# Coursera Statistical Inference Course Project

*Ralph Fehrer*

*Wednesday, November 19, 2014*

In probability theory, the central limit theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution (http://en.wikipedia.org/wiki/Central_limit_theorem).

In this markdown file, the central limit theorem is applied to estimate the mean and the standard deviation of an exponential distribution and the results are compared to the real values of the distribution parameters.

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. For the experiments in this paper, lambda is set to

```
lambda = 0.2
```

The mean and the standard deviation of an exponential distribution are

```
exponentialMean = exponentialSd = 1/lambda
```

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution

In the first experiment, the distribution of 1000 averages, estimated from 40 exponential(0.2)s is simulated:

```
numberOfExponentials = 40
numberOfAverages = 1000
mns=0
for (i in 1 : numberOfAverages) mns = c(mns, mean(rexp(numberOfExponentials,lambda)))
```

The sample mean of the averages distribution is

```
mean(mns)
```

```
## [1] 4.98
```

The difference between the sample mean and the real distribution mean is

```
abs(exponentialMean-mean(mns))
```

```
## [1] 0.01975
```

2. Show how variable it is and compare it to the theoretical variance of the distribution.

In the second experiment, the distribution of 1000 standard deviations, estimated from 40 exponential(0.2)s is simulated:

```r
numberOfExponentials = 40
numberOfStandardDeviations = 1000
sds=0
for (i in 1 : numberOfStandardDeviations) sds = c(sds, sd(rexp(numberOfExponentials,lambda)))
```

The sample mean of the standard deviations distribution is

```r
mean(sds)
```

```
## [1] 4.825
```

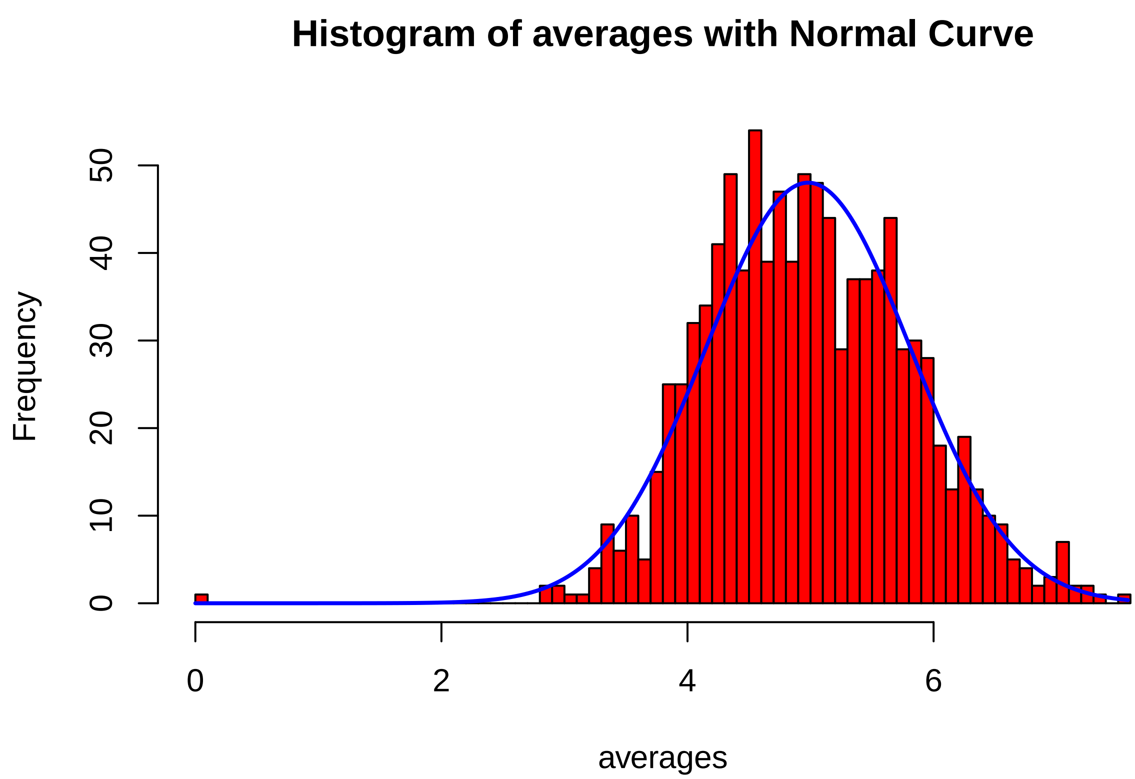
The difference between the sample mean of the standard deviations distribution and the real standard deviation of the exponential distribution is

```r
abs(exponentialSd-mean(sds))
```

```
## [1] 0.1751
```

3. Show that the distribution is approximately normal.

```r
h<-hist(mns, breaks=100, col="red", xlab="averages",
main="Histogram of averages with Normal Curve")
xfit<-seq(min(mns),max(mns),length=1000)
yfit<-dnorm(xfit,mean=mean(mns),sd=sd(mns))
yfit <- yfit*diff(h$mids[1:2])*length(mns)
lines(xfit, yfit, col="blue", lwd=2)
```



```r
h<-hist(sds, breaks=100, col="red", xlab="averages",
main="Histogram of standard deviations with Normal Curve")
xfit<-seq(min(sds),max(sds),length=1000)
yfit<-dnorm(xfit,mean=mean(sds),sd=sd(sds))
yfit <- yfit*diff(h$mids[1:2])*length(sds)
lines(xfit, yfit, col="blue", lwd=2)
```