Math 152 - Statistical Theory - Homework 4

write your name here

Due: Friday, September 18, 2020, midnight PDT

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

1: PodQ

Describe one thing you learned from someone in your pod this week (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 $X_1, \ldots, X_n$ form a random sample from a Poisson distribution for which the mean θ is unknown, $(\theta > 0)$.
a. Determine the MLE of θ, assuming that at least one of the observed values is different from θ. b. Show that the MLE of θ does not exist if every observed value is 0.

5: 7.5.6

Suppose that $X_1, \ldots, X_n$ form a random sample from a normal distribution for which the mean μ is known, but the variance σ^2 is unknown. Find the MLE of σ^2.

6: 7.5.11

Suppose that $X_1, \ldots, X_n$ form a random sample from the uniform distribution on the interval $[\theta_1, \theta_2]$, where both θ_1 and θ_2 are unknown ($-\infty < \theta_1 < \theta_2 < \infty$). Find the MLEs of θ_1 and θ_2.

7: R - MLE of the Cauchy distribution

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 θ are of no use. The MLE of θ is still useful, though not easy to find.

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

As stated in the text, the likelihood is proportional to

$$\prod_{i=1}^{n} [1 + (x_i - \theta)^2]^{-1}$$

a. 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 mentor pod!)
b. Consider trying to find the root of a function f(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

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

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.

c. Generate 50 observations from a Cauchy distribution centered at $\theta = 10$. Based on these 50 observations, use parts a. and b. to estimate θ 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.

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:

```
x=rt(50,1)+10  # 50 random Cauchy variables centered at 10
vect.theta = c()  # placeholder
theta.guess=((pick a starting value, you might try different ones))  # just a number


for (i in 1:10) {  # play around with how many times you loop through.  10 is likely too small.

    f1=((compute first derivative of log-likelihood evaluated at theta.guess))
    f2=((compute second derivative at theta.guess))
```

```
# f1 and f2 need to be written as R functions of theta.guess

    theta.guess=theta.guess - f1/f2
    print(theta.guess)
    vect.theta = c(vect.theta, theta.guess) # keeping track of all your guesses
}

plot(vect.theta)
```

where the actual formulas for f1 and f2 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.

Include plot of your guesses and your final value for the MLE of θ.