STATISTICAL INFERENCE COURSE PROJECT, PART 1: SIMULATION EXERCISES

CHE PAKRUL AZMI

1 - OVERVIEW

The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` λ is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. For this simulation, we set $\lambda = 0.2$. In this simulation, we investigate the distribution of averages of 40 numbers sampled from exponential distribution with $\lambda = 0.2$.

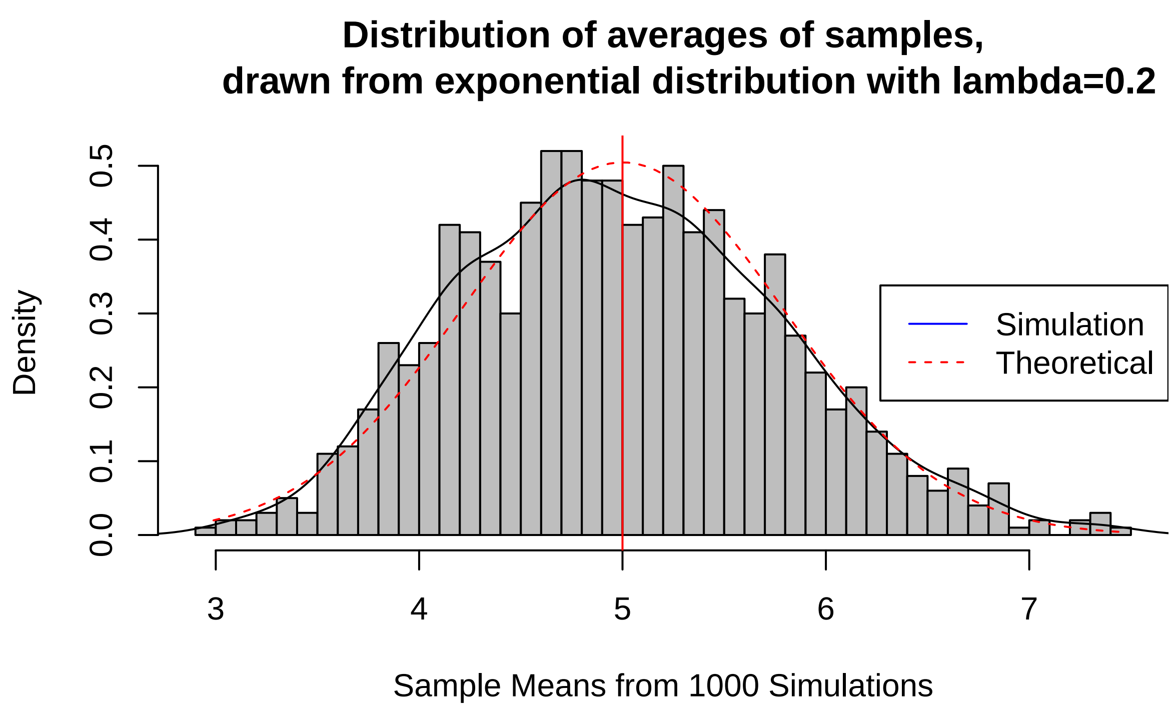
2 - SIMULATIONS

For this, we will do thousand simulated averages of 40 exponentials.

```
set.seed(3)
lambda <- 0.2
num_sim <- 1000
sample_size <- 40
sim <- matrix(rexp(num_sim*sample_size, rate=lambda), num_sim, sample_size)
row_means <- rowMeans(sim)
```

3 - SAMPLE MEAN VERSUS THEORETICAL MEAN

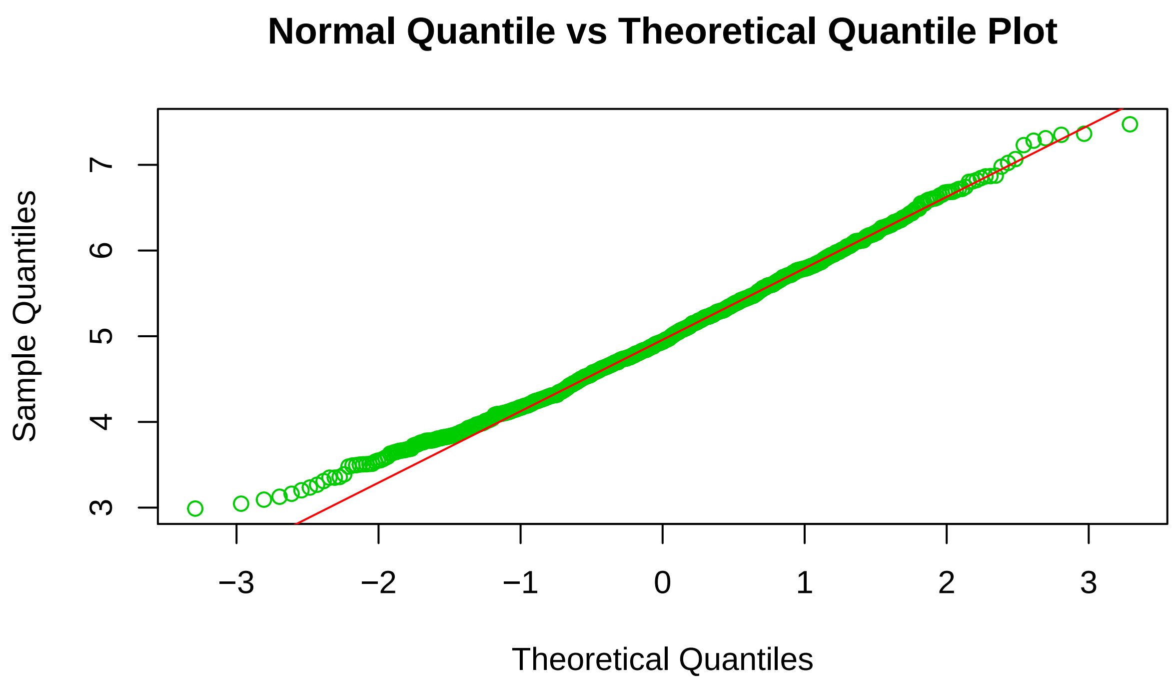
The distribution of sample means is as follows.



The distribution of sample means is centered at 4.9866197 and the theoretical center of the distribution is $\lambda^{-1} = 5$. The variance of sample means is 0.6257575 where the theoretical variance of the distribution is $\sigma^2/n = 1/(\lambda^2 n) = 1/(0.04 \times 40) = 0.625$.

4 - SAMPLE VARIANCE VERSUS THEORETICAL VARIANCE

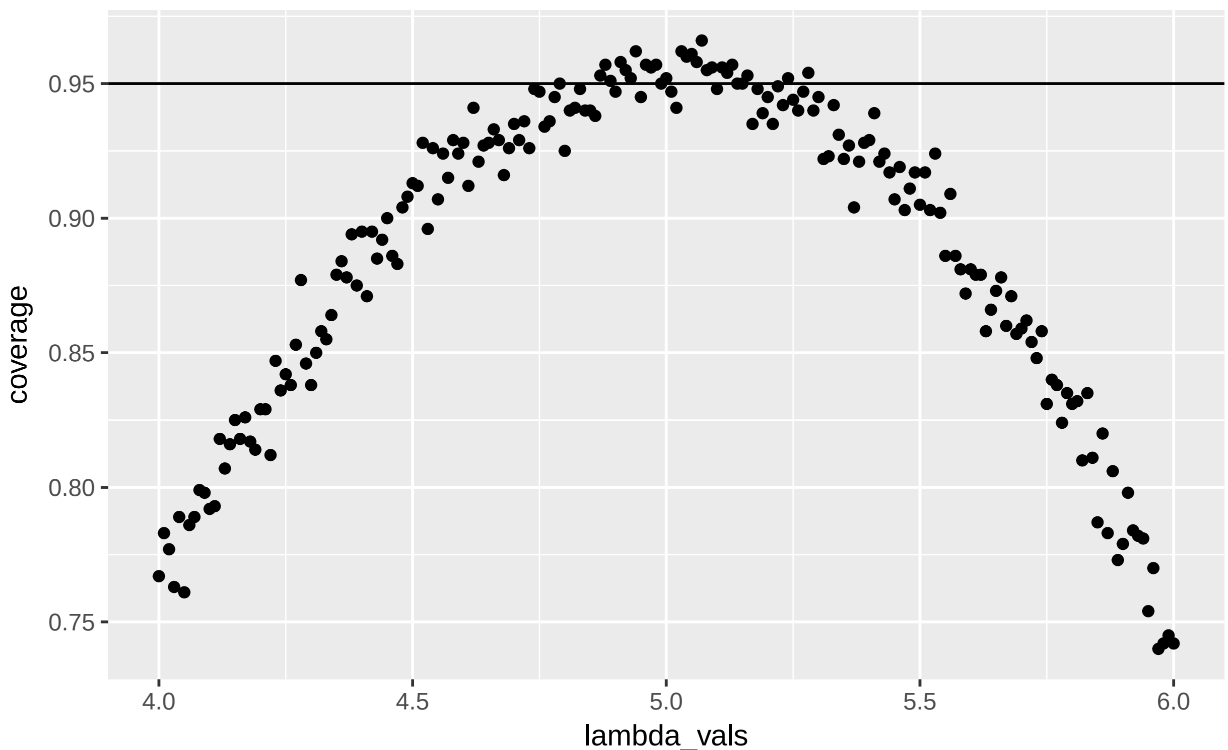
Due to the central limit theorem, the averages of samples follow normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. Also, the Normal Quantile vs Theoretical Quantile Plot below is approximately normal.



5 - Distribution

Finally, let's evaluate the coverage of the confidence interval for $1/\lambda = \bar{X} \pm 1.96\frac{S}{\sqrt{n}}$

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```



The 95% confidence intervals for the rate parameter (λ) to be estimated ($\hat{\lambda}$) are $\hat{\lambda}_{low} = \hat{\lambda}(1 - \frac{1.96}{\sqrt{n}})$ and $\hat{\lambda}_{upp} = \hat{\lambda}(1 + \frac{1.96}{\sqrt{n}})$. As can be seen from the plot above, for selection of $\hat{\lambda}$ around 5, the average of the sample mean falls within the confidence interval at least 95% of the time. Note that the true rate, λ is 5.

The report including the code for plots. Please refer this link https://github.com/chepakrul/stat_inference