

Understanding an Exponential Distribution

Jessica Bohning

Overview

The following will simulate an exponential distribution with 1,000 simulations and a rate of 0.2 and then take a random sample (with replacement) of 40. The goal will be to understand how this simulation compares to the theoretical values and to the normal distribution.

How does the sample mean compare to the theoretical mean?

I will use the rexp function to simulate the data and find the mean

```
#Set up variables
set.seed(248)
lambda<- 0.2
n<-40

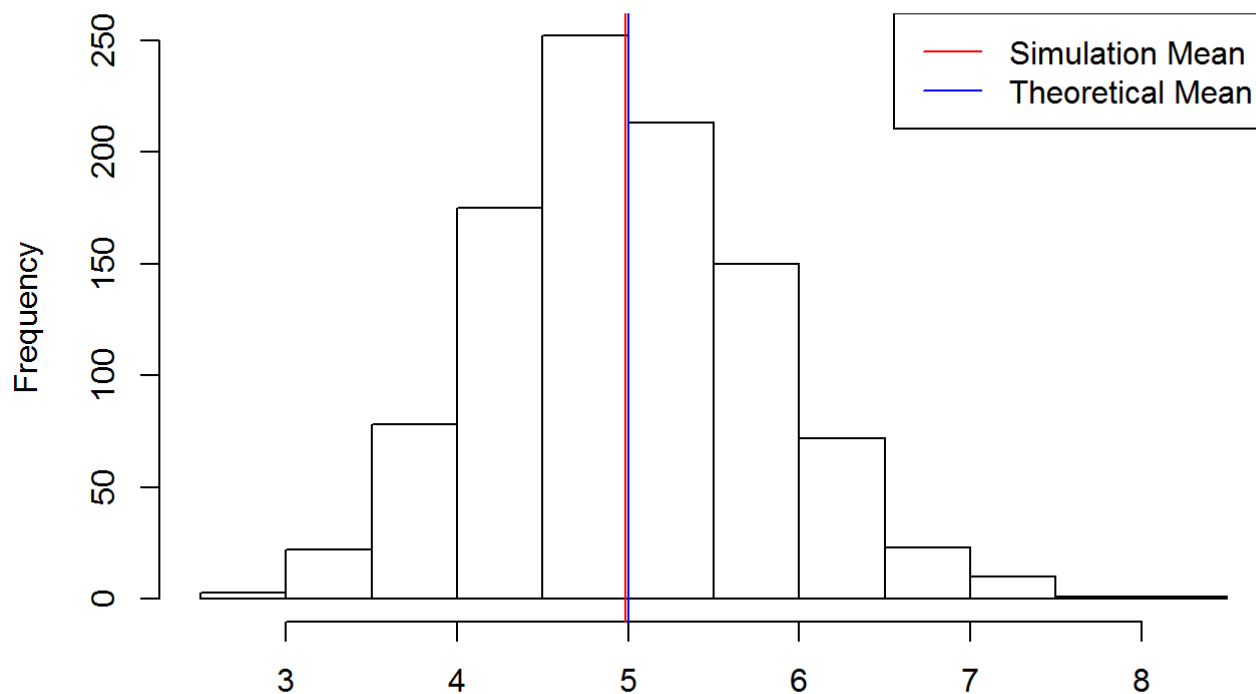
#run simulation 1000 times
datasimulation<-NULL
for (i in 1:1000){
  datasimulation[i]<-mean(rexp(n,lambda))
}

#samplemeans<-NULL
#Sample 40 means (with replacement)
#samplemeans<-datasimulation[sample(1:1000,40,replace=TRUE)]

#Histogram the sample means
#Theoretical mean is 1/Lambda

hist(datasimulation,main="Histogram of Sample",xlab=NULL,ylab="Frequency")
abline(v=mean(datasimulation),col="red")
abline(v=1/lambda,col="blue")
legend("topright",c("Simulation Mean","Theoretical Mean"),col=c("red","blue"),lwd=1)
```

Histogram of Sample



As seen in the previous histogram, the sample distribution's mean lies close to the theoretical mean.

How does the sample variance compare to the theoretical variance?

```
samplevariance<-var(datasimulation)
theoreticalvariance<-(1/lambda)^2/n
```

The variance of the sample was 0.647 and the theoretical variance was 0.625 and they are very close.

Is the distribution approximately normal?

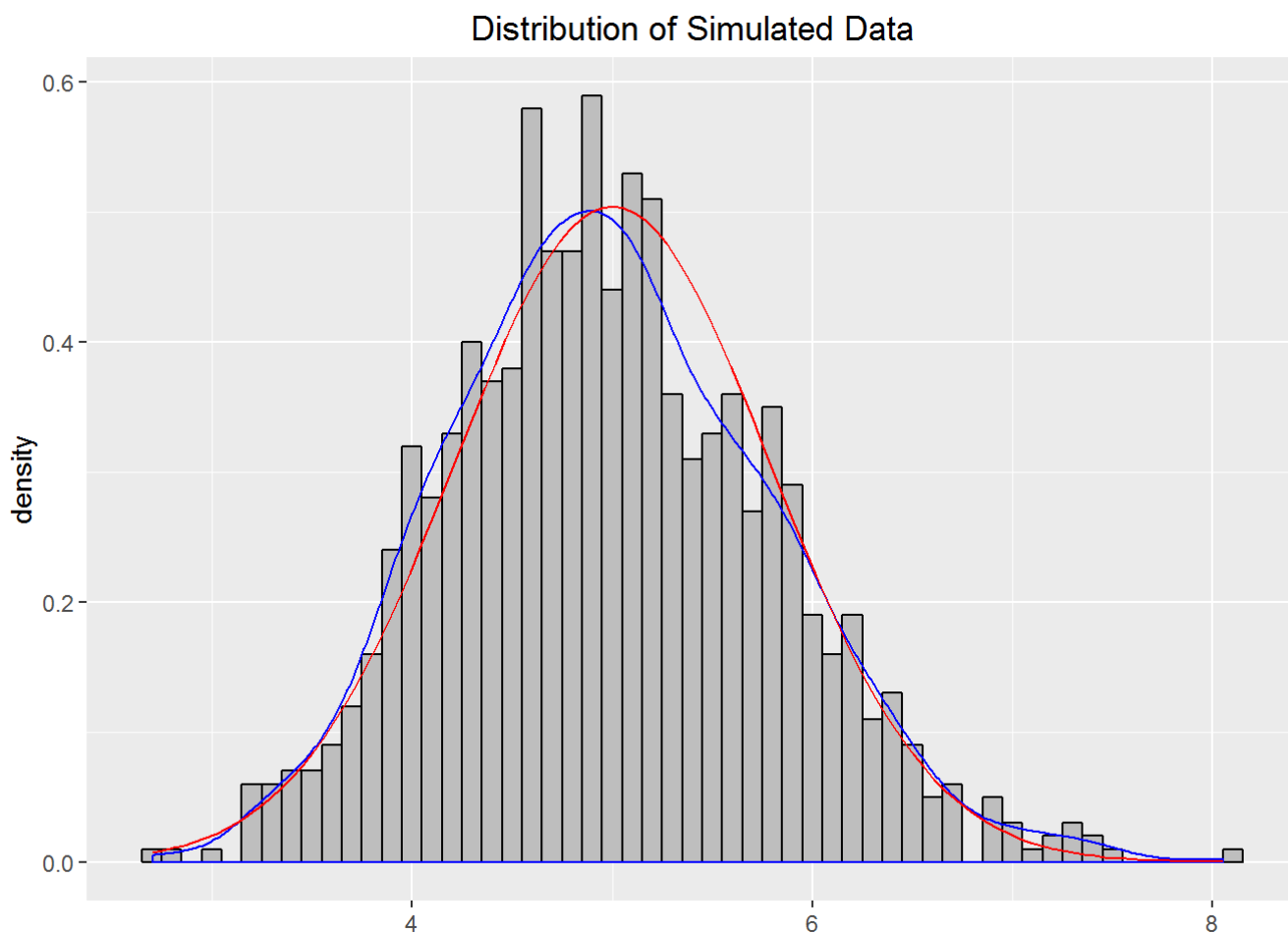
The Central Limit Theorem states that given a large enough sample from a population (with a finite level of variance), the mean of all samples from the same population will be approximately equal to the mean of the population.

Note: that the blue line below shows the distribution of the simulated data and the red line shows the normal distribution

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
set.seed(564)
ggplot(data.frame(datasimulation),aes(datasimulation))+
  geom_histogram(binwidth=.1,fill="gray",col="black"),
  aes(y=..density..))+
  geom_density(col="blue")+
  stat_function(fun=dnorm,args=list(mean=1/lambda,
                                   sd=sqrt(theoreticalvariance)),
               col="red")+
  xlab(NULL)+
  ggtitle("Distribution of Simulated Data")
```



Note: that the blue line above shows the distribution of the simulated data and the red line shows the normal distribution

As predicted by the Central Limit Theorem, the histogram above shows the simulated distribution is approximately normal