Statistical Inference Project One

girldewhirle

6 July 2015

Overview:

Simulation and analysis to demonstrate the Central Limit Theorem.

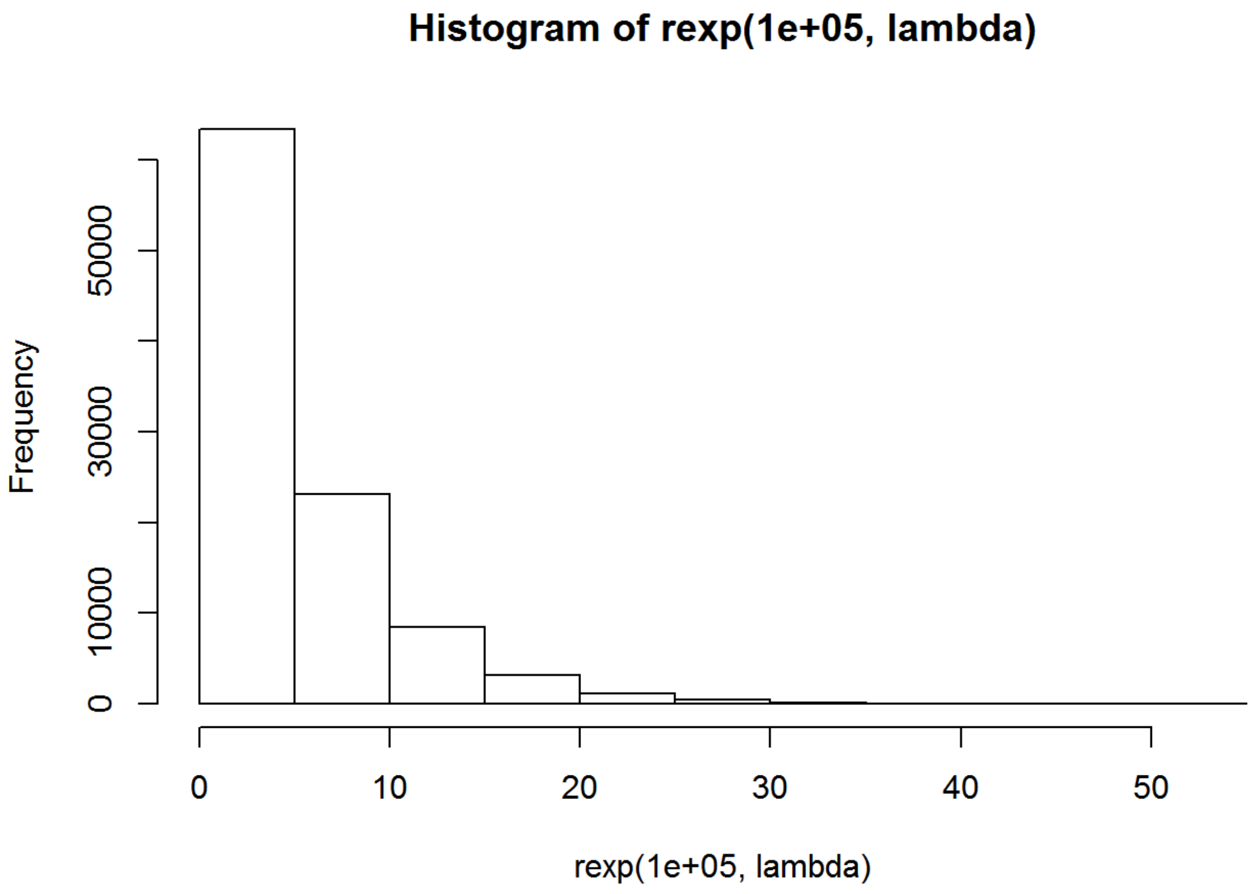
Simulations

In all simulations lambda = 0.2.

```
lambda=0.2
```

An exponential distribution looks like this (n is much greater than 40 to more clearly see the pattern:

```
hist(rexp(100000,lambda))
```



Central Limit Theorem is a statistical theory that the distribution of the means of IID (Independent and Identically Distributed) variables approaches a normal distribution for large sample sizes.

In the context of this exercise the means of 1000 exponential distribution samples (with a sample size of 40) will be simulated. On the basis of CLT we can predict that:

1. the mean of the distribution (of sample means) will approximate the theoretical mean of the

exponential distribution (1/lambda)

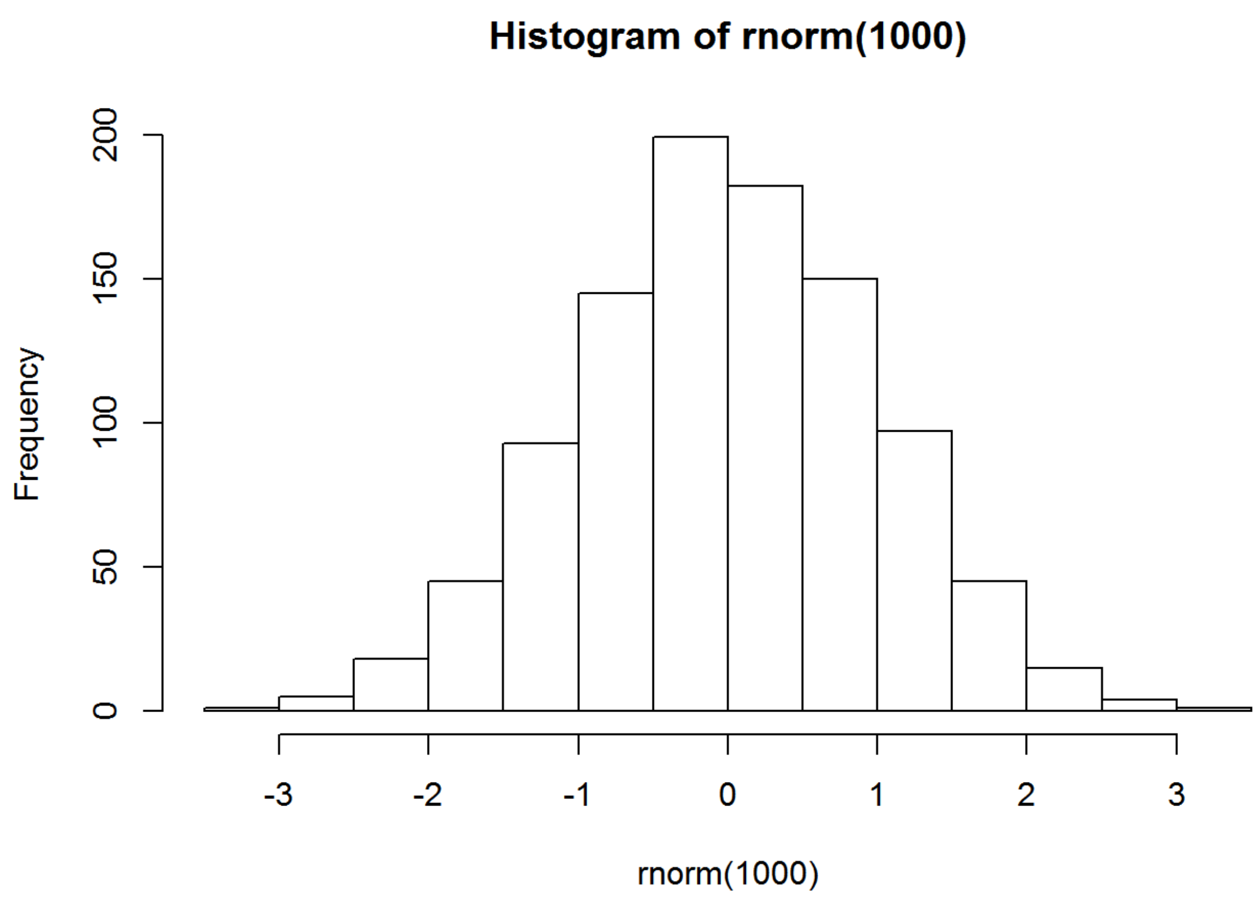
```
## mean of exponential is 1/lambda
meanexp<-1/lambda
## STANDARD DEVIATION of exponential is 1/lambda
sdexp<-1/lambda

## expected mean of distribution
## now taking 1000 means from 1000 samples
## expect mean to be close to population mean therefore
expectedmean<-meanexp
```

2. the variance of the distribution (of sample means) will approximate the theoretical variance of the exponential distribution (1/lambda)^2/n (in this instance n=40)

```
## STANDARD DEVIATION of exponential distribution is 1/lambda
sdexp<-1/lambda
## variance of exponential square of sdexp
varexp<-sdexp^2
## expected variance of distribution of mean of samples
expectedvariance<-25/40
```

3. the distribution (of sample means) is close to a normal distribution

```
## sample normal distribution
hist(rnorm(1000))
```



Running simulation of a sample of 40 values from an exponential distribution 1000 times

```
## run exp simulation 1000 times and take mean
## create variable and set to null
expmeanvalues<-NULL
## calculate mean of sample of 40 from exponential distribution and append to variable
for (i in 1:1000){expmeanvalues<-c(expmeanvalues,mean(rexp(40,lambda)))}
```

Comparison of means:

Expected mean is 1/lambda = 1/0.2 = 5. Mean of distribution of sample means is:

```
meansim<-mean(expmeanvalues)
print(meansim)
```

```
## [1] 4.975804
```

Expected variance = (1/lambda)^2/40 = 5^2/40 = 25/40 = 0.625. Variance of distribution of sample means is:

```
variancesim<-var(expmeanvalues)
print(variancesim)
```

```
## [1] 0.6639881
```

Distribution of sample means approximates normal.

```
library(ggplot2)
```

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

```
# hist(expmeanvalues)
# abline(v=meansim, col="blue")
# abline(v=meanexp,col="red")
## create dataframe so can use ggplot
plotdata<-as.data.frame(expmeanvalues)
## create plot object
plot_object<-ggplot(data = plotdata,aes(x=expmeanvalues))
## plot with histogram and density
plot_object +
  # histogram
  geom_histogram(aes(y=..density..),binwidth=.2,colour="black",fill="white") +
  # density curve
  geom_density(colour="red")+
  # line of expected mean/theoretical mean
  geom_vline(xintercept=expectedmean,colour="green",linetype="longdash") +
  # line of actual mean of sample means
  geom_vline(xintercept=meansim,colour="blue") +
  ## add annotation
  annotate("text",x=7,y=0.45,label="Expected Mean",colour="green") +
  ## add annotation
  annotate("text",x=7,y=0.5,label="Mean of Sample Means",colour="blue") +
  ## label axis
  xlab("Means of Samples") +
  ylab("Frequency of Value")
```