STA 260s20 Assignment Three: Confidence Intervals Part One[1]

These homework problems are not to be handed in. They are preparation for Quiz 3 (Week of Jan. 27) and Term Test 1. **Please try each question before looking at the solution**.

1. The "pth quantile" or "p quantile" of a probability distribution is the point with p of the probability at or below it. That is, if x_p is the p quantile of the distribution of the random variable X, $F_X(x_p) = p$. Here are some numbers we will need to construct confidence intervals. Let z_p denote the p quantile of the standard normal distribution. Let $Z \sim N(0,1)$
   (a) What is $P(-z_{0.975} < Z < z_{0.975})$? I think it helps to draw a picture.
   (b) What is $z_{0.975}$? The answer is a number from the table of the standard normal distribution, now included in the formula sheet.
   (c) What is $P(-z_{0.995} < Z < z_{0.995})$?
   (d) What is $z_{0.995}$? This time you will need to interpolate in the table.
   (e) Let α denote a small probability; the values $\alpha = 0.05$ and $\alpha = 0.01$ are common. What is $P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2})$?

2. Let $X_1, \ldots X_n$ be independent random variables from a distribution with expected value μ and variance σ^2, where μ and variance σ^2 are both unknown. Let $\widehat{\sigma}^2_n$ be a consistent estimator of σ^2. Using the Central Limit Theorem, derive a $(1-\alpha)*100\%$ confidence interval for μ. This yields a 95% confidence interval if $\alpha = 0.05$, or a 99% confidence interval if $\alpha = 0.01$. "Derive" means show all the High School algebra. Your final answer is a pair of formulas, a formula for the lower confidence limit $L(X_1, \ldots X_n)$ and a formula for the upper confidence limit $U(X_1, \ldots X_n)$. Use the notation $z_{1-\alpha/2}$ for the $1-\alpha/2$ quantile of the standard normal distribution.

3. The label on the peanut butter jar says peanuts, partially hydrogenated peanut oil, salt and sugar. But we all know there is other stuff in there too. There is very good reason to assume that the number of rat hairs in a jar of peanut butter has a Poisson distribution with mean λ, because it's easy to justify a Poisson process model for how the hairs get into the jars (technical details omitted). A sample of thirty 500g jars yields $\overline{X}_n = 9.2$.
   (a) Give point estimate and a 95% confidence interval for λ. Indicate why your chosen $\widehat{\sigma}^2_n$ is consistent. Show a little work. Your answer is a pair of numbers, a lower confidence limit and an upper confidence limit.
   (b) There is a government standard that says the true expected number of rat hairs in a 500g jar may not exceed 8. Do these results suggest that the company may be in violation of the regulation?

[1]Copyright information is at the end of the last page.

4. In a coffee taste test, 100 coffee drinkers tasted coffee made with two different blends of coffee beans, the old standard blend and a new blend. We will adopt a Bernoulli model for these data, with θ denoting the probability that a customer will prefer the new blend. Suppose 60 out of 100 consumers preferred the new blend of coffee beans.

   (a) Assuming that the participants in the study are a random sample of coffee drinkers (which they almost certainly are not) give a point estimate and a 95% confidence interval for the *percentage* of coffee drinkers who would prefer the new blend of coffee beans.

   (b) The research department is afraid that management will not be satisfied with a margin of probable error of almost 10%. The would like to say "These results are expected to be accurate within three percentage points, 19 times out of 20."

      i. Suppose that the true proportion of customers who would like the new blend is actually 0.60. What sample size n would be required to achieve the desired margin of error?

      ii. Suppose the true proportion equals 0.7. What sample size is required now?

      iii. Suppose the true proportion equals 0.9. What sample size is required now?

      iv. Suppose the true proportion equals 0.3. What sample size is required now?

      v. Now suppose that the researchers are cautious, and want to guarantee that the margin of error will be 3% or less regardless of what the true value of θ might be. What is the required required sample size?

   Clearly, these people are going to over-spend their budget. They are in deep trouble.

5. A sample of $n = 64$ high school students take a standardized multiple choice vocabulary test. We obtain a sample mean of $\overline{x}_n = 105$ and a sample variance of $s^2 = 256$. Give a 95% confidence interval for μ. Your answer is a pair of numbers, the lower confidence limit and the upper confidence limit.

6. The physics of volcanic eruptions suggests that the intervals between eruptions at a particular site should be independent exponential random variables, with the unknown value of the parameter λ depending on the site. Fifteen eruptions have been recorded at the West Thumb Geyser Basin during historic times. The 14 intervals between eruptions (in years) are: `28.83 9.25 3.00 8.58 0.50 2.08 0.50 26.25 33.08 2.17 3.08 0.58 2.42 5.08`.

   (a) Estimating the expected interval between eruptions is easy. Give a reasonable point estimate. The answer is a number. In terms of λ, what are you estimating?

   (b) We want a confidence interval, but $n = 14$ is too small to be comfortable using the Central Limit Theorem. However, an exact (not asymptotic) confidence interval is within reach.

      i. Find the distribution of $\overline{X}_n$. Show your work.

ii. Show that $Y = 2\lambda n\overline{X}_n$ has a chi-squared distribution. For reasons that may become clear later in the course, the parameter ν of the chi-squared distribution is called the "degrees of freedom." What are the degrees of freedom of Y?

iii. Using the quantiles $\chi^2_{0.025}$ and $\chi^2_{0.975}$, derive a 95% confidence interval for the expected time between eruptions.

iv. Unfortunately, the chi-squared table in our text does not give quantiles for the lower part of the distribution. The free open source R software can do it easily, though. `q` is for quantile.

```
> qchisq(0.025,28)
[1] 15.30786
> qchisq(0.975,28)
[1] 44.46079
```

So with $\nu = 28$ degrees of freedom, $\chi^2_{0.025} = 15.31$ and $\chi^2_{0.975} = 44.46$. Give the 95% confidence interval. Your answer is a pair of numbers, the lower confidence limit and the upper confidence limit.

(c) It's not justified because of the small sample size, but go ahead and use the Central Limit Theorem to produce a 95% confidence interval for

i. The expected time between eruptions.

ii. The parameter λ.

7. This question mirrors the development of confidence intervals based on the Central Limit Theorem, except that most of the technical details are easier. Let $X_1, \ldots X_n$ be independent Uniform $(0, \theta)$ random variables.

(a) Let T_n be the maximum X_i value. Obtain the cumulative distribution function of T_n. "Obtain" could be as simple as copying your answer to Question 5c of Assignment 2.

(b) Let $Y_n = n(1 - \frac{T_n}{\theta})$.

i. Find the support of Y_n, or equivalently, the shortest interval A for which $P(Y_n \in A) = 1$. Show your work.

ii. Derive the cumulative distribution function of Y_n, and write it using indicator functions.

iii. Using the definition of convergence in distribution, show that $Y_n \stackrel{d}{\rightarrow} Y \sim$ Exponential(1).

(c) Using the last result, derive a general $(1 - \alpha)$ confidence interval for θ. Note that generally speaking, it's a good idea to seek confidence intervals that are as short as possible, though for Problem 6 it would have been more trouble than it's worth. Here, it matters. Because the exponential density is decreasing, the shortest interval starts with zero. So, begin the derivation of the confidence interval with $1 - \alpha = P(0 < Y < y_{1-\alpha})$, where $y_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard exponential distribution.

(d) Is it possible for the true value of θ to be below your lower confidence limit? Answer Yes or No.

(e) Unlike the quantiles of the normal and chi-squared distributions, the quantiles of the exponential can be obtained with an ordinary scientific calculator. Give an explicit formula for $y_{1-\alpha}$. Show your work.

(f) In the following R session, I simulate $n = 30$ observations from a Uniform(0,4) distribution, and calculate the mean and maximum. So the true parameter value $\theta = 4$ is actually known. This never happens in real-world applications, but when you're developing an estimation method, it's good to try it on simulated data where you know the truth, to see how the estimates behave. The statistics are rounded to 3 decimal places.

```
> x = runif(30,0,4)
> round(mean(x),3)
[1] 1.785
> round(max(x),3)
[1] 3.905
```

i. Using the simulated data above, calculate your 95% confidence interval. The answer is a pair of numbers, a lower confidence limit and an upper confidence limit.

ii. For comparison, calculate a 95% confidence interval based on the Central Limit Theorem.

iii. Which confidence interval do you like more, and why?

8. In the construction of confidence intervals, a *pivotal quantity*, or *pivot* is a random variable that is a function of the parameter, but whose distribution does not depend on the value of the parameter. You construct an interval based on the random variable, and then manipulate the inequalities until the parameter is alone in the middle. What are the pivots in this assignment?
