Statistics, 408, Fall 2021, 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 13 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.

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

Problem	Points	Score
1	10	
2	10	
3	10	
4	10	
5	10	
6	10	
7	10	
8	10	
Total:	80	

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 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 $j^2 - j < 3$" is:
   "For every integer j, we have $j^2 - j \geq 3$."

   TRUE FALSE (circle one)

   (b) (2 points) Let $\mathbf{P}$ be the uniform probability law on $[0, 1]$. Let $x_1, x_2, \ldots \in [0, 1]$ be a countable set of distinct points. Then

   $$\mathbf{P}\left(\cup_{n=1}^{\infty}\{x_n\}\right) = 0.$$

   TRUE FALSE (circle one)

   (c) (2 points) Let $X_1, \ldots, X_n$ be i.i.d random variables drawn from a family of probability density functions $\{f_\theta \colon \theta \in \mathbf{R}\}$ where $f_\theta \colon \mathbf{R} \to [0, \infty)$ for all $\theta \in \mathbf{R}$. Then there must exist some integer $k \geq 1$, $\exists$ some function $t \colon \mathbf{R}^n \to \mathbf{R}^k$ and there exists some statistic $Y = t(X_1, \ldots, X_n)$ such that Y is a sufficient statistic for θ.

   TRUE FALSE (circle one)

(d) (2 points) Suppose $t(X)$ defined in the definition of p-value is a continuous random variable. Then the p-value satisfies

$$\mathbf{P}_\theta(p(X) \leq c) \leq c, \qquad \forall\, c \in (0,1), \qquad \forall\, \theta \in \Theta.$$

TRUE FALSE (circle one)

(e) (2 points) Let $X_1, \ldots, X_n$ be positive random variables. Then Pearson's chi-squared statistic

$$S := \sum_{j=1}^{n} \frac{\left(X_j - \mathbf{E}X_j\right)^2}{\mathbf{E}X_j}$$

has a chi-squared distribution.

TRUE FALSE (circle one)

2. (10 points) Let $X_1, \ldots, X_n$ be a random sample of size n from a Poisson distribution with unknown parameter $\lambda > 0$. (So, $\mathbf{P}(X_1 = k) = e^{-\lambda}\lambda^k/k!$ for all integers $k \geq 0$.)

- Find an MLE for λ. As usual, justify your answer.
- Is the MLE you found unique? That is, could there be more than one MLE for this problem? Justify your answer.

3. (10 points) Let $X_1, \ldots, X_n$ be a random sample of size n from the uniform distribution on $[\theta - 1/2, \theta + 1/2]$ where $\theta \in \mathbf{R}$ is unknown.

   Show that

   $$(X_{(1)}, X_{(n)})$$

   is a sufficient statistic for θ.

   (Here $X_{(1)} = \min_{1 \le i \le n} X_i$ and $X_{(n)} = \max_{1 \le i \le n} X_i$.)

4. (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.

5. (10 points) Let $X_1, \ldots, X_n$ be a random sample of size n from a Poisson distribution with unknown parameter $\lambda > 0$. (So, $\mathbf{P}(X_1 = k) = e^{-\lambda}\lambda^k/k!$ for all integers $k \geq 0$.)

   Let Y be the estimator $Y = 1_{\{X_1=0\}}$.

   (That is, $Y = 1$ when $X_1 = 0$, and otherwise $Y = 0$.)

   - Explicitly compute $W_n := \mathbf{E}_\lambda(Y \mid \sum_{i=1}^n X_i)$.
   - State an inequality comparing $\text{Var}_\lambda(Y)$ and $\text{Var}_\lambda(W_n)$.
     (Hint: you can freely use that $\sum_{i=1}^n X_i$ is sufficient for λ.)
   - What happens to W_n as $n \to \infty$? Does it converge to something? Justify your answer.

   (Hint: a sum of n independent Poissons with parameter λ is a Poisson with parameter $n\lambda$.)

6. (10 points) Suppose X_1, X_2 is a random sample from a Gaussian random variable X with unknown mean $\mu_X \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$. Suppose Y_1, Y_2 is a random sample from a Gaussian random variable Y with unknown mean $\mu_Y \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$. Assume that X_1, X_2 is independent of Y_1, Y_2, i.e. assume that X, Y are independent.

   Suppose you find that $X_1 = 1$, $X_2 = 3$, $Y_1 = 2$ and $Y_2 = 4$.

   Explicitly construct a confidence interval of the form $[a, b]$ for $\mu_X - \mu_Y$, so that

$$\mathbf{P}\Big(a \leq \mu_X - \mu_Y \leq b\Big) = \frac{1}{2\sqrt{2}} \int_{-3}^{3} \Big(1 + \frac{s^2}{2}\Big)^{-3/2} ds.$$

   Hint: Recall that Student's t-distribution with p degrees of freedom has density

$$f_T(s) := \frac{\Gamma(\frac{p+1}{2})}{\sqrt{p}\sqrt{\pi}\Gamma(p/2)} \Big(1 + \frac{s^2}{p}\Big)^{-(p+1)/2}, \qquad \forall\, s \in \mathbf{R}.$$

   Hint: $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$, $\Gamma(3/2) = \sqrt{\pi}/2$, $\Gamma(2) = 1$, $\Gamma(5/2) = 3\sqrt{\pi}/4$, $\Gamma(3) = 2$.

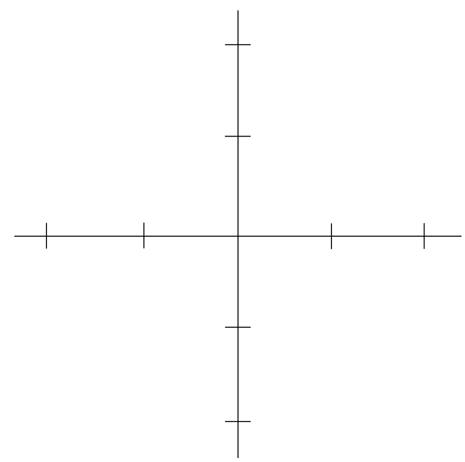
7. (10 points) Suppose you are given the following three data points in (x, y) coordinates:

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

- Find the parabola of the form $y = mx^2 + b$ that best fits these three points. That is, find $m, b \in \mathbf{R}$ that minimizes the quantity.

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

- 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 parabola $y = mx^2 + b$ that best fits the points.



8. (10 points) Consider the following table with turkey data. We have 4 (vegetarian) turkeys, with various temperatures x (Fahrenheit), and the status y of each turkey is cooked (corresponding to a value of $y = 1$) or not cooked (corresponding to a value of $y = 0$). Using logistic regression, we would like to find $a, b \in \mathbf{R}$, i.e. find a function

$$h(ax + b)$$

that best fits your data, where $h(t) = 1/(1 + e^{-t})$ for all $t \in \mathbf{R}$.

That is, given a temperature x, $h(ax+b)$ should be close to 1 when the turkey is cooked, and $h(ax + b)$ should be close to 0 when the turkey is not cooked.

Turkey	Temperature	Done? Yes or no.
1	150	no
2	155	yes
3	160	no
4	165	yes

- Describe in detail how you would find the $a, b \in \mathbf{R}$ that best fit the data using a computer to do logistic regression.
- Make sure to justify why maximization procedure must have at most one global maximum (perhaps by using a result from the homework).

(Scratch paper)

(Extra Scratch paper)