Math 152 - Statistical Theory - Homework 5

write your name here

Due: Friday, September 25, 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: 7.6.5

Suppose that $X_1, \ldots, X_n$ form a random sample from the uniform distribution on the interval [a, b], where both endpoints a and b are unknown. Find the MLE of the mean of the distribution.

3: 7.6.6

Suppose that $X_1, \ldots, X_n$ form a random sample from a normal distribution for which both the mean and the variance are unknown. Find the MLE of the 0.95 quantile of the distribution, that is, of the point θ such that $P(X < \theta) = 0.95$.

4: 7.6.8

Suppose that $X_1, \ldots, X_n$ form a random sample from a gamma distribution for which the pdf is given by

$$f(x|\alpha) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \quad x > 0$$

(Eq. (7.6.2)) . Find the MLE of the ratio $\Gamma'(\alpha)/\Gamma(\alpha)$.

5: 7.6.9

Suppose that $X_1, \ldots, X_n$ form a random sample from a gamma distribution for which both parameters α and β are unknown. Find the MLE of α/β. Hint: use the Gamma distribution pdf from the back of the book. Also, for this particular problem, only one derivative is necessary.

6. 7.6.21

Prove that the method of moments estimators of the mean and variance of a normal distribution are also the MLEs.

7: 7.6.23

Suppose that $X_1, \ldots, X_n$ form a random sample from the beta distribution with parameters α and β. Let $\theta = (\alpha, \beta)$ be the vector parameter.
a. Find the method of moments estimator for θ.
b. Show that the method of moments estimator is not the MLE. (Note: you do not need to actually solve for the MLE.)

8: R - Tanks

Consider the tank problem encountered in class. Your task at hand is to provide the best possible estimator for the true number of tanks. Consider the R code below that analyzes two estimators ($2\overline{X}$ and the maximum of the sample). You should **provide an argument (in words but using the evidence collected here) for your choice of estimator** using the following information:

- a comparison of at least 5 estimators (of your choice!)
- consideration of sample bias, sample variance, sample median, sample mean, sample MSE
- run the entire analysis twice. Once with a sample size of n=5, once with a sample size of n=100

Hint on the R code: create new estimators one by one by adding them on to the two which I wrote.

1. make sure your new estimator has a place holder
2. use the `c()` function inside the `for()` loop to keep the computed estimator for each sample of tanks
3. `estimate` should now hold all the different estimates
4. ... same with `method` and finally `all.estimates`.
5. the rest of the code should just run.

```
# Keep the population size at 447
npop = 447
nsamp = 5  # change this to 100 for the second part of the analysis
reps = 10000

xbar2 = c()  # placeholder for your repeated sample statistics
sampmax = c()

for (i in 1:reps){
  mysample =  sample(1:npop,nsamp,replace=F)  # sample some tanks from the population
    xbar2 = c(xbar2, 2*mean(mysample) )
  sampmax = c(sampmax, max(mysample))
}


estimate <- c(xbar2, sampmax)
method <- c(rep("2xbars", reps), rep("samplemax", reps))
all.estimates <- data.frame(estimate, method)


# below is some syntax with which you may be unfamiliar.
# if you don't have the packages, you may need to run, for example:
# install.packages("tidyverse")

# also, if you create the statistics as above, all of the code below should
```

```r
# work nicely without you doing anything to it.

# please ask if you have question!


all.estimates %>%
  group_by(method) %>%
  summarize(mean = mean(estimate), median = median(estimate), bias = mean(estimate) - npop,
            var = var(estimate), mse = (mean(estimate) - npop)^2 + var(estimate))
```

```
## # A tibble: 2 x 6
##   method     mean median   bias    var    mse
##   <chr>     <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 2xbars     448.   450.   1.49 13408. 13410.
## 2 samplemax  374.   389  -73.3   3884.  9250.
```

```r
ggplot(all.estimates, aes(x = estimate)) + geom_histogram() +
  geom_vline(xintercept = npop) + facet_wrap(~method)
```