STA 711 Exam 2

Due: Monday, April 15, 11am on Canvas.

Instructions: Submit your work as a single PDF. For this assignment, you may include written work by scanning it and incorporating it into the PDF.

Mastery: To master this exam, you will need to master at least 6 of the 8 questions.

Rules: This is an open-book, open-notes exam. You may:

- Use any resources from the course (the textbook, the course website, class notes, previous assignments, etc.)
- Email me, or come to office hours, with specific questions (I may be somewhat less helpful than for regular assignments)
- Use one or two days from your bank of extension days, if you need more time on the exam

You may *not*:

- Use the internet to look up any questions on the exam
- Use any resources outside of the course (other textbooks, notes from other courses or universities, etc.)
- Use WolframAlpha or any generative AI to help solve the problems
- Discuss the exam with anyone else

Showing work: Make sure to show all work for each question. You may use results that we have stated or proved in class or homework, but you must cite the relevant results and explain your steps (for example, if you use WLLN for a step, say so).

Hints: Several questions have hints on strategies for approaching the problem which you may wish to consider. These hints are not complete solutions, and further steps (possibly using other tools) may be needed as well.

Convergence

1. Suppose that $Y_1, ..., Y_n$ are an iid sample from the *log-normal* distribution, with pdf

$$f(y|\mu, \sigma^2) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}(\log(y) - \mu)^2 \right\}.$$

   (a) Show that $\log(Y_i) \sim N(\mu, \sigma^2)$.

   (b) Show that the geometric mean $\left(\prod_{i=1}^{n} Y_i\right)^{1/n} \xrightarrow{p} \exp(\mu)$. (Hint: your answer should involve the continuous mapping theorem...)

2. Let $X_1, ..., X_n$ be a random sample from a distribution with cdf $F(x) = \frac{1}{1+e^{-x}}$, with $x > 0$. Find a sequence a_n such that $X_{(n)} - a_n$ converges in distribution. (Recall that $X_{(n)} = \max\{X_1, ..., X_n\}$).

3. Suppose that $X_1, ..., X_n$ are independent random variables, with $\mathbb{E}[X_i] = \mu_i$ and $Var(X_i) = \sigma_i^2$. Show that if

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 = 0,$$

   then

$$P\left(\frac{1}{n} \left| \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i \right| > \varepsilon \right) \to 0$$

   for all $\varepsilon > 0$.

4. Let $(X_1, Y_1), ..., (X_n, Y_n)$ be an iid sample from the joint distribution of X and Y. We are interested in estimating the *correlation* between X and Y:

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

   The Pearson correlation estimates ρ by

$$\widehat{\rho} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}.$$

   The goal of this question is to show that $\widehat{\rho} \xrightarrow{p} \rho$ as $n \to \infty$. You may use, without proof, the result that $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \xrightarrow{p} Var(X)$.

   (a) Show that $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) \xrightarrow{p} Cov(X, Y)$.

   (b) Show that $\widehat{\rho} \xrightarrow{p} \rho$.

Hypothesis testing

5. The pdf of a bivariate normal distribution is

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\},$$

where $-1<\rho<1$, $\sigma_1,\sigma_2>0$, and $\mu_1,\mu_2\in\mathbb{R}$. Let $(X_1,Y_1),...,(X_n,Y_n)$ be an iid sample from a bivariate normal distribution, and let $\widehat{\mu}_1$, $\widehat{\mu}_2$ denote the MLEs of μ_1 and μ_2 (you do not need to calculate them). It can be shown (but you do not need to derive this) that the MLEs for σ_1^2, σ_2^2, and ρ are then

$$\widehat{\sigma}_1^2=\frac{1}{n}\sum_{i=1}^n(X_i-\widehat{\mu}_1)^2 \qquad \widehat{\sigma}_2^2=\frac{1}{n}\sum_{i=1}^n(Y_i-\widehat{\mu}_2)^2 \qquad \widehat{\rho}=\frac{1}{n}\sum_{i=1}^n\left(\frac{X_i-\widehat{\mu}_1}{\widehat{\sigma}_1}\right)\left(\frac{Y_i-\widehat{\mu}_2}{\widehat{\sigma}_2}\right).$$

We want to test $H_0:\rho=0$ vs. $H_A:\rho\neq 0$. Show that the likelihood ratio test rejects when $\widehat{\rho}^2$ (the unrestricted MLE of ρ) is large. You may use, without proof, that the restricted MLEs $\widehat{\mu}_1$, $\widehat{\mu}_2$, $\widehat{\sigma}_1^2$, and $\widehat{\sigma}_2^2$, under the restriction that $\rho=0$, are the *same* as the unrestricted MLEs.

6. Let $X_1,...,X_n$ be an iid sample from a distribution with pdf

$$f(x|\theta)=\frac{2}{\sqrt{\theta\pi}}\exp\left\{-\frac{x^2}{\theta}\right\} \qquad x>0,\ \theta>0.$$

You may use, without proof, the fact that $\dfrac{2X_i^2}{\theta}\sim\chi_1^2$.

   (a) Show that the likelihood ratio test of $H_0:\theta=\theta_0$ vs. $H_A:\theta\neq\theta_0$ rejects when $\dfrac{2\sum_{i=1}^n X_i^2}{\theta_0}<c_1$ or $\dfrac{2\sum_{i=1}^n X_i^2}{\theta_0}>c_2$, for some $c_1<c_2$.

   (b) Find values c_1,c_2 such that the likelihood ratio test from (a) is a size α test.

7. Let $X_1,...,X_n$ be an iid sample from a continuous distribution with probability function $f(x|\theta)$, and $T\equiv T(X_1,...,X_n)$ a test statistic calculated from the sample. Suppose we have some hypothesis test which rejects $H_0:\theta=\theta_0$ when $T(X_1,...,X_n)>c$, for some threshold c. In class, we showed that the p-value for this test can be written

$$p=P_{\theta_0}(T(X_1^*,...,X_n^*)\geq T(X_1,...,X_n)),$$

where $X_1^*,...,X_n^*\sim f(x|\theta_0)$ is an independent sample under H_0. In other words, the p-value is the probability of "our data or more extreme" under H_0.

   (a) Show that $p=1-F_{0,T}(T(X_1,...,X_n))$, where $F_{0,T}$ denotes the cdf of T under H_0 (i.e. when $\theta=\theta_0$).

   (b) Show that, if H_0 is true, then $p\sim Uniform[0,1]$.

8. (Two-sample test for difference in means) Suppose we observe two independent samples

$$X_1,...,X_{n_1}\overset{iid}{\sim}N(\mu_1,\sigma^2) \qquad Y_1,...,Y_{n_2}\overset{iid}{\sim}N(\mu_2,\sigma^2)$$

from two different Normal populations, with a common variance σ^2. We are interested in testing

$$H_0 : \mu_1 = \mu_2 \qquad H_A : \mu_1 \neq \mu_2$$

A common approach to testing these hypotheses is the *two-sample t-test with pooled variance*, which uses the following test statistic:

$$T = \frac{\overline{X} - \overline{Y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$, and $s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \overline{X})^2$, $s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \overline{Y})^2$.

Show that $T \sim t_\nu$, and give the correct degrees of freedom ν.