Course Project Part1

Casey645

March 6, 2016

Part 1-Simulation Exercise

Simulations

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 $\lambda = 0.2$.

Lets start by doing a thousand simulated averages of 40 exponential numbers.

```
#Set seed
set.seed(3)
lambda <- 0.2
#We perform 1000 simulations with 40 samples
sample_size <- 40
simulations <- 1000

#Lets do 1000 simulations
simulated_exponentials <- matrix(rexp(simulations*sample_size, rate=lambda),
simulations, sample_size)
#Averages of 40 exponentials
row_means <- rowMeans(simulated_exponentials)
```

Results

1.Show where the discribution is centered, and compare to the theotretical center of distribution.

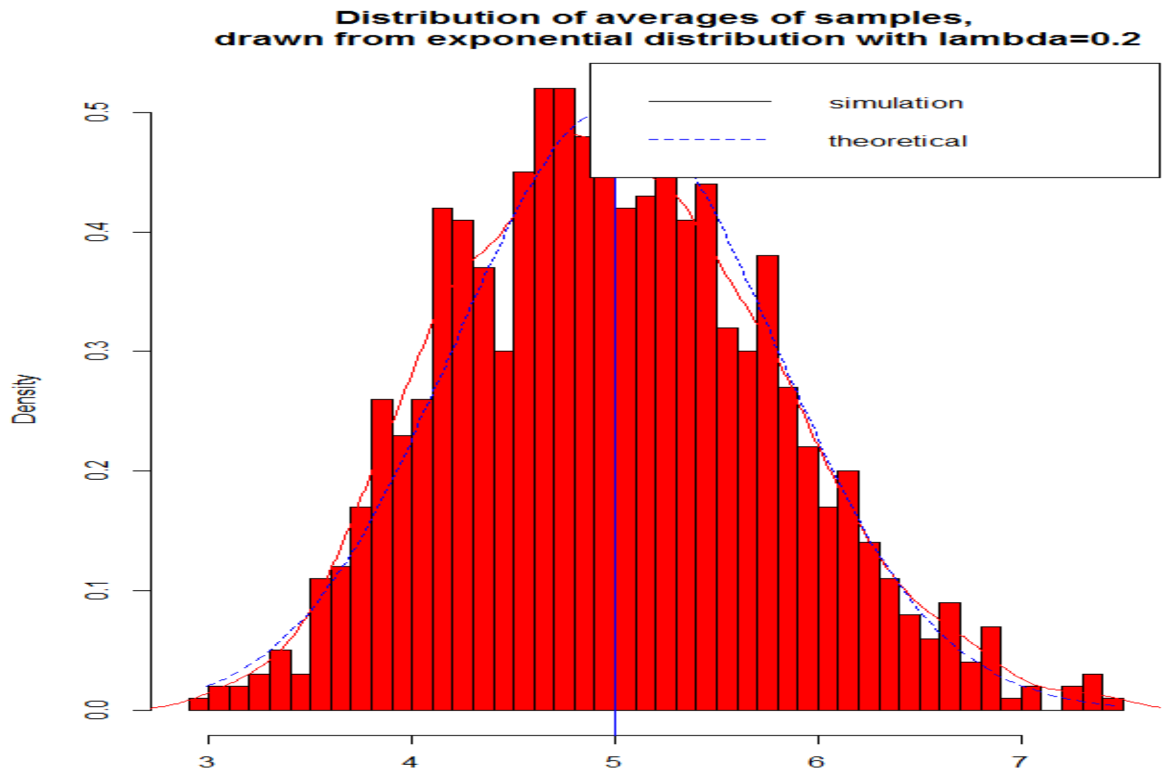
```
#mean of distribution of averages of 40 exponentials
mean(row_means)

##[1] 4.98662

#mean from analytical expression
1/lambda

##[1] 4.98662
```

The distribution of sample means - shown below:



Conclusion - the distribution of averages of 40 exponentials is centered at r mean(row_means) and the same is close to the theoretical center of the distribution, which is λ^{-1} = r 1/lambda.

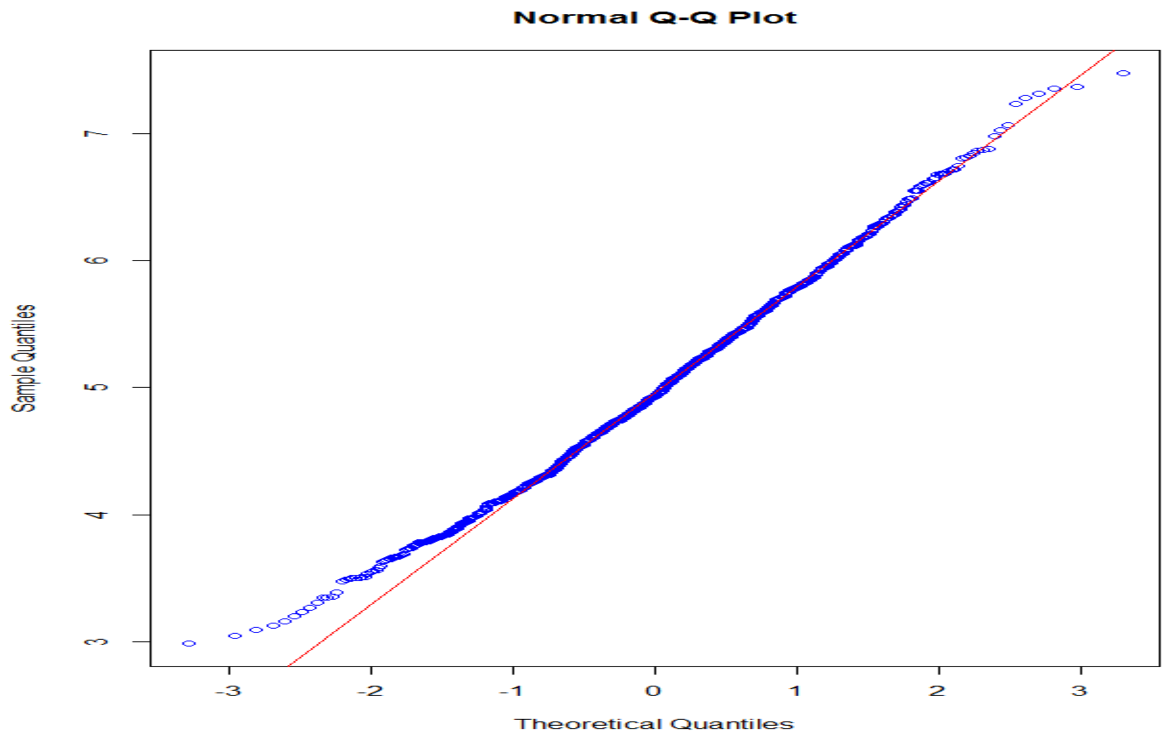
Show how variable it is and compare it to the theoretical variance of the distribution.

```
#standard deviation of distribution of averages of 40 exponentials
sd(row_means)
#standard deviation from analytical expression
(1/lambda)/sqrt(sample_size)

#Variance of the sample mean
var(row_means)
#Theoritcal variance of the distribution
1/((0.2*0.2) * 40
```

Therefore, the variability in distribution of averages of 40 exponentials is close to the theoretical variance of the distribution. The variance of sample means is r var(row_means) where as the theoretical variance of the distribution is $\sigma^2/n = 1/(\lambda^2 n) = 1/(0.04 \times 40)$ = r 1/(0.04 * 40).

Show that the distribution is approximately normal.

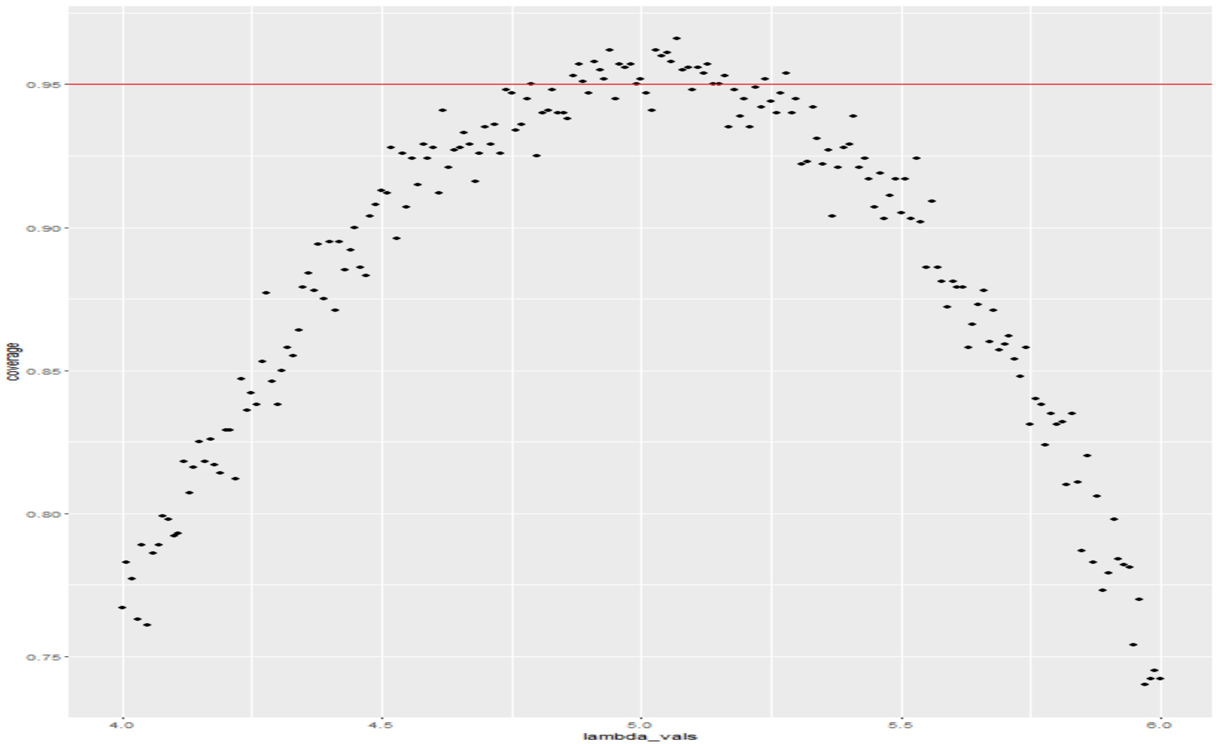


Due to the central limit theorem, the averages of samples follow normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. Also, the q-q plot suggests the distribution of averages of 40 exponentials is very close to a normal distribution.

Evaluate the coverage of the confidence interval for $1/\lambda = \bar{X} \pm 1.96\frac{S}{\sqrt{n}}$.

```
#Evaluate coverage of the confidence interval
lambda_vals <- seq(4, 6, by=0.01)
coverage <- sapply(lambda_vals, function(lamb) {
    mu_hats <- rowMeans(matrix(rexp(sample_size*simulations, rate=0.2),
                               simulations, sample_size))
    ll <- mu_hats - qnorm(0.975) * sqrt(1/lambda**2/sample_size)
    ul <- mu_hats + qnorm(0.975) * sqrt(1/lambda**2/sample_size)
    mean(ll < lamb & ul > lamb)
})

#Plot coverage
library(ggplot2)
{r, echo=FALSE}
qplot(lambda_vals, coverage) + geom_hline(yintercept=0.95, col=2)
```



The 95% confidence intervals for the rate parameter (λ) to be estimated ($\hat{\lambda}$) are $\hat{\lambda}\text{low} = \hat{\lambda}(1 - \frac{1.96}{\sqrt{n}})$ and $\hat{\lambda}\text{upp} = \hat{\lambda}(1 + \frac{1.96}{\sqrt{n}})$. 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.

Since, we consider the distribution of averages of exponentials, the standard deviation of this distribution already incorporates the () term i.e. it is the standard error.

The confidence interval is given by r mean(row_means) + c(-1, 1) * 1.96 * sd(row_means)