Statistical Inference Course Project, Part 1: Simulation Exercises

Sukanya Sen

6th April 2018

## Overview::

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

## Simulations:

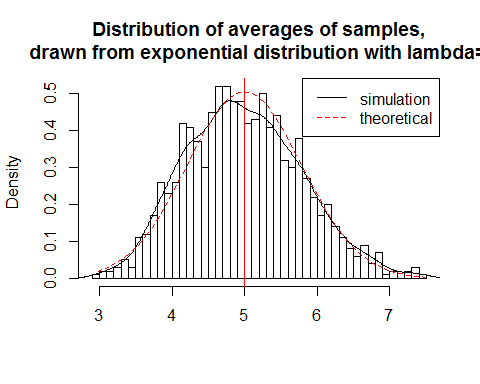
Let’s do a thousand simulated averages of 40 exponentials.

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)

## Sample Mean versus Theoretical Mean:

The distribution of sample means is as follows.

# plot the histogram of averages  
hist(row\_means, breaks=50, prob=TRUE,  
 main="Distribution of averages of samples,  
 drawn from exponential distribution with lambda=0.2",  
 xlab="")  
# density of the averages of samples  
lines(density(row\_means))  
# theoretical center of distribution  
abline(v=1/lambda, col="red")  
# theoretical density of the averages of samples  
xfit <- seq(min(row\_means), max(row\_means), length=100)  
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sample\_size)))  
lines(xfit, yfit, pch=22, col="red", lty=2)  
# add legend  
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("black", "red"))

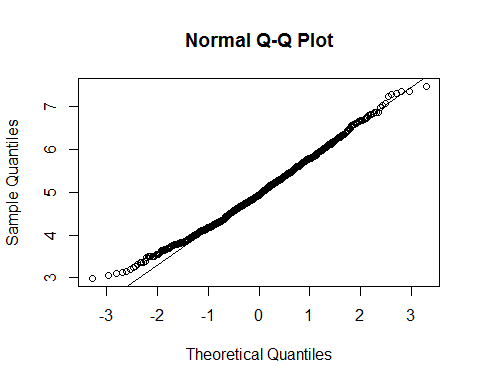


## Sample Variance versus Theoretical Variance:

The distribution of sample means is centered at 4.9866197 and the theoretical center of the distribution is = 5. The variance of sample means is 0.6257575 where the theoretical variance of the distribution is = 0.625.

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)



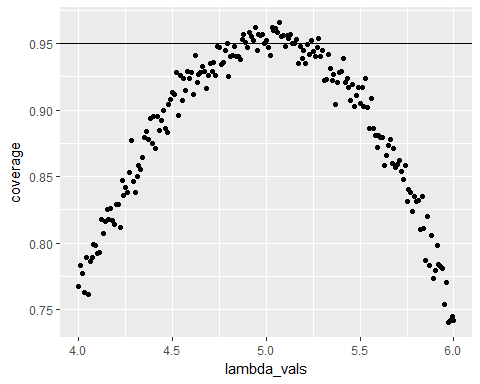
## Distribution:

Finally, let’s evaluate the coverage of the confidence interval for

lambda\_vals <- seq(4, 6, by=0.01)  
coverage <- sapply(lambda\_vals, function(lamb) {  
 mu\_hats <- rowMeans(matrix(rexp(sample\_size\*num\_sim, rate=0.2),  
 num\_sim, 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)  
})  
  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.2

qplot(lambda\_vals, coverage) + geom\_hline(yintercept=0.95)



The 95% confidence intervals for the rate parameter () to be estimated () are agnd . As can be seen from the plot above, for selection of around 5, the average of the sample mean falls within the confidence interval at least 95% of the time. Note that the true rate, is 5.