Experimenting the "CENTRAL LIMIT" theorem

Author: "G ARGENTON"

Overview

This document attempt to illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials, compare it to a normal distribution and doing so, to illustrate the central limit theorem.

1. The sample mean is shown and compared to the theoretical mean of the distribution.
2. The sample variance is shown and compared to the theoretical variance of the distribution.
3. The distribution is plotted (histogram) and compared to normal distribution.

Simulations

We will investigate the exponential distribution with lambda = 0.2.

First we set the seed to make the analysis reproducible, then we initialize parameters lambda (0.2) and numbers (n=40) of explonentials.

```
set.seed(1003)
lambda <- 0.2; n <- 40;
```

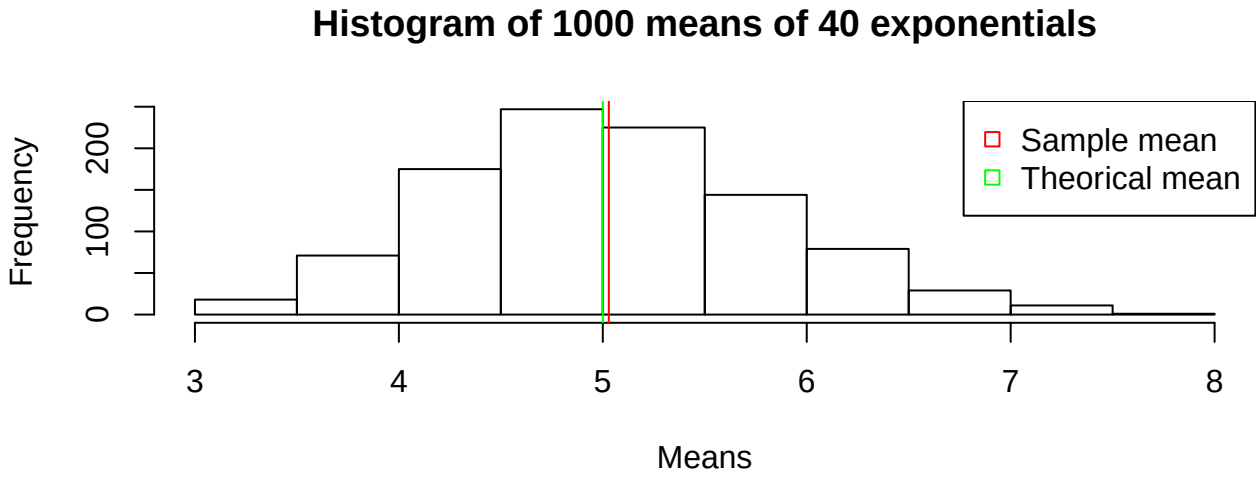
Sample Mean versus Theoretical Mean:

Let's simulate 1000 trials of the mean of 40 exponentials and study the distribution of that mean.

Results are shown using an histogram of those 1000 trial. Sample mean is represented by a verticle red line. Theorical mean (1/lambda) is represented by a vertical green line.

If n is sufficiently large (n=40) those two values should be very close.

```
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(n, lambda)))
hist(mns, main = "Histogram of 1000 means of 40 exponentials",xlab = "Means")
abline(v=mean(mns), col = "red" ,lwd =1); abline(v=(1/lambda), col = "green" ,lwd =1);
legend("topright", pch = 0, col = c("red","green"), legend = c("Sample mean","Theorical mean"))
```



```
print(paste("Sample mean:", round(mean(mns),2)));print(paste("Theorical mean:", round(1/lambda,2)));
```

```
## [1] "Sample mean: 5.03"
```

```
## [1] "Theorical mean: 5"
```

```
print(paste("GAP between sample & theorical mean in %:",round(abs((mean(mns)-1/lambda)*100/(1/lambda)),1
```

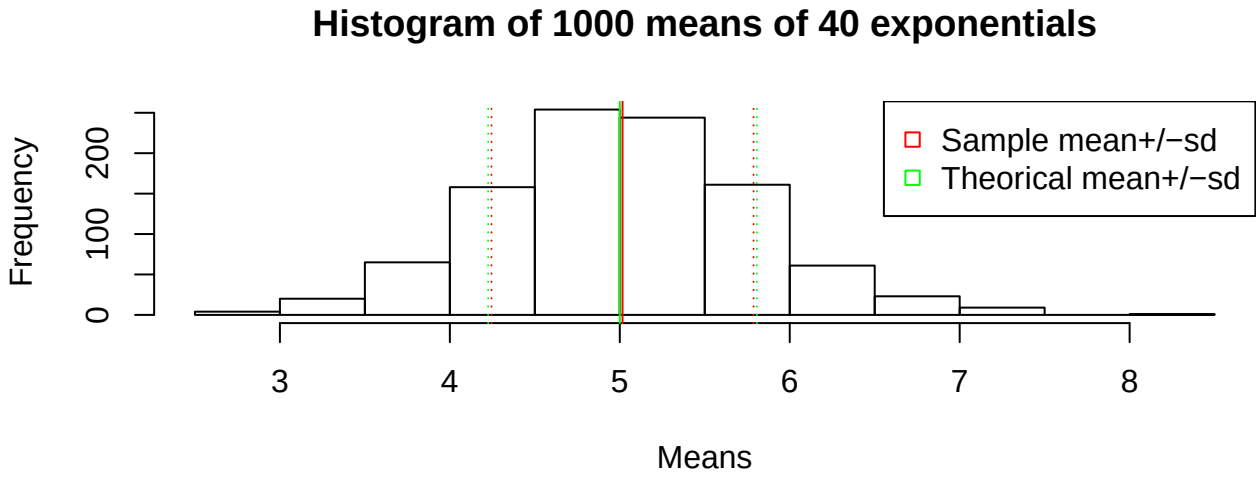
```
## [1] "GAP between sample & theorical mean in %: 0.59"
```

Sample Variance versus Theoretical Variance:

Let's calculate the variance of the sample of 1000 trials of the mean of 40 exponentials and compare it to the theorical variance of that distribution (1/lambda)^2/40.

If n is sufficiently large (n=40) those two values should be close.

```
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(n, lambda)))
hist(mns, main = "Histogram of 1000 means of 40 exponentials", xlab = "Means")
sample.var <- var(mns); th.var <- (1/lambda)^2/40;
abline(v=mean(mns), col = "red" ,lwd =1); abline(v=(1/lambda), col = "green" ,lwd =1);
abline(v=mean(mns)-sqrt(sample.var), col = "red" ,lwd =1, lty = "dotted")
abline(v=mean(mns)+sqrt(sample.var), col = "red" ,lwd =1, lty = "dotted")
abline(v=mean(mns)-sqrt(th.var), col = "green" ,lwd =1, lty = "dotted")
abline(v=mean(mns)+sqrt(th.var), col = "green" ,lwd =1, lty = "dotted")
legend("topright", pch = 0, col = c("red","green"), legend = c("Sample mean+/-sd ","Theorical mean+/-sd
```




```
print(paste("Sample variance:", round(sample.var,2))); print(paste("Theorical variance:", round(th.var,
```

```
## [1] "Sample variance: 0.59"
```

```
## [1] "Theorical variance: 0.62"
```

```
print(paste("GAP between sample variance & theorical variance in %:", round(abs((sample.var-th.var)*100,
```

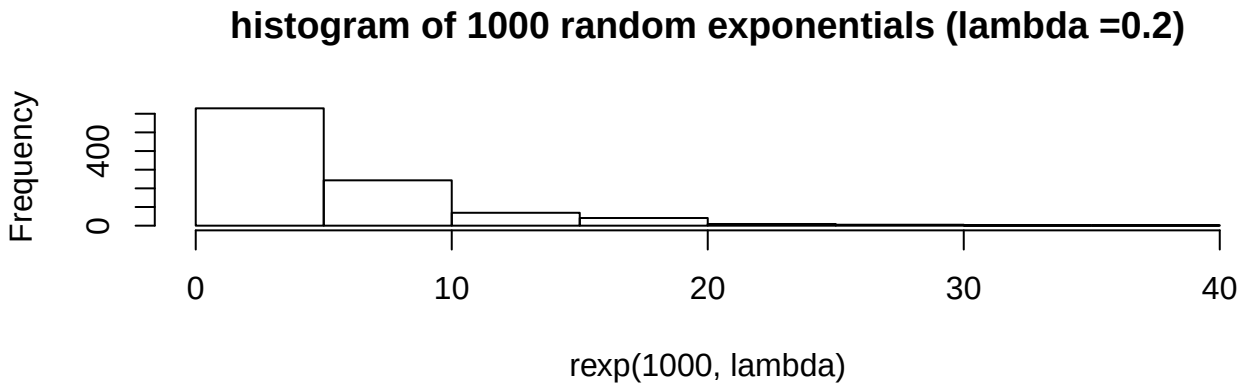
```
## [1] "GAP between sample variance & theorical variance in %: 4.85"
```

Distribution: Via figures and text

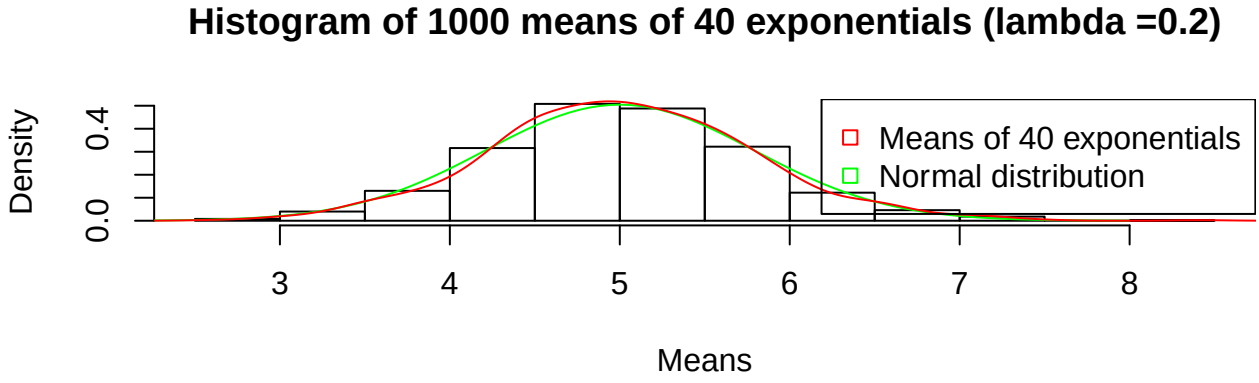
According to central limit theorem, if n is large, distribution of means of n exponentials is close to a normal distribution of mean 1/lambda and standard deviation of (1/lambda)/sqrt(n)

Lets plot an histogram of 1000 random exponentials and an histogram of 1000 means of 40 exponentials and compare its density curve (red curve) to the density curve of a normal distribution (green curve)

```
hist(rexp(1000, lambda), main = "histogram of 1000 random exponentials (lambda =0.2)")
```




```
hist(mns, main = "Histogram of 1000 means of 40 exponentials (lambda =0.2)", prob = TRUE, xlab = "Means
plot(function(x) dnorm(x,1/lambda,1/lambda/sqrt(40)), xlim=c(2,8), col = "green",add = TRUE )
lines(density(mns), col = "red")
legend("topright", pch = 0, col = c("red","green"), legend = c("Means of 40 exponentials","Normal distri
```



The two curves are very close!