EE - 598
Probabilistic Simulation Models
Project 3
USC ID- 6571916065

BOOTSTRAPPING

Bootstrapping can be considered as one of the re-sampling methods and is used to measure the accuracy of a sampling estimate – bias, variance, confidence intervals etc.

Two situations when Bootstrapping is particularly useful are:

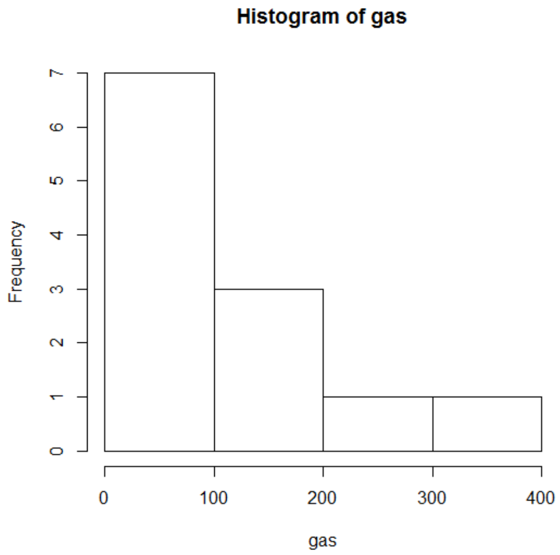
- The distribution model to which data complies isn't known: there is no information about the type of distribution of data and hence Bootstrapping helps to assess the properties of whichever distribution underlies the sample data without making any assumptions about it.
- Sample size is not enough: Since Bootstrapping employs re-sampling; it enables close estimation of a statistic from a limited sample data of a population.

Problem 1:
To find the 95% Confidence Interval of the 'NJGAS' data set.

NJGAS dataset –

No.	Data from 'NJGAS'
1	150
2	367
3	38
4	12
5	11
6	134
7	12
8	251
9	63
10	8
11	13
12	107

Results:



Descriptive statistics about the data using **summary** and **length** in **R** gives the following results –

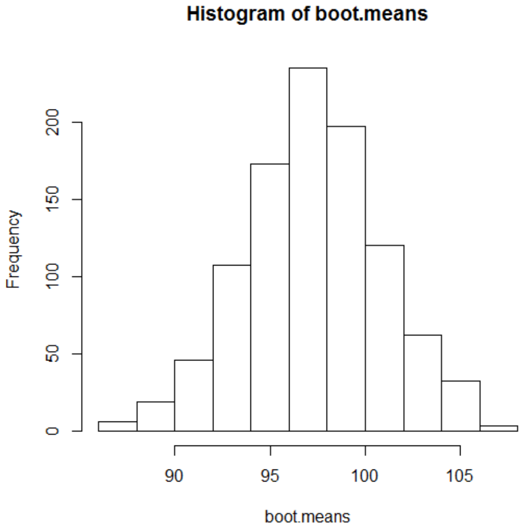
```
> summary(gas)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   8.00   12.00   50.50   97.17  138.00  367.00
> length(gas)
[1] 12
```

We see that the point estimate is **97.17**
We then use Bootstrapping to provide 95% confidence interval of the true mean of the population using the given data.

The number of samples is taken to be **1000.**

```
Quantiles :
      25%       50%     97.5%
 94.92675  97.24850 104.48880
```

The 50% mark lies at **97.24850** lies pretty close to the point estimate, and we can say with 95% confidence interval that the true mean lies in **[94.92675, 104.48880]**



Problem 2:

95% confidence interval by bootstrapping the MC Integral Estimates for $\mathbf{x^{-0.5}}$ **in [0.01, 1]** (from Project 2b)

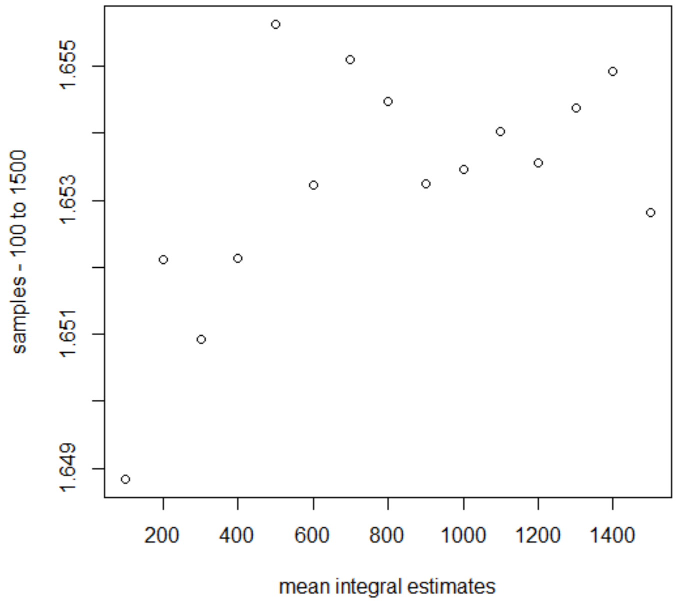
The sample size is varied on each run (k = 15 runs in total) starting from n = 100 to n = 1500, incrementing by 100 on each iteration.

However, the true integral is about 1.666667.

The **95% confidence intervals** for the 15 runs of integral estimates are as below:

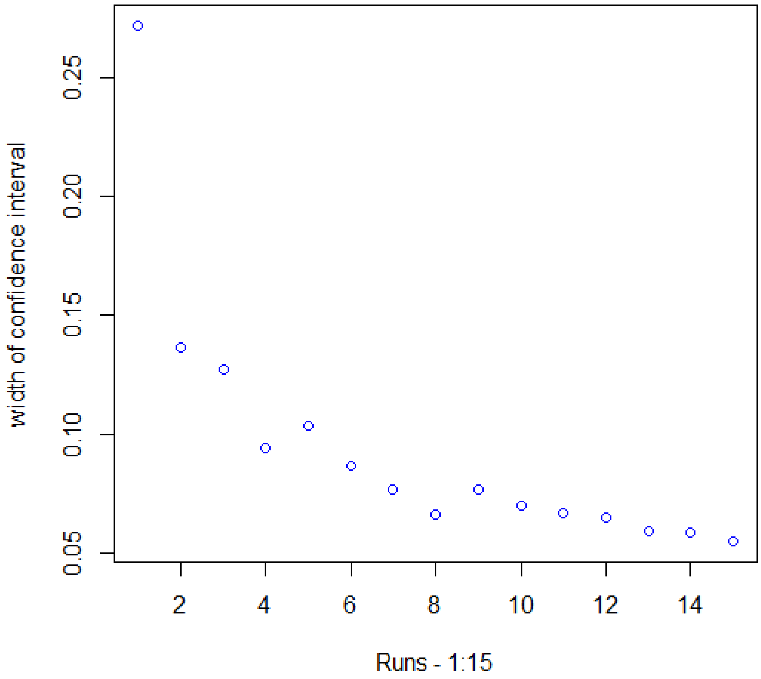
```
     25%     97.5%
1.589855 1.861482
1.613738 1.750318
1.619474 1.746850
1.626481 1.720752
1.629578 1.733110
1.628361 1.715136
1.635745 1.712693
1.638401 1.704553
1.634361 1.710888
1.636553 1.706279
1.636737 1.703396
1.637861 1.702925
1.639098 1.698359
1.639834 1.698426
1.639107 1.694250
```

The integral mean estimates according to Bootstrapping are represented in the scatter plot below:



It is clear that as the sample size is increased, the range in which the estimate lies is elevated. However, the effect of the increase in sample size is more pronounced in analyzing the confidence interval widths.

The variance bias is considerably reduced with the increase in sample size – in other words, the width of the confidence interval decreases. This leads to a better accuracy of the estimate of mean (or the true integration value).


width of confidence interval
Runs - 1:15