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

Martin Cote

April 26, 2015

Overview

Investigation of the exponential distribution in R and comparison with the Central Limit Theorem. This report will test using a simulation the distribution obtained from it, confirming the CLT by validating the distribution as a standard normal distribution.

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

```
num <- 40 # Number of averages
lambda <- 0.2 # By default, our lambda value will be permanently set to 0.2
numsim <- 1000 # Number of simulations

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

Sample Mean versus Theoretical Mean

Comparing the sample mean versus the theoritical mean.

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

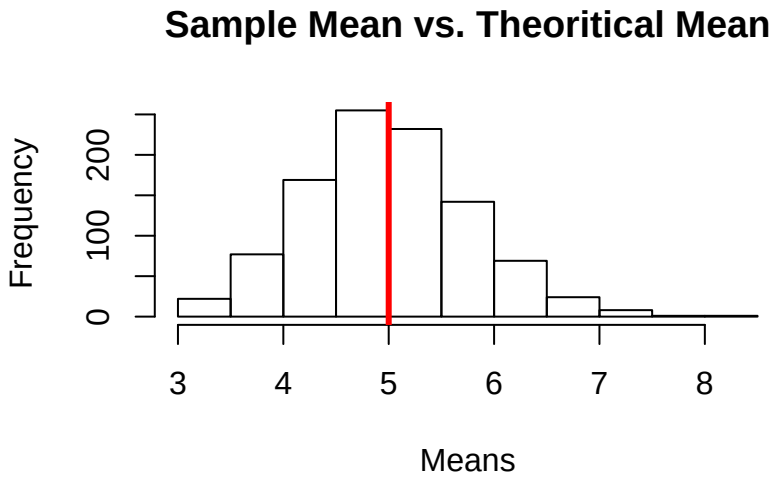
```
## [1] 4.991369
```

```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.

```r
# Displaying the histogram of all simulated means of 40 random exponantials
hist(mns, xlab="Means", ylab="Frequency", main="Sample Mean vs. Theoritical Mean")
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 versus the theoritical variance.

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

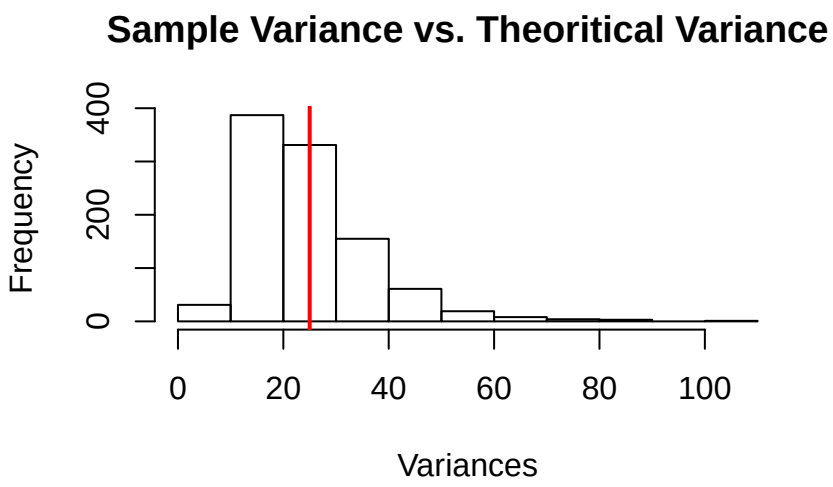
```
## [1] 24.69762
```

```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.

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

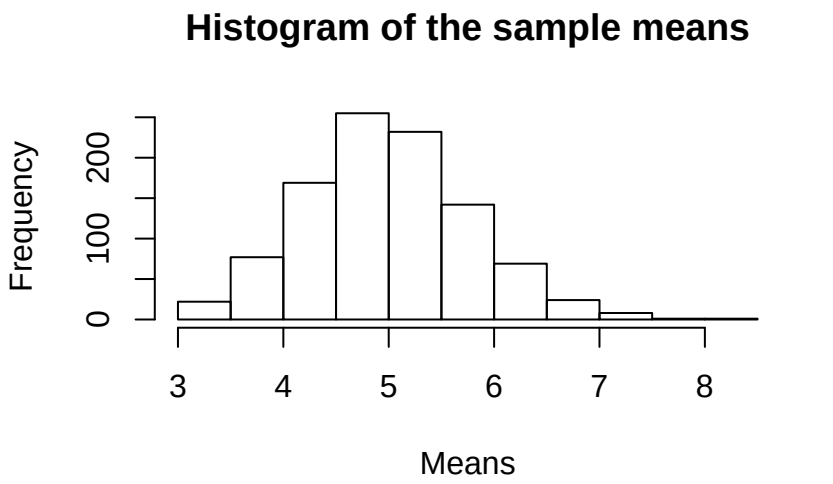


The distribution is centered at or around both the sample variance or the theoritical variance.

Distribution

Investigating if the overall distribution is normal.

```r
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.