Simulation Illustration of Central Limit Theorem

Author: Viktor L.

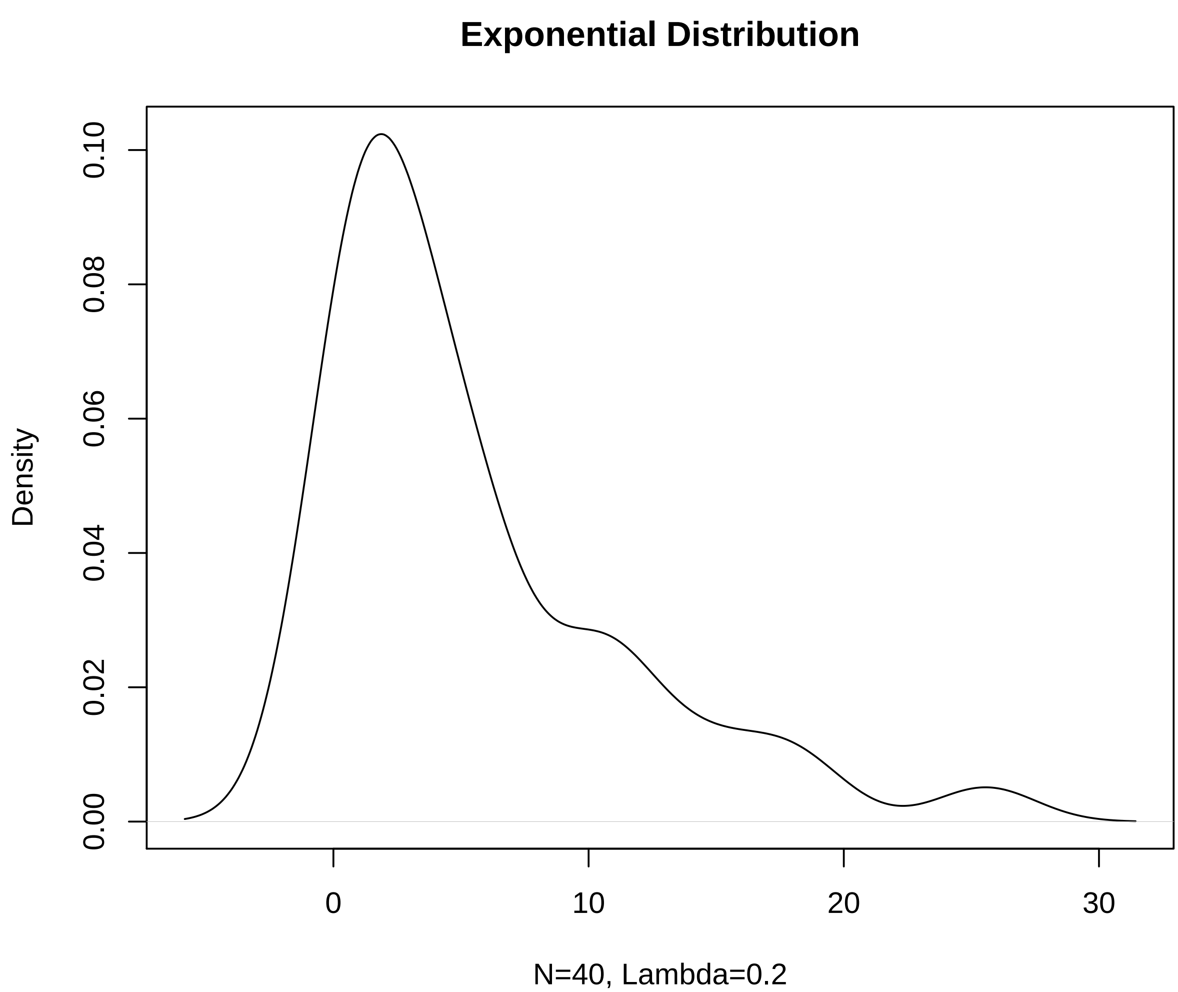
Overview

This article is a short illustration of the central limit theorem using a simulation of random variables from the exponential distribution. The central limit theorem states that the arithmetic mean of a sufficiently large sample of independent draws from any distribution (given it has a well-defined mean and variance) is approximately normally distributed.

Simulations

First, let's take a look at the density of the exponential distribution.

```
plot(density(rexp(40,0.2)), main="Exponential Distribution", xlab="N=40, Lambda=0.2")
```



This looks different from the density of a normal distribution.

We can simulate draws from the exponential distribution with the R command rexp().
Here we simulate 1000 draws of 40 expontial random varables with parameter lamda = 0.2 and take the mean of each draw.

```
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(40, 0.2)))
mean(mns)
```

```
## [1] 5.026412
```

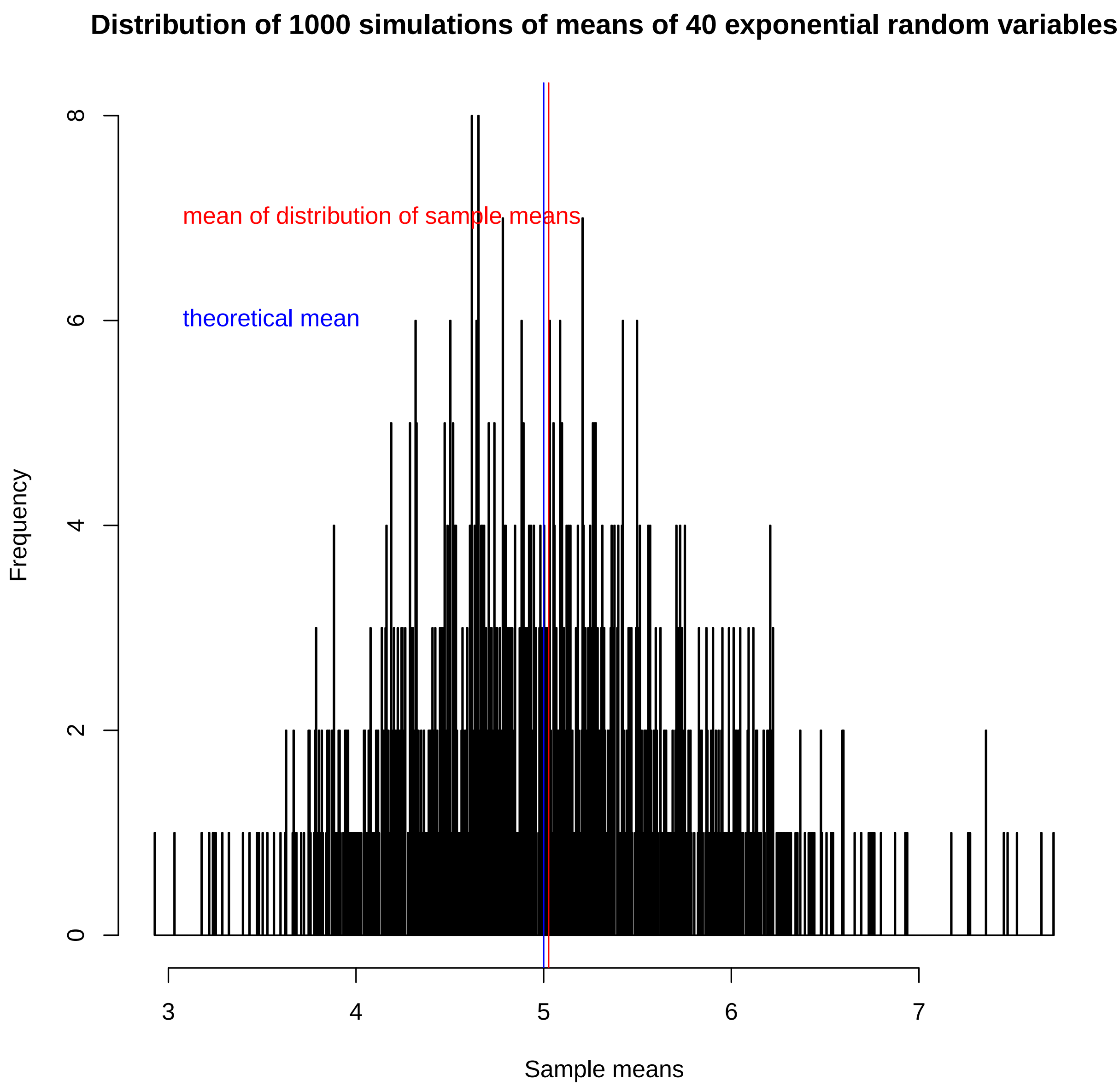
```
summary(mns)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.927   4.502   4.992   5.026   5.512   7.718
```

Sample Mean versus Theoretical Mean

If we plot the means of the simulations, we can see, that the distribution of means looks quite different from exponential distribution. The next plot shows a histogram of simulated means with a red line indicating the sample mean, 5.026 and a blue line indicating the theoretical mean of the exponential distribution, 5 (1/lambda).

```
hist(mns, breaks=1000, main="Distribution of 1000 simulations of means of 40 exponential random variabl
     xlab="Sample means")
abline(v=mean(mns), col="red")
text(3,7, pos=4,"mean of distribution of sample means", cex=1,col="red")
abline(v=5, col="blue")
text(3,6, pos=4,"theoretical mean", cex=1,col="blue")
```



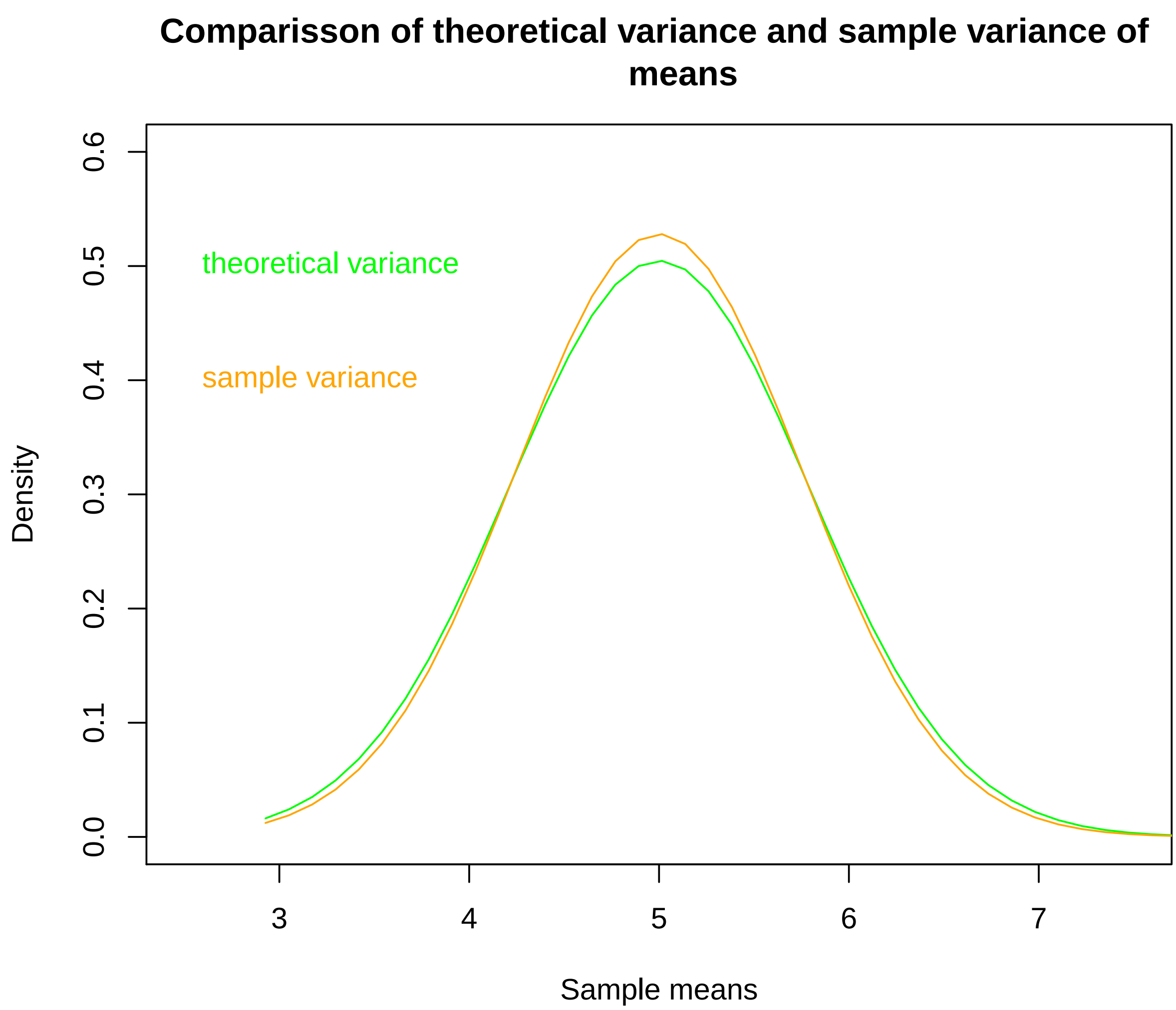
We can see that the mean of 1000 simulations is pretty close to the theoretical mean.

Sample Variance versus Theoretical Variance

```
sample_var <- var(mns)
theor_var <- 1/(0.2^2)/40
```

Just as the mean, the sample variance of the simulations, 0.571 is pretty close the theoretical variance 0.625 (variance/sample size). To illustrate the difference between those variances, the next plot depicts two normal densities with the same mean of 5 and the two different variances, the sample variance and the theoretical variance.

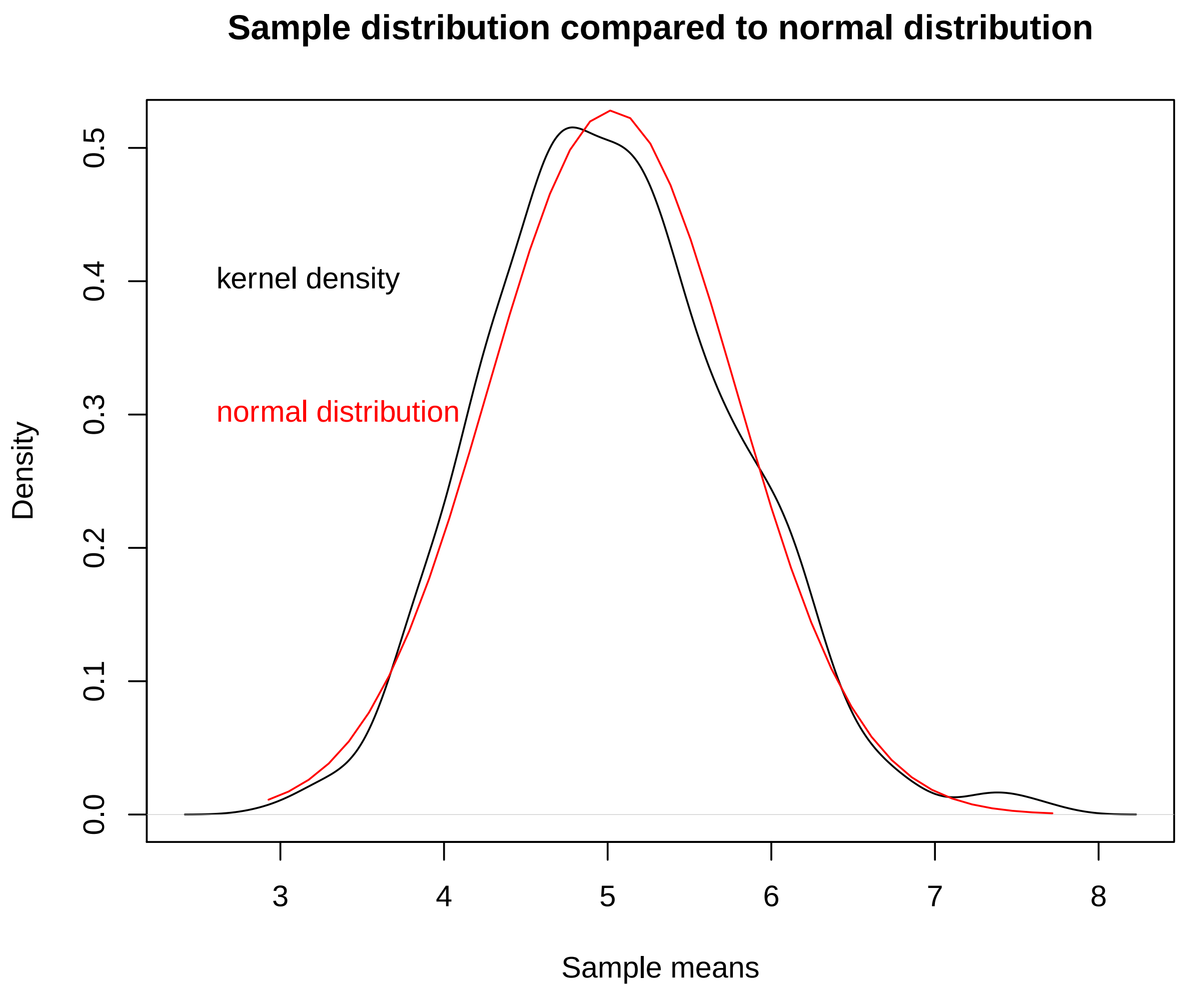
```r
sample_var <- var(mns)
theor_var <- 1/(0.2^2)/40
xfit <- seq(min(mns), max(mns), length=40)
yfit <- dnorm(xfit, 5, sqrt(theor_var))
plot(1, type="n", xlim=c(2.5,7.5), ylim=c(0,0.6),
     main="Comparisson of theoretical variance and sample variance of
     means", xlab="Sample means", ylab="Density")
lines(xfit, dnorm(xfit, 5, sqrt(theor_var)), col = "green")
lines(xfit, dnorm(xfit, 5, sqrt(sample_var)), col = "orange")
text(2.5,0.5, pos=4,"theoretical variance", cex=1,col="green")
text(2.5,0.4, pos=4,"sample variance", cex=1,col="orange")
```



Distribution

To check the normality of the simulated means, we plot the kernel density of the simulations next to a normal density with the same mean and standard deviation.

```r
plot(density(mns), main="Sample distribution compared to normal distribution",
     xlab="Sample means")
xfit <- seq(min(mns), max(mns), length=40)
lines(xfit, dnorm(xfit, mean=mean(mns), sqrt(sample_var)), col = "red")
text(2.5,0.4, pos=4,"kernel density", cex=1,col="black")
text(2.5,0.3, pos=4,"normal distribution", cex=1,col="red")
```



Just as the central limit theorem suggests for large sample sizes, the distribution of means looks very similar to a normal distribution.