R Lab: Large Sample CI (solutions)

STAT 343: Mathematical Statistics

Spatial Organization of Chromosome (Rice Problem 8.45)

The R code below reads in the data and calculates the maximum likelihood estimate:

```
library(readr)
library(ggplot2)
chromatin <- read_csv("http://www.evanlray.com/data/rice/Chapter%208/Chromatin/data05.txt",
  col_names = FALSE)
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   X1 = col_double()
## )
```

```
colnames(chromatin) <- "distance"
theta_hat <- 1/(2 * nrow(chromatin)) * sum(chromatin$distance^2)
theta_hat
```

```
## [1] 4.527792
```

1. Find the variance of a large-sample normal approximation to the sampling distribution of the MLE based on the observed Fisher information. (You should be able to calculate a number.)

In our in-class example, we found that

$$J(\theta_0) = -\left[\frac{n}{\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n} x_i^2\right]\Bigg|_{\theta=\theta_0}$$

Here we are asked to find the variance of the MLE based on the observed Fisher information. Our best guess for θ is $\hat{\theta}^{MLE}$ here, so we will plug that in for θ:

$$J\left(\hat{\theta}^{MLE}\right) = -\left[\frac{n}{\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n} x_i^2\right]\Bigg|_{\theta=\hat{\theta}^{MLE}}$$

We also know that

$$Var\left(\hat{\Theta}^{MLE}\right) = \frac{1}{J\left(\hat{\theta}^{MLE}\right)}$$

```
est_var_theta_hat_J <- 1 / (1/theta_hat^3 * sum(chromatin$distance^2) - nrow(chromatin) / theta_hat^2)
est_var_theta_hat_J
```

```
## [1] 0.1151736
```

2. Find the variance of a large-sample normal approximation to the sampling distribution of the MLE based on the Fisher information. (You should be able to calculate a number.)

In our in-class example, we found that

$$I(\theta_0) = \frac{n}{\theta^2}\bigg|_{\theta=\theta_0}$$

Our best guess for θ here $\hat{\theta}^{MLE}$, so we will plug that in for θ:

$$I\left(\hat{\theta}^{MLE}\right) = \frac{n}{\left(\hat{\theta}^{MLE}\right)^2}$$

The inverse of the information will give the variance:

```
est_var_theta_hat_I <- theta_hat^2 / nrow(chromatin)
est_var_theta_hat_I
```

```
## [1] 0.1151736
```

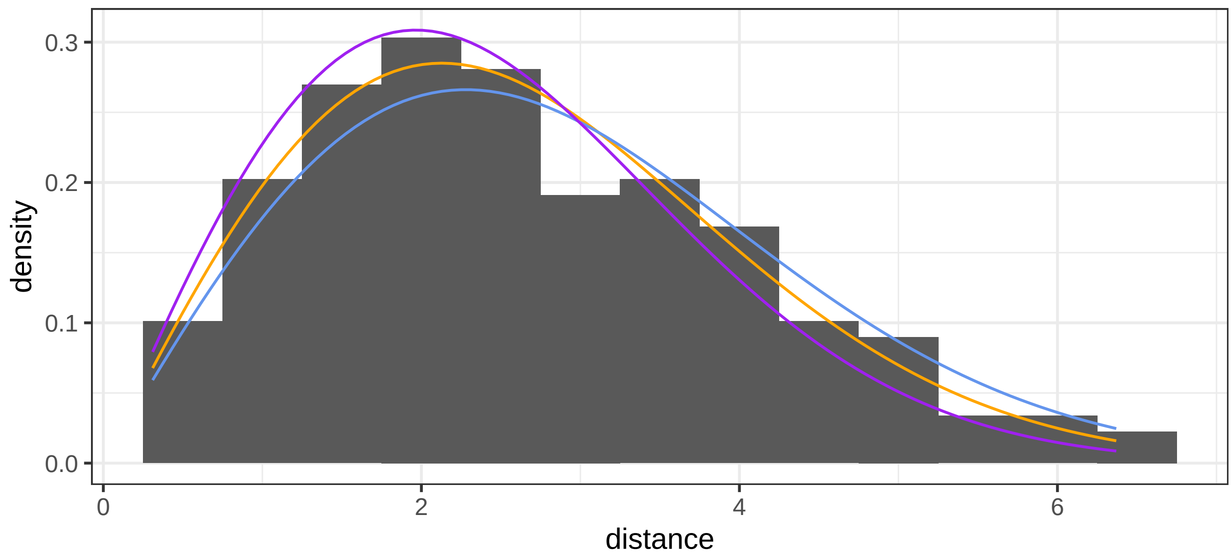
3. The plot below shows a representation of the pdf of the Rayleigh(θ) distribution based on the maximum likelihood estimate of θ. Add two more curves corresponding to the values of θ at the upper and lower end points of an approximate 95% confidence interval for θ based on the Fisher information from part 2.

R does not come with a function to calculate the pdf of the Rayleigh distribution, so I have defined one below for you to use.

```
drayleigh <- function(x, theta) {
  log_value <- rep(-Inf, length(x))
  inds <- which(x > 0)

  log_value[inds] <- log(x[inds]) - log(theta) - (x[inds]^2)/(2 * theta)

  return(exp(log_value))
}
ggplot(data = chromatin, mapping = aes(x = distance)) +
  geom_histogram(binwidth = 0.5, mapping = aes(y = ..density..)) +
  stat_function(fun = drayleigh, args = list(theta = theta_hat), color = "orange") +
  stat_function(fun = drayleigh, args = list(theta = theta_hat - qnorm(0.025) * sqrt(est_var_theta_hat_I)),
                color = "cornflowerblue") +
  stat_function(fun = drayleigh, args = list(theta = theta_hat - qnorm(0.975) * sqrt(est_var_theta_hat_I)),
                color = "purple") +
  theme_bw()
```


density
0.0
0.1
0.2
0.3
distance
0
2
4
6