Statistical Inference Project, Part 1: Simulation

Overview

The goal of this project is to investigate the exponential distribution and compare it with the Central Limit Theorem. This report will describe the simulation and discuss the reuslts.

Simulations

In each simulation, 40 random samples were chosen from the exponential distribution with $\lambda = 0.2$. The mean of these 40 samples is calculated. This process is repeated 1000 times. To accomplish this in R, the following code is executed:

```
means<- array(dim=1000)
lambda<- 0.2
for(i in 1:length(means)){means[i]<- mean(rexp(40,lambda))}
```

The Results

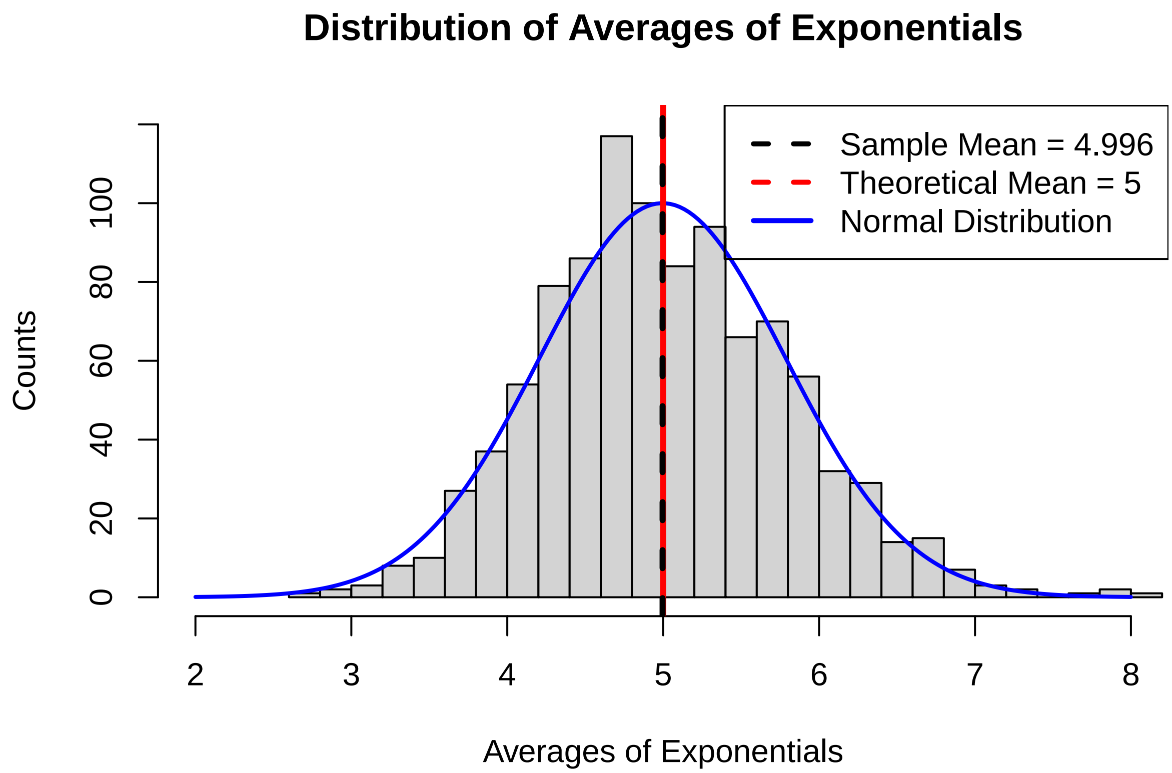
The exponential distribution has a mean $\mu = 1/\lambda = 1/0.2 = 5$ and variance $1/\lambda = 5$. If we take 1000 averages of n=40 samples from this distribution, theory tells us our sample distribution will be centered on the population mean: $\mu = \bar{x} = 5$, with a sample variance of $Var(\bar{x}) = \sigma^2/n = (1/\lambda)^2/n = 0.625$.

The following code is used to calculate the sample mean and variance of these simulations:

```
mean(means)
var(means)
```

The sample mean evaluates to $\bar{x} = 4.996$ with sample variance $Var(\bar{x}) = 0.616$, just as we expect from theory. The distribution is plotted on the next page, overlayed with a bell curve. The plot demonstrates that the sample mean is centered on the theoretical mean, and the distribution is approximately normal. This is in agreement with the Central Limit Theorem. Although we are sampling from an exponential distribution, the distribution of averages of our samples is normally distributed, centered on the theoretical mean of the population.

To summarize, our theoretical mean is $\mu = 5$ and the sample mean from the simulation is $\bar{x} = 4.996$. The theoretical variance we expected from the simulation is $(1/\lambda)^2/n = 0.625$, and the variance we calculated was $Var(\bar{x}) = 0.616$. We conclude that this simulation has successfully demonstrated the Central Limit Theorem.



The Plot can be reconstructed with the following code. Note that the probability density curve has been scaled so the maximum is approximately aligned with the maximum counts of the histogram.

```
hist(means, breaks= 20, col = "lightgrey",
     xlab="Averages of Exponentials", ylab = "Counts",ylim=c(0, 120), xlim=c(2,8),
     main="Distribution of Averages of Exponentials")
curve(100*exp(-(x-mean(means)^2/(2*sd(means)^2)), from = 2, to=8, n=200,
      col="blue", lwd=2, add=TRUE)
abline(v=5, col="red", lwd=3,lty=1)
abline(v=mean(means), col="black", lwd=3,lty=2)
legend("topright",
       c("Sample Mean = 4.996", "Theoretical Mean = 5",
         "Normal Distribution"),
       lty=c(2,2,1), lwd=c(2.5,2.5,2.5),
       col= c("black","red","blue"))
```