Math 541a, Spring 2022, 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 7 pages (including this cover page) and 4 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.

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

Problem	Points	Score
1	5	
2	10	
3	10	
4	15	
Total:	40	

[a]April 5, 2022,

Reference sheet

Below are some definitions that may be relevant.

We say that a sequence of random variables $X_1, X_2, \ldots : \Omega \to \mathbf{R}$ **converges in probability** to a random variable $X : \Omega \to \mathbf{R}$ if: for all $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}(|X_n - X| > \varepsilon) = 0.$$

We say that a sequence of real-valued random variables $X_1, X_2, \ldots$ **converges in distribution** to a real-valued random variable X if, for any $t \in \mathbf{R}$ such that $\mathbf{P}(X \leq t)$ is continuous at t,

$$\lim_{n\to\infty} \mathbf{P}(X_n \leq t) = \mathbf{P}(X \leq t).$$

We say that a sequence of random variables $X_1, X_2, \ldots : \Omega \to \mathbf{R}$ **converges almost surely** to a random variable $X : \Omega \to \mathbf{R}$ if

$$\mathbf{P}(\lim_{n\to\infty} X_n = X) = 1.$$

Suppose $X = (X_1, \ldots, X_n)$ is a random sample of size n from a distribution f where $f \in \{f_\theta : \theta \in \Theta\}$ is a family of densities (such as an exponential family). Let $t : \mathbf{R}^n \to \mathbf{R}^k$, 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.

We say Y is **minimal sufficient** for θ if Y is sufficient for θ and, for every statistic $Z : \Omega \to \mathbf{R}^m$ that is sufficient for θ, there exists a function $r : \mathbf{R}^m \to \mathbf{R}^k$ such that $Y = r(Z)$.

We say Y is **complete** for $\{f_\theta : \theta \in \Theta\}$ if the following holds:

For any $f : \mathbf{R}^m \to \mathbf{R}$ such that $\mathbf{E}_\theta f(Y) = 0 \quad \forall\, \theta \in \Theta$, it holds that $f(Y) = 0$.

We say Y is **ancillary** for θ if the distribution of Y does not depend on θ.

Let $X, Y, Z : \Omega \to \mathbf{R}$ be discrete or continuous random variables. Let A be the range of Y. Define $g : A \to \mathbf{R}$ by $g(y) := \mathbf{E}(X|Y = y)$, 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 $g(Y)$.

Let $\{f_\theta : \theta \in \Theta\}$ be a family of multivariable probability densities or probability mass functions. Assume $\Theta \subseteq \mathbf{R}$. Let X be a random vector with distribution f_θ. Define the **Fisher information** of the family 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.

1. (5 points) Let Y, Z be a statistics, and suppose Z is sufficient for $\{f_\theta \colon \theta \in \Theta\}$. Show that $W := \mathbf{E}_\theta(Y|Z)$ does not depend on θ. That is, there is a function $t \colon \mathbf{R}^n \to \mathbf{R}$ that does not depend on θ such that $W = t(X)$, where X is the random sample. (You may assume that X, Y, Z are all discrete.)

2. (10 points) Let X, Y be random variables such that (X, Y) is uniformly distributed in the region

$$\{(x, y) \in \mathbf{R}^2 \colon y \geq 0,\ x + y \leq 1,\ -x + y \leq 1\}.$$

Compute

$$\mathbf{E}(X|Y).$$

3. (10 points)

- Let $X := (X_1, \ldots, X_n)$ be a random sample of size n from a Gaussian distribution with unknown mean $\mu \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$. (That is, $X_1, \ldots, X_n$ are i.i.d. and X_1 is a Gaussian with mean μ and variance σ^2.)

  Computer the Fisher information $I_X(\sigma)$ for any $\sigma > 0$.
  (Consider $\mu \in \mathbf{R}$ to be fixed.)

- Give an example of a random variable Z whose distribution depends on a parameter $\theta \in \Theta$ such that $I_Z(\theta)$ does not exist for some $\theta \in \Theta$.

4. (15 points) Let $X_1, \ldots, X_n$ be a random sample from the Bernoulli distribution with unknown parameter $0 < p < 1$, so that, for all $1 \leq i \leq n$,

$$\mathbf{P}(X_i = 1) = p, \qquad \mathbf{P}(X_i = 0) = 1 - p.$$

- Find a complete sufficient statistic for p. (As usual, justify your answer.)
- Find the UMVU for p^3. (You may assume $n \geq 3$.)
  (Hint: $X_1 X_2 X_3$ is an estimator for p^3.)

(Scratch paper)