

Bootstrap function (sample, method, bootstrap.resamples):

- Check inputs
- Different methods: nonparametric percentile, nonparametric BCa, nonparametric smooth, parametric percentile

For parametric, need to ask user for distribution to sample from, or estimate it from the data? The latter might be too general? Maybe user inputs the distribution but we calculate our own MLEs for the parameters? Will be a nice way to do this, will figure it out once the other methods are coded up.

- Calculate statistic on resampled data
- If BCA method : change quantiles to be evaluated
- Return CI using bootstrap distribution
- Try implementing as many methods as possible using replicate() to make the code go faster?

Simulation function (samples.sizes, bootstrap.resamples, bootstrap.methods, simulations.per.case) :

- Uses nested for loops to call the bootstrap function simulations.per.case times for each setting
- Produce all the results in a 4 dimensional array with dimensions [sample.sizes, bootstrap.resamples, bootstrap.methods, confidence intervals produced]

Summary function(simulation function output):

- Replaces each vector of confidence intervals in the array with the observed coverage
- Maybe other statistics of interest too if there's enough time?

Plotting function(summary function output):

- Takes a summary object and plots the coverage for each method
- Maybe using matplotlib? Shown by Eric for 5751 Mark-Recapture simulation project – Would be nice to have all methods on the same graph...
- Do two plots : One with sample.size on x axis and one with bootstrap.resamples
- What to do with the other dimension? Could either set it to its highest value, or take the average across the other dimension? Highest value will be easier, start with that?
- If there's time, do a 3D plot to have sample.size, bootstrap resamples AND coverage all in a single plot? – Try a few different packages?

Driving code:

- Set seed.
- Call simulation function on normal and poisson data
- Get the summaries using the summary function
- Do some plots using the plot function
- (try a few different settings for the comparisons to illustrate the difference that seem to be the most significant). i.e make sure to have a larger range of values at small sample sizes where the BCa is expected to do better, have some values with larger resamples (999?) to see which methods do better as computation increases, BCa will probably be better than percentile at any given sample size when resamples is high? Percentile

should be pretty good when sample size is high, given it approximates the true distribution ok? Include some sample sizes which are quite large, 500 or 1000?

Other code:

- Do pilot simulations to decide on how many simulations is reasonable, want computing time to be < 5 hours as will probably have to be run 4 or 5 times before the deadline as settings get tweaked and I get more of an idea of which settings are better?