Overview of estimation of parameters:

When we estimate a parameter, we want to

- Find an estimator (a function of the data, but not the parameter)

which meets the following criteria:

- The values of the estimator are, on the average, about the same as the parameter .
- The variability of the values are as small as is reasonable/possible.

These are mathematical statistics topics, because, to do them we need multivariate distributions and/or the distribution OF THE STATISTIC, not just the distribution of the values of the individual data points.

The information we have to find the distribution of the various possible statistics so far isn't much. We have the Central Limit Theorem, which gives us (for moderately large sample sizes,) for many distributions, that the sums and means are normally distributed. We also have some work with order statistics in Chapter 8 (slides 21 and 26.)

The most straightforward of these is to find an estimator – a maximum likelihood estimator. For that we need the multivariate distribution, which is the product of the distributions of the individual independent variables. You are expected to be able to derive maximum likelihood estimators for whatever distribution of the data you are given.

What is more difficult in most situations is to find the distribution of the summary statistic in order to find the bias and the variability of the estimator.

BUT, we can approach this by simulation. For this, you need to program simulations yourself to simulate the sampling distribution of a statistic. With that distribution, you can simply look at the center and variability to understand much of the important information about how good the statistic is to estimate the parameter.

This week's work is to give you enough experience with the mathematical statistics part so that you can

- Derive maximum likelihood estimators for most parameters of most distributions AND for functions of the parameters.
- Understand what the expected value and variance of a summary statistic are, including why and how they are both important in assessing the usefulness of an estimator of the parameter.
- Become confident that you could create a simulation to investigate the expected value and variance of any summary statistic you can think of for any distribution that you can program. (In your work, most of you will investigate these much more often by simulation than by theoretical distributions.)

Homework 12.

Work out and discuss at least Part 1 problems 1-3 before you start the homework to be turned in. Problems 4-7 help reinforce the ideas we're covering here and also are the underpinnings of the theoretical methods of statistics we covered in the previous two weeks.

Part 1. Not to be turned in. Discuss on Discussion Board. (Unlike our typical homework, there are not short answers to these.)

1. Let $X_1, X_2, \ldots, X_n$ be independent random variables from a distribution with pdf $f(x\,|\,\theta) = \begin{cases} (\theta+1)\,x^{\theta} & \text{for } 0<x<1 \text{ and } \theta>-1 \\ 0 & \text{elsewhere} \end{cases}$

   Find the maximum likelihood estimator of θ. How do you know it gives a maximum?

For problems 2-7, consider $X_1, X_2, \ldots, X_n$ to be independent random variables from a $\text{Normal}\left(\mu, \sigma^2\right)$ where both parameters are unknown. (As you might imagine, the solutions to this are available in almost any mathematical statistics text.)

2. Find maximum likelihood estimators of both parameters. Call these two estimators $\hat{\mu}$ and $\hat{\theta}^2$ (You may skip the step of showing that the critical values are a maximum, because that's somewhat harder for these two-dimensional problems than I intend for you to try to do now.)

Recall the definition of Bias. $\text{Bias}(U) = E(U) - \theta$.

An unbiased estimator of a parameter θ is an estimator for which the bias is zero.

3. Is the MLE of μ an unbiased estimator? Show your work.
4. Show that the MLE of σ^2 is **not** an unbiased estimator of σ^2.
5. Find a linear function of that MLE of σ^2 which IS an unbiased estimator of σ^2.

   (If we call our MLE of the name $\hat{\theta}^2$, this says: find some constants *a* and *b* so that $E\left(a\cdot\hat{\theta}^2 + b\right) = \sigma^2$) We will call this estimator "theta-squared-tildle" $\tilde{\theta}^2$. (We also know it as S^2.)

   (One can't do that for all similar problems, but for this problem you can.)

   The estimator obtained here is one you are quite used to using. Do you recognize it? (We know it as S^2.)

6. What is the distribution of the estimator $\hat{\mu}$?
7. One can prove that $\dfrac{(n-1)\tilde{\theta}^2}{\sigma^2} \sim \chi^2_{n-1}$ (That proof is beyond the scope of this course. It is, however, the basis of the last half of the theoretical dist'ns on our statistical formulas document.)
8. Read and consider the "Enemy Tank Problem" project. This project is to illustrate the use of simulation instead of theoretical calculations to determine the variance and bias of estimators in order to determine whether one estimator is preferable to another.

Part 2. Work on these problems in the usual way you work on graded homework problems. Here are the three datasets to use for problems 9, 11, and 13. Each has 10 data values.

Problem 9	Problem 11	Problem 13
2.85	17.47	0.7
11.07	9.1	0.83
3.84	13.74	2.39
7.92	30.68	1.98
5.83	41.66	3.76
8.81	16.53	2.24
12.33	43.44	12.44
9.74	21.18	7.34
3.38	43.16	2.76
21.91	74.17	14.03

9. Let $X_1, X_2, \ldots, X_n$ be independent random variables from a distribution with pdf
$$f(x \mid \theta) = \begin{cases} \left(\dfrac{1}{\theta^2}\right) x e^{-x/\theta} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$
Find the maximum likelihood estimator of θ ~~if α is known.~~ Use the data given above to find the numerical value for our estimate of θ and then choose the closest number among the choices given.

   Choices: 3 6 9 12 15 18

10. The invariance property of MLEs tells us that the MLE of θ^2 is the square of the MLE of θ. (A very general result for all distributions and parameters.)

    For the distribution in the previous problem, using the same data as for that problem, find the numerical value for our estimate of θ^2. Then choose the closest number among the choices given.

    Choices: 4 9 25 121 200

11. Let $X_1, X_2, \ldots, X_n$ be independent random variables from a distribution

with pdf $f(x\,|\,\theta) = \begin{cases} \left(\dfrac{1}{\Gamma(\alpha)\theta^{\alpha}}\right) x^{\alpha-1} e^{-x/\theta} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$ where α is a known value.

Find the maximum likelihood estimator of θ. The data given above was generated for $\alpha = 3.7$. Find the numerical value for our estimate of θ and then choose the closest number among the choices given.

Choices 2 30 60 90

12. The invariance property of MLEs tells us that the MLE of $\sqrt{\theta}$ is the square root of the MLE of θ. (A very general result for all distributions and parameters.)

Find the numerical value for our estimate of $\sqrt{\theta}$ Then choose the closest number among the choices listed.

Choices: 1 2 3 4 5 6 7 8 9

13. Let $X_1, X_2, \ldots, X_n$ be independent random variables from a distribution

with pdf $f(x\,|\,\theta) = \begin{cases} \dfrac{1}{2\theta+1} & \text{for } 0 \le x \le 2\theta+1 \\ 0 & \text{elsewhere} \end{cases}$

Find the maximum likelihood estimator of θ. Find the numerical value for our estimate of θ using the data given above. Then choose the closet number among the choices listed.

Choices: 1 4 10 12 14

14. For the same distribution as in the previous question, the variance of the distribution is $\dfrac{(2\theta+1)^2}{12}$.

Find the MLE of the variance of this distribution for the same data as in the previous problem. Find the numerical value for our estimate of the variance. Then choose the closest number among the choices listed.

Choices: 10 12 14 16 20 25