ST 502 R project 1

Overview

For this project you will work in groups of 2. The project involves involves a Monte Carlo simulation study in R (generating data in R to investigate properties of estimators/CIs) and the creation of a small report.

You will either create

- a .R file and a .pdf report
- or a .Rmd file that has both your R code and text along with an outputted .pdf (preferred)

You should not use generative AI for this project for any purpose other than debugging a program.

Project Goal

Consider the common problem of estimating the probability of success, p, from a Binomial random sample. Our goal will be to compare different confidence interval procedures that attempt to capture p.

Methods to Compare

To conduct inference on p, we could use many methods. For instance, we could use $\underset{\sim}{\hat{p}} = Y/n$ (both the maximum likelihood and method of moments estimator) and construct a confidence interval using the CLT (as done in extra practice 3). The observed interval comes out to be

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

This interval (referred to as a Wald interval) is very easy to use but is not the best! The article on the assignment link compares a number of competing confidence interval procedures. We'll recreate parts of their simulations!

Methods to compare (you may need to see the paper for more info):

- Wald interval
- Adjusted Wald interval (section 3 of the paper)
- Clopper-Pearson (exact) interval (top right of page 122, section 1 - for clarity given below)
- Score interval (top left of page 123, section 1)
- Raw percentile interval using a parametric bootstrap
- Bootstrap t interval using a parametric bootstrap

Clopper-Pearson:

$$\left[1+\frac{n-y+1}{yF_{2y,2(n-y+1),1-\alpha/2}}\right]^{-1} < p < \left[1+\frac{n-y}{(y+1)F_{2(y+1),2(n-y),\alpha/2}}\right]^{-1}$$

where $F_{a,b,c}$ is the value from the F-distribution with a numerator and b denominator degrees of freedom that has c to the right of it.

Simulations

Which intervals should we use in practice?

We really need to understand what good properties of a confidence are. The main things to consider are:

- proportion of intervals that capture the true value (hopefully $1 - \alpha$)
- proportion that miss above and proportion that miss below
- average length of the interval

We'd like to compare the performance of these intervals for making inference for various values of p and n.

We want to generate data for combinations of p and n and compare the above items.

General Procedure for Simulations

1. Create a function that calculates and returns the Wald interval. Another function to create the adjusted Wald, another to find the Clopper-Pearson, and another to find the Score interval. Then create a function that finds the two bootstrap based intervals (putting these together saves computational time!)

- For example

```
waldCI <- function(y, n, alpha = 0.05){
  c(y/n - qnorm(1-alpha/2)*sqrt(((y/n)*(1-y/n))/n),
y/n + qnorm(1-alpha/2)*sqrt(((y/n)*(1-y/n))/n))
}
```

- (R functions by default return the last thing done. Here we create a vector with the lower and upper end point as our last (and only) thing. You can use `if else` logic to do the special cases if needed.)

2. Generate $N = 1000$ random samples from a binomial where n varies across 15, 30, to 100 and p varies from 0.01 to 0.99 (15 total values of p). (`rbinom(1000, size = 15, prob = 0.01)` does the trick for one combination of n and p!)
3. Create and save 95% CIs for p using the six methods mentioned above. For the bootstrap methods, use $B = 100$ resamples.

   - Note: For some intervals you will need to have a special case for when $y = 0$ or $y = n$ is observed as the intervals will not be able to be calculated for those observed values. Instead, set those intervals to $[0, 0]$ and $[1, 1]$, respectively.
   - For the bootstrap, occasionally your bootstrap resamples will give sample proportions of 0 or 1. This leads to a t-statistic that is either -Inf or Inf. In those cases, just throw out the values. That is, don't use them in the quantile calculation for the bootstrap t interval.

   - I'd get your code working with a $N = 10$ and $B = 20$. Then up those to the correct values and let the computations go wild - it may take 10-20 minutes or so to run.

4. Calculate the properties mentioned above for all combinations of n and p using the simulation results.

Report

The code to do the above should be turned in with the report that is outlined below.

Create a small report (2-3 pages or so - not counting code/graphs/tables). **Place the code and plots directly into the report where it makes sense to talk about them.** Your report should discuss

- the goals of your simulation study
- the methods used and how they are calculated
- your creation of data process and code
- your code to execute the Monte Carlo study
- the results of your study
- the conclusions about which intervals you favor

Feel free to source some of that from the paper mentioned above (with a citation to be clear about which parts are your ideas and which are theirs).

You should report the coverage results for each sample size/method in plots similar to those on the bottom of page 123 of the paper.

Similary, report the average length using similar plots (this wasn't done in the paper but is useful). Produce standard errors of these average lengths as well.

Rubric for Grading (total = 100 points)

Item	Points	Notes
Report introduction/goals	15	Worth either 0, 2.5, . . . , 15
Methods section	25	Worth either 0, 2.5, . . . , 25
Data Creation	10	Worth either 0, 2.5, . . . , 10
Confidence Interval Functions	30	Worth either 0, 2.5, . . . , 30
Results/graphs/conclusions	20	Worth either 0, 2.5, . . . , 20

Notes on grading:

- For each item in the rubric, your grade will be lowered one level for each each error/issue.
- You should use Good Programming Practices when coding (see wolfware). If you do not follow GPP you can lose up to 30 points on the project.