Name: HUAN Q. BUI

Statistics 482 Spring 2020
Exam 1

Wednesday, 1 April 2020

- This is an open-note and open-book exam.
- All work must be your own. You may not give or receive any kind of aid, either verbally, visually, or otherwise, during this exam. No other sources may be consulted, except as specified above.
- The exam has 90 possible points. There are 6 questions and 10 pages, including this cover page.
- A standard normal table and information concerning common distributions' densities, expectations, variances, and moment-generating functions is available on Moodle.
- You have 4 hours to complete the exam so plan your time accordingly. I have included the possible points next to each problem.
- Some questions are more difficult than others, and the questions may not be in order of difficulty.
- Don't spend too much time on any one question; if you get stuck, go on and try another part.
- Whenever possible, show your work and explain your reasoning. In case you make a mistake, I can more easily give you partial credit if you explain your steps.
- Please type answers directly into questions, or provide a document with answers neatly written. Use the file submission option in Moodle to upload your document.

Question 1 (12 points total)

A. (3 points) Let Y be a random variable that has an F-distribution with v_1 numerator and v_2 denominator degrees of freedom. Find the distribution of $U = 1/Y$.

B. (9 points) Suppose that Y_1, Y_2, Y_3, Y_4, and Y_5, are a random sample from a normal population with mean 0 and standard deviation 1. Also let $\bar{Y} = \frac{1}{5}\sum_{i=1}^{5} Y_i$, and let Y_6 be another independent observation from the same distribution. What is the distribution of each of the following? Explain your rationale briefly.

a. $\sum_{i=1}^{5} \boldsymbol{Y_i^2}$

b. $\sum_{i=1}^{5} (\boldsymbol{Y_i} - \boldsymbol{\bar{Y}})^2$

c. $\sum_{i=1}^{5} (\boldsymbol{Y_i} - \boldsymbol{\bar{Y}})^2 + \boldsymbol{Y_6^2}$

Question 2 (18 points total)

A random sample of 37 second graders who participated in sports had manual dexterity scores with mean 32.19 and standard deviation 4.34. An independent sample of 37 second graders who did not participate in sports had manual dexterity scores with mean 31.68 and standard deviation 4.56.

A. **(6 points)** Perform an appropriate hypothesis test to see if there is sufficient evidence to indicate that second graders who participated in sports have a higher mean dexterity score. Use $\alpha = 0.05$. You may assume that the standard deviations given above are *population* values and that the distribution of dexterity scores is approximately normal. Make sure to report all necessary components: hypotheses, test statistic, p-value, and conclusion.

Question 2 continued…

B. (6 points) For the test in part A, calculate the power when $\mu_{sports} - \mu_{no\ sports} = 3$.

C. (6 points) Write an algorithm (not formal R code, but just the process) to perform a bootstrap-based test for the scenario described in part A.

Question 3 (20 points total)

Let Y be a single randomly drawn observation from a distribution with density given by:

$$f_Y(y) = \frac{2(\theta - y)}{\theta^2}; \qquad 0 < y < \theta$$

A. (8 points) Show that $U = Y/\theta$ is a pivotal quantity.

B. (4 points) Use the pivotal quantity in part A to find a 90% lower confidence limit for θ. Note that you should start with the expression: $P(U < a) = 0.90$, where a is the lower confidence limit. (Hint: you may need the quadratic formula: $y = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

Question 3 continued…

C. (8 points) If you wanted to create a random sample of observations having the distribution of *Y* above (assume a known, but fixed, θ) from *n* random uniform variables on the interval (0,1). Describe the algorithm you would use if you wanted to utilize inverse transform sampling. Be as specific as possible.

Question 4 (20 points)

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample drawn from a distribution with density given by,

$$\boldsymbol{f_Y(y)} = \begin{cases} \left(\dfrac{\mathbf{1}}{\boldsymbol{\theta^2}}\right)\boldsymbol{y e^{-y/\theta}}, & \boldsymbol{y > 0} \\ \mathbf{0}, & \textbf{elsewhere} \end{cases}$$

Note that *Y* has a gamma distribution with parameters $\alpha = 2$ and $\beta = \theta$.

A. **(6 points)** Find the maximum likelihood estimator, $\hat{\theta}$, of θ.

B. **(2 points)** What is the maximum likelihood estimator of $V(Y_i)$?

C. **(6 points)** Find expected value, $E(\hat{\theta})$, and variance, $V(\hat{\theta})$, of the estimator you found in part A.

Question 4 continued…

D. **(6 points)** Is $\hat{\theta}$ from part (A) a consistent estimator of θ? Support your statement.

Question 5

Consider a sequence of random variables $\{X_n\}$, where each has density given by:

$$f_{X_n}(x) = \begin{cases} \left(1 - \frac{x}{n}\right)^{n-1}; & 0 \le x \le n \\ 0; & \text{elsewhere} \end{cases}$$

A. (4 points) Find the cumulative distribution function of X_n.

B. (6 points) What is the limiting distribution of this sequence of random variables (i.e., what does this sequence converge to in distribution)?

Question 6 (10 points total)

For each of the statements below, indicate whether it is true or false.

1. We are unable to calculate the power of a hypothesis test unless we have both a simple null hypothesis and a simple alternative hypothesis.

True **False**

2. The sum of *n* i.i.d random exponential(2) random variables has a exponential (2n) distribution.

True **False**

3. We can compare the p-value to alpha because they are both conditional probabilities that are both conditional on the null being true.

True **False**

4. A bootstrap sample is taken with replacement because we are simulating the population by pretending that we have an infinite number of copies of our original sample.

True **False**

5. One of the main advantages of the accept-reject algorithm (acceptance sampling), as compared with inverse transform sampling, is that you do not need to know the density function of the random sample you are trying to generate.

True **False**