Statistical Inference Final Project, Part 1

Dwight Shackelford

July 29, 2016

Statistical Inference Class Project

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter.

The mean of exponential distribution is 1/lambda and

the standard deviation is also (1/lambda)/sqrt(sample size).

lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials by running 1000 simulations.

We will illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. We will:

1 Show the sample mean and compare it to the theoretical mean of the distribution.

```
sim_cnt<-1000
lambda<-.2
samp_sz<-40
set.seed(222)
mn<-1/lambda
st_dv<-1/lambda/sqrt(samp_sz)
## create 1000 means to create the distribution
mns<-sapply(1:sim_cnt,function(x){mean(rexp(samp_sz,lambda))})

## sample mean
mean(mns)

## [1] 4.97924

## theoretical mean
mn

## [1] 5
```

The sample mean [4.979] compares very closely to the theoretical mean [5] of the distribution.

2 Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
## sample variance
var(mns)

## [1] 0.6542521

## theoretical variance
((1/lambda)/sqrt(samp_sz))^2

## [1] 0.625
```

The variance of the simulation [0.6542521] is very close to the theoretical variance [.625]

2a Show the sample sd and compare it to the theoretical sd.

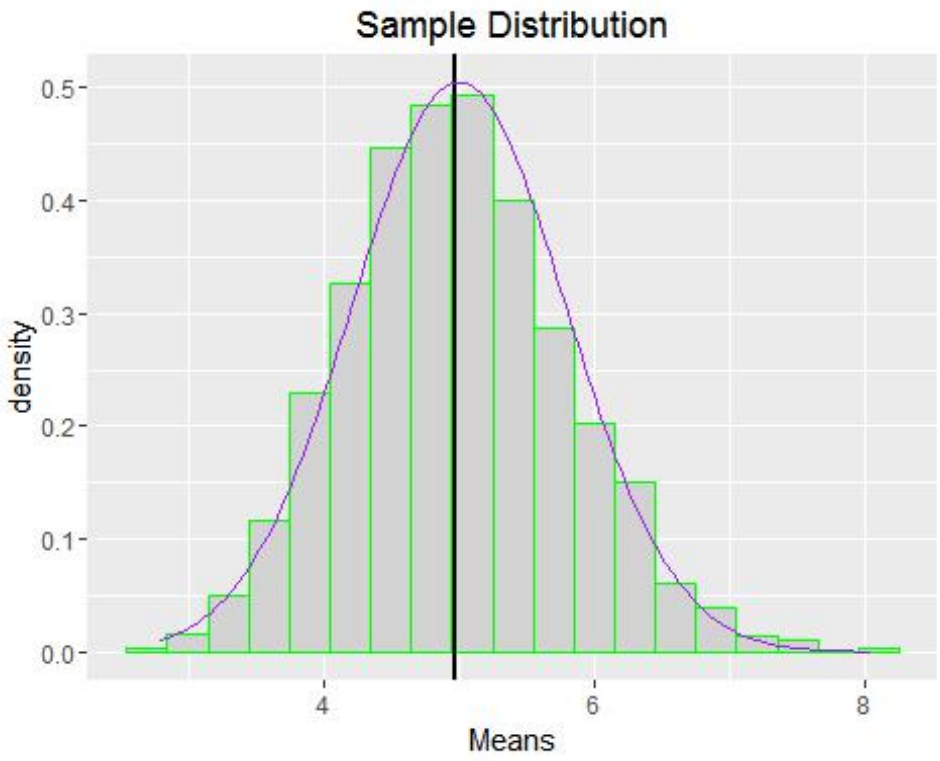
```
## sample sd
sd(mns)

## [1] 0.8088585

## theoretical sd
st_dv

## [1] 0.7905694
```

Sample standard deviation [0.8088585] is very close to the theoretical standard deviation [0.7905694]

3 Show that the distribution is approximately normal.

we will overlay a normal distribution curve onto the histogram, with parameters equaling the theoretical values, and compare the two graphs.

```
library(ggplot2)
g<-ggplot(data=data.frame(mns), aes(x=mns))+ geom_histogram(binwidth=.3,
 aes(y=..density..), fill="lightgrey", color="green")+geom_vline(xinter
cept=mean(mns),size =1, col="black")+labs(x="Means")+ggtitle("Sample Di
stribution")+stat_function(fun=dnorm, color="purple",args=list(sd=st_dv,
mean=mn) )
print(g)
```



According to the overlay of the dnorm line, the sample distribution is approximately normal.