Math 152 - Statistical Theory - Homework 9

write your name here

Due: Friday, October 23, 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: 8.8.4

Suppose that a random variable has the normal distribution with mean 0 and unknown standard deviation $\sigma > 0$. Find the Fisher information $I(\sigma)$ in X.

3: 8.8.5

Suppose that a random variable X has the normal distribution with mean 0 and unknown variance $\sigma^2 > 0$. Find the Fisher information $I(\sigma^2)$ in X. Note that in this exercise the variance σ^2 is regarded as the parameter, whereas in Exercise 4 the standard deviation σ is regarded as the parameter.

Also, show that the unbiased estimator of σ^2 is not efficient.

Note to self: there is no simple way to get from problem 8.8.4 to 8.8.5. That is, we know the MLE of a function is that function of the MLE. Not true for information.

4: 8.8.16

Suppose that $X_1, \ldots, X_n$ form a random sample from the Bernoulli distribution with unknown parameter p, and the prior pdf of p is a positive and differentiable function over the interval $0 < p < 1$. Suppose, furthermore, that n is large, the observed values of $X_1, \ldots, X_n$ are $x_1, \ldots, x_n$, and $0 < \overline{x} < 1$. Show that the posterior distribution of p will be approximately a normal distribution with mean $\overline{x}$ and variance $\overline{x}(1 - \overline{x})/n$.

Note to self: this is a Bayesian problem. See the connection between Fisher Information and asymptotic / approximate Bayesian distributions in the text.

5: 8.9.15

Suppose that $X_1, \ldots, X_n$ form a random sample from a distribution for which the pdf is as follows:

$$f_x(x|\theta) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where the value of θ is unknown ($\theta > 0$). Determine the asymptotic distribution of the MLE of θ. (Note: The MLE was found in Exercise 9 of Sec. 7.5.)

6: 8.9.18

Suppose that $X_1, \ldots, X_n$ form a random sample from the exponential distribution with unknown parameter β (use the book's parametrization of the exponential distribution). Construct an efficient estimator that is not identically equal to a constant, and determine the expectation and the variance of this estimator.

7: R - $N(\theta, \theta^2)$

We are going to study almost, but not exactly, the same model as from the exam. The model for this problem is normal with mean and standard deviation both θ (i.e., variance θ^2, not θ as in the example from the exam). Therefore, we know $\theta \geq 0$.

Note that the Fisher Information in θ is $3/\theta^2$.

The results from class about properties of MLEs are asymptotic. What happens in small samples?

The estimators of θ we wish to compare are:

- the sample median
- the sample mean
- the sample standard deviation times the sign of the sample mean
- the MLE

a. Show (with pencil / LaTeX) that the MLE of θ is

$$\hat{\theta} = -\overline{x}/2 + \sqrt{\left(\sum x_i^2\right)/n + \overline{x}^2/4}$$

.

b. Use a simulation to compare the four estimators above with respect to bias, variance, and MSE. Answer the following questions in your comparison:

   i. Which estimator is (empirically) least biased?
   ii. Which estimator has lowest empirical variability? Do any of the estimators reach the CRLB (assume unbiasedness)?
   iii. Which estimator has lowest empirical MSE?
   iv. Are you comfortable with the idea of using a normal distribution to describe the sampling distribution for any/all of the estimators? Explain.
   v. Which estimator would you recommend for the given setting? Explain.

```r
# Use sample size n = 15. keep this set-up in the first 3 lines
n.obs <- 15
n.samps <- 10000
theta <- exp(1)
```

```r
means <- numeric(n.samps)
medians <- numeric(n.samps)
sds <- numeric(n.samps)
MLEs <- numeric(n.samps)


for (i in 1:n.samps){
  # generate some data
  # means[i] <- mean(the data you generated)
  # etc.
}
```

You can write alternative code for calculating relevant characteristics of the distribution and displaying it, but I choose to put it together in a tidy framework like this:

Note: in the code below I've created smoothed histograms (called density plots) so as to plot the empirical distributions on top of one another.

```r
library(tidyverse)

est.info <- data.frame(value = c(means , ___, ...),
                       type = c(rep("mean", n.samps), ____, ...) )

est.info %>%
  group_by(type) %>%
  summarize(est.bias = mean(value) - theta, est.mean = mean(value),
            est.var = var(value), est.sd = sd(value)) %>%
  mutate(est.mse = est.var + est.bias^2)


est.info %>%
  ggplot(aes(x=value, color = type)) + geom_density() +
  geom_vline(xintercept = exp(1))
```