Homework 2
STA4322/STA5328
Introduction to Statistics Theory
Spring 2020, MWF 3:00pm
Due date: Friday, January 31st, 2020

All work must be shown for complete credit. W.M.S. denotes the course textbook, *Mathematical Statistics with Applications*. Problem 3 is worth 2 points of extra credit.

1. If $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed normal random variables with mean μ and variance σ^2, then $\bar{Y} \sim \mathrm{N}(\mu, \sigma^2/n)$ where $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. Prove this statement using moment generating functions.

2. Let X have probability density function

$$f_X(x) = \begin{cases} \frac{2(\theta - x)}{\theta^2} & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}.$$

   (a) Show that X has distribution function

$$F_X(x) = \begin{cases} 0 & x \le 0 \\ \frac{2x}{\theta} - \frac{x^2}{\theta^2} & 0 < x < \theta \\ 1 & x \ge \theta \end{cases}.$$

   (b) Show that X/θ is a pivotal quantity.

   (c) Use the pivotal quantity from (b) to find a 90% lower confidence limit for θ.

3. (Extra credit) Suppose that $Y_1, Y_2, \ldots, Y_n$ form a random sample from the exponential distribution with unknown mean μ. Propose a confidence interval for μ with confidence level γ $(0 \le \gamma \le 1)$. Note that this should be an **exact** interval, i.e., does not use a CLT approximation.

   Hint: Determine constants c_1 and c_2 such that $P(c_1 < \frac{1}{\mu}\sum_{i=1}^n Y_i < c_2) = \gamma$. Note that the constants c_1 and c_2 can be left as quantiles of some particular distribution.

4. The downtime per day for a computing facility has mean 4 hours and standard deviation 0.8 hours.

   (a) Suppose we want to compute probabilities about the average daily downtime for a period of 30 days.

      i. What assumptions must be true to use the central limit theorem to obtain a valid approximation for probabilities about the average daily downtime?

      ii. Under the assumptions you stated in the previous part, what is the approximate probability that the average daily downtime for a period of 30 days is between 1 and 5 hours? Your answer may be left in terms of Φ, the distribution function of a standard normal random variable.

(b) Under the assumptions you stated in the previous part, what is the approximate probability that the *total* downtime for a period of 30 days is less than 115 hours? Your answer may be left in terms of Φ, the distribution function of a standard normal random variable.

5. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean zero and variance σ^2. Construct a 95% lower confidence limit for σ^2. Your anwser may be left in terms of quantiles of some particular distribution.

6. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables. Let each X_i be uniformly distributed on the interval $(-1, 1)$. Similarly, let $Y_1, \ldots, Y_n$ be independent and identically distributed random variables also uniformly distributed on the interval $(-1, 1)$. Assume all Y_i and X_i are independent. Let us then define a new variable Z_i: the variable $Z_i = 1$ if (X_i, Y_i) lies within the unit-disk (disk centered at 0 with radius 1); the variable $Z_i = 0$ if (X_i, Y_i) lies outside the unit-disk.

   (a) What is the distribution of Z_i?

   (b) Let $\bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i$. What is the mean and variance of $\bar{Z}$?

   (c) What is the approximate, large-sample distribution of $4\bar{Z}$ (assuming n is very large)?

   (d) Approximate $P(|4\bar{Z} - \pi| < .01)$ using the central limit theorem.