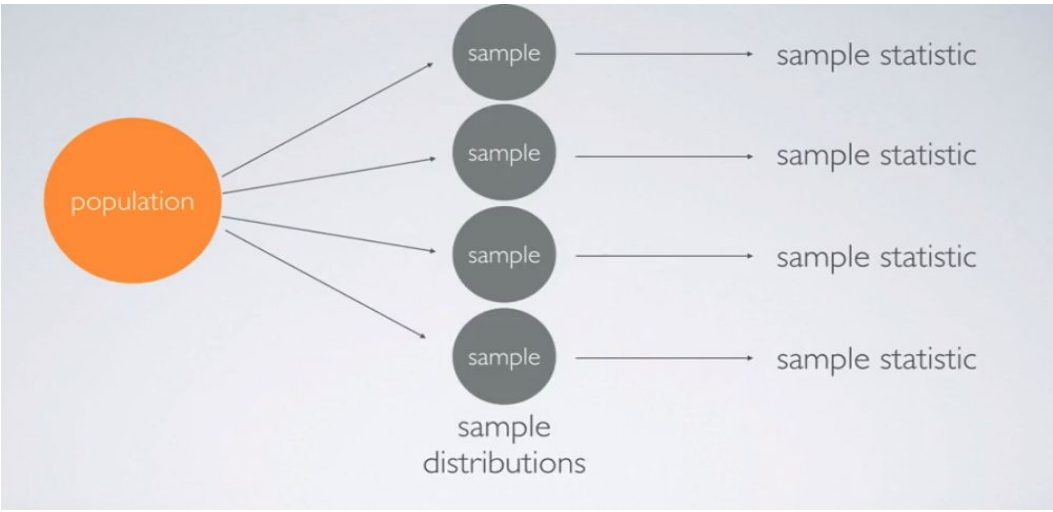
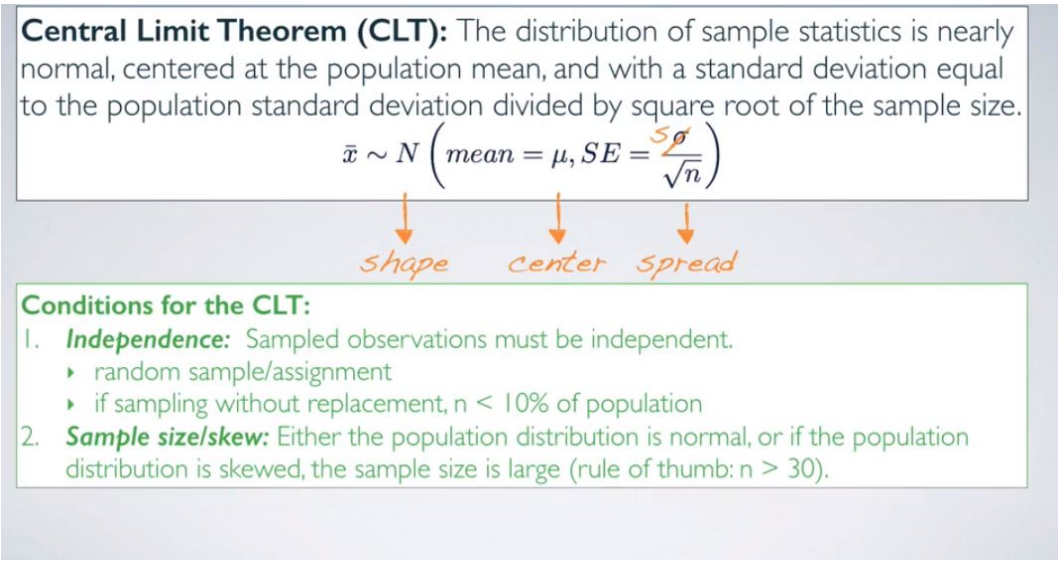
Notes on Sampling Vs Bootstrap Distributions

Made by Aparna Shastry

Sampling Distribution: Consider a population of size N, and samples of size n taken from that population. There are two ways to take samples (also known as sampling the population): with replacement or without replacement. Without replacement sampling requires that n < 0.1N or preferably, < 0.05N

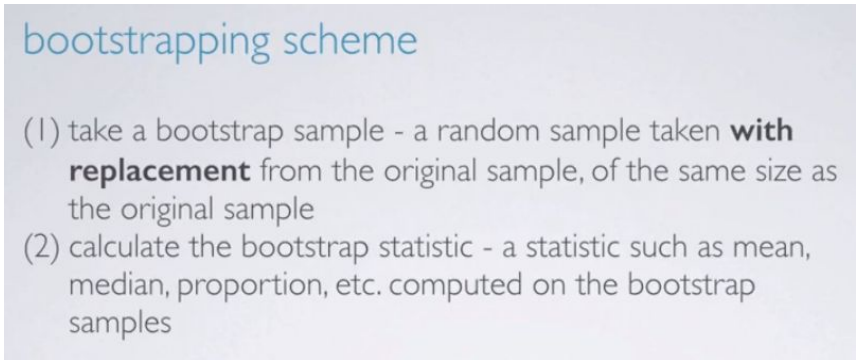


Sample statistics in this case is mean. CLT says how the mean (also called sample statistic) of these samples is distributed. Note that sample distribution and sampling distributions are different. Sample distribution is same as original population distribution, sampling distribution is distribution of mean of samples taken from the population.

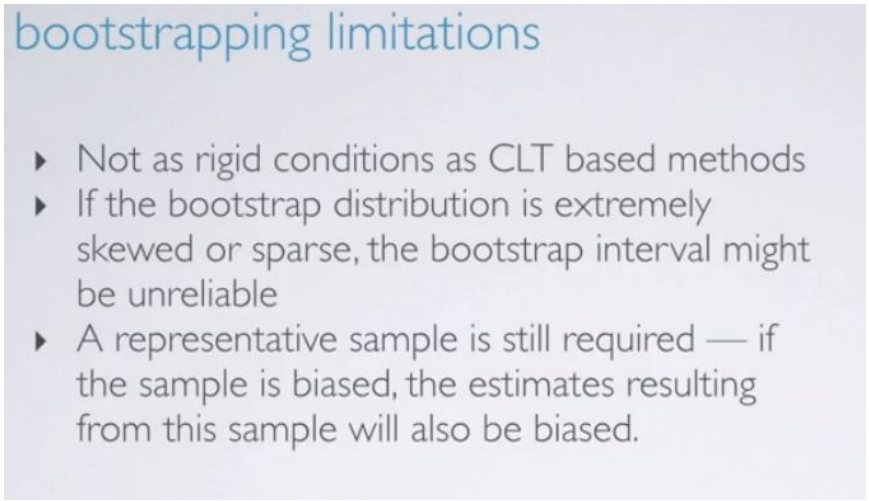


If population is not nearly normal, the more skewed it is, the larger n we need, for the CLT to apply.

Bootstrap Distribution: We don't have access to population almost always. We have access to one sample of the population. Then we do bootstrapping to simulate the effect of taking multiple samples from the population.



Distribution of sample statistic of bootstrap sample taken from a sample with replacement, of same size as original sample



Summary:

- Sampling Distribution is created by sampling with or without from the population. The statistics under consideration is mean or standard deviation
- Bootstrap Distribution is created by sampling with replacement from one sample of the population. The statistics under consideration can be mean, median, proportion or standard deviation
- Both are distributions of sample statistics
- Both are useful in estimating true population statistics.

Images: Screenshots from slides of Duke university coursera course on inferential statistics