Math 408, Fall 2021, USC — Instructor: Steven Heilman

Name: ____________ USC ID: ____________ Date: ____________

Signature: ____________. Discussion Section: ____________
(By signing here, I certify that I have taken this test while refraining from cheating.)

Exam 2

This exam contains 9 pages (including this cover page) and 5 problems. Enter all requested information on the top of this page.

You may *not* use your books, notes, or any calculator on this exam.

You are required to show your work on each problem on this exam. The following rules apply:

- You have 50 minutes to complete the exam, starting at the beginning of class.
- **Organize your work**, in a reasonably neat and coherent way, in the space provided. Work scattered all over the page without a clear ordering will receive very little credit.
- **Mysterious or unsupported answers will not receive full credit**. A correct answer, unsupported by calculations, explanation, or algebraic work will receive no credit; an incorrect answer supported by substantially correct calculations and explanations might still receive partial credit.
- If you need more space, use the back of the pages; clearly indicate when you have done this. Scratch paper appears at the end of the document.

Problem	Points	Score
1	12	
2	10	
3	10	
4	10	
5	10	
Total:	52	

Do not write in the table to the right. Good luck![a]

Reference sheet

Below are some definitions that may be relevant.

Let $\{f_\theta \colon \theta \in \Theta\}$ be a family of multivariable probability density functions (PDFs) or probability mass functions (PMFs). Suppose $X = (X_1, \ldots, X_n)$ is a random sample of size n, so that X has distribution f_θ (i.e. f_θ is the joint distribution of $X_1, \ldots, X_n$). Let $t \colon \mathbf{R}^n \to \mathbf{R}^k$, so that $Y := t(X_1, \ldots, X_n)$ is a statistic.

We say that Y is a **sufficient statistic** for θ if, for every $y \in \mathbf{R}^k$ and for every $\theta \in \Theta$, the conditional distribution of $(X_1, \ldots, X_n)$ given $Y = y$ (with respect to probabilities given by f_θ) does not depend on θ.

Let $g \colon \Theta \to \mathbf{R}$. Let $t \colon \mathbf{R}^n \to \mathbf{R}$ and let Y be an unbiased estimator for $g(\theta)$. We say that Y is **uniformly minimum variance unbiased (UMVU)** for $g(\theta)$ if, for any other unbiased estimator Z for $g(\theta)$, we have

$$\mathrm{Var}_\theta(Y) \leq \mathrm{Var}_\theta(Z), \qquad \forall\, \theta \in \Theta.$$

Let $X, Y, Z \colon \Omega \to \mathbf{R}$ be discrete or continuous random variables. Let A be the range of Y. Define $h \colon A \to \mathbf{R}$ by $h(y) := \mathbf{E}(X|Y = y)$, for any $y \in A$. We then define the **conditional expectation** of X given Y, denoted $\mathbf{E}(X|Y)$, to be the random variable $h(Y)$.

Assume $\Theta \subseteq \mathbf{R}$. Define the **Fisher information** of X to be

$$I(\theta) = I_X(\theta) := \mathbf{E}_\theta\Big(\frac{d}{d\theta} \log f_\theta(X)\Big)^2, \qquad \forall\, \theta \in \Theta,$$

if this quantity exists and is finite.

A **maximum likelihood estimator** (MLE) is a statistic $Y = Y_n$ taking values in Θ satisfying

$$f_Y(X) \geq f_\theta(X), \qquad \forall\, \theta \in \Theta.$$

Suppose $\Theta_0 \subseteq \Theta$ and $\Theta_1 = \Theta_0^c$. The **power** of a hypothesis test with rejection region $C \subseteq \mathbf{R}^n$ is defined to be

$$\beta(\theta) := \mathbf{P}_\theta(X \in C), \qquad \forall \theta \in \Theta.$$

The **significance level** of a hypothesis test with rejection region C is defined to be

$$\alpha := \sup_{\theta \in \Theta_0} \beta(\theta).$$

Here sup denotes the supremum, i.e. the least upper bound (the upper bound with the smallest value). The **p-value** of a (family of) hypothesis tests with rejection regions $C = \{x \in \mathbf{R}^n \colon t(x) \geq c\}$ is the statistic $p(X)$ where

$$p(x) := \sup_{\theta \in \Theta_0} \mathbf{P}_\theta(t(X) \geq t(x)), \qquad \forall\, x \in \mathbf{R}^n$$

Here $t \colon \mathbf{R}^n \to \mathbf{R}$.

1. Label the following statements as TRUE or FALSE. If the statement is true, **explain your reasoning**. If the statement is false, **provide a counterexample and explain your reasoning**.

   (a) (2 points) The negation of the statement
   "There exists an integer j such that $j^2 - j < 8$" is:
   "For every integer j, we have $j^2 - j \geq 8$."

   TRUE FALSE (circle one)

   (b) (2 points) We have

   $$\sup_{\lambda \in (0,1)} \lambda^2 = 1.$$

   TRUE FALSE (circle one)

   (c) (2 points) Let Z be a sufficient statistic for $\{f_\theta \colon \theta \in \Theta\}$ and let Y be an unbiased estimator for θ. Define $W := \mathbf{E}_\theta(Y|Z)$. Let $\theta \in \Theta$ with $\mathrm{Var}_\theta(Y) < \infty$. Then

   $$\mathrm{Var}_\theta(W) \leq \mathrm{Var}_\theta(Y).$$

   TRUE FALSE (circle one)

(d) (2 points) There can be at most one maximum likelihood estimator. That is, if a maximum likelihood estimator exists, it is unique.

TRUE FALSE (circle one)

(e) (2 points) Suppose X is a UMVU for $\theta \in \Theta$, and Y is an estimator for θ. Then

$$\mathbf{E}_\theta(X - \mathbf{E}_\theta X)^2 \leq \mathbf{E}_\theta(Y - \mathbf{E}_\theta Y)^2, \qquad \forall\, \theta \in \Theta.$$

(In answering this question, you can freely use a result from the homework/quizzes.)

TRUE FALSE (circle one)

(f) (2 points) Let $X\colon \Omega \to \mathbf{R}^n$ be a random variable with distribution f_θ. Let $I_X(\theta)$ denote the Fisher Information of X. Let Y be an unbiased estimator for $\theta \in \Theta$. Suppose $I_X(\theta) > 0$ and

$$\mathrm{Var}_\theta(Y) = \frac{1}{I_X(\theta)}, \qquad \forall\, \theta \in \Theta,$$

Then Y is UMVU for θ.

TRUE FALSE (circle one)

2. (10 points) Let $X_1, \dots, X_n$ be i.i.d continuous random variables with $\mathbf{E}\,|X_1| < \infty$.

   In each question below, simplify your answer to the best of your ability.

   Unlike other questions, in this question you can freely use a result from a previous homework concerning general properties of conditional expectation.

   - Compute $\mathbf{E}(X_1 \,|\, X_1)$.
   - Compute $\mathbf{E}(X_1 \,|\, X_2)$.
   - Compute $\mathbf{E}(X_1 \,|\, X_1 + \cdots + X_n)$.

3. (10 points) Let $\theta \in \mathbf{R}$ be an unknown parameter. Consider the density

$$h_\theta(x) := \begin{cases} e^{-(x-\theta)}, & \text{if } x \geq \theta \\ 0, & \text{if } x < \theta. \end{cases}$$

Suppose $X_1, \ldots, X_n$ is a random sample of size n, such that X_i has density h_θ for all $1 \leq i \leq n$.

Show that $X_{(1)} = \min_{1 \leq i \leq n} X_i$ is a sufficient statistic for θ.

4. (10 points) Suppose $X_1, \ldots, X_n$ are i.i.d. random variables, and X_1 has density

$$h_\theta(x) = \theta x^{\theta-1}, \qquad \forall\, 0 < x < 1,$$

where $\theta > 0$ is unknown. (If $x \notin (0,1)$, $h_\theta(x) = 0$.)

Find an MLE Y_n of θ. As usual, you should justify your answer.

5. (10 points) Let X be a geometric distributed random variable with unknown parameter p where $p \in \{1/3, 2/3\}$. (So, $\mathbf{P}(X = k) = (1-p)^{k-1}p$ for all integers $k \geq 1$.) Suppose we want to test the hypothesis H_0 that $p = 1/3$ versus the alternative H_1 that $p = 2/3$. (The hypothesis test uses only X itself, you don't need to take more samples.)

   - Explicitly describe the rejection region C of the UMP (uniformly most powerful) test among all hypothesis tests with significance level less than or equal to $5/9$. Justify that your test is UMP.
   - Suppose we observe that $X = 1$. Report a p-value for this observation, for the UMP test you found.

(Scratch paper)