Statistical Inference Course Project: Part 1

Juan Agustin Melendez

July 21, 2015

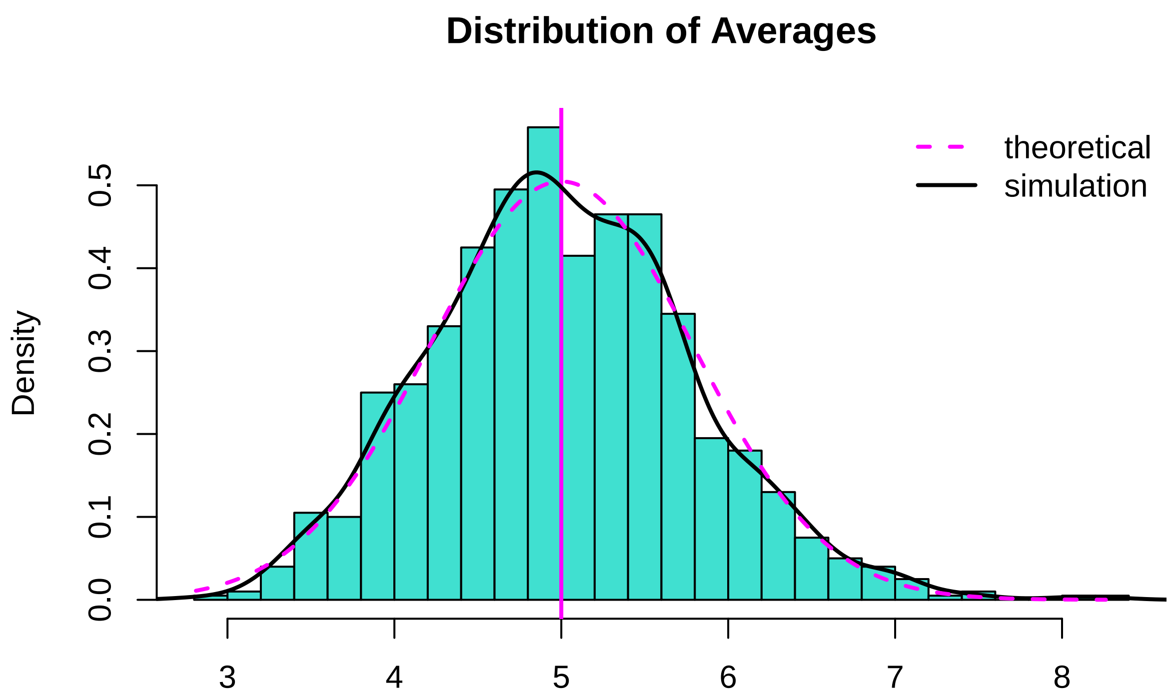
Description

The purpuse of this project is to investigate the exponential distribution in R and to demonstrate that the distribution of the averages of samples drawn from 40 exponential distributions are approximately normal. The simulation will be inspected visually by plotting the distribution of the sample averages with the actual and theoretical curve lines showing the distribution is centered around the mean. The simulation and theoretical means and variances will be compared and the Central Limit Theorem will be used to evaluate the coverage of the confidence intervals.

Simulaiton

The simulation consists of using the rexp() function in R and passing number of simulations times the number of exponentials and a lambda value of 0.2. Results form the simulation are stored in a matrix format to allow analysis.

```
nosim <- 1000
lambda <- 0.2
n <- 40
set.seed(12)
sim <- matrix(rexp(nosim*n, rate=lambda), nosim, n)
sim.means <- rowMeans(sim)
#Plotting the distribution of averages
hist(sim.means, col="turquoise", prob=TRUE, xlab = "", main="Distribution of Averages", breaks=30)
abline(v=1/lambda, col="magenta", lwd=2)
lines(density(sim.means), lwd=2)
# Normal distribution curve
xnorm <- seq(min(sim.means), max(sim.means), length=1000)
ynorm <- dnorm(xnorm, mean=1/lambda, sd=(1/lambda/sqrt(n)))
lines(xnorm, ynorm, col="magenta", lty=2, lwd=2)
legend("topright", c("theoretical","simulation"), lty=c(2,1), lwd=(2), col=c("magenta", "black"), bty="r
```



The plot shows the distribution of the averages of samples drawn from an exponential distribution with a lambda value of 0.2. The solid black line shows the simulated density curve while the magenta color line represents the normal distribution curve with lower and upper limit set to the minimum and maximum values of the similation averages, a mean of 1/lambda and a standard deviation of 1/lambda/sqrt(n).

This plot shows that the average of samples drawn from an exponential distribution is approximate to a normal distribution and its centered around its mean of 5. Further investigation is performed below by calculating and comparing the theoretical and simulation means and variances.

Mean and Variance: Theoretical vs Simulation

```
simulation.mean <- round(mean(sim.means), 2)
theoretical.mean <- 1/lambda
sample.variance <- round(var(sim.means), 2)
theoretical.var <- (1/lambda)^2/n
```

As can be observed from the calculations above, the theoretical and simulation means are very close to each other with values of **5** and **5.01** respectively. The calculated variance and the theoretical variance are fairly close to each other as well with a value of **0.61** and **0.625** respectively.

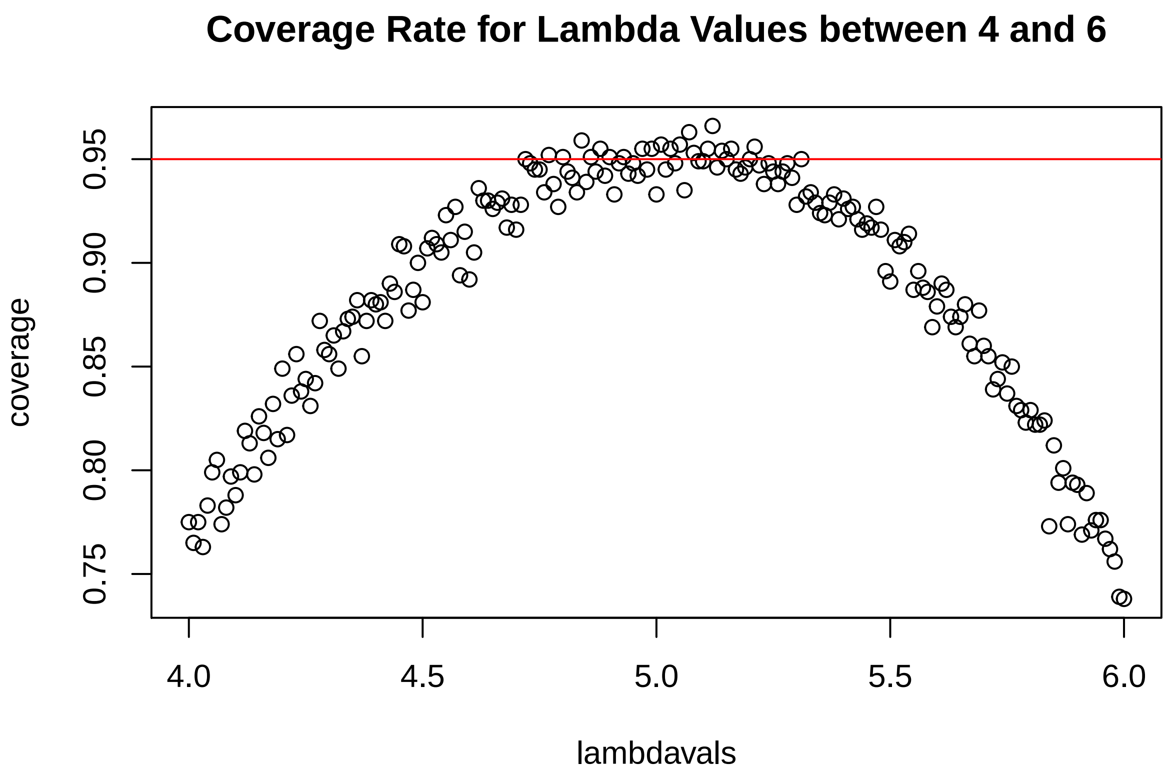
Simulating the Exponential Coverage Rate

The coverage rate for lambda values between 4 and 6 is evaluated below using confidence intervals based on the Central Limit Theorem. Confidence intervales took the form:

$Est \pm ZQ \times SE_{Est}$

where Est is the estimated mean, ZQ is the quantile from a normal distribution (qnorm(0.975) for 95% confidence interval), and SE_{Est} is the estimated standard error.

```
lambdavals <- seq(4, 6, by=0.01)
coverage <- sapply(lambdavals, function(lvals) {
    mhats <- rowMeans(matrix(rexp(nosim*n, rate=0.2), nosim, n))
    ll <- mhats - qnorm(0.975) * sqrt(1/lambda**2/n)
    ul <- mhats + qnorm(0.975) * sqrt(1/lambda**2/n)
    mean(ll < lvals & ul > lvals)
})
plot(lambdavals, coverage, main="Coverage Rate for Lambda Values between 4 and 6")
abline(h=0.95, col="red")
```



As can be seen from the plot above, the average of sample means fall withing the confidence interval at least 95% of the times. This reinforces that the distribution of the averages of samples of an exponential distribution is approximately normal.