Statistical Inference Project

David Thayer

Saturday, January 24, 2015

In this report we will investigate the exponential distribution and the Central Limit Theorem. We will draw 1000 samples of a vector of 40 values drawn from the exponential function and compare the sample mean with the theoretical mean. We will also investigate the sample variance compared to the theoretical variance.

```
mns <- sapply(1:1000, function(i) { mean(rexp(40, 0.2))})
stddevs <- sapply(1:1000, function(i) { sd(rexp(40, 0.2))})
smplMean <- mean(mns)
smplStdDev <- mean(stddevs)

delta <- abs(5.0 - smplMean)
deltaVar <- abs(5.0 - smplStdDev)
```

The following summary of the sample means indicates that the mean of the collection of samples is very close to the theoretical mean of 5.0.

```
summary(mns)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.805   4.511   4.985   5.018   5.490   7.815
```

```
## sample variance
sd(mns)
```

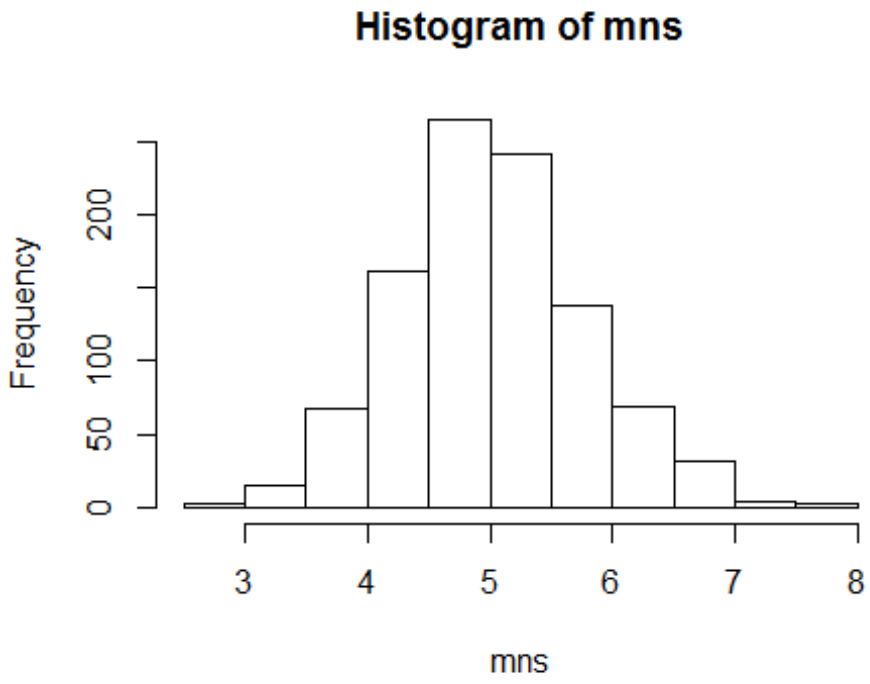
```
## [1] 0.764529
```

```
## theoretical variance
5*5/40
```

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

The sample variance and mean are very close to the theoretical values.

Examining a histogram of the sample means we see that it looks very close to a bell shaped normal curve.



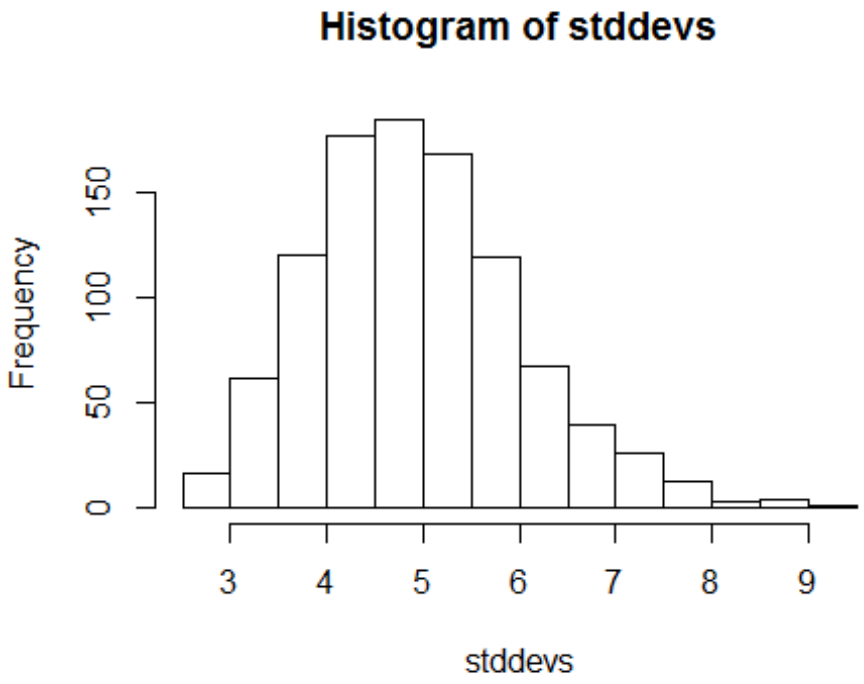
The following summary of the sample variance indicates that the variance of the collection of samples is very close to the theoretical mean of 5.0.

```
summary(stddevs)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.511   4.146   4.830   4.937   5.561   9.034
```

```
sd(stddevs)
```

```
## [1] 1.086881
```



Examining the quantiles of the sample distribution versus the quantiles of the normal(5,5) distribution we find the differences between corresponding quantiles are very small.

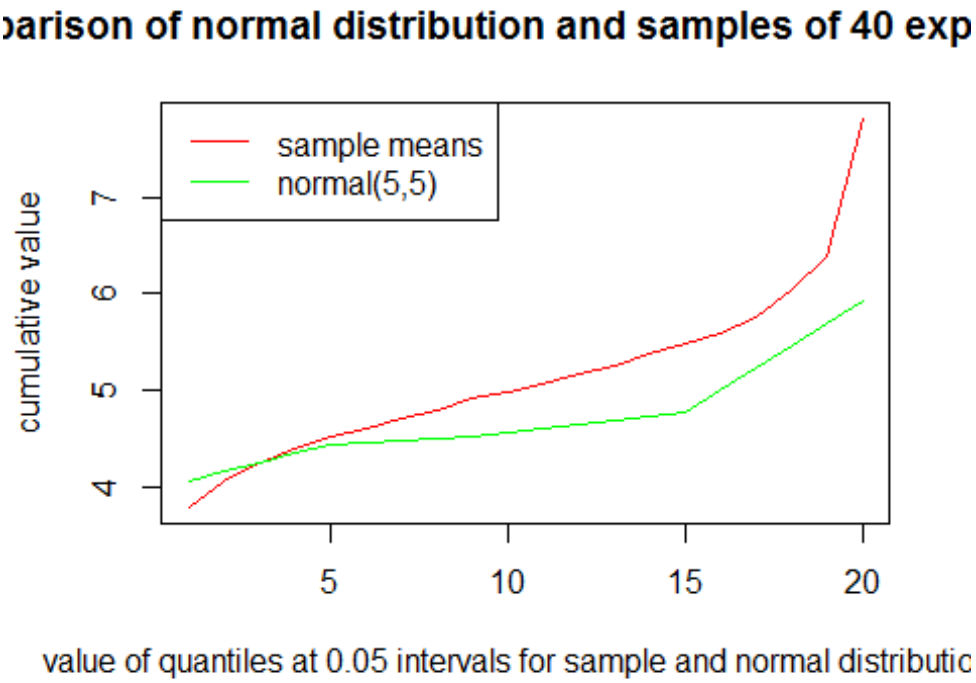
```
d1 <- quantile(mns, seq(from = 0.05, to = 1.0, by = 0.05))
d2 <- quantile(rnorm(5.0,5.0), seq(from = 0.05, to = 1.0, by = 0.05))
```

The differences between the sets of quantiles are listed below.

```
d2 - d1

##           5%          10%          15%          20%          25%
##  0.277795889  0.091973585  0.003858828 -0.047248325 -0.069358583
##          30%          35%          40%          45%          50%
## -0.140468621 -0.221784183 -0.279395327 -0.391665660 -0.431611397
##          55%          60%          65%          70%          75%
## -0.476050255 -0.525617344 -0.575454532 -0.651391243 -0.709025213
##          80%          85%          90%          95%         100%
## -0.587760270 -0.525582002 -0.594335836 -0.701380771 -1.894501146
```

Plotting the two sets of quantiles we can visually compare them.



Comparing the plots of the quantile values of the sample means and the quantiles of the normal distribution with the theoretical mean and variance are close to each other. This is a good indicator that the distribution of the sample means is normal.