Statistical Inference Course Project

MLauden

February 22, 2017

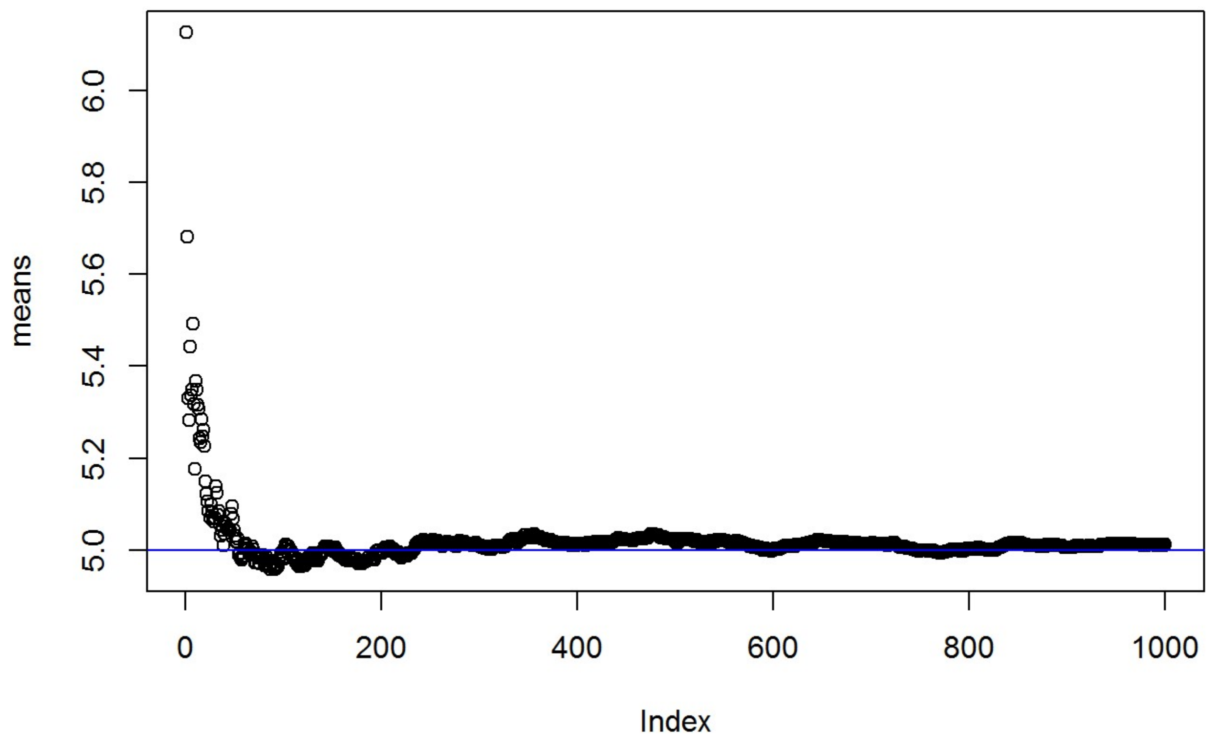
Overview

The following section compares the distribution of the means against a random sampling of data from the exponential function. The actual data appears to take the form of an exponentially decaying function. The sampling of the means appears to be symmetric about the theoretical mean of five. This suggests the data are normally distributed and follow the central limit theorem.

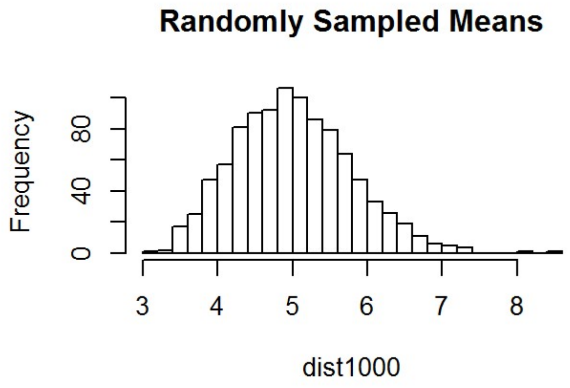
The theoretical mean will be added using a blue line to some of these graphs to better help visually interpret the weight of the distribution. The mean of the exponential distribution is:

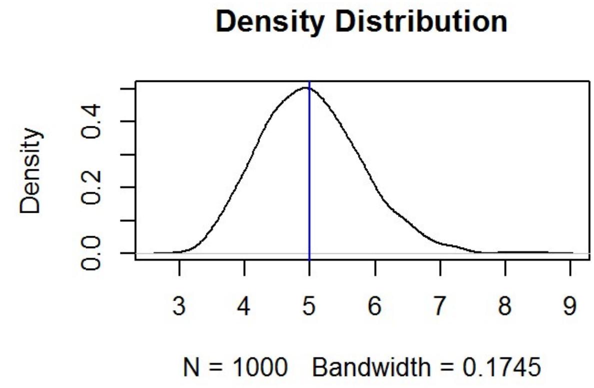
mean=1/lambda =1/.2=5

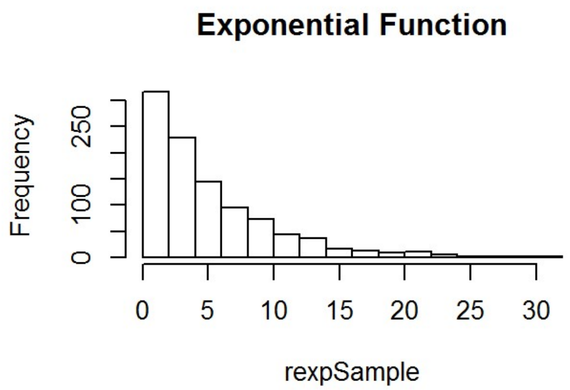
```
#Set variables
lamda=.2
means<-NULL
dist1000<-NULL
dist<-NULL

for(i in 1:1000){
  dist1000<-rbind(dist1000,mean(rexp(40,lamda)))  #assign output data of expone
ntial distribution to variable
}

for(i in 1:1000){
  means<-rbind(means,mean(dist1000[1:i]))
}

plot(means)
 abline(h=5, col="blue")
```




```
par(mfrow=c(2,2)) #Sets structure of graph panel
 hist(dist1000, #plot the data
      breaks = 20,
       main="Randomly Sampled Means")
 plot(density(dist1000),
      main="Density Distribution"
 )
 abline(v=5, col="blue")

 rexpSample<-rexp(1000,lamda) # calculate values for function
 hist(rexpSample,     # Plot the values of the rexp function
      breaks=20,
      main="Exponential Function"
      )
```







Distribution

The distribution of the means appears symmetric around the mean in the shape of a bell curve. This suggests that the data is normally distributed.

The mean of the calculated values from the rexp function is similar to the theoretical distribution mean of 5:

```
mean(rexpSample)
```

```
## [1] 5.12807
```

Calculate the variance of the samples

The theoretical variance will be added using a blue line to some of these graphs to better help visually interpret the weight of the distribution. As with the case of the mean, the variability is high for the initial sample points. As n increases, the data start to center around the expected mean/variance.

The theoretical value of the variance of the exponential distribution is:

variance=(1/lambda)^2 /n =(1/.2)^2/40= 0.625

This theoretical value will be plotted as a blue line in the simulated values in the following graph.

```
#initialize variables
 distVar1000<-NULL
 variances<-NULL
 #Same code for creating a dataset of means, but switching to the "var" functio
n to calculate variances


for(i in 1:1000){
  variances<-rbind(variances,var(dist1000[1:i]))
}
plot(variances)
 abline(h=0.625, col="blue")
```