Graduate Statistics, 541A, Spring 2019, USC — Instructor: Steven Heilman

Name: ______________________ USC ID: ______________ Date: ___________

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

Mid-Term 2

This exam contains 7 pages (including this cover page) and 4 problems. Check to see if any pages are missing. Enter all requested information on the top of this page.

You may *not* use your book or any calculator on this exam. You *cannot* use your homeworks. You are required to show your work on each problem on the exam. The following rules apply:

- You have 50 minutes to complete the exam, starting at the beginning of class.
- **If you use a theorem or proposition from class or the notes or the book you must indicate this** and explain why the theorem may be applied. It is okay to just say, "by some theorem/proposition from 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	10	
2	10	
3	10	
4	10	
Total:	40	

[a]April 4, 2019,

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 in L_2** to a random variable $X : \Omega \to \mathbf{R}$ if

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

We say that a sequence of random variables $X_1, X_2, \ldots : \Omega \to \mathbf{R}$ **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 $g(Y)$.

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. In any case, you must **justify your answer** as stated on the front cover page of this exam.

   (a) (2 points) Let $n \geq 2$ be an integer. Let $X_1, \ldots, X_n$ be a random sample of size n from the Gaussian distribution with mean $\mu \in \mathbf{R}$ and variance $\sigma^2 > 0$. Let $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$. and let $S := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$. Then $\overline{X}$ and S are independent random variables.

   TRUE FALSE (circle one)

   (b) (2 points) A complete sufficient statistic always exists. (In your answer here, you are allowed to cite a result from the homework, though try to be specific in your response.)

   TRUE FALSE (circle one)

   (c) (2 points) A constant function is both ancillary and complete.

   TRUE FALSE (circle one)

   (d) (2 points) A complete sufficient statistic is minimal sufficient.

   TRUE FALSE (circle one)

   (e) (2 points) Let $X_1, \ldots, X_n$ be a random sample of size n from the Gaussian distribution with mean $\mu \in \mathbf{R}$ and variance $\sigma = 1$, so that μ is unknown. Let Y be a complete sufficient statistic for μ. Let Z be an ancillary statistic for μ. Then Y and Z are independent.

   TRUE FALSE (circle one)

2. (10 points) Give an example of a statistic Y that is complete and nonconstant, but such that Y is not sufficient.

3. (10 points) Suppose $X_1, \ldots, X_n$ is a random sample of size n from the Gaussian distribution with unknown mean $\mu \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$. You may freely use that the sample mean $\overline{X}$ is UMVU for μ and $(\overline{X}, S^2)$ is complete sufficient for (μ, σ^2), where $S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$.

   Fix $\sigma^2 > 0$. Give an explicit expression for a statistic Y that is UMVU for μ^2, and prove that Y is UMVU for μ^2.

4. (10 points) Prove the Rao-Blackwell Theorem:

   Let Z be a sufficient statistic for $\{f_\theta \colon \theta \in \Theta\}$ and let Y be an estimator for $g(\theta)$. Define $W := \mathbf{E}_\theta(Y|Z)$. Let $\theta \in \Theta$ with $r(\theta, Y) < \infty$ and such that $\ell(\theta, y)$ is convex in $y \in \mathbf{R}$. Then

   $$r(\theta, W) \leq r(\theta, Y).$$

   (Recall that $\ell \colon \Theta \times \mathbf{R}^k \to \mathbf{R}$, and $r(\theta, Y) := \mathbf{E}_\theta \ell(\theta, Y)$.)

   (In your response, you can cite results from a homework, if you want.)

(Scratch paper)