Validation of Central Limit Theorem

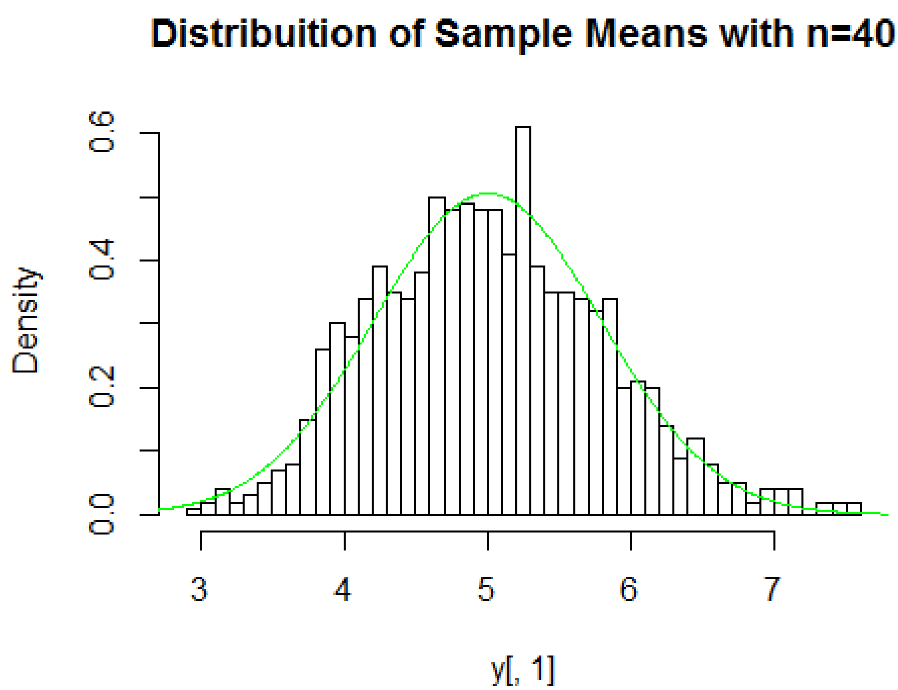
Coursera: Statistical Inference Project 1

Eric Hullander

Exponential Distribution

- $X \sim \exp(\lambda)$.
- $\mu = E[X] = 1/\lambda$
- $\sigma = [X] = 1/\lambda$

The mean of 40 samples are taken from an exponential distribution described above. A simulation is run and the mean is calculated 10000 times. The mean and standard deviation of the means are then calculated from the distribution of sample means. As can be seen, the histogram of sample means approximates the normal curve with $\mu = 1/\lambda$ and sd, $\sigma = (1/\lambda)/\sqrt{n}$.



Mean of Sample Means

$E[\overline{X}_i] =$

```
## [1] 5.019785
```

Theoretical Mean

$1/\lambda =$

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

Standard Deviation of Sample Means (aka the standard error)

$\sqrt{E[(\bar{X}_i - \mu)^2]} =$

```
## [1] 0.8044541
```

Theoretical Standard Deviation of Sample Means

$(1/\lambda)/\sqrt{n} =$

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

The sample mean is a good predictor of the population mean and with a standard error of about .8. This value agrees with the theoretical mean and calculated standard error.

```
N=10000
n=40
m=5
ypop=rexp(N, 1/m)

y=data.frame()
for(i in 1:1000)
  {
    ysamp=sample(ypop,n, replace=FALSE)
    y[i,1]=mean(ysamp)
    y[i,2]=var(ysamp)
    ##print(x)
   }
hist(y[,1], breaks = 50, prob=TRUE, main='Distribuition of Sample Means with
n=40')
x <- seq(0, 10, length=1000)
#calculate mean and sd of sample means
ms=mean(y[,1])
sds=sd(y[,1])
#plot the normal curve with mu and sd = (1/lambda)/sqrt(n)
norm <- dnorm(x,m,m/sqrt(n))
par(col="green")
lines(x,norm)
```