

Bootstrapping

- Bootstrapping is a statistical procedure that resamples a dataset to create many simulated datasets.
- Uses technique called resampling with replacement
- Allows us to estimate confidence intervals, standard errors, hypothesis testing of numerous sample statistics
- First published by Bradley Efron in 1979

Technical Jargon:

- Resampling
- Sample Statistic
- Sampling Distribution

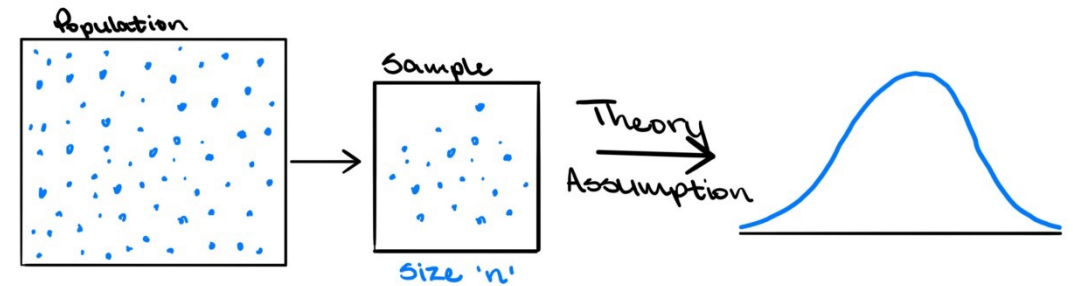
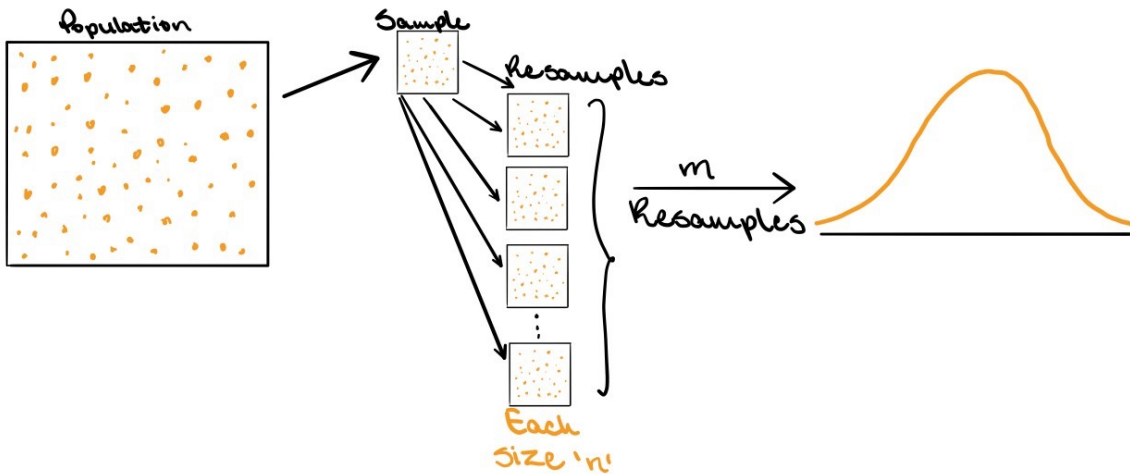
Resampling with Replacement

Toy Sample = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

10 bootstrapped samples:

```
[1, 2, 3, 3, 4, 5, 7, 8, 10, 10]
[2, 3, 3, 4, 4, 4, 4, 7, 8, 10]
[1, 1, 2, 3, 3, 4, 4, 6, 9, 10]
[1, 2, 3, 4, 5, 6, 7, 10, 10, 10]
[2, 4, 4, 5, 7, 8, 8, 9, 9, 10]
[3, 4, 4, 7, 7, 8, 9, 9, 10, 10]
[1, 1, 2, 2, 3, 5, 5, 6, 8, 10]
[1, 2, 2, 2, 4, 4, 6, 7, 10, 10]
[1, 1, 3, 4, 5, 5, 7, 8, 10, 10]
[1, 4, 5, 8, 8, 8, 9, 9, 10, 10]
```

Bootstrapping vs Traditional Approach



Advantages of Bootstrapping:

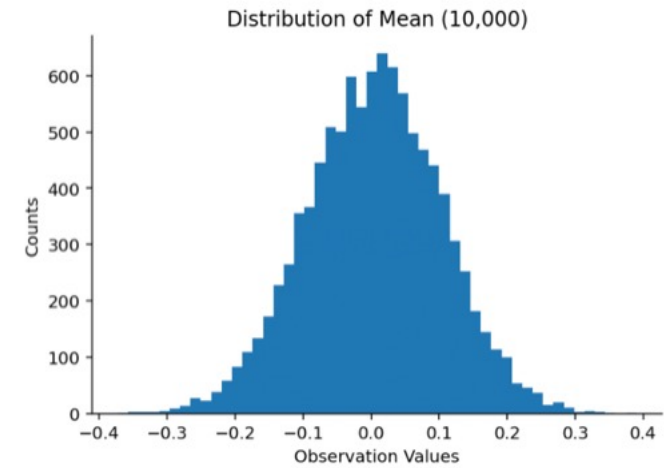
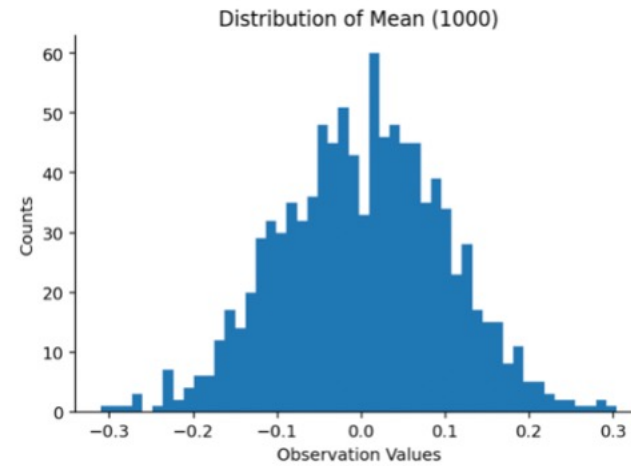
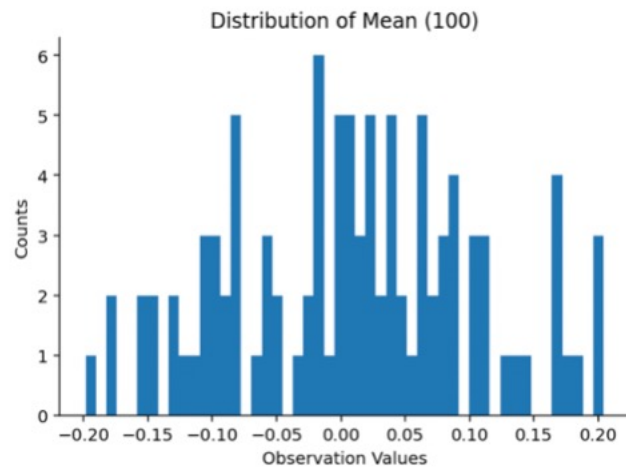
1. Easy to understand
2. Doesn't make any assumptions about the distribution of the data

Applications of Bootstrapping

- Means
- Median
- Decision Trees (Bagging)

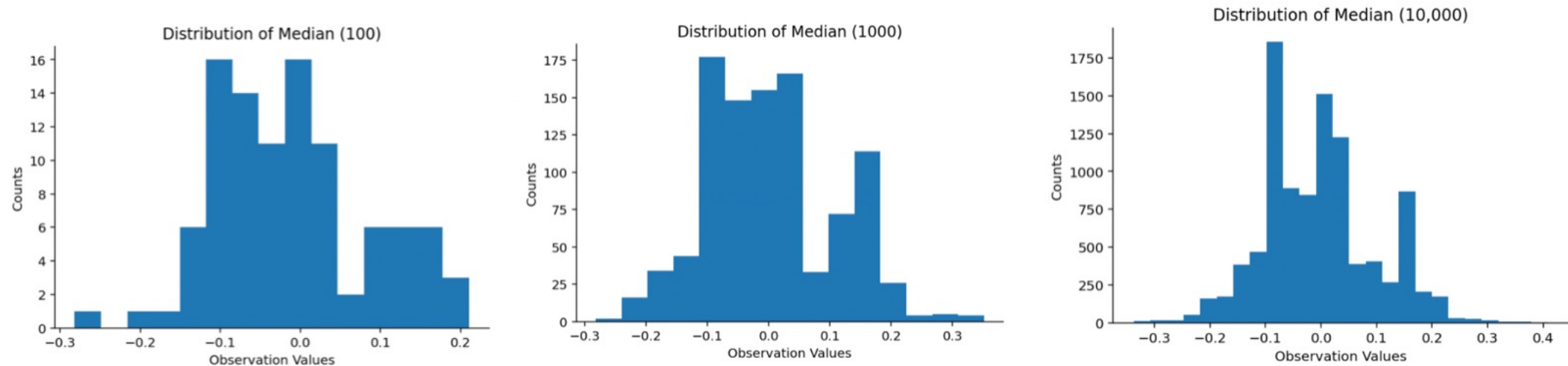
Means

Sample: 100 random samples from a standard normal

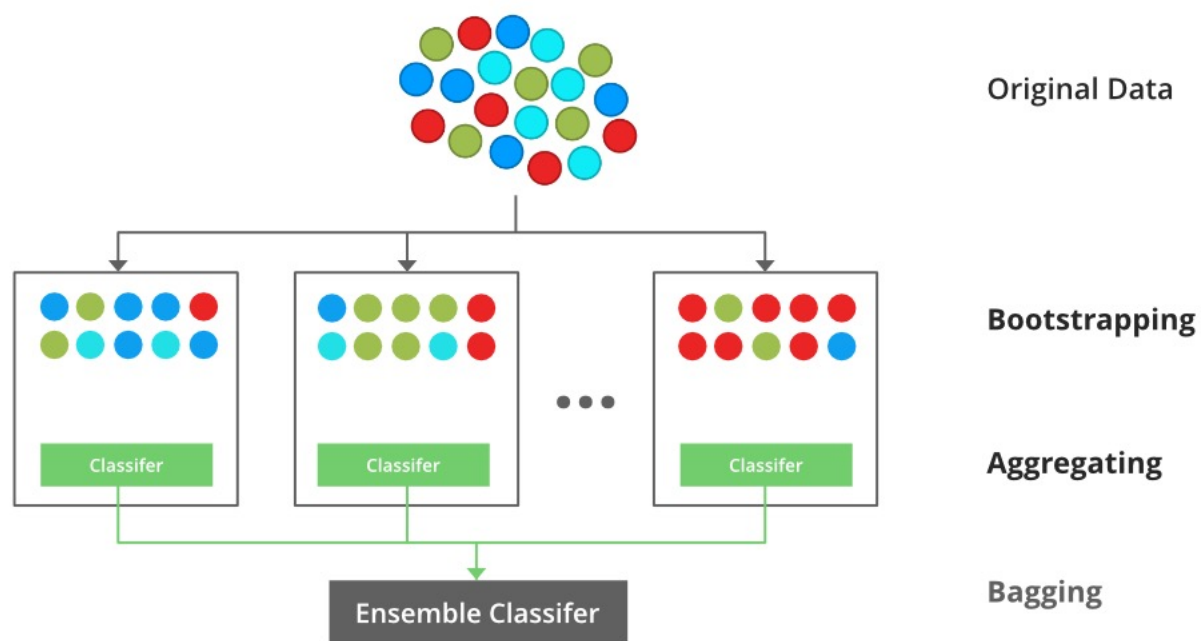


Medians

Sample: 100 random samples from a standard normal



Bagging or Bootstrap Aggregation



- Questions?