]] <- dbinom(k_values, size = n, prob = p)

  # Poisson distribution
  lambda <- n * p
  poisson_probs[[as.character(n)]] <- dpois(k_values, lambda)
}

# Plot Binomial PMFs
par(mfrow = c(1, 2))  # Set up a 1x2 grid of plots

colors <- rainbow(length(n_values))

plot(NULL, xlim = c(-0.5, 10.5), ylim = c(0, 0.2), xlab = "k", ylab = "Probability", main = "Binomial Distribution", cex.main = 0.9)

for (i in seq_along(n_values)) {
  lines(k_values, binomial_probs[[as.character(n_values[i])]], type = "b", pch = 19, col = colors[i], lty = 1, lwd = 2)
}

legend("topright", legend = paste("n =", n_values), col = colors, lty = 1, lwd = 2, pch = 19, title = "Sample Size")

# Plot Poisson PMFs
plot(NULL, xlim = c(-0.5, 10.5), ylim = c(0, 0.2), xlab = "k", ylab = "Probability", main = "Poisson Distribution", cex.main = 0.9)

for (i in seq_along(n_values)) {
  lines(k_values, poisson_probs[[as.character(n_values[i])]], type = "b", pch = 19, col = colors[i], lty = 2, lwd = 2)
}

legend("topright", legend = paste("n =", n_values), col = colors, lty = 2, lwd = 2, pch = 19, title = "Sample Size")

par(mfrow = c(1, 1))  # Reset the plotting layout
```

- result



They are almost(exactly) the same.

Problem 5

- 4.28

4.28 Playing darts. Calculate the following probabilities and indicate which probability distribution model is appropriate in each case. A very good darts player can hit the bull's eye (red circle in the center of the dart board) 65% of the time. What is the probability that he

(a) hits the bullseye for the 10^{th} time on the 15^{th} try?

(b) hits the bullseye 10 times in 15 tries?

(c) hits the first bullseye on the third try?

```
prob_a <- dbinom(x = 9, size = 14, prob = 0.65) * 0.65
prob_b <- dbinom(x = 10, size = 15, prob = 0.65)
prob_c <- 0.65 * (1 - 0.65)^2
```

Ans:

```
(a):  0.1415591
(b):  0.2123387
(c):  0.079625
```

- 4.32

4.32 Stenographer's typos. A very skilled court stenographer makes one typographical error (typo) per hour on average.

(a) What probability distribution is most appropriate for calculating the probability of a given number of typos this stenographer makes in an hour?

(b) What are the mean and the standard deviation of the number of typos this stenographer makes?

(c) Would it be considered unusual if this stenographer made 4 typos in a given hour?

(d) Calculate the probability that this stenographer makes at most 2 typos in a given hour.

Ans:

(a) Poisson distribution

(b) mean: 1, standard deviation: 1(parameter λ = 1)

(c) Yes, dpois(4, 1) = 0.01532831, it's quite low.

(d) ppois(2, 1) = 0.9196986

- 4.35

4.35 Roulette winnings. In the game of roulette, a wheel is spun and you place bets on where it will stop. One popular bet is that it will stop on a red slot; such a bet has an 18/38 chance of winning. If it stops on red, you double the money you bet. If not, you lose the money you bet. Suppose you play 3 times, each time with a \$1 bet. Let Y represent the total amount won or lost. Write a probability model for Y.

Ans:

$$P(Y=3)=(18/38)^3=0.1063$$
$$P(Y=1)=3*(18/38)^2*(20/38)=0.3543$$
$$P(Y=-1)=3*(18/38)*(20/38)^2=0.3936$$
$$P(Y=-3)=(20/38)^3=0.1458$$

- 4.36

4.36 Speeding on the I-5, Part I. The distribution of passenger vehicle speeds traveling on the Interstate 5 Freeway (I-5) in California is nearly normal with a mean of 72.6 miles/hour and a standard deviation of 4.78 miles/hour.[40]

(a) What percent of passenger vehicles travel slower than 80 miles/hour?

(b) What percent of passenger vehicles travel between 60 and 80 miles/hour?

(c) How fast do the fastest 5% of passenger vehicles travel?

(d) The speed limit on this stretch of the I-5 is 70 miles/hour. Approximate what percentage of the passenger vehicles travel above the speed limit on this stretch of the I-5.

```r
p_slower_than_80 <- pnorm(80, mean = 72.6, sd = 4.78)
p_between_60_and_80 <- pnorm(80, mean = 72.6, sd = 4.78) - pnorm(60, mean = 72.6, s
fastest_5_percent <- qnorm(0.95, mean = 72.6, sd = 4.78)
p_above_70 <- 1 - pnorm(70, mean = 72.6, sd = 4.78)

cat("(a): ", p_slower_than_80 * 100, "\n")
cat("(b): ", p_between_60_and_80 * 100, "\n")
cat("(c): ", fastest_5_percent, "\n")
cat("(d): ", p_above_70 * 100, "\n")
```

Ans:

```
(a):  93.9203
(b):  93.50083
(c):  80.4624
(d):  70.67562
```

- 4.39

4.39 Auto insurance premiums. Suppose a newspaper article states that the distribution of auto insurance premiums for residents of California is approximately normal with a mean of \$1,650. The article also states that 25% of California residents pay more than \$1,800.

(a) What is the Z-score that corresponds to the top 25% (or the 75^{th} percentile) of the standard normal distribution?

(b) What is the mean insurance cost? What is the cutoff for the 75th percentile?

(c) Identify the standard deviation of insurance premiums in California.

```
z_score_75th_percentile <- qnorm(0.75)
mean_insurance_cost <- 1650
sd_insurance_premiums <- (1800 - mean_insurance_cost) / z_score_75th_percentile
cutoff_75th_percentile <- mean_insurance_cost + z_score_75th_percentile * sd_insura

cat("(a): ", z_score_75th_percentile, "\n")
cat("(b): ", cutoff_75th_percentile, "\n")
cat("(c): ", sd_insurance_premiums, "\n")
```

Ans:

```
(a):  0.6744898
(b):  1800
(c):  222.3903
```