

Math 152 - Statistical Theory - Homework 9

write your name here

Due: Thursday, November 10, 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://m152-stat-theory.netlify.app/syllabus.html> for resources.

Assignment

Goals: In this assignment, the fun will include:

- calculating Fisher information
- setting up hypotheses
- calculating power and error rates.

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: 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 8.8.4 the standard deviation σ is regarded as the parameter.

Also, show that the unbiased estimator of σ^2 $\left(s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right)$ 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, \dots, 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, \dots, X_n are x_1, \dots, x_n , and $0 < \bar{x} < 1$. Show that the posterior distribution of p will be approximately a normal distribution with mean \bar{x} and variance $\bar{x}(1 - \bar{x})/n$.

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, \dots, 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 .

6: 9.1.1 Let X have the exponential distribution with parameter β . Suppose that we wish to test the hypotheses $H_0 : \beta \geq 1$ versus $H_1 : \beta < 1$.

$$f(x|\beta) = \beta e^{-\beta x} \quad x > 0$$

Consider the test procedure δ that rejects H_0 if $X \geq 1$.

- Determine the power function of the test.
- Compute the size of the test.

7: 9.1.6 Suppose that a single observation X is to be taken from the uniform distribution on the interval $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$, and suppose that the following hypotheses are to be tested:

$$\begin{aligned} H_0 &: \theta \leq 3, \\ H_1 &: \theta \geq 4. \end{aligned}$$

Construct a test procedure δ for which the power function has the following values: $\pi(\theta|\delta) = 0$ for $\theta \leq 3$ and $\pi(\theta|\delta) = 1$ for $\theta \geq 4$.

8: 9.1.9 Assume that X_1, \dots, X_n are i.i.d. with the normal distribution that has mean μ and variance 1. Suppose that we wish to test the hypotheses:

$$\begin{aligned} H_0 &: \mu \geq \mu_0, \\ H_1 &: \mu < \mu_0. \end{aligned}$$

Find a test statistic T such that, for every c , the test δ_c that rejects H_0 when $T \geq c$ has power function $\pi(\mu|\delta_c)$ that is decreasing in μ .

9: 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 $\theta \geq 0$.

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. The MLE of θ is

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

. Show (with pencil / LaTeX) that the Fisher Information in θ is $3/\theta^2$.

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.

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

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 from the given model
  # 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.

```
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))
```