STAT 641: BOOTSTRAPPING METHODS

Jiyoun Myung

Department of Statistics and Biostatistics
California State University, East Bay

Spring 2021, Day 8

Review

Jackknife Resampling

- The bootstrap method is not always the best one. One main reason is that the bootstrap samples are generated from $\hat{F}$ and not from F.
- The jackknife method is a leave-one-out strategy of the estimation of a parameter. Thus, if there are n data points in the sample, the jackknife sampling technique will consist of n samples each with $n - 1$ data points in each sample subset analysis.
- The bootstrap involves drawing random samples from the sample observations while the jackknife does not involve any randomness.
- Cons: still fairly computationally intensive, does not perform well for non-smooth (e.g. the median) and nonlinear statistics (e.g. Correlation coefficient), requires observations to be independent of each other - meaning that it is not suitable for time series analysis.

delete-d-jackknife

- The jackknife method of estimation can fail if the statistic $\hat{\theta}$ is not smooth. Smoothness implies that relatively small changes to data values will cause only a small change in the statistic.
- The jackknife is not a good estimation method for estimating percentiles (such as the median), or when using any other non-smooth estimator.
- **delete-d-jackknife** is to handle cases of non-smooth estimators.
- There will be $\binom{n}{d}$ jackknife samples of size $n - d$ and therefore $\binom{n}{d}$ jackknife replications, and the estimate of the standard error is

$$\left(\frac{n-d}{d\binom{n}{d}} \sum \left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)} \right)^2 \right)^{1/2}$$

  where the summation is take over all $\binom{n}{d}$ subsets of points.
- In practice, if n is large and d is chosen such that $\sqrt{n} < d < n$, then the problems of non-smoothness are removed.

Resource: https://math.montana.edu/jobo/thainp/jack.pdf

Review

Comparing Bootstrap CIs

CI	Symmetric	Range Resp	Trans Resp	Accuracy	Normal Samp Dist?	Other
BS SE	Yes	No	No	1^{st} order	Yes	param assump $F(\hat{\theta})$
BS-t	No	No	No	2^{nd} order	Yes/No	computer intensive
perc	No	Yes	Yes	1^{st} order	No	small $n \rightarrow$ low accuracy
BCa	No	Yes	Yes	2^{nd} order	No	limited param assump

Review

infer package

- Part of tidymodels
- 4 main verbs: specify(), hypothesize(), generate(), calculate()
- Example: comparing two proportions

Presentation today

rsample package

- Part of tidymodels
- Description of **bootstraps** function A bootstrap sample is a sample that is the same size as the original data set that is made using replacement. This results in analysis samples that have multiple replicates of some of the original rows of the data. The assessment set is defined as the rows of the original data that were not included in the bootstrap sample. This is often referred to as the “out-of-bag” (OOB) sample.

Double Bootstrap Confidence Interval

- Whether the basic or percentile bootstrap method is used to calculate confidence intervals, there is a possibly non negligible difference between the nominal $1 - \alpha$ coverage and the actual probability coverage of the interval in repeated sampling, even if B is very large.
- One common use of the double bootstrap is to estimate the standard error of estimate for each bootstrap resample when no formula is known for it. This nests the bootstrap estimation of the standard error within the bootstrap resampling for the desired statistic.

Presentation today

Bootstrapped Timeseries Forecast Confidence Intervals