Statistical Inference: A simulation exercise

pierre attey

Sunday, January 25, 2015

Background

The goal of this project is to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations

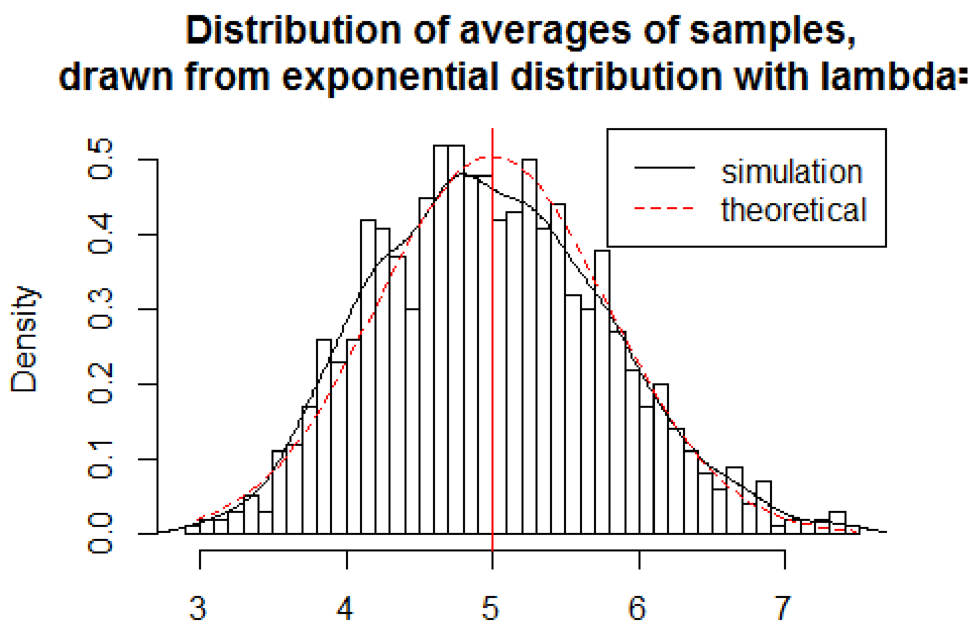
`Show the sample mean and compare it to the theoretical mean of the distribution`

- To ensure reproducibility of the output on the machine we use R command set.seed

```r
set.seed(3)
lambda <- 0.2
Num.simulation <- 1000
size <- 40
simulation <- matrix(rexp(Num.simulation*size, rate=lambda), Num.simulation,size)
row_means <- rowMeans(simulation)
```

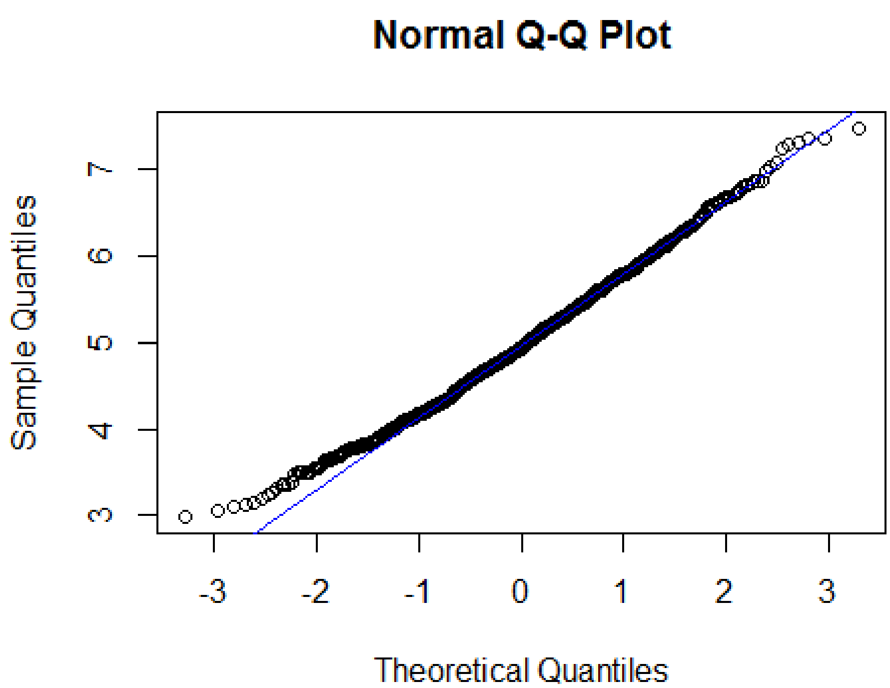
- The contains of this plot is: Sample means distributions with a plot for histogram of averages, density of the averages of samples,theoretical center of distribution.

```r
hist(row_means, breaks=50, prob=TRUE,
     main="Distribution of averages of samples,
     drawn from exponential distribution with lambda=0.2",
     xlab="")
lines(density(row_means))
abline(v=1/lambda, col="red")
xfit <- seq(min(row_means), max(row_means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(size)))
lines(xfit, yfit, pch=22, col="red", lty=2)
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("black", "red"))
```



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 below suggests the normality

```
qqnorm(row_means); qqline(row_means,col = "blue")
```



`Evaluate the coverage of the confidence interval for 1/lambda` - calculate the confidence interval,mean of distribution of averages of 40 exponentials and mean from analytical expression

```
mean(row_means) + c(-1, 1) * 1.96 * sd(row_means)
```

```
## [1] 3.436165 6.537074
```

```r
mean(row_means)
```

```
## [1] 4.98662
```

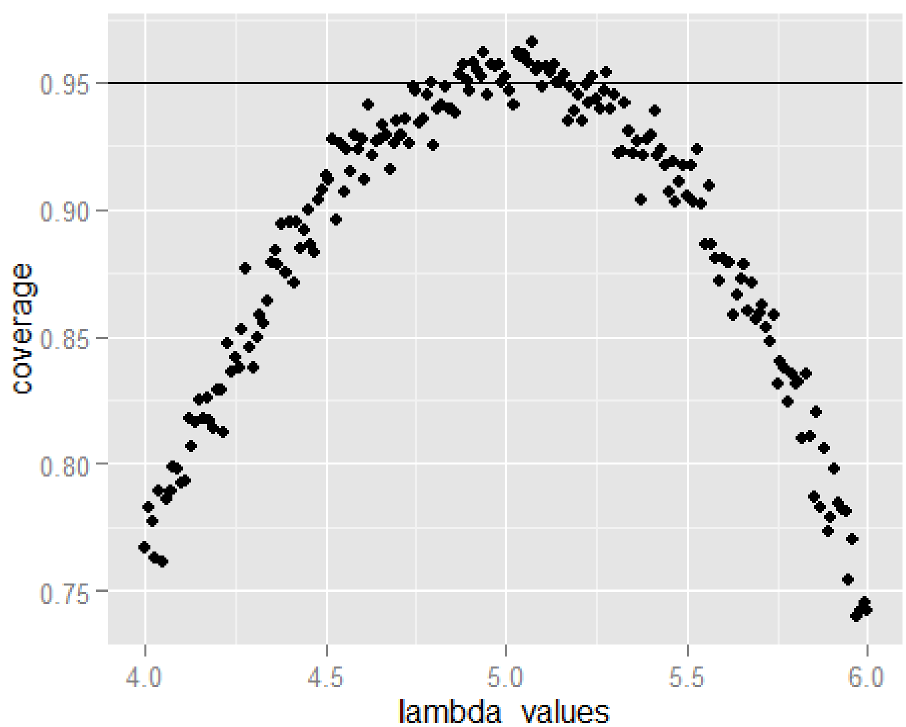
```r
1/lambda
```

```
## [1] 5
```

- The confidence interval is given by: [3.436165, 6.537074]
- mean from analytical expressionngigen by:5
- mean of distribution of averages of 40 exponentials given by: 4.98662

```r
lambda_values <- seq(4, 6, by=0.01)
coverage <- sapply(lambda_values, function(lamb) {
    mu_hats <- rowMeans(matrix(rexp(size*Num.simulation, rate=0.2),
                               Num.simulation, size))
    ll <- mu_hats - qnorm(0.975) * sqrt(1/lambda**2/size)
    ul <- mu_hats + qnorm(0.975) * sqrt(1/lambda**2/size)
    mean(ll < lamb & ul > lamb)
})
```

```r
library(ggplot2)
qplot(lambda_values, coverage) + geom_hline(yintercept=0.95)
```



- The average of the sample mean falls within the confidence interval at least 95% of the time