

# Math 152 - Statistical Theory - Homework 4

write your name here

Due: Thursday, September 22, 11:59pm

## Important Note:

You should work to turn in assignments that are clear, communicative, and concise. Part of what you need to do is not print pages and pages of output. Additionally, you should remove these exact sentences and the information about HW scoring below.

Click on the *Knit to PDF* icon at the top of R Studio to run the R code and create a PDF document simultaneously. [PDF will only work if either (1) you are using R on the network, or (2) you have LaTeX installed on your computer. Lightweight LaTeX installation here: <https://yihui.name/tinytex/>]

Either use the college's RStudio server (<https://rstudio.pomona.edu/>) or install R and R Studio on to your personal computer. See: <https://research.pomona.edu/johardin/math152f20/> for resources.

## Assignment

**Goals:** In this assignment, the fun will include:

- using calculus (e.g., derivatives) to find maximum likelihood estimators
- using constraints (e.g., bounds) to limit maximum likelihood estimators
- reflecting on when maximum likelihood estimators don't exist

## Book problems

- Feel free to do the book problems with a pencil or in LaTeX (RMarkdown supports writing mathematics using LaTeX).
- If you use a pencil, you can take a picture of the problem(s), and include the image(s) using (remove the tick marks to make it work):

![] (myimage.jpeg)

- Note that myimage.jpeg needs to live in the same folder as the relevant .Rmd file (maybe you called the folder "math 152 hw" and put it on your desktop?)
- Saving as jpg, jpeg, png, or pdf should work, but make sure to specify the exact name of the file.
- If you have the 3rd edition of the book, the problems will be the same unless they don't exist – that is, the 4th edition *added* problems but didn't change the order of them. Ask me if you want to see the 4th edition problems.

## Assignment

**1: Community Q** Describe one thing you learned (not from lecture, maybe from working in pairs during class) from a member of the class (student, mentor, professor) – it could be: content, logistical help, background material, R information, etc. 1-3 sentences.

**2: 7.5.2** It is not known what proportion  $p$  of the purchases of a certain brand of breakfast cereal are made by women and what proportion are made by men. In a random sample of 70 purchases of this cereal, it was found that 58 were made by women and 12 were made by men. Find the MLE of  $p$ .

**3: 7.5.3** Consider again the conditions in Exercise 2, but suppose also that it is known that  $\frac{1}{2} \leq p \leq \frac{2}{3}$ . If the observations in the random sample of 70 purchases are as given in Exercise 7.5.2, what is the MLE of  $p$ ?

**4: 7.5.5** Suppose that  $X_1, \dots, X_n$  form a random sample from a Poisson distribution for which the mean  $\theta$  is unknown, ( $\theta > 0$ ).

a. Determine the MLE of  $\theta$ , assuming that at least one of the observed values is different from  $\theta$ . b. Show that the MLE of  $\theta$  does not exist if every observed value is 0.

**5: 7.5.6** Suppose that  $X_1, \dots, X_n$  form a random sample from a normal distribution for which the mean  $\mu$  is known, but the variance  $\sigma^2$  is unknown. Find the MLE of  $\sigma^2$ . (Hint: the problem will be much easier if you set  $\sigma^2 = \theta$  and then take the derivative with respect to  $\theta$  instead of with respect to  $\sigma^2$ .)

**6: 7.5.11** Suppose that  $X_1, \dots, X_n$  form a random sample from the uniform distribution on the interval  $[\theta_1, \theta_2]$ , where both  $\theta_1$  and  $\theta_2$  are unknown ( $-\infty < \theta_1 < \theta_2 < \infty$ ). Find the MLEs of  $\theta_1$  and  $\theta_2$ .

**7: R - MLE of the Cauchy distribution** We get to use R as a way to understand a very wonky MLE!!! Isn't it cool that we can use the computer to understand situations that don't have easy closed form solutions?

Example 7.6.5 in the text looks at the MLE for the center of of a Cauchy distribution. The Cauchy distribution is interesting because the tails decay at a rate of  $1/x^2$ , so that when you try to take the expected value of  $X$ , you end up integrating something that looks like  $1/x$  over the real line. Hence, the expected value **does not exist**. Thus, method of moments estimators for  $\theta$  are of no use. The MLE of  $\theta$  is still useful, though not easy to find.

$$f(x_i|\theta, \sigma) = \frac{1}{\pi\sigma} \cdot \frac{1}{1 + (\frac{x_i - \theta}{\sigma})^2}$$

As stated in the text, the likelihood is proportional to

$$f(\underline{x}|\theta) \propto \prod_{i=1}^n [1 + (x_i - \theta)^2]^{-1}$$

- Compute the first and second derivative of the **log** likelihood. (I've specifically left off the information you might find useful: derivative with respect to *what*? If you don't know, think about it, look back at the notes, talk to your peers in mentor sessions!)
- Consider trying to find the root of a function  $h(x)$ . Suppose your current guess is some value  $x_0$ . We might approximate the function by the tangent line (first order Taylor approximation) at  $x_0$  and take our next guess as the root of that line. Use the Taylor expansion to find the next guess of  $x_1$  as

$$x_1 = x_0 - \frac{h(x_0)}{h'(x_0)}$$

Continually updating the guess via this method is known as Newton's Method or the Newton-Raphson Method.

Using your work from a., find  $x_1$  as a function of  $x_0$ .

- Generate 50 observations from a Cauchy distribution centered at  $\theta = 10$ . Based on these 50 observations, use parts a. and b. to estimate  $\theta$  with a maximum likelihood approach. Remember, we're trying to maximize the likelihood, so the function that we are trying to find the root of is derivative of the log-likelihood.

What is your estimate of  $\theta$ ? Include plot(s) of your guesses and your final value for the MLE of  $\theta$ .

Note: Once you have completed a. and thought about c., feel free to message me for help on the R syntax. If you tell me exactly what you need and the R error you are getting (i.e., why it doesn't work), I will tell you how to fix the R code.

The R code might look something like this:

where the actual formulas for l1 and l2 will come from part a. Everything in double parentheses needs to be replaced by proper R syntax. The rest of it will run in R. In case you are curious, `sum()` is the R function for summing a vector of values.

```
library(tidyverse)

# Use Newton's method to find the MLE for a Cauchy distribution
cauchy_mle <- function(guess, data){
  l1 <- ((compute 1st derivative of log-likelihood evaluated at "guess")) # 1st deriv
  l2 <- ((compute 2nd derivative of log-likelihood evaluated at "guess")) # 2nd deriv
  guess <- guess - l1/l2
  guess
}

# set up the initial conditions (including the dataset to use)
set.seed(54321) # to get the same answer each time the .Rmd is knit

n_obs <- 50
observations <- rcauchy(n_obs, location = 10) # 50 random Cauchy, centered at 10
theta_guess <- ((pick a starting value, you might try different ones)) # just a number
```

**Running `cauchy_mle` recursively** Run the function repeatedly:

```
theta_guess <- ((pick a starting value, you might try different ones)) # just a number
guess_vec <- numeric()
reps <- 10 # play around with how many times you loop through.
# you probably don't need very many times through to see what happens.
# what happens might seem unsettling!

for(i in 1:reps){
  theta_guess <- cauchy_mle(____, ____ ) # what changes for each rep?
  guess_vec[i] <- theta_guess
}
guess_vec

data.frame(guess_vec) %>% # ggplot needs a data frame instead of a vector
  ggplot(aes(y = guess_vec, x = 1:reps)) +
  geom_line() # line plot connecting the guesses for each rep
```