An Overview of Bootstrapping: Properties and Variations

RTG: Modern Tools in Statistics and Applications

Dmitriy Izyumin

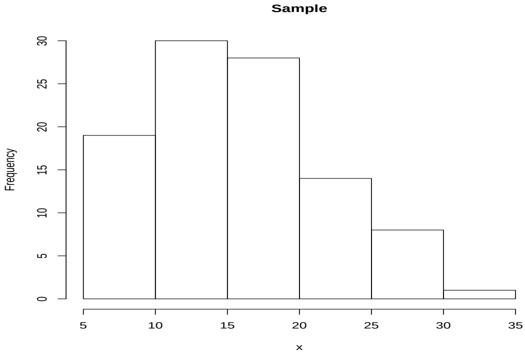
June 1, 2018

Overview

1. Basic idea and methodology of the bootstrap.
2. Some variations and extensions of the bootstrap.

Motivating Example

- I took a random sample from a population, and I would like to use that sample to estimate the population mean.
- The sample mean is $\bar{x} = 15.66$, and the histogram is below.



Motivating Example

- Since the sample mean is a random variable, I now need to know how the value of the sample mean would vary from sample to sample.
- Ideally, I'd like to see the sample mean values of other samples from the same population. Then I would have some sense of the sampling distribution. This is usually not feasible.
- In some cases, it's possible to derive the sampling distribution using theoretical results.

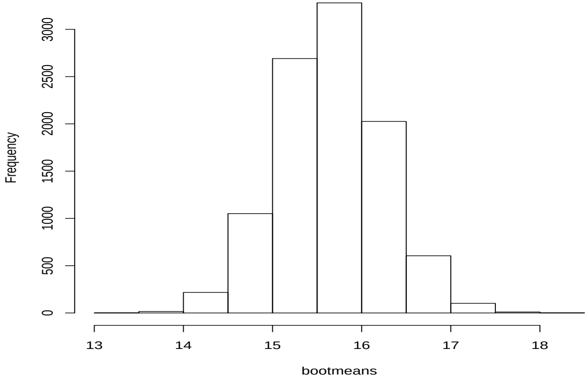
Motivating Example

- I can't draw more samples from the population.
- However, I can *resample* the observed values to create more samples that resemble the original sample. I can then compute the sample means of those samples, and see how they vary from sample to sample.
- Idea: If my original sample is a good representation of the population, then the statistics obtained from resampled data will behave similarly to statistcs obtained by sampling from the population.

Motivating Example

- I constructed 10000 new samples of size 100 by resampling from the original sample *with replacement*.

- Below is a histogram of 10000 sample means obtained from these samples.



What is Bootstrapping?

- Introduced by Bradley Efron in 1979
- Data-driven resampling procedure
- Useful for estimating the sampling distribution of a statistic
- Can help to obtain measures of estimator quality (bias, variance, confidence intervals, percentiles, etc.)

Setting

- Population with distribution F (possibly unknown)
- We are interested in a parameter θ of F.
- Take a random sample $(x_1, \cdots, x_n)$ of size n, and compute a sample estimate T_n of θ.
- T_n is based on a random sample, so we observe only one of many potential possible values.
- Now we would like to know how T_n varies relative to θ. In other words, we would like to know the sampling distribution of T_n.

Setting

- Population: distribution F, parameter θ
  Sample: statistic $T_n = \hat{\theta}$
- We would like to know how T_n varies relative to θ.
  - It's not feasible to take more samples, and compute more realizations of T_n.
  - There may be theoretical results that specify the sampling distribution of T_n under some assumptions.
  - Bootstrapping is another alternative.

Standard Bootstrap Approach

- Population: distribution F, parameter θ
  Sample: statistic $T_n = \hat{\theta}$

- For some large number B, repeat the following B times:
  1. Obtain a sample of size n by sampling with replacement from the original sample.
  2. Compute T_n^* from the sample.

- Observe how the bootstrap estimates $T_{n,1}^*, \cdots, T_{n,B}^*$ vary around T_n.

- This is an approximation of how the possible values of T_n vary around θ.

Another Viewpoint

- The empirical distribution based on the sample is

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \leq x\}}.$$

- Bootstrap samples are samples of size n taken from $\hat{F}_n$.
- Idea: If $\hat{F}$ is a good approximation of F, and T_n is a smooth enough function, then the bootstrap distribution of T_n^* will be similar to the sampling distribution of T_n.

Favorable Properties

- Simple to implement
- Nonparametric (though parametric versions exist)
- Data-driven and automatic
- Many types of bootstrapping for different scenarios

When to Bootstrap

Bootstrapping is useful when...

1. The true sampling distribution of the statistic is hard to derive.
2. The assumptions needed for usual inference are clearly violated

When NOT to Bootstrap

- The bootstrap distribution of T_n^* will approximate the sampling distribution of T_n if:
  1. $\hat{F}_n$ is close enough to F
  2. T_n is a smooth enough mapping
- If these assumptions do not hold, bootstrapping without some corrections may produce inaccurate results.
- See Hall, 1991 for technical details.
- See Canty, *et al*, 2006 for examples.

Animated Examples

Here are some neat animated examples of the bootstrap.

▸ Animations

Animation design led by Chris Wild. Animations built in R by Keng Hao (Danny) Chang.

Parametric Bootstrap

- Suppose the population distribution $F(\theta)$ is known (up to parameter θ).
- Note: this does not mean we know the sampling distribution of T_n, which may depend on F in some complicated way.
- Non-parametric bootstrap: Take samples from $\hat{F}_n$.
- Parametric bootstrap: Take samples from $F(T_n)$.

Block Bootstrap

- If the data has a dependence structure (time or spatial), then standard bootstrap samples will not retain that structure.

- Instead, take "blocks" of $b < n$ consecutive observations at a time.

- Paste those blocks together to create bootstrap samples.

- The value b needs to be large enough to capture local dependence structure.

Bag of Little Bootstraps

- If n is very large, the standard bootstrap becomes computationally expensive or unfeasible.
- BLB is a scalable extension of the bootstrap to massive data.
- Retains favorable properties of the bootstrap.
- Less demanding computationally, and easy to parallelize.
- Kleiner, A. *et al*, 2012

BLB - Steps

1. Start with sample of size n.
2. Obtain s subsamples of size $b < n$ without replacement.
3. Carry out bootstrap on each subsample using r bootstrap samples of size n.
4. Compute bootstrap results for each subsample.
5. Average over the results from the subsamples.

Why Size b and not n?

- Each resample has *at most* b unique points.
- Scales in b with respect to computation time and storage space.
- Well-suited for parallel computing.
- Under standard assumptions, BLB is consistent.
- If b and s increase reasonably fast with n, then BLB has the same higher-order correctness as the bootstrap; i.e. same convergence rate!

Bagging - Bootstrap Aggregating

- Machine learning ensemble approach.
- Average predictions of models trained on bootstrap resamples.

Motivation

- Some predictors (e.g. CART, neural nets) are inherently unstable, and sensitive to perturbations in training data.
- Fitting a predictor to different arrangements of training data can help us understand and correct the instability.
- "The vital element is the instability of the prediction method. If perturbing the learning set can cause significant changes in the predictor constructed, then bagging can improve accuracy." - Leo Breiman, 1994
- Reduces estimator variance and tendency to overfit.

References I

B. Efron.
Bootstrap methods: Another look at the jackknife.
The Annals of Statistics, 7(1):126, 1979.

Peter Hall.
The bootstrap and Edgeworth expansion.
Springer, 1995.

Angelo J. Canty, Anthony C. Davison, David V. Hinkley, and Valrie Ventura.
Bootstrap diagnostics and remedies.
Canadian Journal of Statistics, 34(1):527, 2006.

Hans R. Kunsch.
The jackknife and the bootstrap for general stationary observations.
The Annals of Statistics, 17(3):12171241, 1989.

References II

Ariel Kleiner, Ameet Talwalkar, Purnamrita Sarkar, and Michael I. Jordan.
A scalable bootstrap for massive data.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(4):795816, 2014.

Leo Breiman.
Bagging predictors.
Machine Learning, 24(2):123140, 1996.