Statistical Inference Class Project 1

Mohamamd Yaqoob

This is the project for the statistical inference class. In it, we will use simulation to explore inference and do some simple inferential data analysis. In this first part we will perform simple simulation and illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s.

Requirement

The exponential distribution can be simulated in R with rexp(nosim, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Pre-Processing

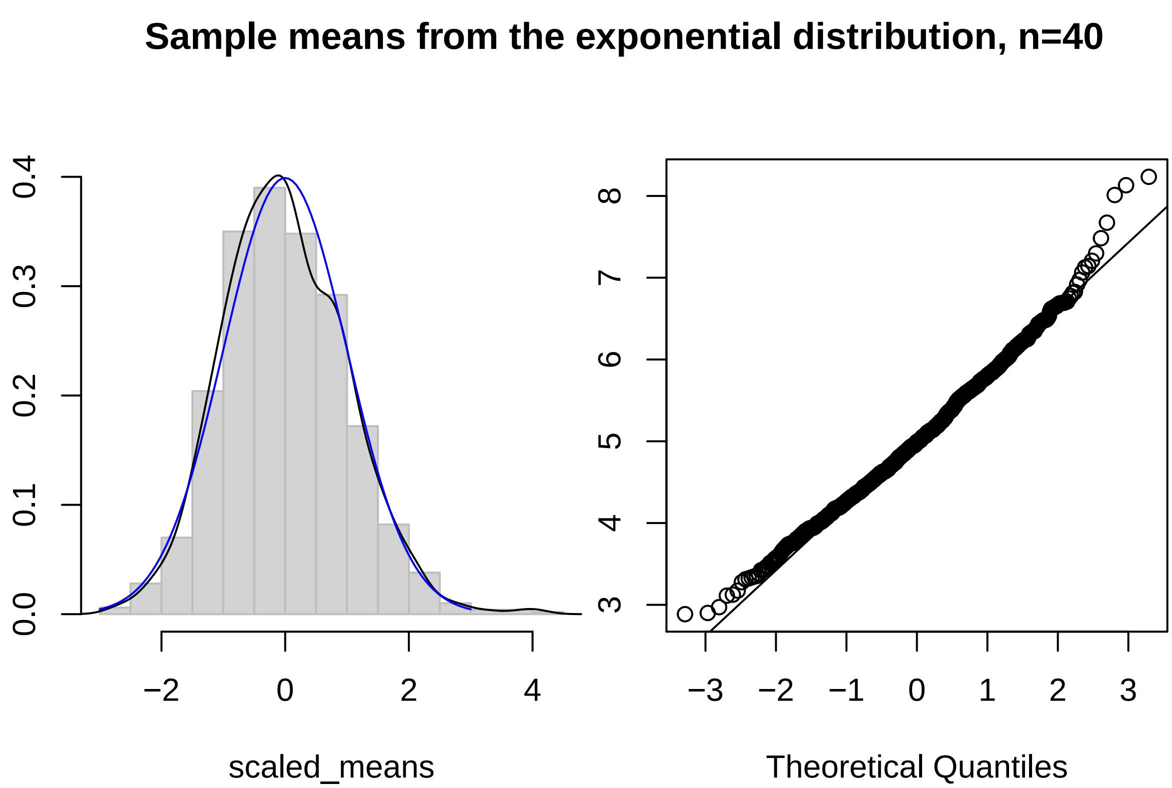
Set seed, intial values, and Plot sample function.

```
set.seed(1891)
lambda <-0.2
n <- 40
```

Question 1:
Show where the distribution is centered at and compare it to the theoretical center of the distribution.

I ran simulation and plot the distribution and compare to normal distribution curve.

```
sim=plot_sample_means(rexp, n=40, title="Sample means from the exponential distribution", rate=0.2)
```



```r
cat("Theoretical Center(Mean)=",1/lambda,"Actual Center(Mean)=",mean(sim),"\n")
```

```
## Theoretical Center(Mean)= 5 Actual Center(Mean)= 5.03
```

Our sample distribution is centered at 5.03 and it is approximately equal to Theoretical center of the distribution which is 5.

Question 2: Show how variable it is and compare it to the theoretical variance of the distribution.

```r
cat("Theoretical Variance = ",((1/lambda)^2)/n,"Actual Variance = ",var(sim),"\n")
```

```
## Theoretical Variance =  0.625 Actual Variance =  0.6083
```

```r
cat("Difference between Theoretical variance and Actual Variance=",((1/lambda)^2)/n - var(sim))
```

```
## Difference between Theoretical variance and Actual Variance= 0.01669
```

The calculation above shows when large number of simlations, the variability from standrard normal decreases.

Question 3: Show that the distribution is approximately normal.

```r
print(shapiro.test(sim))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  sim
## W = 0.9925, p-value = 5.654e-05
```

Shapiro-Wilk normality test shows p value greater than 5% thus the distribtuion is approximately normal. We can use Q-Q plots coupled with histograms in Answer 1 to check the assumption that the underlying population is normally distributed. Histogram shows the shape of a distribution and shows it follows the bell curve i.e. normal distribution. By overlaying both sample curve and bell cruve confirms our assumption the sample curve is pretty normal. QQ plot shows the observations follow approximately a normal distribution, the resulting plot is roughly a straight line with a positive slope, thus showing that data does seem to follow approximately a normal distribution.

Question 4: Evaluate the coverage of the confidence interval for 1/lambda:

```r
print(t.test(sim,conf.level = 0.975)$conf)
```

```
## [1] 4.974 5.085
## attr(,"conf.level")
## [1] 0.975
```

From the output, we can see that the mean for the sample is 5.029736. The two-sided 97.5% confidence interval tells us that mean is likely to be less than 5.085102 and greater than 4.974371. The p-value of 2.2e-16 tells us that if the mean were to 5, the probability of selecting a sample with a mean less than or equal to this one would be approximately 0%. Since the p-value is less than the significance level of 0.025 means that there is no evidence that the mean is outside the CI.