# Demonstration of the Central Limit Theorem

## Overview

We compare distribution of means 1000 samples (n=40) of exponentially distributed (lambda = 0.2) random numbers with the theoretical expectation based on the Central Limit Theorem

## Simulation

1000 iterations: generating a random sample (sample size = 40) of exponentially distributed data points with lambda = 0.2

```
## mns is the array of the sample means - 1000 in total
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(40,0.2)))
```

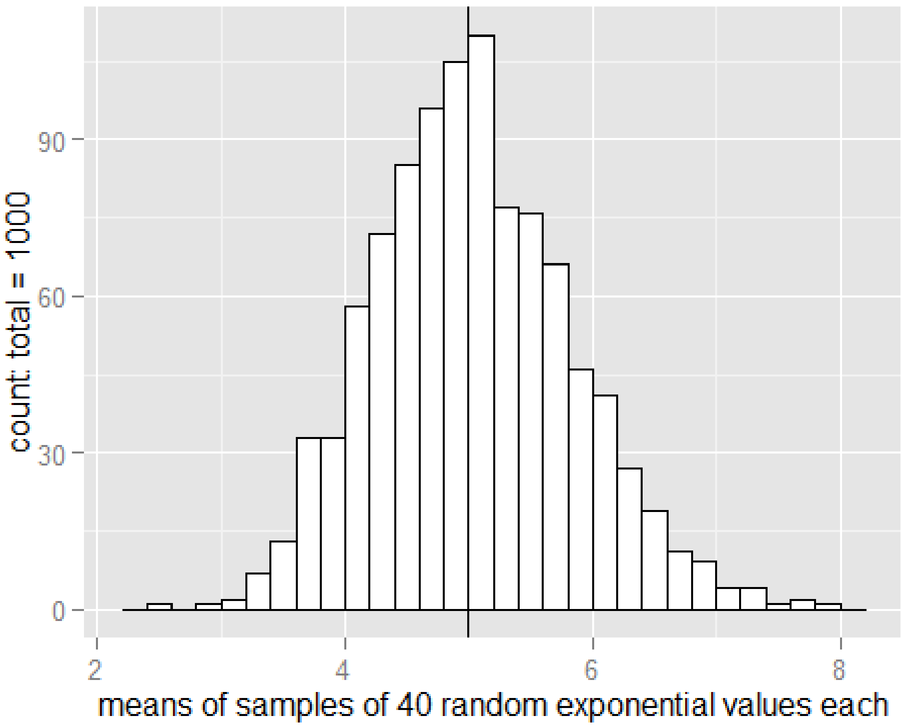
## Sample Mean vs. Theoretical Mean

Theoretically, we expect the sample means to be centred on the real mean of the exponential distribution (1/lambda = 5). This is close approximated by the simulated data:

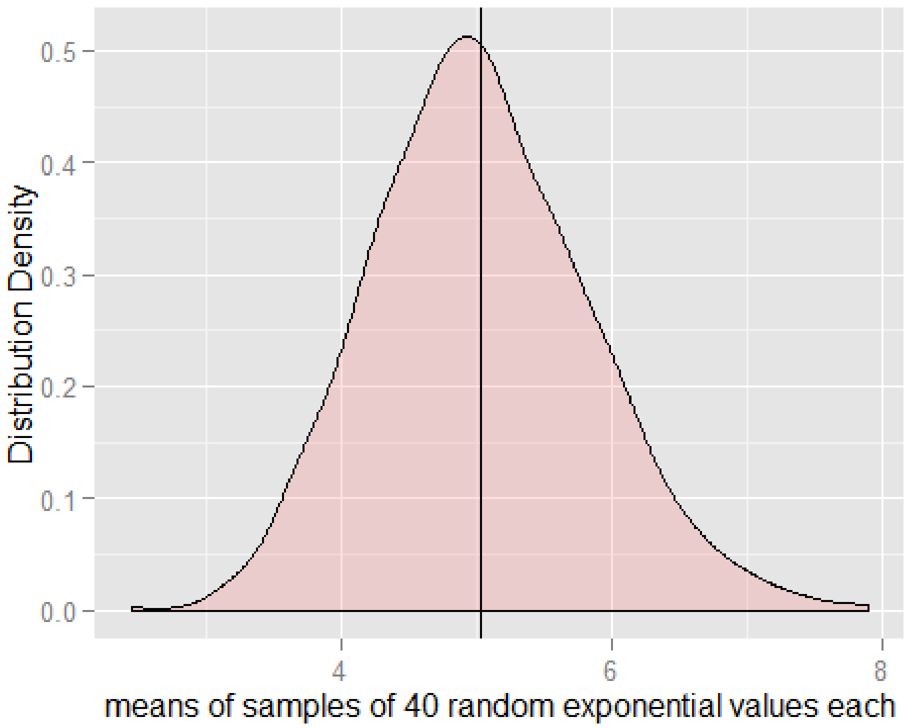
```
mean(mns)

## [1] 5.029584
```

We plot a histogram of the sample means - to show a 'bell' type (Gaussian) distribution, centred on x = 5

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.1.3

df_mns <- data.frame(mns)

g <- ggplot(df_mns, aes(x=mns)) + geom_histogram(binwidth=.2,
colour="black", fill="white")
g <- g + geom_vline(xintercept = 5)
g <- g + xlab("means of samples of 40 random exponential values each")
g <- g + ylab("count: total = 1000")
g
```




```
h <- ggplot(df_mns, aes(x=mns)) + geom_density(alpha=.2, fill="#FF6666") +
geom_vline(xintercept=mean(mns))
h <- h + xlab("means of samples of 40 random exponential values each")
h <- h + ylab("Distribution Density")
h
```



## Sample Variance vs. Theoretical variance

Theoretically, we expect the distribution of the sample means to have variance = sigma^2 / n = 25 /40 = 0.625 This is closely approximated by the actual variance of the sample means:

```
var(mns)

## [1] 0.6432676
```

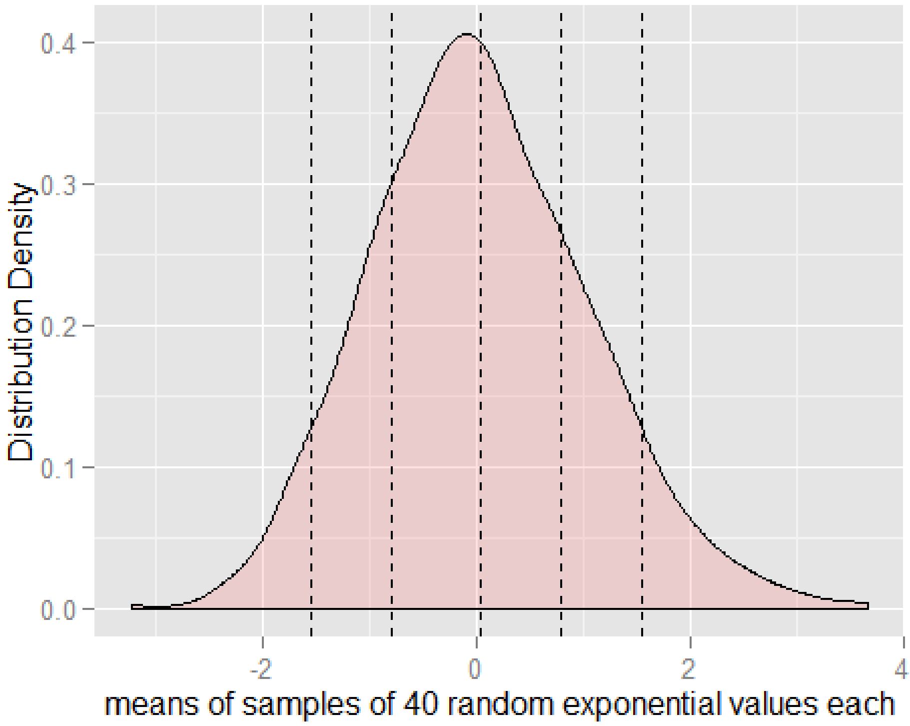
## Comparing with the Standard Normal Distribution

Standardising the distribution of the sample means

```
nl_mns <- (mns-5)/(5/sqrt(40))
df_nl_mns <- data.frame(nl_mns)
```

We plot density curve of the standardised sample means, and compare with some known quantiles and cumulative densities of the Standard Normal distribution

```
j <- ggplot(df_nl_mns, aes(x=nl_mns)) + geom_density(alpha=.2,
fill="#FF6666") + geom_vline(xintercept=c(-1.96*0.7905694, -0.7905694,
mean(nl_mns), 0.7905694, 1.96*0.7905694), linetype="dashed")
j <- j + xlab("means of samples of 40 random exponential values each")
j <- j + ylab("Distribution Density")
j
```




The above exploratory graph shows that the standardised distribution of the 1000 sample means is Gaussian, centred on zero, and its bell-curve distribution resembles the

Standard Normal   Demonstrating the probability distribution of sample means within population mean +- 1.0*sd (theoretical expectation: 0.68*1000 = 680)

```
s1_nl_mns <- nl_mns[(nl_mns <= 1) & (nl_mns >= -1)]
length(s1_nl_mns)

## [1] 683
```

Demonstrating the probability distribution of sample means within population mean +- 1.96*sd (theoretical expectation: 0.95*1000 = 950)

```
s2_nl_mns <- nl_mns[(nl_mns <= 1.96) & (nl_mns >= -1.96)]
length(s2_nl_mns)

## [1] 948
```