Exponential Distribution Investigation in R - Statistical Inference: Course Project - Question 1

Martin Cote

April 26, 2015

Overview

This report provides a comparison between the values obtained directly via the sample (and by the Central Limit Theorem) and to those of the entire population.

Required tools

Loading the required libraries.

Simulation

Preparing the simulation using: 1. num: number of exponentials per simulations 2. lambda: the rate parameter used within this investigation 3. numsim: number of simulations used in this investigation

By: 1. running a 'numsim' times the 'rexp' (the "random exponantial distribution") with a 'lambda' rate for n=num. 2. saves the results into matrix for further manipulation

```
# Number of averages
num <- 40
# By default, our lambda value will be permanently set to 0.2
lambda <- 0.2

# Number of simulations
numsim <- 1000

# Simulating the random generation of an exponential distribution
simulateddata <- matrix(data=replicate(numsim, rexp(n=num, rate=lambda)), nrow=numsim, ncol=num, byrow='
```

Sample Mean versus Theoretical Mean

Comparing the sample mean (i.e. the mean for each rows) versus the theoritical mean (i.e. the mean for the entire simulation).

```
# Calculating the overall means
mns <- rowMeans(simulateddata)
mean(mns)
```

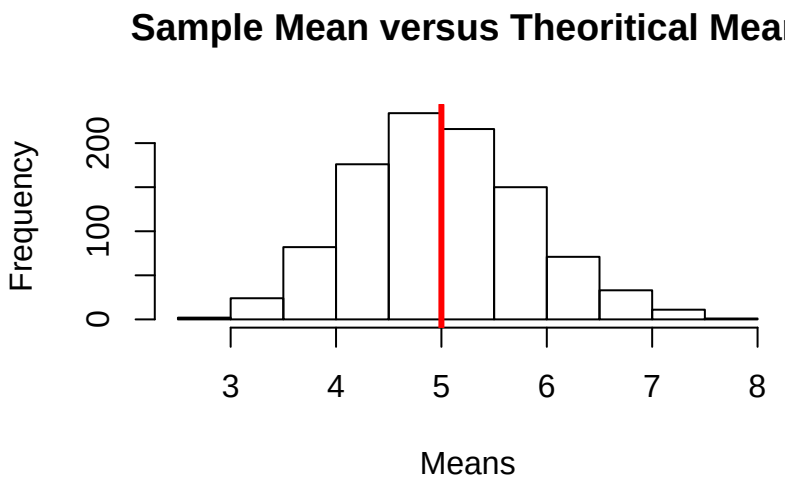
```
## [1] 5.000664
```

```r
# Calculating the mean
1 / lambda
```

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

As observed, the sample mean calculated is fairly closed to the averages of 40 means exponantials simulated, hence proving the CLT.

```r
# Displaying the histogram of all simulated means of 40 random exponantials
hist(mns, xlab="Means", ylab="Frequency", main="Sample Mean versus Theoritical Mean")
#abline(v=mean(mns), col="blue", lwd=3)
abline(v=1/lambda, col="red", lwd=3)
```



The distribution is centered at or around both the sample mean or theoritical mean.

Sample Variance versus Theoretical Variance

Comparing the sample variance (i.e. the variance for each rows) versus the theoritical variance (i.e. the variance for the entire simulation).

```r
# Calculating the variances and averaging them:
vrs = NULL
for (i in 1:numsim) {
  vrs = c(vrs, var(simulateddata[i, ]))
}
mean(vrs)
```

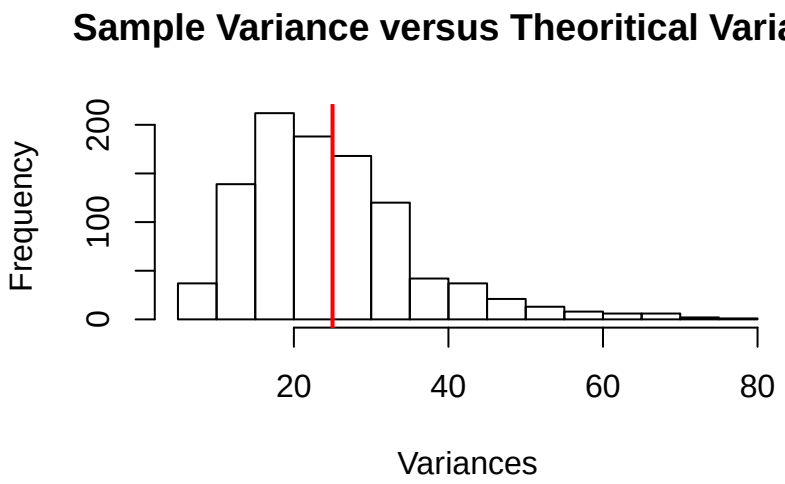
```
## [1] 24.78703
```

```r
# Calculating the variance
(1 / lambda)^2
```

```
## [1] 25
```

As observed, the averages of all variances of the simulated 40 random exponantials is fairly closed to the calculated variance, hence proving the CLT.

```
hist(vrs, xlab="Variances", ylab="Frequency", main="Sample Variance versus Theoritical Variance")
abline(v=(1/lambda)^2, col="red", lwd=2)
```

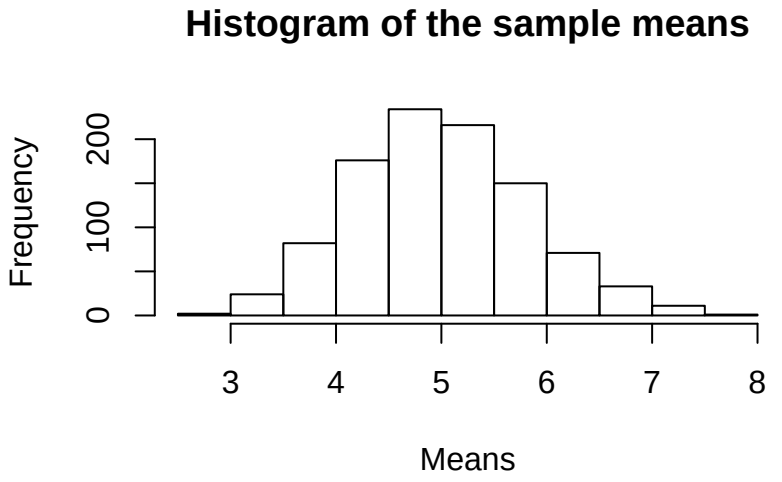


The distribution is centered at or around both the sample variance or the theoritical variance.

Distribution

Investigating if the overall distribution is normal.

```
hist(mns, xlab="Means", ylab="Frequency", main="Histogram of the sample means")
```



Since the histogram follows/is closed to a normal distribution, we can assume the distribution is approxamitively normal.