18.05 Problem Set 11, Spring 2022

Problem 1. (20: 10,5,5 pts.) **(Height)**
Suppose μ is the average height of a college male. You measure the heights (in inches) of twenty college men, getting data $x_1, \ldots, x_{20}$, with sample mean $\overline{x} = 69.55$ in. and sample variance $s^2 = 14.26$ in^2. Suppose that the x_i are drawn from a normal distribution with unknown mean μ and unknown variance σ^2.

(a) Using $\overline{x}$ and s^2, construct a 90% t–confidence interval for μ.

(b) Now suppose you are told that the height of a college male is normally distributed with standard deviation 3.77 in. Using the same data as in part (a), construct a 90% z–confidence interval for μ.

(c) In (b), how many people in total would you need to measure to bring the **width** of the 90% z–confidence interval down to 1 inch?

Problem 2. (10 pts.) **Confidence intervals from standardized statistics**
The Beta distribution arises in a surprising way: draw a sample of size n from a uniform(0,1) distribution and let w_2 be the second smallest value. Then it turns out that

$$w_2 \sim \text{Beta}(2, n-1).$$

Now suppose you draw a sample of size n from a uniform$(0, a)$ distribution, where a is unknown. If we let y_2 be the second smallest data value then the standardized order statistic

$$y_2/a \sim \text{Beta}(2, n-1).$$

Use y_2 and `qbeta` in R, to define a 95% confidence interval for a. (Because n and y_2 are not given this will be a general formula not numbers.)

Finally, supposing $n = 9$ and $y_2 = 1.5$, give the 95% confidence interval for a.

Problem 3. (35: 15,10,10 pts.) **Various variances**
Consider a sample of size n drawn from a Bernoulli(θ) distribution. (That is, a draw from a binomial(n, θ) distribution.) In constructing a confidence interval the conservative estimate is that the variance of the underlying Bernoulli distribution is $\sigma^2 = 1/4$ –this is conservative because for any θ we know that $\sigma^2 \leq 1/4$.

(a) In this problem we want to compare how well normal distribution using the conservative estimate matches the one using the true variance

(i) Let $\theta = 0.5$ and $n = 250$. Make a plot that includes

• the pmf $p(x|\theta)$ of the binomial(n, θ) distribution.

• the pdf of the normal distribution with the same mean and variance as the binomial((n, θ)) distribution

• the normal distribution with the same mean (as the binomial distribution) and conservative variance to your plot.

Note: The conservative variance for a Bernoulli(θ) distribution is 1/4. So the conservative variance for a binomial(n, θ) is $n/4$.

Note. It is not reasonable to compare the probabilities in a pmf to the densities in a pdf. In order to make them comparable, but also to make the graphs readable, you should plot the pmf as points, but think of them as the top of a density histogram with bin width 1 and breaks on the half integers.

(ii) Make a similar plot for $\theta = 0.3$ and $n = 250$.

(iii) Make a similar plot for $\theta = 0.1$ and $n = 250$.

In each case, how close are each of the normal distributions to the binomial distribution? How do the two normal distributions differ? Based on your plots, for what range of θ do you think the conservative normal distribution is a reasonable approximation for binomial(n, θ) with large n?

(b) Suppose θ is the probability of success, and that the result of an experiment was 140 successes out of 250 trials. Find 80% and 95% confidence intervals for θ using the conservative variance. (For the 95% interval use the rule-of-thumb that $z_{0.025} = 2$.)

(c) Using the same data as in part (b), find an 80% **posterior** probability interval for θ using a flat prior, i.e. Beta$(1, 1)$. Center your interval between the 0.1 and 0.9 quantiles. Compare this with the 80% confidence interval in part (b).

Hint: Use `qbeta(p, a, b)` to do the computation.