## Sampling and Bootstrapping, MLE

Before you leave lab, make sure you click here so that you're marked as having attended this week's section. The CA leading your discussion section can enter the password needed once you've submitted.

## 1 Warmups

### 1.1 Sample and Population Mean

Computing the sample mean is similar to the population mean: sum all available points and divide by the number of points. However, sample variance is slightly different from population variance.

1. Consider the equation for population variance, and an analogous equation for sample variance.

$$
\begin{gather*}
\sigma^{2}=\frac{1}{N} \sum_{i=1}^{N}\left(x_{i}-\mu\right)^{2}  \tag{1}\\
S_{\text {biased }}^{2}=\frac{1}{n} \sum_{i=1}^{n}\left(X_{i}-\bar{X}\right)^{2} \tag{2}
\end{gather*}
$$

$S_{\text {biased }}^{2}$ is a random variable to estimate the constant $\sigma^{2}$. Because it is biased, $E\left[S_{\text {biased }}^{2}\right] \neq \sigma^{2}$. Is $E\left[S_{\text {biased }}^{2}\right]$ greater or less than $\sigma^{2}$ ?
2. Consider an alternative Random Variable, $S_{\text {unbiased }}^{2}$ (known simply as $S^{2}$ in class). The technique of un-biasing variance is known as Bessel's correction. Write the $S_{\text {unbiased }}^{2}$ equation.
a. $E\left[S_{\text {biased }}^{2}\right]<\sigma^{2}$. The intuition is that the spread of a sample of points is generally smaller than the spread of all the points considered together. This becomes more clear when we consider the unbiased version and how it makes the expression evaluate to a larger number.
b. $S_{\text {unbiased }}^{2}=S^{2}=\frac{1}{n-1} \sum_{i=1}^{n}\left(X_{i}-\bar{X}\right)^{2}$

### 1.2 MLE

Suppose $x_{1}, \ldots, x_{n}$ are iid (independent and identically distributed) values sampled from some distribution with density function $f(x \mid \theta)$, where $\theta$ is unknown. Recall that the likelihood of the data is

$$
L(\theta)=f\left(x_{1}, x_{2}, \ldots, x_{n} \mid \theta\right)=\prod_{i=1}^{n} f\left(x_{i} \mid \theta\right)
$$

Recall we solve an optimization problem to find $\hat{\theta}$ which maximizes $L(\theta)$, i.e., $\hat{\theta}=\arg \max _{\theta} L(\theta)$.

1. Write an expression for the $\log$-likelihood, $L L(\theta)=\log L(\theta)$.
2. Why can we optimize $L L(\theta)$ rather than $L(\theta)$ ?
3. Why might we optimize $L L(\theta)$ rather than $L(\theta)$ ?
a. $\sum_{i=1}^{n} \log f\left(x_{i} \mid \theta\right)$
b. Logarithms are strictly increasing functions. For any strictly increasing function $f$ and any function $g$, the following holds: $\arg \max g(x)=\arg \max f(g(x))$
c. Finding the max of a function requires you take derivatives. The original expression consists of a product of many functions of $\theta$. This leads to a difficult derivation because of the chain rule of calculus.
By taking the log, we convert the product into sums, and the derivative of the sum of many functions of $\theta$ is much easier to compute.
That is, $\frac{\partial}{\partial \theta} \sum_{i} f_{i}(\theta)=\sum_{i} \frac{\partial}{\partial \theta} f_{i}(\theta)$ which is easy.
However, $\frac{\partial}{\partial \theta} \prod_{i} f_{i}(\theta)=$ a complicated expression

## 2 Problems

### 2.1 Variance of Hemoglobin Levels

A medical researcher treats patients with dangerously low hemoglobin levels. She has formulated two slightly different drugs and is now testing them on patients. First, she administered drug A to one group of 50 patients and drug B to a separate group of 50 patients. Then, she measured all the patients' hemoglobin levels post-treatment. For simplicity, assume that all variation in the patient outcomes is due to their different reactions to treatment.

The researcher notes that the sample mean is similar between the two groups: both have mean hemoglobin levels around $10 \mathrm{~g} / \mathrm{dL}$. However, drug B's group has a sample variance that is $3(\mathrm{~g} / \mathrm{dL})^{2}$ greater than drug A's group. The researcher thinks that patients respond to drugs A and B differently.

Specifically, she wants to make the scientific claim that drug A's patients will end up with a significantly different spread of hemoglobin levels compared to drug B's.

You are skeptical. It is possible that the two drugs have practically identical effects and that the observed different in variance was a result of chance and a small sample size, i.e. the null hypothesis. Calculate the probability of the null hypothesis using bootstrapping. Here is the data. Each number is the level of an independently sampled patient:
Hemoglobin Levels of Drug A's Group $\left(S^{2}=6.0\right): 13,12,7,16,9,11,7,10,9,8,9,7,16,7,9,8$, $13,10,11,9,13,13,10,10,9,7,7,6,7,8,12,13,9,6,9,11,10,8,12,10,9,10,8,14,13,13,10$, 11, 12, 9

Hemoglobin Levels of Drug B's Group ( $S^{2}=9.1$ ): 8, $8,16,16,9,13,14,13,10,12,10,6,14,8$, $13,14,7,13,7,8,4,11,7,12,8,9,12,8,11,10,12,6,10,15,11,12,3,8,11,10,10,8,12,8,11,6$, $7,10,8,5$

Complete the prompts in this Colab notebook to investigate this question using bootstrapping.

You can visit this notebook to see how we fleshed out all of the different functions expected of you for this problem.

### 2.2 Parameter Estimation and Wealth Distribution

The broader field of economics also relies on likelihood estimation and parameter estimation. One continuous probability distribution-one with a long tail as $x$ approaches infinity-is used to model wealth inequality and the socioeconomic problems that stem from it. This probability distribution is given as:

$$
f(x \mid \omega)=\frac{\omega 3^{\omega}}{x^{\omega+1}}, \text { where } x \geq 3, \omega>1
$$

Assume that you've observed a sample of iid random variables ( $X_{1}, X_{2}, X_{3}, \ldots, X_{n}$ ), where each of the $X_{i}$ is modeled according to the above probability distribution function.
a. What is the log-likelihood function $L L(\omega)$ of the sample $\left(X_{1}, X_{2}, X_{3}, \ldots, X_{n}\right)$ ? Simplify using properties of logarithms wherever possible.

$$
\begin{aligned}
L(\omega) & =\prod_{i=1}^{n} \frac{\omega 3^{\omega}}{x_{i}^{\omega+1}} \\
L L(\omega) & =\sum_{i=1}^{n} \log \frac{\omega 3^{\omega}}{x_{i}^{\omega+1}} \\
& =\sum_{i=1}^{n} \log \omega+\sum_{i=1}^{n} \log 3^{\omega}-\sum_{i=1}^{n} \log x_{i}^{\omega+1} \\
& =\log \omega \sum_{i=1}^{n} 1+\omega \sum_{i=1}^{n} \log 3-\omega \sum_{i=1}^{n} \log x_{i}-\sum_{i=1}^{n} \log x_{i} \\
& =n \log \omega+n \omega \log 3-\omega \sum_{i=1}^{n} \log x_{i}-\sum_{i=1}^{n} \log x_{i}
\end{aligned}
$$

b. Set up the equation that would need to be solved in order to compute $\hat{\omega}_{M L E}$. Once you arrive at the equation and have worked through any calculus, you can stop and simply present the equation that can be solved via simple algebraic manipulation.

Computing $\hat{\omega}_{M L E}$ amounts to finding the value of $\omega$ that solves the equation $\frac{\delta L L(\omega)}{\delta \omega}=0$. Fortunately, the derivative of interest is easily computed, because our $L L(\omega)$ is a sum of logarithms and linear terms.

$$
\begin{aligned}
L L(\omega) & =n \log \omega+n \omega \log 3-\omega \sum_{i=1}^{n} \log x_{i}-\sum_{i=1}^{n} \log x_{i} \\
\frac{\delta L L(\omega)}{\delta \omega} & =\frac{n}{\omega}+n \log 3-\sum_{i=1}^{n} \log x_{i}=0
\end{aligned}
$$

You were welcome to stop at the above equation, since it can be solved via algebraic manipulation you needn't show. We do, however, present the solution below for completeness, since this is a a real probability distribution used in many fields of statistical analysis.

$$
\begin{aligned}
& \frac{n}{\hat{\omega}_{M L E}}=S_{\log X}-n \log 3, \text { where } S_{\log X}=\sum_{i=1}^{n} \log x_{i} \\
& \hat{\omega}_{M L E}=\frac{n}{S_{\log X}-n \log 3}
\end{aligned}
$$

