Homework 4: 2/13 - 2/17

STA 336 Due 11:59pm Monday, February 20 Name:

Instructions: Write-up complete solutions to the following problems and submit answers on Gradescope. 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 rubric for homework problems appears on the final page of this assignment.

- Unless otherwise noted, problem numbers are taken from the 4th edition of DeGroot and Schervish's *Probability and Statistics.*

Monday 2/13

Section 7.6: 11
Section 8.7: 6

Additional Problems

(For those who took 335 in Fall 2022, these may look familiar)

AP1. Suppose $X \sim \text{Pois}(\theta)$ with θ unknown. Let $\psi = P(X = 0)^2 = e^{-2\theta}$.

(a) Show that the estimator $\delta(X) = (-1)^X$ is the **only** unbiased estimator of ψ. *Hint: Use LOTUS, which says that*

$$E[g(X)] = \sum_x g(x)P(X = x)$$

(b) Explain why $\delta(X)$ is not a very helpful estimator.

(c) Find the MLE for ψ. *Hint: Calculate the MLE of θ and use the invariance property.*

(d) (Optional Challenge Problem; not graded) Compute the MSE for both the unbiased estimator and the MLE and show that in general, the MLE is a much better estimator than the unbiased estimator.

AP2. In this problem, we show that every non-trivial Bayes estimator for a parameter θ is biased. We'll adopt the Bayesian perspective and assume that both θ and $\mathbf{X}$ are random variables.

Recall that the Bayes estimator is defined to be

$$\delta(\mathbf{X}) = E[\theta|\mathbf{X}]$$

And that an estimator $\hat{\theta}$ for θ is said to be unbiased if

$$E[\hat{\theta}|\theta] = \theta$$

[Technical note: In the following parts, assume that both the mean and variance of the prior distribution of θ are finite.]

(a) Suppose an estimator $\hat{\theta}$ is unbiased. Show that

$$E[(\hat{\theta} - \theta)^2] = E[\hat{\theta}^2] - E[\theta^2]$$

(b) Show that for the Bayes estimator,

$$E[(\delta(\mathbf{X}) - \theta)^2] = E[\theta^2] - E[(\delta(\mathbf{X}))^2]$$

Hint: Condition on θ and use the Law of Total Expectation.

(c) Show that the Bayes estimator $\delta(\mathbf{X})$ is unbiased if and only if $\delta(\mathbf{X}) = \theta$ with probability 1. (That is, the value of θ can be perfectly predicted using the observed data $\mathbf{X}$.)

Wednesday 2/15

Section 7.6: (20, 21), 22
Note: Problems enclosed in parentheses will be graded as a single problem.

Friday 2/17

Additional Problems

The following two additional problems have been moved to Homework 5. These problems DO NOT need to be completed for Homework 4. I've kept them listed on this assignment in case you want to start working on them in advance.

AP3. **(This problem has been moved to HW 5).** In this problem and the one that follows, you will compare two different estimators for θ in the model $X \sim \text{Bin}(n, \theta)$. The classic (frequentist) estimator for θ is the sample proportion

$$\hat{\theta}_c = \frac{X}{n}.$$

An alternative estimator is called the **Agresti-Coull** estimator, taking the form

$$\hat{\theta}_a = \frac{X+2}{n+4}$$

Informally, this estimator is obtained by adding two imaginary successes and two imaginary failures to the observed data. But it can also be viewed as the Bayes estimator for a certain choice of prior distribution. In some settings, the Agresti-Coull estimator can be used to create a more precise confidence interval than the one created from the classic estimator.

One way to measure the performance of an estimator is to calculate the Mean Squared Error. The Bayesian Mean Squared Error (MSE_b) for a Bayesian estimator is the expected squared loss, conditional on the data:

$$\text{MSE}_b(\hat{\theta}(\mathbf{X})) = E[(\hat{\theta} - \theta)^2 | \mathbf{X}]$$

That is, the expected value is taken by treating $\theta | \mathbf{X}$ as random, but $\mathbf{X}$ as fixed.

(a) Verify for any Bayesian estimator $\hat{\theta}$, the Bayesian Mean Squared Error of $\hat{\theta}$ is equal to variance of the posterior distribution of θ.

(b) Show that the Agresti-Coull estimator $\hat{\theta}_a$ is a Bayes estimator by choosing an appropriate prior distribution for θ.

(c) Show how to obtain the classic estimator $\hat{\theta}_c$ by choosing an appropriate *improper* prior distribution for θ.

(d) Compute the Bayesian Mean Squared Error for $\hat{\theta}_a$, as well as for $\hat{\theta}_c$.

(e) Suppose $n = 10$. Plot graphs of the Bayesian Mean Squared for $\hat{\theta}_a$ and for $\hat{\theta}_c$, as functions of $x \in \{0, 1, \ldots, 10\}$.

(f) Suppose $n = 10$. For what values of x are the Bayesian Mean Squared Error for $\hat{\theta}_a$ and $\hat{\theta}_c$ maximal, and for what values the Bayesian MSE minimal? For what values of x is the Bayesian MSE smaller for $\hat{\theta}_a$ than for $\hat{\theta}_c$?

AP4. **(This problem has been moved to HW 5).** This problem continues the discussion of the classic estimator and the Agresti-Coull estimator from the previous problem. In practice, the classic estimator is usually considered as a frequentist estimator, and so its performance should be evaluated as such.

The frequentist's Mean Squared Error (MSE_f) is the expected difference between a parameter θ and its estimator $\hat{\theta}(\mathbf{X})$, where we treat the parameter θ as fixed, and the data $\mathbf{X}$ as random:

$$\text{MSE}_f(\hat{\theta}(\mathbf{X})) = E[(\hat{\theta} - \theta)^2]$$

Note that $\text{MSE}_f(\hat{\theta})$ is a function of the parameter θ, but not of the data $\mathbf{X}$.

(a) Verify for any frequentist estimator $\hat{\theta}$, the frequentist Mean Squared Error can be decomposed as the sum of the variance of the estimator and the squared bias of the estimator:

$$\text{MSE}_f(\hat{\theta}(\mathbf{X})) = \text{Var}(\hat{\theta}) + (\text{Bias}(\hat{\theta}))^2$$

(b) Show that the classic estimator $\hat{\theta}_c$ is the MLE for the model $X \sim \text{Bin}(n, \theta)$.

(c) Show that the classic estimator $\hat{\theta}_c$ is an unbiased estimator of θ, while the Agresti-Coull estimator $\hat{\theta}_a$ is a biased estimator.

(d) Compute the variance of $\hat{\theta}_a$ and of $\hat{\theta}_c$. Which estimator has lower variance?

(e) Calculate the frequentist MSE for the the classic estimator $\hat{\theta}_c$. For what values of θ is MSE minimal? Maximimal? Then calculate the frequenstist MSE for the Agresti-Coull estimator $\hat{\theta}_a$. For what values of θ is MSE minimal? Maximal?

(f) Suppose $n = 10$. Create plots of the frequentist Mean Squared Error for $\hat{\theta}_a$ and $\hat{\theta}_c$, as functions of θ. For what values of θ is frequentist MSE for $\hat{\theta}_a$ lower than for $\hat{\theta}_c$?

(g) Give an intuitive explanation for the preceding result. *Hint: think as a Bayesian and consider the prior distributions corresponding to* $\hat{\theta}_c$ *and* $\hat{\theta}_a$

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.