STAT430 Homework #4: Due Friday, February 21, 2020.

Name: KEY

Submit this homework as a pdf file to Canvas. I have created boxes after each problem for you to write you solution.

Question 1

Let Y have a probability density function

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

a) Find the cumulative distribution function (cdf) of Y.

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= \int_{-\infty}^{y} \frac{2(\theta - u)}{\theta^2} I(0 < y < \theta) du \\
&= \int_{0}^{y} \frac{2(\theta - u)}{\theta^2} dy \\
&= \frac{2u\theta - u^2)}{\theta^2}\Big|_0^y \\
&= \frac{2y\theta - y^2}{\theta^2}, \quad 0 < y < \theta
\end{aligned}$$

b) Show that Y/θ is a pivot.

Let $X = Y/\theta$ with support $\mathcal{X} = \{x : 0 < x < 1\}$. Then

$$\begin{aligned}
F_X(x) &= P(X \leq x) = P(Y/\theta \leq x) = P(Y \leq x\theta) \\
&= F_Y(x\theta) \\
&= \frac{2x\theta^2 - x^2\theta^2}{\theta^2} \\
&= 2x - x^2, \quad 0 < x < 1
\end{aligned}$$

Neither the distribution of X nor the support of X depends on θ, so $X = Y/\theta$ is a pivotal quantity.

c) Use the pivotal quantity to create <u>three</u> different 90% confidence intervals for θ.

We can use the pivot Y/θ to find $P(a \leq Y/\theta \leq b) = 0.90$. Consider $\alpha = 1 - 0.9 = 0.1$. We can distribute this quantity to either side of the confidence interval in any way we want. For example, let 0.05 probability be on the left side of the interval and 0.05 probability be on right side. Then we compute:

$$0.05 = P(X \leq a) = F_X(a) = 2a - a^2.$$

So $a = 0.025$ (note that in solving this, there are two solutions for a. But because of the support of X, we know that a should be between 0 and 1. Only one of the solutions satisfies this. Now we can find b in a similar manner:

$$0.95 = P(X \leq b) = F_X(b) = 2b - b^2.$$

So $b = 0.293$ Now we have that the probability $Y/\theta \geq .025$ and $Y/\theta \leq .293$ is 0.90. Solving the inequalities for θ, we get that there is a 0.90 probability that

$$\theta \leq Y/.025 \quad \text{and} \quad \theta \geq Y/.293.$$

So a 90% confidence interval for θ is $(Y/.293, Y/.025)$.

Note that the 90% does NOT refer to the probability that θ is in this interval. It refers to the probability of obtaining a value y such that $y/\sqrt{0.95}$ is less than θ and $y/\sqrt{0.05}$ is greater than θ.

d) Provide a 90% confidence interval for θ^2.

I'm not entirely sure how to do this, but my thought is to find the distribution of $T = X^2 = (Y/\theta)^2$ and use this in the same way as we did above to find a confidence interval for θ^2. So, first finding the distribution of T, we have

$$F_T(t) = P(T \leq t) = P(X^2 \leq t) = P(-\sqrt{t} \leq X \leq \sqrt{t})$$

The support (range) of X is between 0 and 1, though, so this is equivalent to finding $P(X \leq t)$. Note that the support of T is also between 0 and 1. Continuing on, we have

$$F_T(t) = F_X(\sqrt{t}) = 2\sqrt{t} - t, \quad 0 < t < 1.$$

Now, we want $0.90 = P(a < T < b)$. This time, lets distribute the remaining .10 probability differently. Lets look for a and b such that $P(T \leq a) = 0.02$ and $P(T \leq b) = 0.92$. Then we have

$$0.02 = 2\sqrt{a} - a, \text{ and } 0.98 = 2\sqrt{b} - b$$

Then solving, we have $a = 0.0001$ and $b = 0.737$. So we have that there is a 90% probability that

$$Y^2/\theta^2 \geq 0.0001 \text{ and } Y^2/\theta^2 \leq 0.737,$$

so the corresponding 90% confidence interval for θ^2 would be

$$Y^2/0.737 \leq \theta^2 \leq Y^2/0.0001.$$

Question 2

Suppose $Y_1, ..., Y_n$ are i.i.d. $N(\mu, \sigma^2)$. Define

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad \widetilde{Y} = \frac{1}{n-1}\sum_{i=2}^{n} Y_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \qquad \widetilde{S}^2 = \frac{1}{n-2}\sum_{i=2}^{n}(Y_i - \widetilde{Y})^2.$$

For each of the following, state the distribution of the quantity, whether the quantity is a pivot, and if the quantity be used to create a confidence interval for μ.

Note that $\widetilde{Y}$ is the sample mean of $n-1$ independent, identically distributed random variables and $\widetilde{S}^2$ is the sample variance of the same $n-1$ iid random variables.

a) $\overline{Y} \sim N(\mu, \sigma^2/n)$. This is not a pivot since the distribution of the quantity depends on the unknown parameter μ.

b) $(\overline{Y} - \mu)/\sigma \sim N(0, 1/n)$. This is a pivotal quantity since the distribution does not depend on the unknown parameter.

The given quantity can be rewritten as

$$\frac{\bar{Y} - \mu}{\sigma} = \frac{\sum_{i=1}^{n} Y_i - n\mu}{n\sigma} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - \mu}{\sigma}$$

This is the mean of n standard-normal random variables, so the distribution is $N(0, \frac{1}{n})$

c) $(\overline{Y} - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$. This is a pivot.

d) $(\overline{Y} - \mu)/(S/\sqrt{n}) \sim t_{n-1}$. This is a pivot.

Recall that $\frac{n-1}{\sigma^2}S^2 \sim \chi^2_{n-1}$ and that $\bar{Y}$ and S^2 are independent. Then the given quantity can be rewritten as

$$\frac{\bar{Y} - \mu}{S/\sqrt{n}} = \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}$$

The numerator is a standard normal random variable, and the denominator is the square root of an independent χ^2_{n-1} random variable divided by its degrees of freedom. This is the form for a t_{n-1} distributed random variable.

e) $(Y_1 - \mu)/\sigma \sim N(0, 1)$. This is a pivot.

f) $(Y_1 - \mu)/\widetilde{S} \sim t_{n-2}$. This is a pivot.

Note that $\frac{n-2}{\sigma^2}\widetilde{S}^2 \sim \chi^2_{n-2}$. Using this fact, we can rewrite the given quantity:

$$\frac{Y_1 - \mu}{\widetilde{S}} = \frac{Y_1 - \mu}{\sigma\sqrt{\frac{(n-2)\widetilde{S}^2}{(n-2)\sigma^2}}} = \frac{\frac{Y_1-\mu}{\sigma}}{\sqrt{\frac{(n-2)\widetilde{S}^2}{\sigma^2}/(n-2)}}$$

In the numerator, we have $\frac{Y_1-\mu}{\sigma} \sim N(0, 1)$ and in the denominator, we have the square root of a χ^2_{n-2} random variable divided by it's degrees of freedom. Since $\widetilde{S}^2$ does not depend on Y_1, it is independent of Y_1. So this is the form of a t_{n-2} distributed random variable.

Question 3

The quality of computer disks is measured by the number of missing pulses. Brand X is such that 80% of the disks have no missing pulses. In this problem we will investigate the probability that 15 or more contain missing pulses if 100 disks of brand X are inspected.

a) Approximate this probability using the CLT and a continuity correction.

Let X be the number of disks with missing pulses. Then $X \sim \text{Bin}(100, 0.2)$, and X is the sum of 100 iid random variables with distribution Bernoulli(0.2). Using the CLT, we can approximate this with

$$P(X \geq 15) = 1 - P\left(\frac{X - n\mu}{\sigma\sqrt{n}} \leq \frac{15 - n\mu}{\sigma\sqrt{n}}\right)$$

and use a standard normal distribution to evaluate. WITH a continuity correction, we make a slight alteration by subtracting an additional 0.5 from the original quantile, 15. So we have

$$1 - P\left(\frac{X - n\mu}{\sigma\sqrt{n}} \leq \frac{14.5 - n\mu}{\sigma\sqrt{n}}\right) = 1 - P\left(\frac{X - 20}{4} \leq \frac{14.5 - 20}{4}\right)$$

```r
1-pnorm((14.5 - 20)/4)
```

```
## [1] 0.9154343
```

b) Calculate the exact probability using `R`.

```r
1-pbinom(14, 100, .2)
```

```
## [1] 0.9195563
```