Simulation and Comparison on A Exponential Distribution

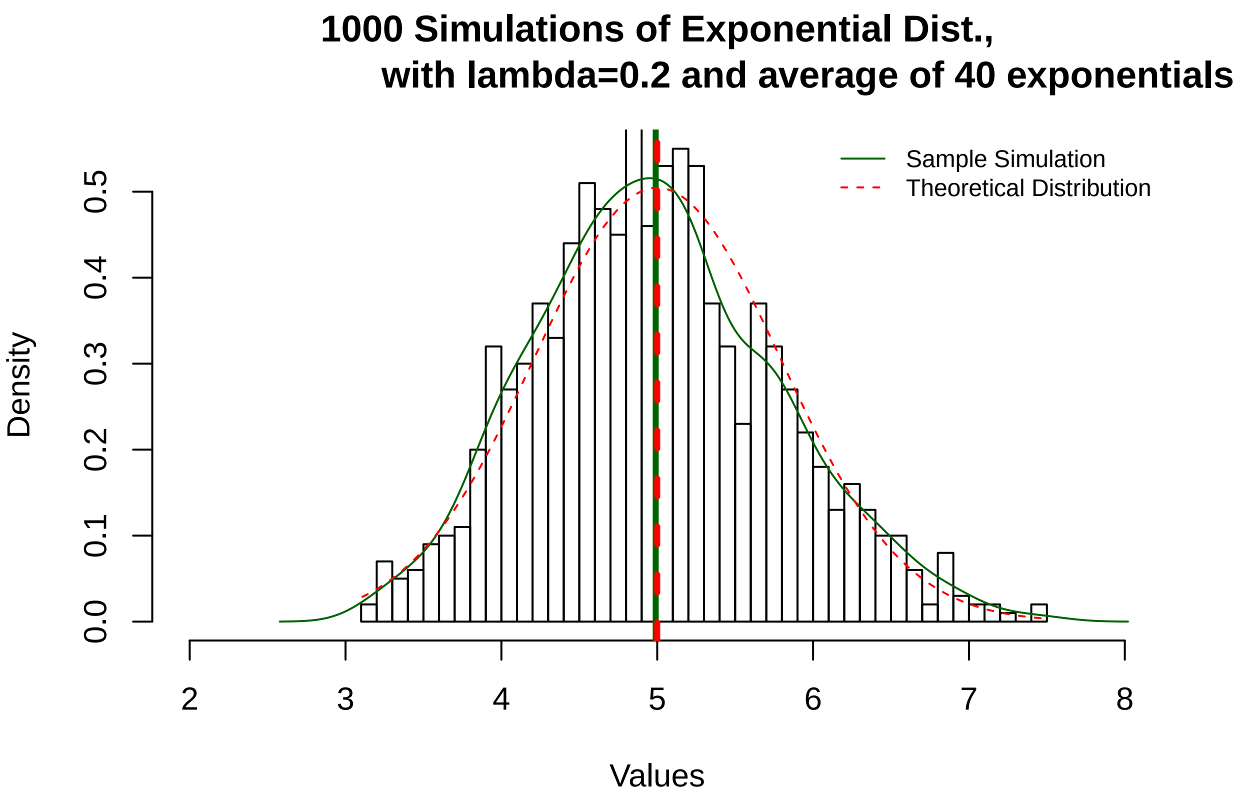
by Mingzhu Ye

Overview

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. Set $\lambda = 0.2$ for all of the simulations. We will investigate the distribution of averages of 40 exponentials. Note that we will need to do a thousand simulations.

Simulation

(Code seen in Appendix.1)

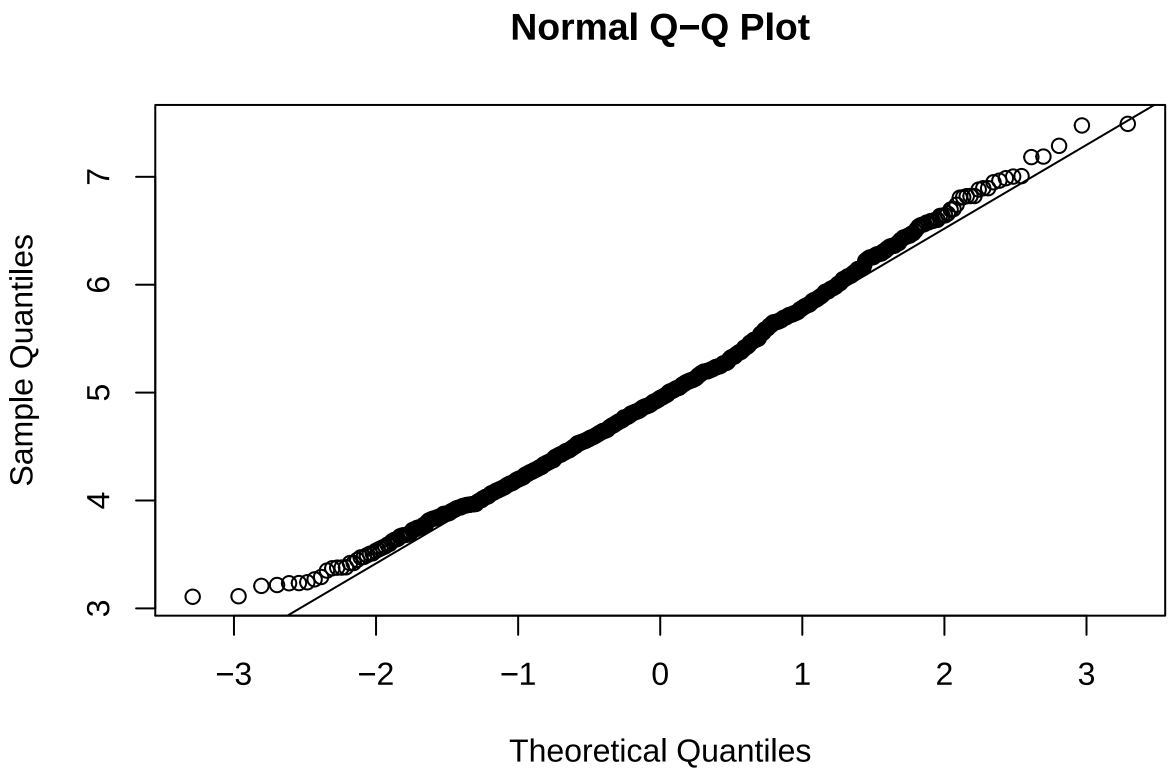


Sample Mean vs Theoretical Mean Based on the simulation and histgram above, we can calculate that sample mean is $mu = 5.02$, theoretical mean is $1/\lambda = 5$. They are quite near.

Sample Variance vs Theoretical Mean Based on the simulation and histfram above, sample variance is $sd(mns)^2 = 0.615$, theoretical variance is $1/\lambda^2/n = 0.625$. They are quite near.

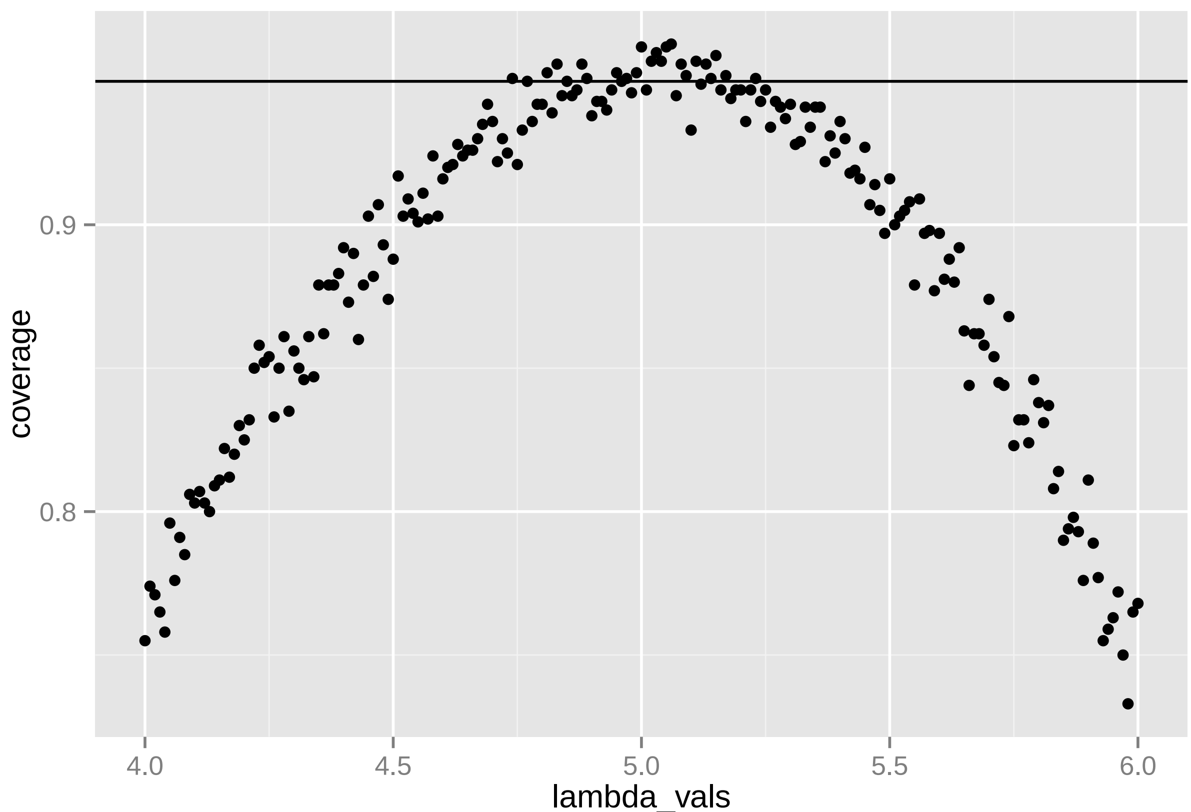
The Distribution is Approximately Normal Due to the Central Limit Theorem(CLT) and simulation above, the means and variances of sample and theoretical are quite near, the averages of samples follow normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. As well, the q-q plot below suggests the normality. Take a look at it:

```
qqnorm(mns)
qqline(mns)
```



Check the Coverage of Each Simulation for 95% Confidence Interval For visualization purposes, the simulation for 10000 simulations was rerun and 95% confidence interval of each simulation was worked out using the interval's own standard deviaton and mean according to the equation $\bar{X} \pm 1.96\sigma/\sqrt{n}$. As shown in the figure below, 95% coverage is clearly seen. (Code seen in Appendix.2)

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```



Conclusion

By investigating the distribution of averages of 40 exponentials via 10000 simulations, due to the Central Limit Theorem(CLT) and simulation above, the means and variances of sample and theoretical are quite near, 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. The q-q plot below suggests the normality. The Coverage of Each Simulation for 95% Confidence Interval is clealy observed.

Appendix

Appendix.1 set.seed(1)
lambda <- .2
n <- 40
no_sim <- 1000
mns <- NULL
sds <- NULL
for(i in 1: no_sim){
values <- rexp(n, lambda)
mns <- c(mns, mean(values))
sds <- c(sds, sd(values))
}
myhist <- hist(mns , freq = FALSE, xlim = c(2, 8), ylim = c(0, .55),
breaks = 50, prob = TRUE,
main = paste(no_sim,“Samples Simulation of Exponential Distribution with lambda=.2”),
xlab = c(“Values”))
mu <- mean(mns)

lines(density(mns), col = “grey”)
abline(v = mu , col = “grey”, lwd = 3, lty = 2)

abline(v = 1/lambda, col = “salmon”, lwd =3, lty = 9)
x <- seq(min(mns), max(mns), length = 100)
y <- dnorm(x, mean = 1/lambda, sd = (1/lambda/sqrt(n)))
lines(x, y, pch=22, col=“salmon”, lty=2)
legend(‘topright’, c(“Sample Simulation”, “Theoretical Distribution”),
lty =1, col = c(“lightblue”, “salmon”), bty = ‘n’, cex = .75)

Appendix.2 lambda_vals <- seq(4, 6, by=0.01)
n <- 40
no_sim <- 1000
coverage <- sapply(lambda_vals, function(lamb) {
mu_hats <- rowMeans(matrix(rexp(n*no_sim, rate=0.2),*
no_sim, n))
ll <- mu_hats - qnorm(0.975) sqrt(1/lambda^2/n)
ul <- mu_hats + qnorm(0.975) * sqrt(1/lambda^2/n)
mean(ll < lamb & ul > lamb)
})

library(ggplot2)
qplot(lambda_vals, coverage) + geom_hline(yintercept=0.95)