# Central Limit Theorem Investigation

Omar Nooreddin

11/21/2018

## Overview

This paper will investigate the Central Limit Theorem (CLT) and demonstrate its properties. This will be done through a series of simulations and figures, along with explanatory text.

## Definition

The following is a definition of the CLT taken from Wikipedia:

"In probability theory, the central limit theorem establishes that, in some situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed."

We're going to investigate the above definition by:

1. Simulating a non-normal population (in our example it will be exponential)
2. Take 40 random samples and calculate their mean
3. Compare the samples' mean with the simulated population's mean
4. We will repeat the above for variance too

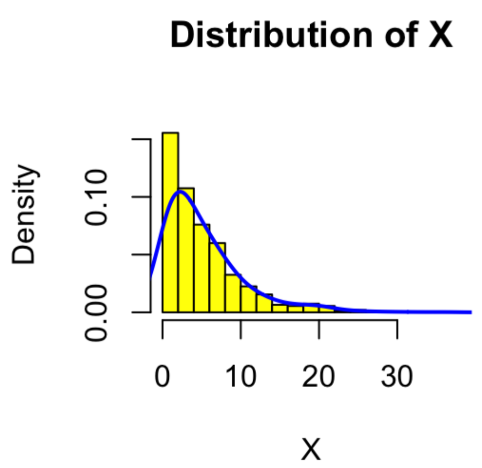
## Simulations

We are going to simulate an exponential population of a 1000 using the `rexp` function with **λ** or "rate" set to 0.2:

```
set.seed(123)
X<-rexp(1000,rate=0.2)
```

The above code has generated a random population of 1000 with an exponential distribution with a mean of 1/λ

```
hist(X,prob=T,ylim = c(0,0.17), breaks=20,col = "yellow",main="Distribution of X")
lines(density(X, adjust = 2),col="blue", lwd = 2)
```

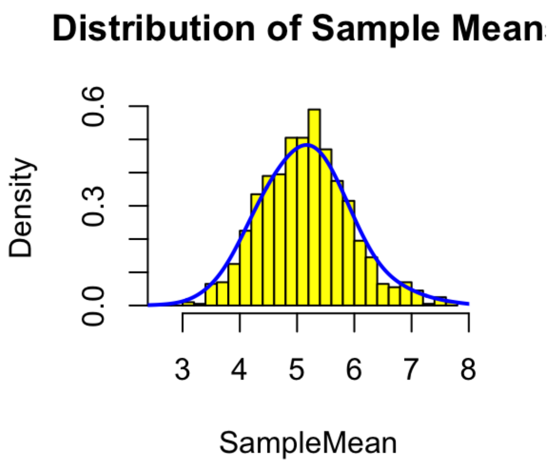



So it can be ascertained from figure above that the distribution is not a normal distribution, but rather an exponential one. This will be important later on to demonstrate even of the population is not normally distributed, the mean and variance of a random sample from this distribution will converge towards a normal distribution.

## Sample Mean

Following our population (X) creation, we are going to take a random sample of 40 from our exp population X and calculate its mean. We're going to repeat this 1000 times and demonstrate how the distribution of those means forms a normal distribution and their mean converges towards the population's mean (a consequence of CLT):

```
SampleMean=NULL
for (i in 1:1000) SampleMean<-c(SampleMean, mean(X[sample(1:length(X),40)]))
hist(SampleMean,prob=T, breaks=20,col = "yellow",main="Distribution of Sample Means")
lines(density(SampleMean,adjust=2), col = "blue", lwd=2)
```




Following our sampling process above, we can see that the sample mean has a normal distribution and a mean of

```
mean(SampleMean)
```

```
## [1] 5.152031
```

Which is almost equal to the population/theoretical mean of X:

```
mean(X)
```

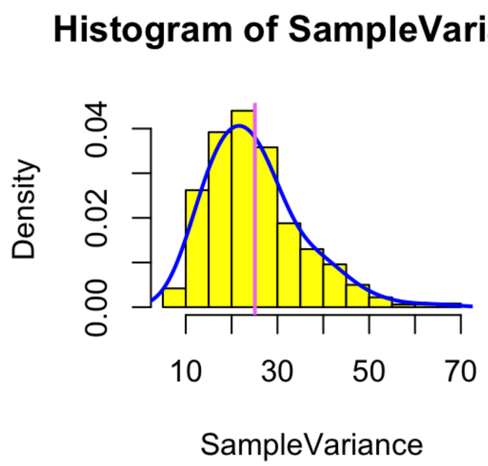
```
## [1] 5.149896
```

## Sample Variance

Similarly to the Sample Mean, the Sample Variance too will have a similar distribution to its population, and its mean will be almost equal to the population's variance. As per the previous section, we're going to demonstrate this by:

1. Taking a sample of 40 from our exponentially distributed population
2. Measure the variance of that sample of 40 and record it
3. Repeat the above 1000 times
4. Measure the mean of the these variances

```
SampleVariance=NULL
for (i in 1:1000) SampleVariance<-c(SampleVariance, var(X[sample(1:length(X),40)]))
hist(SampleVariance, prob=T, col = "yellow")
lines(density(SampleVariance, adjust=2), col="blue", lwd = 2)
abline(v=mean(SampleVariance),col="violet", lwd = 2)
```




As with Sample Mean we will compare the mean of the Sample Variances with the theoretical Variance of population X:

```
mean(SampleVariance)
```

```
## [1] 25.07877
```

With population X's variance:

```
var(X)
```

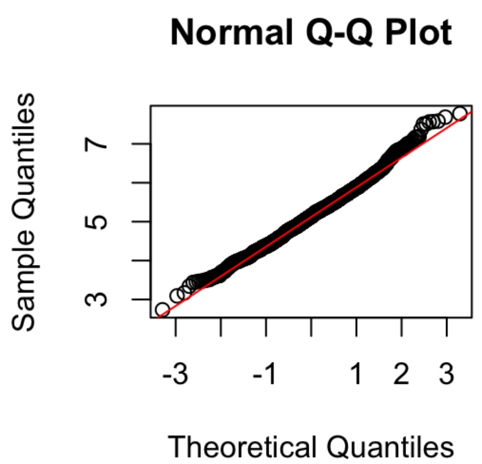
```
## [1] 25.2207
```

It can be noted that they are almost identical.

## Distribution

To check the normality of our Sample Mean data we can create Q-Q Plot which can show us whether our data is normally distributed or not:

```
qqnorm(SampleMean)
qqline(SampleMean,col="red")
```




Given our data falls on our Q-Q line this indicates that the distribution of our data is normal.