Course Assignment – A Simulation Exercise

Author: Sahil Grover

Overview: In this report, we will do 1000 simulations of a fixed data of exponentials. We will:

- Sample the data with replacement
- Compare the statistic properties i.e. mean and variance (theoretical vs. sample)
- Also we will check if the distribution of averages is approximately a normal distribution.

R Code (Simulations):

```
# Producing data of 40 exponentials with lambda=0.2
        data <- rexp(40,0.2)
# Sampling the data for 1000 simulations
        sampled_data <- matrix(sample(data,40*1000,replace = TRUE),nrow=1000,ncol=40)
# Finding the averages of simulations
        sample_averages <- apply(sampled_data, 1, mean)
# Comparing statistic properties
        compare_average<- c(mean(data),mean(sample_averages))
        compare_variance<- c(var(data),var(sample_averages))
# Plotting the sampled_average
        hist(sample_averages,col="PINK")
```

As mentioned in the above R code, the averages and variance are stored in compare_average, compare_variance.

Theoretical Mean versus Sampled Mean

```
> compare_average
[1] 5.111837 5.097333
```

We can see that the average remain the same as the sampled data represent the original data.

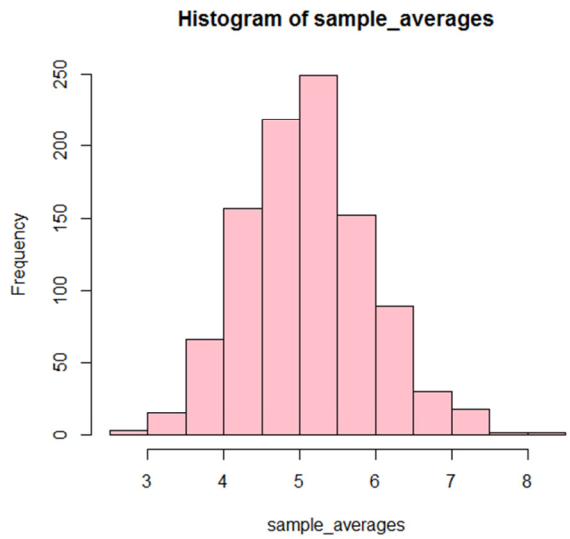
Theoretical Variance versus Sample variance

```
> compare_variance
[1] 27.8007031  0.6891256
```

The theoretical variance (27.8007031) and sampled variance (0.6891256) has a huge difference. This is primarily because of population size. There were only 40 observations in data while sampling was done with 1000 simulations effectively taking 40*1000 = 40,000 sample observations in account.

Distribution

Now, see below the histogram of the distribution of sample_averages.



Since the distribution is centered at “5.1” which is the mean of the sample_averages and 95% of the area is covered within 3.5 and 6.5 which is almost equal to (3.439728 to 6.760272) 95% confidence intervals (with variance = 0.69) assuming normal distribution.

*In the above analysis, the confidence intervals are calculated by using the expression: mean + c(-1,1) * 2 * sqrt(variance).*