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 1

This exam contains 8 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	10	
2	5	
3	5	
4	10	
5	10	
Total:	40	

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

[a]February 25, 2022,

Reference sheet

Below are some definitions that may be relevant.

We say that a sequence of random variables $X_1, X_2, \ldots \colon \Omega \to \mathbf{R}$ **converges in probability** to a random variable $X \colon \Omega \to \mathbf{R}$ 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 $X_1, X_2, \ldots \colon \Omega \to \mathbf{R}$ **converges in** L_2 to a random variable $X \colon \Omega \to \mathbf{R}$ if

$$\lim_{n\to\infty} \mathbf{E}\,|X_n - X|^2 = 0.$$

We say that a sequence of random variables $X_1, X_2, \ldots \colon \Omega \to \mathbf{R}$ **converges almost surely** to a random variable $X \colon \Omega \to \mathbf{R}$ 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 $f \in \{f_\theta \colon \theta \in \Theta\}$ is a family of densities (such as an exponential family). 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.

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

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

For any $f \colon \mathbf{R}^m \to \mathbf{R}$ 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 $X, Y, Z \colon \Omega \to \mathbf{R}$ be discrete or continuous random variables. Let A be the range of Y. Define $g \colon A \to \mathbf{R}$ 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 $g(Y)$.

1. (10 points)

- Let (X, Y) be a vector in $\mathbf{R}^2$ that is uniformly distributed in the set

$$\{(x, y) \in \mathbf{R}^2 \colon x^2 + y^2 = 1\}.$$

  Compute $\mathbf{E}\,|X|$.

- Let $X_1, \ldots, X_n$ be i.i.d. standard Gaussian random variables. (So, each random variable has PDF $t \mapsto \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$, $t \in \mathbf{R}$.) Compute

$$\lim_{n \to \infty} \mathbf{P}(X_1^2 + \cdots + X_n^2 \leq n).$$

  (Hint: use the Central Limit Theorem. Recall also that $\mathbf{E}X_1^2 = 1$ and $\mathbf{E}X_1^4 = 3$.)

2. (5 points) Recall that a Gaussian density with mean $\mu \in \mathbf{R}$ and standard deviation $\sigma > 0$ has the following form:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad \forall\, x \in \mathbf{R}.$$

Write this density as a two-parameter exponential family $\{f_\theta \colon \theta \in \Theta\}$, where $\Theta = \{(\mu, \sigma^2) \colon \mu \in \mathbf{R},\ \sigma > 0\}$, and $\theta = (\theta_1, \theta_2) \in \Theta$. That is, write this density as

$$f_\theta(x) := h(x) \exp\Big(\sum_{i=1}^{2} w_i(\theta) t_i(x) - a(w(\theta))\Big), \qquad \forall\, x \in \mathbf{R},$$

for some $t_1, t_2 \colon \mathbf{R} \to \mathbf{R}$, $w_1, w_2 \colon \Theta \to \mathbf{R}$, $w := (w_1, w_2)$, $a \colon \mathbf{R}^2 \to \mathbf{R}$, $h \colon \mathbf{R} \to [0, \infty)$.

3. (5 points) Let X be a minimal sufficient statistic. Let Z be another sufficient statistic. Let u be a function such that

$$Z = u(X).$$

Show that Z is a minimal sufficient statistic.

4. (10 points) Let $X = (X_1, \ldots, X_n)$ be a random sample of size n, so that X_1 is a sample from the uniform distribution on the three element set $\{\theta, \theta+1, \theta+2\}$, where $\theta \in \mathbf{R} =: \Theta$. (The distribution of X is then $f_\theta(x) = \prod_{i=1}^{n} (1/3) 1_{x_i \in \{\theta, \theta+1, \theta+2\}}$, $\forall\ x \in \mathbf{R}^n$, $\forall\ \theta \in \mathbf{R}$.)

   Let Y be a statistic that is sufficient for θ. (Let $t \colon \mathbf{R}^n \to \mathbf{R}^k$ such that $Y = t(X_1, \ldots, X_n)$.)

   Is it true that we can write

$$f_\theta(x) = g_\theta(t(x))h(x), \qquad \forall\, \theta \in \Theta, \quad \forall\, x \in \mathbf{R}^n,$$

   for some $g_\theta \colon \mathbf{R}^k \to [0, \infty)$, $h \colon \mathbf{R}^n \to [0, \infty)$?

   Either prove this factorization can be done, or prove that it cannot be done.

5. (10 points) Let $X_1, \ldots, X_n$ be a random sample of size n, so that X_1 is a sample from the uniform distribution on the interval $[\theta - 1/2, \theta + 1/2]$, where $\theta \in \mathbf{R}$ is unknown.

   Show that $(X_{(1)}, X_{(n)})$ is minimal sufficient but not complete.

(Scratch paper)