Statistical Inference Course Project Pt.1

Omer Yavin

March 2, 2018

Simulation exercise

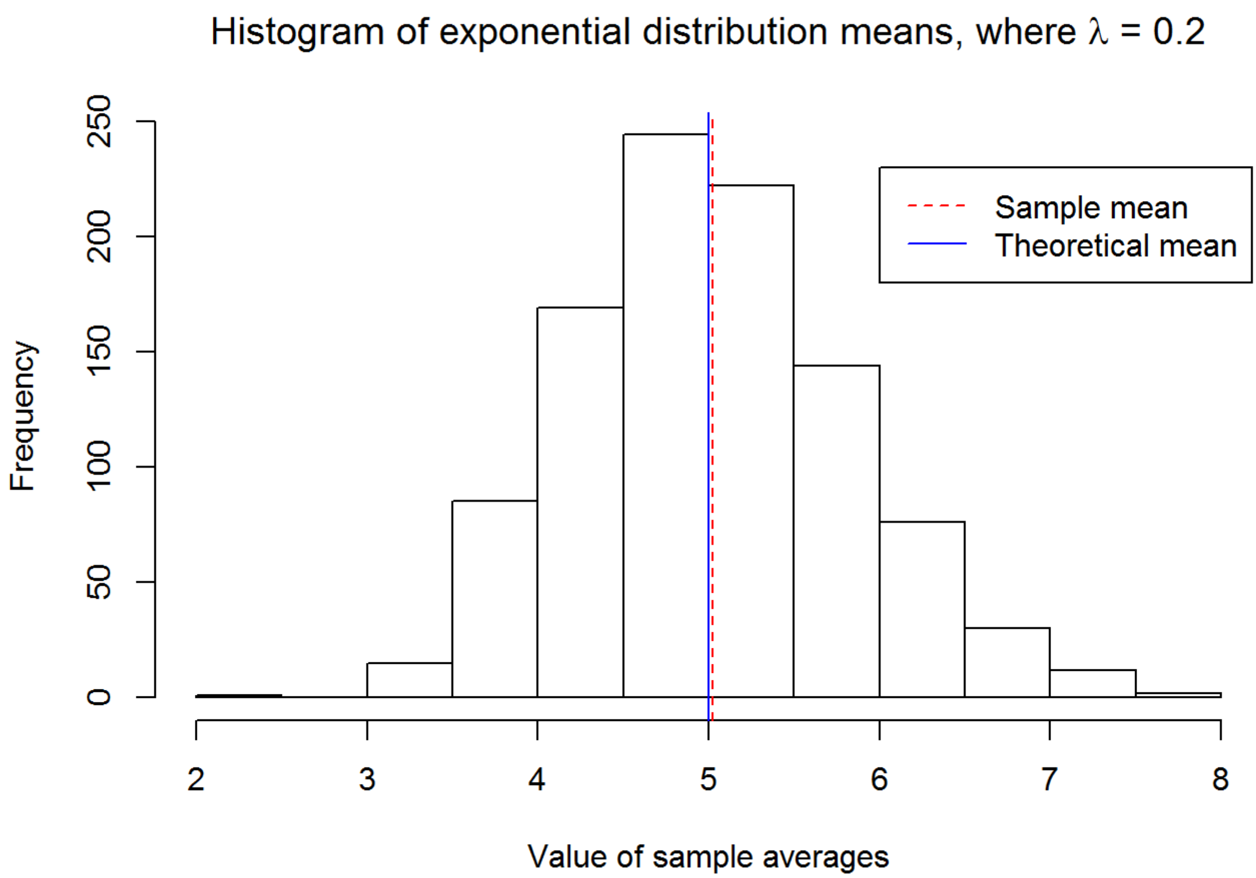
In this part of the project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The mean of exponential distribution is 1/and the standard deviation is also 1/. We will set = 0.2. We will investigate the distribution of averages of 40 exponential distribution simulations.

This will be done in 3 stages:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Part 1: Sample Mean versus Theoretical Mean

To get a sample mean, 1000 iterations (as instructed) of 40 exponential distribution* 'experiments' each will be performed. For each experiment the mean will be calculated, then a histogram created of the aggregate results.




```
## [1] 5.025289
```

Visually one can see the sample mean is very close to the theoretical mean, also the number itself is close to 5. But how close really? Running r's t.test (in the appendix) one can examine this (null) hypothesis - $H_0 : \mu = 5$.

In the result - zero (and the acquired sample mean) is well within the 95% confidence range. Also p-value is around 0.33, which means even if is allowed to go up to 0.3, the null hypothesis (=5) would still not be rejected.

Part 2: Sample Variance versus Theoretical Variance

The same sample of a thousand iterations of means of 40 'experiments' of exponential distributions will be used here, this time to compare the sample variance to the theoretical variance.

The theoretical variance for an exponential distribution is $\sigma^2 = \frac{1}{\lambda^2}$, hence:

```
#Theoretical variance
1/lambda^2
```

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

```
#Sample variance
var(esm)
```

```
## [1] 0.6519395
```

It seems as though the variances are very far apart, but in reality the sample variance we are calculating is not variance of the mean of 40 samples at a time, not of the samples themselves. The theoretical variance of the sample mean shrinks in relation to the sample size (N) ($\sigma^2_{\bar{X}} = \frac{\sigma^2}{N}$), hence:

```
#Theoretical mean variance
1/(sample_sz*lambda^2)
```

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

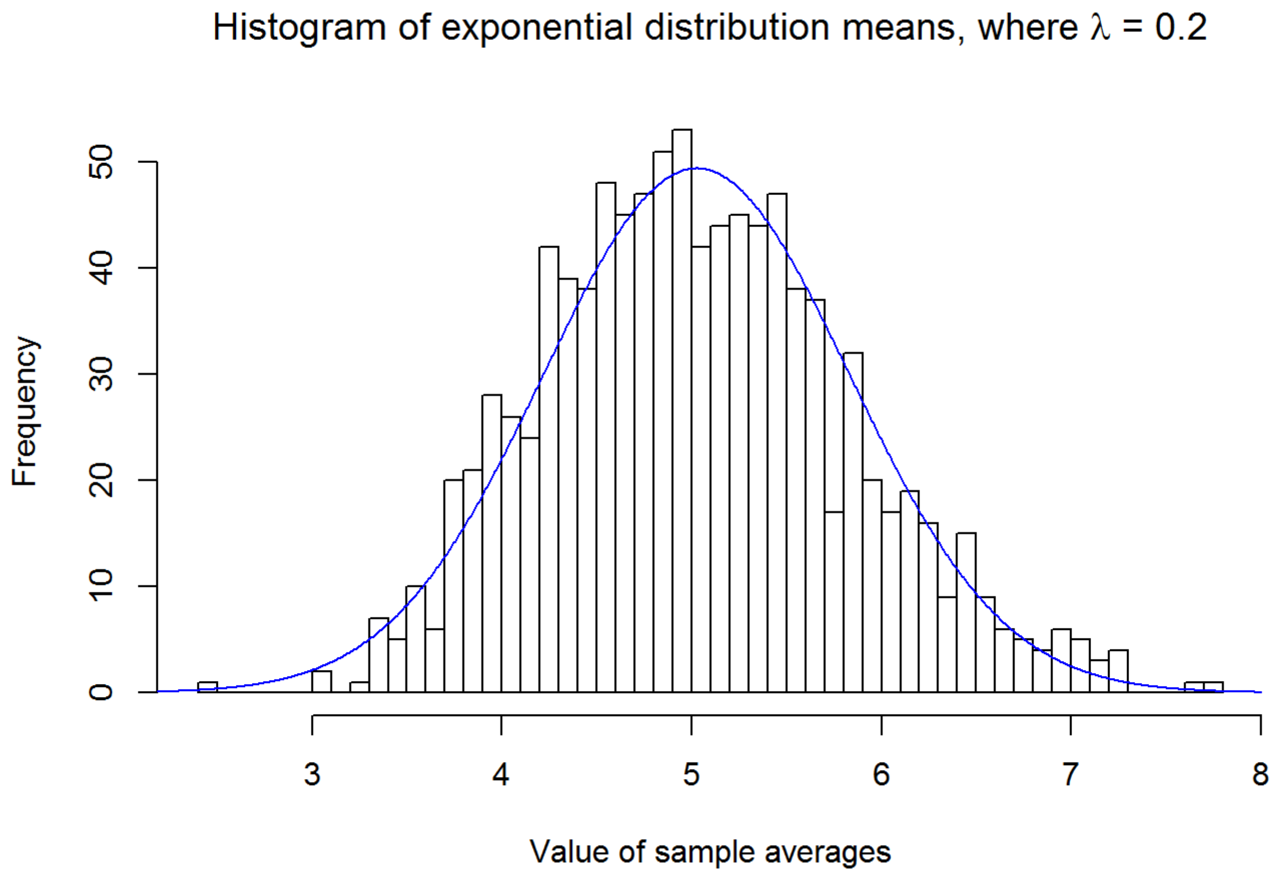
It is clear that the sample variance calculated earlier is actually very close to this result.

Part 3: Distribution

It is visible in the histogram itself that the distribution is bell shaped, as a normal distribution. To make this even more visible, an actual normal distribution (with the same mean and variance we've calculated) can be simulated and plotted on top of the obtained histogram.

The histogram slots will be made denser here to try and create a more continuous histogram plot.

```
h<-hist(esm,breaks = 60, xlab = "Value of sample averages", main = expression(paste("Histogra
m of exponential distribution means, where ", lambda, " = 0.2")))
x<-seq(2,8,length = 1000)
norm_fit<-dnorm(x,mean=sample_mn,sd=sd(esm))
norm_fit<-norm_fit*diff(h$mids[1:2])*length(x)
lines(x,norm_fit,col = "blue",lty=1)
legend(6,230,legend = "Normal distribution",col = "blue",lty=1)
```



With the ‘theoretical plot’ on top of the histogram it is visible just how similar the distribution is, in shape, to a normal distribution. It can be concluded that they are quite similar.

(*) - While each experiment is random, the seed was set to a constant (4) for reproducible results, and the ability to comment on the nature of the results.

Appendix

1. The code used to generate the histogram with the means marked over it (plot suppressed here):

```
sample_sz = 40
iters = 1000
lambda = 0.2
esm = NULL
for (i in 1 : iters) esm = c(esm, mean(rexp(sample_sz,lambda)))
sample_mn = mean(esm)
#hist(esm, xlab = "Value of sample averages", main = expression(paste("Histogram of exponenti
al distribution means, where ", lambda, " = 0.2")))
#abline(v=sample_mn,col="red",lty=2)
#abline(v=1/lambda,col="blue",lty=1)
#legend(6,230,legend = c("Sample mean","Theoretical mean"),col = c("red","blue"),lty=c(2,1))
sample_mn
```

2. Two-sided t.test for determining if the simulated mean is equal to the theoretical one:

$H_0 : \mu = 5$

$H_1 : \mu \neq 5$

```
t.test(esm,mu=1/lambda)
```

```
##
##  One Sample t-test
##
## data:  esm
## t = 0.50648, df = 999, p-value = 0.6126
## alternative hypothesis: true mean is not equal to 5
## 95 percent confidence interval:
##  4.963014 5.062720
## sample estimates:
## mean of x
##  5.012867
```