Statistics, 408, Fall 2023, USC Instructor: Steven Heilman

Name: ____________________ USC ID: ______________ Date: ____________

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

Final Exam

This exam contains 12 pages (including this cover page) and 8 problems. Check to see if any pages are missing. Enter all requested information on the top of this page.

You may use your books and notes on this exam. You cannot use a calculator or any other electronic device (or internet-enabled device) on this exam. You are required to show your work on each problem on the exam. The following rules apply:

- You have 120 minutes to complete the exam.
- **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 is at the end of the exam.

Problem	Points	Score
1	12	
2	10	
3	10	
4	10	
5	10	
6	10	
7	10	
8	10	
Total:	82	

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

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

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

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

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 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 $C = \{x \in \mathbf{R}^n \colon t(x) \ge c\}$ is the statistic $p(X)$ where

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

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

1. (a) (4 points) Let X be a random variable with
$\mathbf{P}(X = -3) = 1/3$ and $\mathbf{P}(X = 3) = 2/3$.
Compute $\mathbf{E}X$ and $\mathbf{E}(X^2)$.

Your Answer: $\boxed{}$

(b) (4 points) Let Y, Z be i.i.d. random variables.
Assume that Y is uniformly distributed in $[0, 1]$.
Compute $\mathbf{P}(Y + Z \leq 1/2)$.

Your Answer: $\boxed{}$

(c) (4 points) State the Central Limit Theorem. Make sure to include **all** assumptions.

[This was repeated from a previous exam]

2. (10 points) Let $X_1, \ldots, X_n$ be i.i.d. random variables. Let $0 < p < 1$ be an unknown parameter. Assume that

$$\mathbf{P}(X_1 = k) = (1-p)^{k-1}p,$$

for all positive integers $k \geq 1$.

You can freely use the following facts: $\mathbf{E}X_1 = 1/p$, $\mathrm{Var}(X_1) = (1-p)/p^2$

- Find a statistic Z such that: Z is a method of moments estimator of p, and Z is a function of $X_1, \ldots, X_n$. Justify your answer.
- Find a statistic W such that W is an MLE for p, and W is a function of $X_1, \ldots, X_n$. (Make sure to justify that an MLE exists. Is an MLE unique in this case?) (Here MLE refers to a Maximum Likelihood Estimator.)

[this was a repeated homework question]

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

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

Suppose $X_1, \ldots, X_n$ is a random sample of size n, such that X_i has PDF f_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.

[This was repeated from a previous exam]

4. (10 points) Let $X_1, \ldots, X_n$ be i.i.d. random variables. Let $\lambda > 0$ be an unknown parameter. Assume that

$$\mathbf{P}(X_1 = k) = (e^{-\lambda})\frac{\lambda^k}{k!},$$

for all nonnegative integers $k \geq 0$.

You can freely use the following facts: $\mathbf{E}X_1 = \lambda$, $\mathrm{Var}(X_1) = \lambda$, $I_{X_1}(\lambda) = \frac{1}{\lambda}$.

You may assume that the MLE Y_n for λ exists and is unique, and it is given by:

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- Is Y_n unbiased for λ?
- Is the sequence of estimators $Y_1, Y_2, \ldots$ consistent? That is, does this sequence of random variables converge in probability to λ as $n \to \infty$? Justify your answer.
- What happens to the quantity $\mathrm{Var}(Y_n)$ as $n \to \infty$? More specifically, does this quantity asymptotically achieve the lower bound in the Cramér-Rao Inequality (for unbiased estimators of λ)? Justify your answer.
- Describe all unbiased estimators of λ that achieve equality in the Cramér-Rao inequality. (Assume such an estimator is a function of $X_1, \ldots, X_n$, for fixed n.)

5. (10 points) Suppose X is a binomial distributed random variable with parameters 2 and $\theta \in \{1/2, 3/4\}$. (So, X has the distribution of the number of heads that appears from flipping a coin twice, where θ is the probability that a heads appears in a single coin flip.)

   We want to test the hypothesis H_0 that $\theta = 1/2$ versus the hypothesis H_1 that $\theta = 3/4$.

   - Explicitly describe the rejection region C of the UMP (uniformly most powerful) test among all hypothesis tests with significance level at most $1/4$.
   - Explicitly describe the rejection region C of the UMP (uniformly most powerful) test among all hypothesis tests with significance level at most $3/4$.
   - Suppose we observe that $X = 2$. Report a p-value for this observation, for the UMP tests you found.

   [This was repeated from a previous exam]

6. (10 points) Suppose you flip a coin 1000 times, resulting in 600 heads and 400 tails. Is it reasonable to conclude that the coin is fair (i.e. it has one half probability of heads and one half probability of tails)? Justify your answer. (Your answer should compute a p-value for a sensible hypothesis test.)

   [This was a modified homework question.]

7. (10 points) Suppose X_1, X_2, X_3 are i.i.d. Gaussian random variables with variance one and unknown mean $\mu \in \mathbf{R}$. (A Gaussian with mean μ and variance $\sigma^2 > 0$ has PDF $\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/[2\sigma^2]}$.) Suppose you find that $X_1 = 1$ and $X_2 = 3$ and $X_3 = 0$.

   Explicitly construct a confidence interval of the form $[a, b]$ for μ, so that

$$\mathbf{P}(a \leq \mu \leq b) = \int_{-1}^{1} \frac{1}{\sqrt{3}\sqrt{2\pi}} e^{-y^2/6} dy.$$

   (Hint: What is the PDF of $X_1 + X_2 + X_3 - 3\mu$?)

   [This was a modified homework question.]

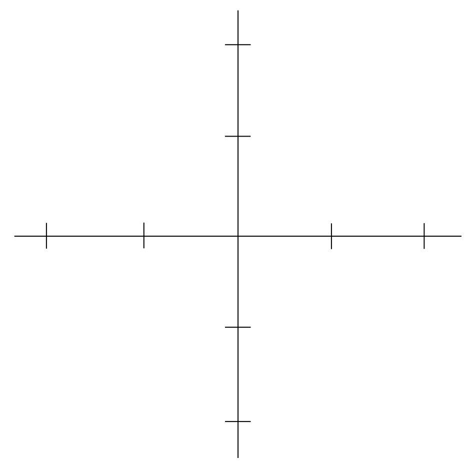
8. (10 points) Suppose you are given the following three data points in (x, y) coordinates:

$$(x_1, y_1) = (-1, 0), \qquad (x_2, y_2) = (0, 0), \qquad (x_3, y_3) = (0, 2).$$

- Find the line of the form $y = mx + b$ that best fits these three points. That is, find $m, b \in \mathbf{R}$ that minimizes the quantity.

$$h(m, b) := \frac{1}{2} \sum_{i=1}^{3} \Big(y_i - (mx_i + b) \Big)^2.$$

- Make sure to prove that the minimal m, b that you find actually minimizes $h(m, b)$.
- Finally, plot the points (x_1, y_1), (x_2, y_2), (x_3, y_3) along with the line $y = mx + b$ that best fits the points.



(Scratch paper)

(Extra Scratch paper)