This problem set covers material from Week 8, dates 4/09- 4/12. Unless otherwise noted, all problems are taken from the textbook. Problems can be found at the end of the corresponding subsection.

Instructions: Write or type complete solutions to the following problems and submit answers to the corresponding Canvas assignment. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A general rubric for homework problems appears on the final page of this assignment.

Tuesday 4/09

None!

Thursday 4/11

1. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from a population with distribution F that has unknown mean μ and variance σ^2. Let $\mathbf{X}^* = (X_1^*, \ldots, X_n^*)$ be a bootstrap sample from the ECDF for $\mathbf{X}$; that is, we have the conditional distribution $X_i^*|\mathbf{X} \overset{\text{iid}}{\sim} \text{DUnif}\{X_1, \ldots, X_n\}$. Let $\bar{X}^*$ denote the sample mean of $\mathbf{X}^*$:

$$\bar{X}^* = \frac{X_1^* + \cdots + X_n^*}{n}$$

   (a) Show that unconditionally, X_1^* has CDF F. *Hint:* To do so, compute the $P(X_1^* \leq x)$ by conditioning on the events that "$X_1^* = X_i$" for $1 \leq i \leq n$, and using the Law of Total Probability.

   (b) Use part (a) to calculate $E[X_1^*]$ and $\text{Var}(X_1^*)$.

   (c) Show that

$$\mathbb{E}[\bar{X}^*|\mathbf{X}] = \bar{X} \qquad \text{and} \qquad \text{Var}(\bar{X}^*|\mathbf{X}) = \frac{\hat{\sigma}^2}{n},$$

   where $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$.

   (d) Note that both $E[\bar{X}^*|\mathbf{X}]$ and $\text{Var}(\bar{X}^*|\mathbf{X})$ are **statistics**, since they are both functions of the random sample $\mathbf{X}$. Moreover, $\text{Var}(\bar{X}^*|\mathbf{X})$ can be used as an estimator of the variance of the sample mean $\bar{X}$. Show that $\text{Var}(\bar{X}^*|\mathbf{X})$ is a *biased* estimator.

2. The skewness of a random variable X with mean μ and variance σ^2 is defined as

$$\mu_3 = \text{Skew}(X) = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right]$$

   and is one measurement of the asymmetry of the distribution. Suppose that $\mathbf{X} = (X_1, \ldots, X_n)$ are an iid sample with common CDF F. The sample skewness is the statistic

$$M_3(\mathbf{X}) = \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^3}{\left(\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2\right)^{3/2}}$$

It is reasonable to use M_3 as the estimator of the skewness μ_3 of a random variable with CDF F. In order to assess the quality of this estimator, we need to approximate the bias and variance of this estimator.

(a) Show that if $\mathbf{x}$ are the observed values of the sample, then the skewness estimate $M_3(\mathbf{x})$ is equal to the skewness of a random variable Y which has the DUnif$\{x_1, \ldots, x_n\}$ distribution.

(b) Write a function in R which will take a vector `x` as input and output the value of the skewness $M_3(\mathbf{x})$ for this vector.

(c) The following sample of size 20 was generated from a skewed distribution with unknown CDF F. Copy-and-paste the following to load the data into R.

```
x <- c(6, 7, 4, 4, 4, 4, 4, 5, 9, 4, 5, 3, 5, 7, 2, 5, 6, 4, 4, 2)
```

Generate 5000 bootstrap samples from `x`. For each sample, compute the sample skewness. Then use these 5000 bootstrap statistics to estimate the **bias** and **standard deviation** of the sample skewness estimator M_3.

Friday 4/12

TBD

General rubric

Points	Criteria
5	The solution is correct *and* well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.
Notes:	For problems with code, well-written means only having lines of code that are necessary to solving the problem, as well as presenting the solution for the reader to easily see. It might also be worth adding comments to your code.