This problem set covers material from Week 4, dates 3/05- 3/08. Unless otherwise noted, all problems are taken from the textbook. Problems can be found at the end of the corresponding subsection. "AP" stands for additional problems not found in the book.

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 3/05

1. Let $X_1, \dots, X_n|\boldsymbol{\theta} \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ where $\boldsymbol{\theta} = (\mu, \sigma^2)$. Find the method of moments estimator for $\boldsymbol{\theta}$ (simplify as much as possible), and compare it to $\hat{\boldsymbol{\theta}}_{MLE}$.
In your comparison, it might be worth remembering the following identity you showed on HW 2: if $X_1, \dots, X_n iid$, then $\sum_{i=1}^n (X_i - a)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - a)^2$ for any $a \in \mathbb{R}$.

2. Let $X_1, \dots, X_n|\theta \stackrel{\text{iid}}{\sim} f(x|\theta) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}$ for $x \in \mathbb{R}$ and unknown parameter $\theta > 0$. Find a method of moments estimator for θ. It may be useful to remember the following facts:
   - If ϕ is a continuous even function, then $\int_{-a}^{a} \phi(x)dx = 2\int_0^a \phi(x)dx$
   - If ϕ is a continuous odd function, then $\int_{-a}^{a} \phi(x)dx = 0$

Thursday 3/07

3. Let $X_1, \dots, X_n|\theta \stackrel{\text{iid}}{\sim} \text{Exp}(\theta)$. We found that that the MLE and MoM estimators for θ are the same: $\hat{\theta} = \frac{1}{\bar{X}} = \frac{n}{\sum X_i}$. Let us try to say something about the bias of this estimator for θ. The WRONG thing to do is to say $\mathbb{E}[\hat{\theta}] = \frac{1}{\mathbb{E}[\bar{X}]}$; remember the function of the expectation and the expectation of the function are rarely equal. Instead, let's try to use the following important theorem:

   Jensen's inequality for probability theory states that if X is a random variable and $g()$ a convex function, then $g(\mathbb{E}[X]) \le \mathbb{E}[g(X)]$.

   Remember that a convex function is one where a linear segment between any two distinct points on the graph of the function always lies above the graph between the two points. We can demonstrate convexity of a univariate function $g(x)$ on an interval I if the second derivative $g''(x) > 0$ for all $x \in I$.

   Using Jensen's inequality, what can you say about the bias of $\hat{\theta}_{MLE}$ for θ in this problem? Does the MLE tend to overestimate or underestimate θ, or is it unbiased?

4. For $X_1, \ldots, X_n|\theta \overset{\text{iid}}{\sim} \text{Unif}[0, \theta]$, we found the estimators of θ under the method of moments and the method of maximum likelihood to be:

$$\hat{\theta}_{MM} = 2\bar{X} \qquad \text{and} \qquad \hat{\theta}_{MLE} = \max\{X_i\}$$

   (a) Find the bias of $\hat{\theta}_{MM}$ for θ.

   (b) Find the bias of $\hat{\theta}_{MLE}$ for θ.

   (c) For the estimators that you found to biased in (a) and/or (b), propose a new *unbiased* estimator $\tilde{\theta}_a$ and/or $\tilde{\theta}_b$ that is a modification of the estimator from (a) and/or (b), respectively.

5. Suppose we have a single $X|\theta \sim \text{Poisson}(\theta)$, and we are interested in estimating $g(\theta) = P(X = 0)^2 = e^{-2\theta}$.

   (a) Find an unbiased estimator of $g(\theta)$. *Hint: LoTUS and* $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

   (b) Comment on at least two reasons why the unbiased estimator in (a) is not very useful.

Friday 3/08

6. In certain problems, we can find unbiased estimators that have the smallest possible MSE among all unbiased estimators. In these cases, such an estimator might be considered "best". We will explore that here.

   Suppose $X_1, \ldots, X_n|\boldsymbol{\theta} \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ where $\boldsymbol{\theta} = (\mu, \sigma^2)$. Further suppose we want to focus on estimators for μ of the following form:

$$\delta(\mathbf{X}) = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n, \tag{1}$$

   where the $a_i \in \mathbb{R}$ are constants.

   (a) Come up with a general rule for the a_i so that $\delta(\mathbf{X})$ is unbiased for μ.

   (b) Assuming that the a_i satisfy the rule you obtained in (a) such that $\delta(\mathbf{X})$ is unbiased for μ, write out a formula for $MSE_\mu(\delta(\mathbf{X}))$.

   (c) Now, suppose we want to find the estimator $\delta^*(\mathbf{X})$ that has the *smallest* MSE among all the unbiased estimators of this particular form in eq. (1). Write an optimization problem for the a_i whose solution would lead to this $\delta^*(\mathbf{X})$. *Example: we seek the set of* a_i *that* $< blank >$.

   (d) Now consider the following Quadratic-Arithmetic mean inequality, which holds for all $a_1, \ldots, a_n \geq 0$:

$$\sqrt{\frac{a_1^2 + \ldots + a_n^2}{n}} \geq \frac{a_1 + \ldots + a_n}{n}$$

   (Proof easily possible via Cauchy-Schwarz inequality, for interested folks). Use this inequality to find the set of a_i that yields $\delta^*(\mathbf{X})$.

(e) How does your estimator $\delta^*(\mathbf{X})$ relate to the typical estimator, $\hat{\mu}_{MLE} = \bar{X}$?

(f) True of false: the estimator $\delta^*(\mathbf{X})$ obtained with the a_i you found in part (d) has lowest MSE among all unbiased estimators for μ. Briefly explain.

7. Let's return to the scenario where $X_1, \ldots, X_n|\theta \overset{\text{iid}}{\sim} \text{Unif}[0, \theta]$.

   (a) Find $MSE_\theta(\hat{\theta}_{MM})$. Simplify as much as possible.

   (b) Find $MSE_\theta(\hat{\theta}_{MLE})$ Simplify as much as possible.

   (c) Suppose we observe $n = 3$ data points from this distribution:

   $$X_1 = 1, X_2 = 9, X_3 = 2$$

   What are the estimates of θ based on this data?

   (d) Based on your answers above, which estimator do you think is better? Does your answer depend on n and/or θ and/or the observed data? Why or why not?

8. **R problem 1**.
   Set-up: During World War II, the Allied forces were interested in estimating the total number of tanks produced and owned by the Germans. These tanks were manufactured/labelled with an integer-valued serial number, where each tank was numbered in ascending, consecutive numbers (i.e. the first tank produced was numbered 1, the second tank produced was numbered 2, etc.). Every now and then, after certain battles or by luck, the Allied forces would capture a German tank and thus gain access to the serial number of that tank.

   Let N denote the unknown number of tanks produced. At any given time, the Allied forces had n captured tanks in their possession. Let the data $X_1, \ldots, X_n \in \{1, 2, \ldots, N\}$ be the serial numbers of each captured tank $i = 1, \ldots, n$. Note that the data are *not* iid Discrete Uniform$(1, n)$ because once a tank was captured, it was not returned back to the Germans (for obvious reasons). Consider the following three estimators of N:

   $$\hat{N}_1 = 2\bar{X} - 1 \qquad \hat{N}_2 = \max\{X_i\} \qquad \hat{N}_3 = \frac{n+1}{n}\max\{X_i\} - 1$$

   While we could obtain the MSEs of each of these three estimators analytically, the calculations are not trivial. So we will turn to simulations to compare these three estimators! Head over to the .Rmd template!

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.