Final Exam: Spring 2022

STAT 343: Mathematical Statistics

- This exam is **due to Gradescope at 11:59 PM ET on Sunday, May 8, 2021.** This is a firm deadline.
- There are five exam questions. You must show all your work to receive full credit. You may typeset your answers in LaTeX or hand write them; everything should be well-organized and legible.
- This is a **open** notes, **open** book, **open** Moodle exam (including homework solutions, etc., which are linked in Moodle). Other resources, besides R, **are not permitted.** Make sure to show all your work/thought process and include any R code (where appropriate) for full credit.
- Once you have accessed the exam, you may only ask for clarification on questions from the instructor. Do not communicate with any other students about the exam until after the exams have been returned. Any such communication is a violation of the Honor Code.
- If you become aware of any honor code violations, you have a duty to report them to the Honor Code Council.

Problem 1

In 2-3 sentences, explain why in a hypothesis test involving simple hypotheses, the p-value is calculated as the probability of obtaining a likelihood ratio test statistic at least as small as the value of the likelihood ratio statistic that was observed in our data set given that the null hypothesis is true. (Why at least as small rather than at least as large?)

Problem 2

Suppose for random variables $Y_1, ..., Y_n$, $Y_i = \beta x_i + \epsilon_i$. In this case, $x_1, ..., x_n$ are fixed **constants** and $\epsilon_i \overset{i.i.d.}{\sim} Normal(0, \sigma^2)$, where σ^2 is known.

(a) Find $E(Y_i)$.

(b) Find $Var(Y_i)$.

(c) Find the maximum likelihood estimator of β, $\hat{\beta}^{MLE}$.

(d) Is $\hat{\beta}^{MLE}$ an unbiased estimator of β?

(e) Find a Wald-based 95% confidence interval for β.

Problem 3

TRUE or FALSE. Justify your choice in 1-2 sentences.

(a) If two different estimators for the same parameter are unbiased, they are equally good and it does not matter which one we use.

(b) The expected value of the sample mean is a random variable.

(c) If two estimators for the same parameter θ, $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are such that $Var(\hat{\Theta}_1) = 0.2$ and $Var(\hat{\Theta}_2) = 0.25$, then $\hat{\Theta}_1$ definitely is preferred to $\hat{\Theta}_2$.

(d) Suppose H_0 and H_A are simple hypotheses and the test that rejects H_0 whenever the likelihood ratio test (LRT) statistic, W, is less than w^* has size α. Then, any other test has power that is less than or equal to the power of the LRT.

Problem 4

Let $X_1, ..., X_n \overset{i.i.d.}{\sim} Uniform(0, \theta)$ be continuous random variables. The following are some helpful facts that you can use as you answer the questions below.

- The MLE of θ is $\hat{\theta} = X_{\max}$, the maximum value in the sample.
- Therefore, it must always be the case that $0 \leq \hat{\theta} \leq \theta$, since $X_{\max} \in [0, \theta]$.
- For any observed value $x_{\max} \in [0, \theta]$, the cumulative distribution function of $\hat{\theta} = X_{\max}$ is

$$
\begin{aligned}
F_{X_{\max}}(x_{\max}|\theta) &= P(X_{\max} \leq x_{\max}|\theta) \\
&= P(\text{all } X_i \leq x_{\max}|\theta) \\
&= \prod_{i=1}^{n} P(X_i \leq x_{\max}|\theta) \\
&= \prod_{i=1}^{n} \frac{x_{\max}}{\theta} \\
&= \left(\frac{x_{\max}}{\theta}\right)^n
\end{aligned}
$$

For $x_{\max} > \theta$, the cumulative distribution function of $\hat{\theta}$ is $F_{X_{\max}}(x_{\max}|\theta) = 1$.

(a) Find the form of the (generalized) likelihood ratio statistic for a test of $H_0 : \theta = \theta_0$ versus $H_A : \theta < \theta_0$. You may assume that the largest value of x in the data set is less than or equal to θ_0.

(b) Suppose that $n = 10$, $\theta_0 = 10$, and we observe $x_{\max} = 5$. What is the p-value for the test?

(c) Suppose we want to make an accept/reject decision for the test using a significance level of α. Find the rejection region for the test of $H_0 : \theta = \theta_0$ versus $H_A : \theta < \theta_0$. Write your answer for this question for a generic θ_0; your answer should involve α, n, and θ_0.

(d) From your introductory statistics class, you may recall that there is duality between confidence intervals and hypothesis tests. In other words, a $100(1 - \alpha)\%$ confidence interval for θ consists of all those values of θ_0 for which the hypothesis $H_0 : \theta = \theta_0$ will not be rejected at level α. Using this information, obtain a confidence interval for θ. Your answer should involve α, n, and θ_0.

Problem 5

Take one homework problem you have worked on this semester that you struggled to understand and solve and explain how the struggle itself was valuable. In the context of this question, describe the struggle and how you overcame the struggle. You might also discuss whether struggling built aspects of character in you (e.g., endurance, self-confidence, competence to solve new problems) and how these virtues might benefit you in later ventures.