Statistical Inference

Yuani

23 August 2015

In this project, we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can will 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. Lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials, with 1000 simulations.

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

```
#defininig variables
lambda <- 0.2
n <- 40
simulations <- 1:1000
#Set seed for reproducibility
set.seed(120)
#run simulation
sim.means <- data.frame(x=sapply(simulations,function(x) {mean(rexp(n,lambda))}))
head(sim.means)
```

```
##          x
## 1 5.091177
## 2 4.323641
## 3 5.139638
## 4 4.980668
## 5 3.864485
## 6 4.496848
```

```
samplemean <- mean(sim.means$x)
samplemean
```

```
## [1] 5.03946
```

```
expectedmean <- 1/lambda
expectedmean
```

```
## [1] 5
```

Based on the above simulation, we see that the sample mean was very close to the theoretical mean of 5.

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

```
samplesd <- sd(sim.means$x)
samplesd
```

```
## [1] 0.7865754
```

```
samplevar <- var(sim.means$x)
samplevar
```

```
## [1] 0.6187008
```

```
expectedsd <- (1/lambda)/sqrt(40)
expectedsd
```

```
## [1] 0.7905694
```

```
expectedvar <- expectedsd^2
expectedvar
```

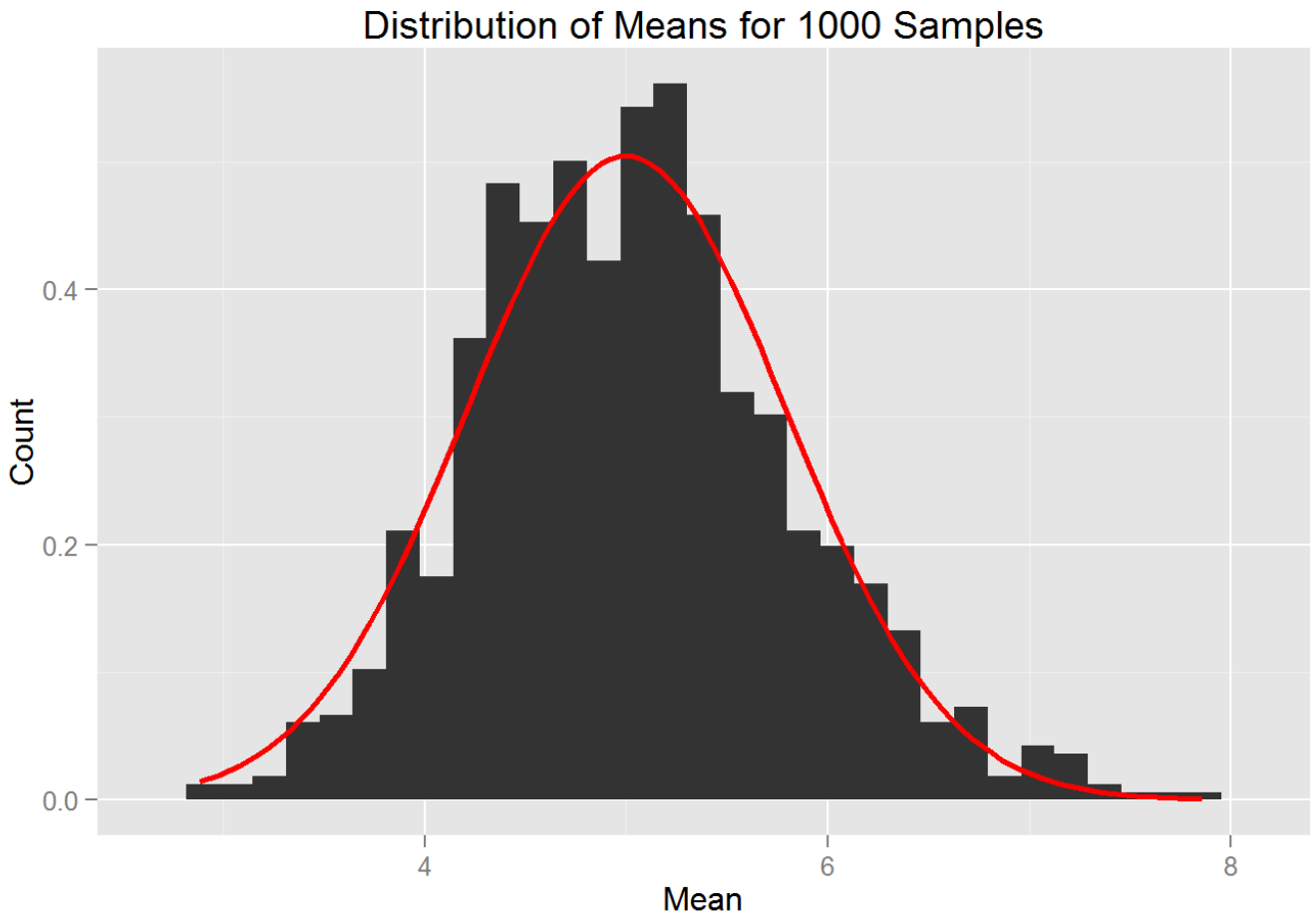
```
## [1] 0.625
```

Based on the above simulation, we see that the sample variance and theoretical variance are close as well.

3. Show that the distribution is approximately normal.

```
library(ggplot2)
ggplot(data=sim.means,aes(x=x)) + geom_histogram(aes(y=..density..))+labs(title="Distri
bution of Means for 1000 Samples") +  labs(x="Mean", y="Count") +
stat_function(fun = dnorm, arg=list(mean=expectedmean,sd=expectedsd),color="red",siz
e=1)
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



The histogram shows the distribution of the sample means from the simulation while the red line shows a normal distribution. By overlaping the 2 graphs together, we can see that the distribution is approximately normal.