HOMEWORK 3
Due Friday, Oct 23, start of class
This homework covers Chapters 6 and 7. You should be working on your homework throughout these two weeks. If you can’t solve some of the problems, please come to office hours. Email is fine only for very short questions.

THEORETICAL PORTION
The theoretical problems should be neatly numbered, written out, and solved. Do not turn in messy work.

1. At time $t = 0$, a lab technician starts an experiment in which the lifetime of 18 identical, independent components are tested over a 48 hour period. The lifetime distribution of each component is exponential with mean lifetime of $w$ hours. The technician leaves the test facility and comes back in 48 hours to find 12 components still running (and 6 of them failed). Based on this sample:
   (a) Define the log-likelihood function.
   (b) In your log-likelihood function, replace the probability that a component will fail within 48 hours with $p_F$, and then find the MLE for $p_F$.
   (c) Use your answer from part (a) to find the MLE of $w$ (Hint: Use the invariance property of the MLE to solve for $w$.)

2. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.
   (a) Show (without integration), that $\text{Var}(\bar{X}) = \sigma^2/n$.
   (b) What is $E(\bar{X}^2)$?

3. Let $X \sim N(\mu_x, \sigma^2_x)$, and $Y \sim N(\mu_y, \sigma^2_y)$. Assume X and Y are independent. Show all steps in solving the problems below:
   (a) What is $E(X+Y)$? $\text{Var}(X+Y)$?
   (b) What is the distribution of $X+Y$ and how do you know?
   (c) What is the distribution of $5X+0.5Y+2$? Include not only the type of distribution but also the values of any parameters of that distribution.

4. APPM 5570 students only: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Uniform}(a, b)$. (i.e. $f(x) = \frac{1}{b-a}$.)
   (a) Write down the log-likelihood function for parameters $a$ and $b$.
   (b) Can you use the log-likelihood to obtain the MLE for $a$ and $b$?
   (c) Show that the MOM estimator for $a = \bar{X} - \sqrt{3 (1/n \sum X_i^2 - \bar{X}^2)}$, and the MOM estimator for $b = \bar{X} + \sqrt{3 (1/n \sum X_i^2 - \bar{X}^2)}$

COMPUTATIONAL PORTION
The computational portion of your homework should be neatly done and include all graphs, code, and comments, labeled and in order based on the problem you are addressing. Do not put graphs in at the end, stick code in random locations, or do anything else that will make this homework difficult to read and grade. LABELS ARE YOUR FRIEND, USE THEM. If you turn in something that is messy or out of order, it will be returned to you with a zero. All computations should be done using R, which can be downloaded for free at https://cran.r-project.org/.

1. Aphid infestation of fruit trees is usually controlled via pesticides or via ladybug inundation. In a particular area, 2 different (and well isolated) groves, with 15 fruit trees each, are selected for an experiment. The trees in both groves are of the same age, roughly the same size and can be assumed to be independent. One grove is sprayed with pesticides, and one is infested with ladybugs. The fruit yield (in pounds) for each tree is given below:
   Treatment #1, Grove with pesticide:
   55.57109, 36.50319, 47.80090, 33.34822, 36.16251,
   35.28337, 41.50154, 44.18931, 40.81439, 33.88648,
   44.90427, 49.97089, 22.85414, 27.84301, 38.49843
Treatment #2, Grove with ladybugs:

45.44505, 35.52320, 46.97865, 45.76921, 41.66216,
54.69599, 58.77678, 49.08538, 48.53812, 70.17137,
51.86253, 39.59365, 42.10194, 47.39945, 39.04648

You can read in this data using the “HW3TreeData.txt” data set.

(a) Plot a relative frequency histogram of the yields in the two groves. Make sure both histograms have the same range on the x-axis.

(b) Comment on the histogram shapes. Which densities do they resemble? In particular do they appear normal?

(c) Find the sample mean of yields for the two groves.

(d) Assuming both samples come from a normal population and using your answer from part (c), provide the two 95% confidence intervals for the true mean yields for trees under the two treatments.

(e) Interpret the confidence interval you constructed for the grove treated with pesticides (i.e. what does a confidence interval really signify?)

(f) Find(or approximate) the sample mean variance, i.e. variance of the sample mean, $\overline{X}$, for the yields in the two groves.

(g) Using a chi-square distribution, construct the 95% confidence interval for the true variance of yield for both groves.

2. Generate 500 samples (each with $n = 300$) from an exponential distribution with a mean equal to 1/2. Retain the entire first sample and store it in “expSample”, and then in “means.of.expSamples” retain only the means of the remaining 499 samples.

(a) Report the mean and standard deviation for “expSample” and “means.of.expSamples”.

(b) Create a histogram for “expSample” and for “means.of.expSamples”.

Both of these data sets originate from the same distribution. Why are they so different? What do you notice about the the means and the standard deviations of the samples?

3. Make a function to calculate a confidence interval, and call it CIcalculator(). The function should allow the user to specify the following in the function call:

(a) “interval.type”, which takes either “p” for a CI for a proportion, or “mu” for a CI for the mean.

(b) “alpha.value”, which sets the $\alpha$-level for the confidence interval.

(c) “data.values”, which is either a vector of numbers if the confidence interval is for the mean, or is a percentage (i.e. single number) if the confidence interval is for a proportion.

(d) “n”, a single number for the sample size of the data set.

In addition, CIcalculator() should also determine if $z$ or $t$ should be used in the calculation. You can assume the data being entered is normally distributed.

If you would like, you can double check your results of your calculations for the mean by comparing them to CI(), found in in the Rmisc package in R. You must email me your function by the time class starts on the homework due date. Only send the function, no other code. I will not do any de-bugging, so be sure it works. This is the line of code I will use to check your function:

CIcalculator(interval.type = "p", alpha.value = 0.10, data.values = 0.42, n = 150)

with variations on the alpha-values, percentages (or vector of data for the mean confidence interval), and $n$.

4. APPM 5570 students only: Refer back to problem 4 in the theoretical section. Generate 1000 samples, each of size 200, from a Uniform distribution with parameter $a = 2$ and parameter $b = 5$. Within each of the 1000 samples, keep track of:
• The MOM estimators for $a$ and $b$.
• The MLEs for $a$ and $b$. (*HINT: For the Uniform distribution, the MLE for $a$ is $X_{(1)}$, the smallest value in the sample, and the MLE for $b$ is $X_{(200)}$, the largest value in the sample.*)

After you have sampled all the data and recorded all the parameter estimates, do the following:

(a) Create four histograms, one for the MLE of each parameter and one for the MOM estimator of each parameter. What do you notice about these histograms?

(b) Calculate the mean of the estimates obtained through random sampling (do each parameter and each method separately). Does either estimator (MOM or MLE) appear to be unbiased?

(c) Calculate the mean-square error (MSE) for each of the four samples. The MSE for $\hat{b}$, the MOM estimate of $b$, is equal to
\[
\frac{1}{1000} \sum_{i=1}^{1000} (\hat{b}_i - \bar{b})^2.
\]
Smaller values of the MSE are better. Does either estimator appear to be “better” on the MSE scale?

(d) Which estimator would you pick and why?