

# Statistics Assignment

Myself

4/17/2021

## Simulation Exercise for Central Limit Theorem

```
set.seed(100)
n <- 40
m <- 1000
lambda <- 0.2
sdom = NULL #sdom = sampling.distribution.of.means
for (i in 1 : m) sdom = c(sdom, mean(rexp(n,lambda)))

theoretical.mean <- 1/lambda
sample.mean <- mean(sdom)
diff.mean <- abs(sample.mean - theoretical.mean)
pc.diff.mean <- round(diff.mean*100/theoretical.mean,2)

population.variance <- (1/lambda)**2
theoretical.variance <- population.variance/(n)
sample.variance <- var(sdom)
diff.variance <- abs(sample.variance - theoretical.variance)
pc.diff.variance <- round(diff.variance*100/theoretical.variance,2)
```

## Analysis of Means

Theoretical Mean of Distribution = Population Mean =  $1/\text{Lambda} = 5$

Sample Mean = 4.9997019

Difference (Absolute) =  $2.9807313 \times 10^{-4}$

Percentage Difference (Absolute) = 0.01

## Analysis of Variance

Theoretical Variance of Distribution = (Population Variance)/n =  $(1/\text{Lambda}^2)/n = 0.625$

Sample Variance = 0.6432442

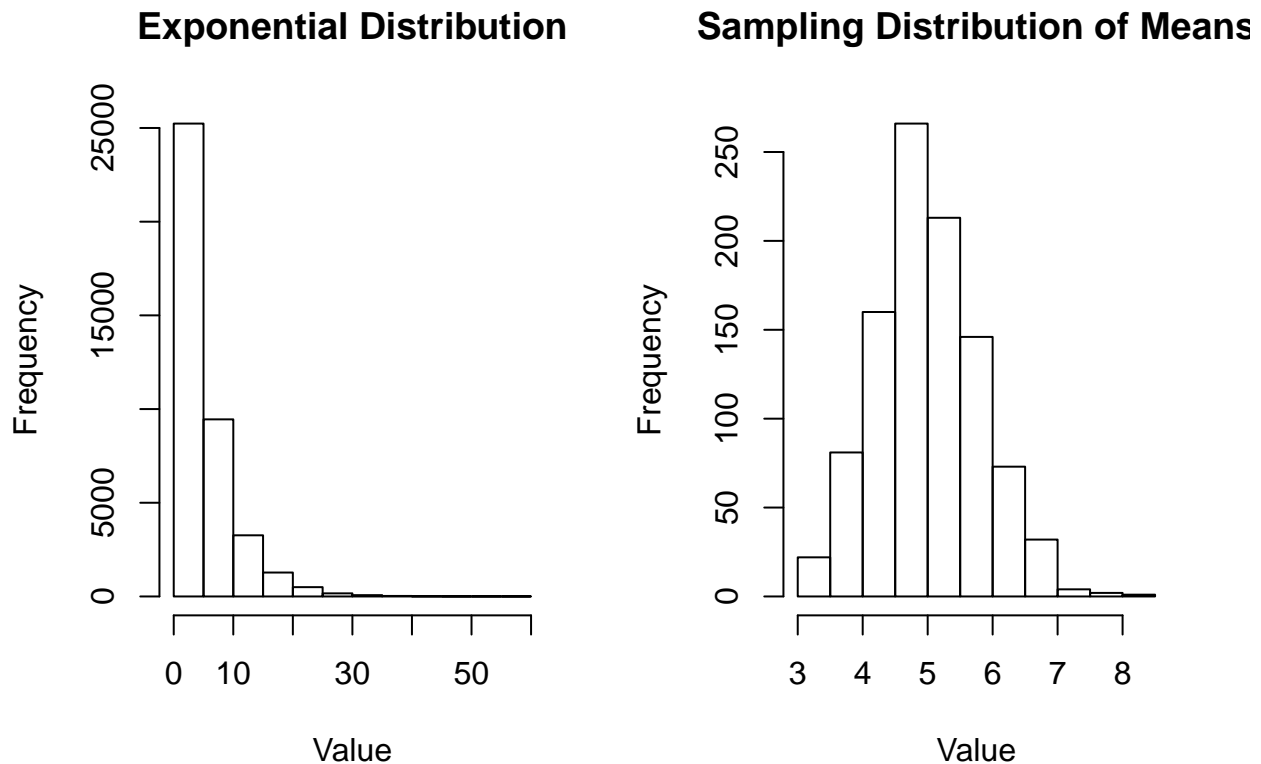
Difference (Absolute) = 0.0182442

Percentage Difference (Absolute) = 2.92

## Analysis of Distribution

Now that we have the sampling distribution of means, let us create a histogram of it and compare it with a distribution of equivalent exponential distribution of population.

```
par(mfrow = c(1, 2))
hist(rexp(m*n,lambda),main="Exponential Distribution",xlab="Value")
hist(sdom,main="Sampling Distribution of Means",xlab="Value")
```



Now to me this right side figure looks like a Normal Distribution. But without going into any elaborate normality tests (like Anderson Darling Test), let us first do some basic checks. What we know for a fact that if the sampling distribution of means is normally distributed, it should follow the 68-95-99 rule. Let us test that out first.

```
sample.sd <- sd(sdom)
sdom.sd.1 <- sdom[(sdom >= (sample.mean - sample.sd)) & (sdom <= (sample.mean + sample.sd))]
sdom.sd.2 <- sdom[(sdom >= (sample.mean - 2 * sample.sd)) & (sdom <= (sample.mean + 2 * sample.sd))]
sdom.sd.3 <- sdom[(sdom >= (sample.mean - 3 * sample.sd)) & (sdom <= (sample.mean + 3 * sample.sd))]

fraction.sd.1 <- round(length(sdom.sd.1)*100/length(sdom),2)
fraction.sd.2 <- round(length(sdom.sd.2)*100/length(sdom),2)
fraction.sd.3 <- round(length(sdom.sd.3)*100/length(sdom),2)
```

The percentages of observations in plus/minus 1, 2 and 3 sigma range are 68, 95.4 and 99.7

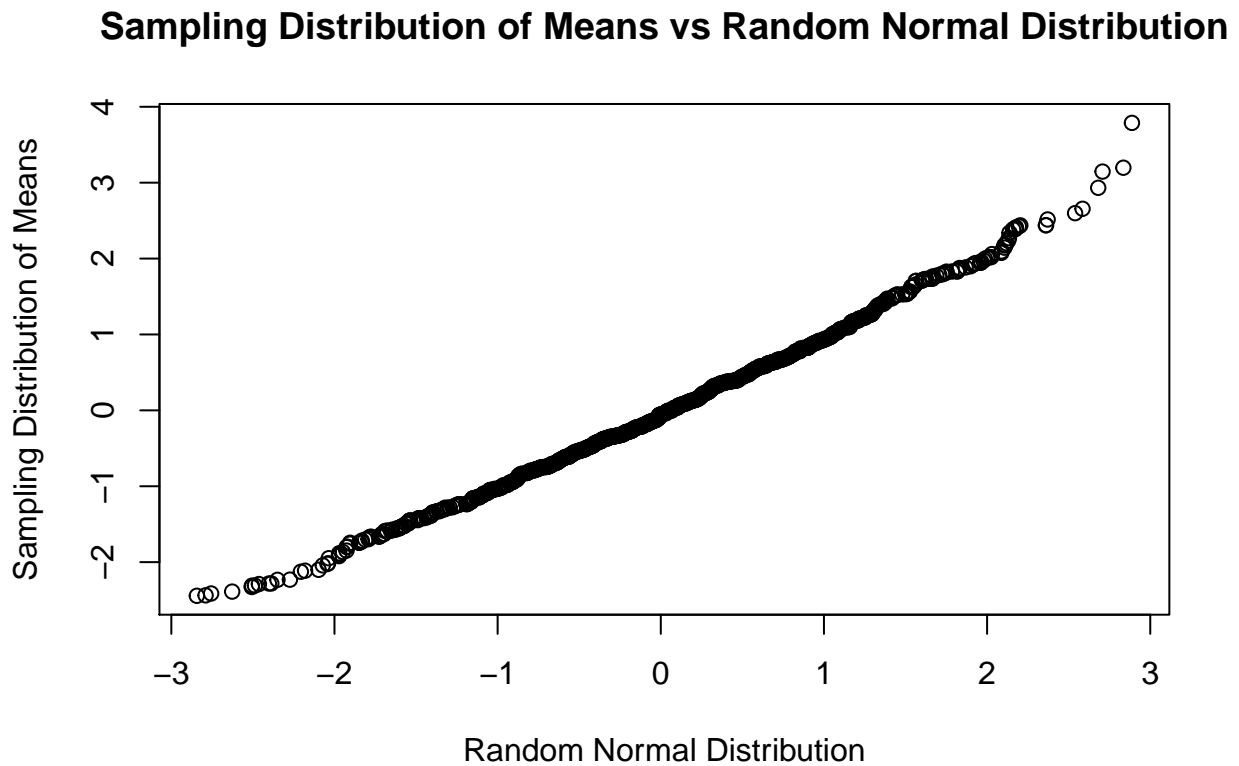
This seems to be very much in conformance to Normal Distribution. Let us do one more thing.

```
z.sdom <- (sdom - mean(sdom))/sd(sdom)
z.sdom <- sort(z.sdom)

x <- sort(rnorm(m))
```

What we have done here is scaled our distribution close to Standard Normal Distribution (we believe so) so that we can compare with one. For reference we take a standard normal distribution of 1000 points. Now we plot with the reference standard normal distribution in X axis and our distribution (scaled). If that is close to a straight line, then we can safely conclude that our Distribution is Normal.

```
plot(x,z.sdom,main="Sampling Distribution of Means vs Random Normal Distribution",xlab="Random Normal D
```



This graph is self evident. **We can safely conclude that the distribution is NORMAL**