Math 662: Probability Distributions

Kevin Rodriguez

Bootstrapping and Resampling

What is Bootstrapping?

How it works?

Advantages of Bootstrapping

Application of Bootstrapping

Bootstrap Distribution

Boot Library in R

Goal

- Our goal is to find a confidence interval for μ
- But our sample size is less than 30 and there is no assumption of Normality

Two Solutions:

- One solution is to find or identify a distribution that is suitable for the population
- Do not assume any Distribution for the population

Two Versions of Bootstrapping:

- Parametric Bootstrapping: Simulating multiple samples from the assumed distribution
- Non-parametric Bootstrapping: Simulating multiple samples directly from the data

What is Bootstrapping?

Bootstrapping is a statistical technique that falls under resampling. Just like standard ways to find confidence intervals, we can use Bootstrapping to estimate a population parameter.

- Bootstrapping is relatively new
- The method was first used in a 1979 paper by Bradely Efron
- The name "Bootstrapping" comes from the phrase "To lift himself up by his bootstraps."
- It is a metaphor for accomplishing an "impossible" task without any help

4/16

How it works?

- Bootstrapping is just sampling with replacement
- Suppose we take a random sample of 10
- Then we consider our random sample as our "population"
- From this we take B samples of 10 from the random sample with replacement
- We generally want B to be between 10,000 and 100,000

In General

- We are interested in estimating a population parameter say, θ
- Let $X = (X_1, X_2, \ldots, X_n)$ be a random sample from an unknown distribution
- Let $\hat{\theta}$ be the estimate for θ
- We then take random sample with replacement of lenght n
- We call these samples, Bootstrap samples = $(X_1^*, X_2^*, \ldots, X_B^*)$
- From the Bootstrap Samples when can calculate estimates for each
- Let $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$ denote the estimates for each Bootstrap samples.

6/16

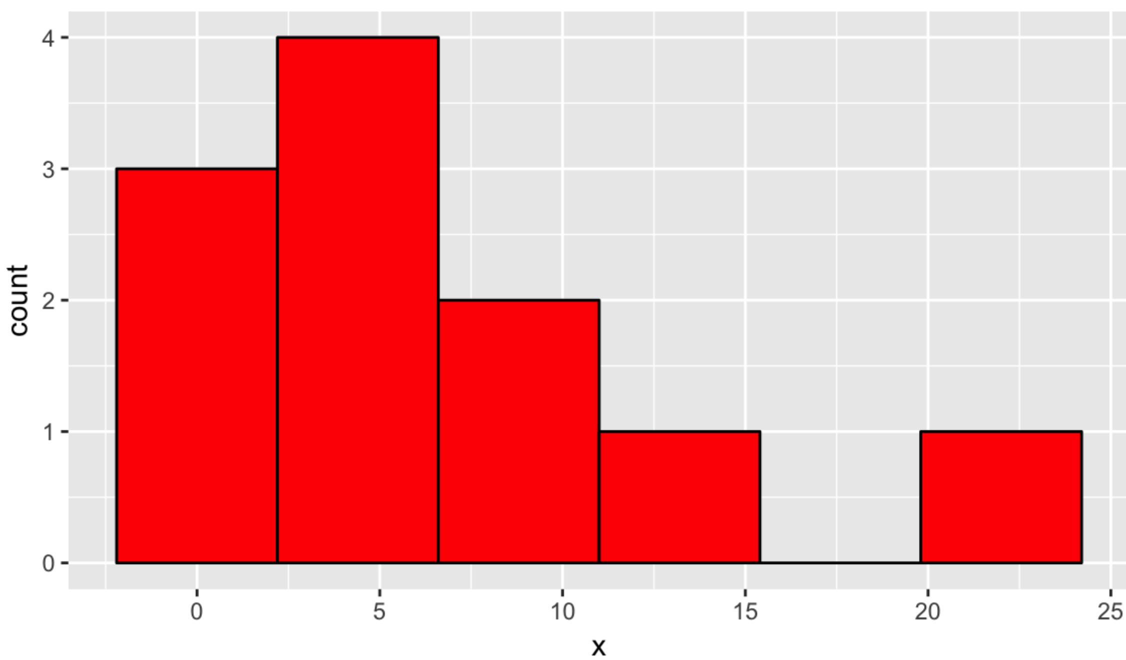
Advantages

- Very quick to perform
- No assumptions about Distribution
- Samples need not be large
- Typically only done when samples are small
- Can be used to estimate any population parameter

Application

```
library(ggplot2)
x <- c(3,13,7,5,6,0,2,4,1,22,9)
mean(x)
```

```
[1] 6.545455
```



Bootstrapping

```
B <- 100000

B_samples <- replicate(B ,sample(x, replace = TRUE))
dim(B_samples)
```

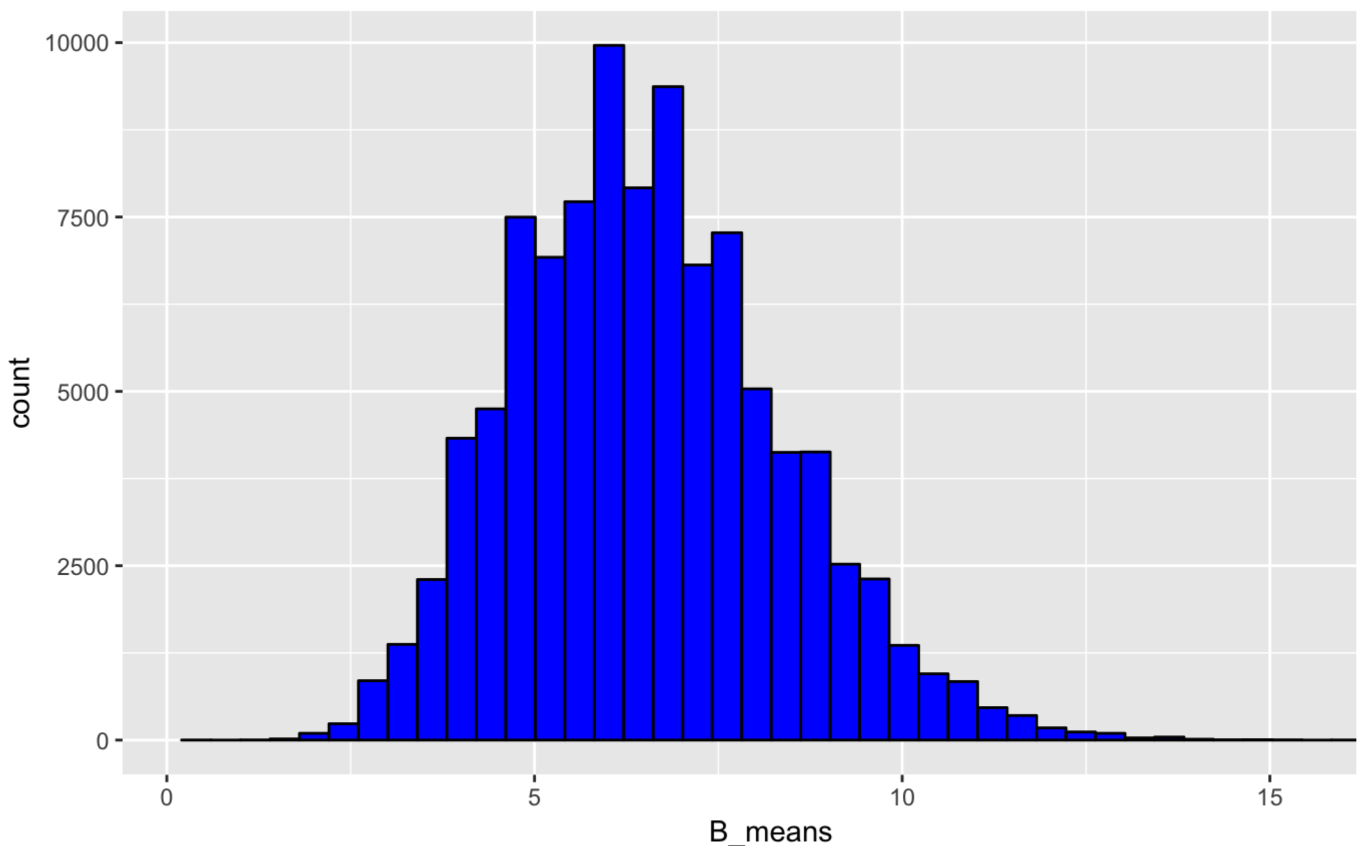
```
[1]     11 100000
```

```
B_means <- apply(B_samples,2,mean)

B_means <- sort(B_means)
q<- quantile(B_means, c(.025,.975))
q
```

```
     2.5%     97.5%
 3.363636 10.454545
```

Bootstrap Distribution




```
## [1] 6.540482
```

10/16

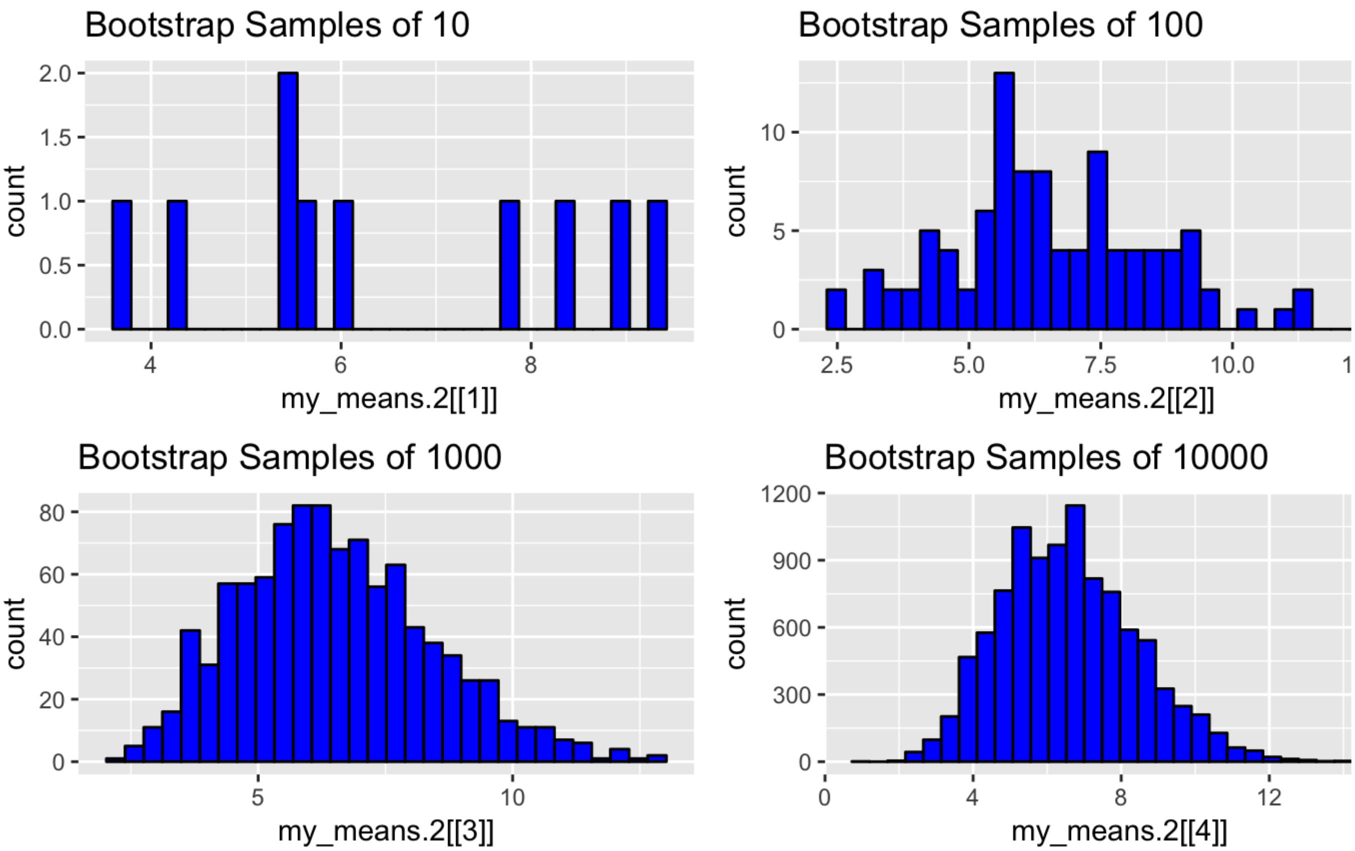
Theorem 4.2.1 (Central Limit Theorem):

Let $X_1, X_2, \ldots, X_n$ denote the observation of a random sample from a distribution that has a mean μ and finite variance σ^2. Then the distribution function of the random variable $W_n = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ converges to Φ, the distribution of the $N(0, 1)$ distribution as $n \to \infty$.

Why is the Central Limit Theorm Important?:

- Since we are taking repeated samples of our "population" then our Bootstrap Distribution will be approximately normal
- The mean of the Bootstrap Distirbution will approximate the mean of the sampling distribution
- The standard deviation of the Bootstrap Distribtuion will be an estimate for the standard error

Comparison of Bootstrap Distributions



Boot Function

```
ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = x, statistic = f1, R = B)


Bootstrap Statistics :
    original      bias    std. error
t1* 6.545455 0.001081818    1.823193
```

boot.ci Function

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 100000 bootstrap replicates

CALL :
boot.ci(boot.out = boot.out, conf = 0.95, type = c("perc", "bc

Intervals :
Level     Percentile            BCa
95%   ( 3.364, 10.455 )   ( 3.818, 11.455 )
Calculations and Intervals on Original Scale
```

References:

- Efron, B. Bootstrap Methods: Another Look at the Jackknife. Ann. Statist. 7 (1979), no. 1, 1–26. doi:10.1214/aos/1176344552. https://projecteuclid.org/euclid.aos/1176344552

- Haukoos, Jason S., and Roger J. Lewis. “Advanced Statistics: Bootstrapping Confidence Intervals for Statistics with ‘Difficult’ Distributions.” Academic Emergency Medicine, Blackwell Publishing Ltd, 28 June 2008, onlinelibrary.wiley.com/resolve/openurl?genre=article&sid=nlm%3Apubmed&issn=1069-6563&date=2005&volume=12&issue=4&spage=360.

- Hogg, Robert V., et al. “Chapter 4: Some Elementary Statistics, Section 2: Confidence Interval.” Introduction to Mathematical Statistics, 7th ed., Pearson, 2013.

- Singh, Kesar, and Minge Xie. “Bootstrap: A Statistical Method.” www.stat.rutgers.edu/home/mxie/stat586/handout/B