

Central Limit Theory applied to Exponential Distribution

Rajesh Sinha

Contents

1	Overview	1
2	Simulations	1
3	Results	2
3.1	Sample Mean versus Theoretical Mean	2
3.2	Sample Variance versus Theoretical Variance	3
3.3	Is the sampling Distribution of Means Normal?	3
4	Conclusions	5

1 Overview

Central Limit Theorem (for our purposes) states that distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases.

Exponential distribution is the “continuous” counterpart of geometric distribution. It describes the time between events in a Poisson Process. So if λ is the rate parameter associated with Poisson distribution, exponential distribution function for exponential distribution is $P(x) = \lambda e^{-\lambda x}$. The expected value $E[X]$ or the mean of an exponential distributed random variable is $\frac{1}{\lambda}$ and its Variance $V[X] = \frac{1}{\lambda^2}$. So the standard deviation is same as mean.

In this simulation, we will generate sample size of 40 Random variables and take its mean and do this 1000 times. The sampling distribution of mean will be used in context of the central limit theory to show as to how well it compares in terms of mean, variance to the theoretical counterpart as well as we will show how well does this observed sampling distribution compared with expected normal distribution.

2 Simulations

Using a present rate parameter $\lambda = 0.2$, we will simulate 40 sized sample drawn 1000 times from a population which has an exponential distribution with the given λ .

```
set.seed(3)
lambda <- 0.2 #Set the lambda to 0.2 as required for this exercise
sample.size <- 40
no.of.simulations <- 1000
# Generate 40,000 exponential distributed numbers, distribute them into a matrix of size 1000 x 40
simulated <- matrix(rexp(no.of.simulations*sample.size, rate=lambda), no.of.simulations, sample.size)
row.Means = rowMeans(simulated) # averages of sample of 40 calculated for each of 1000 rows
```

The row.Means variable has now the averages of 40 sized sample for each of 1000 draws.

3 Results

3.1 Sample Mean versus Theoretical Mean

The Theoretical mean for the sampling distribution of averages is still centered at the population mean = $\frac{1}{\lambda}$.

```
mu <- 1/lambda; mu;
```

```
## [1] 5
```

```
observed.mean <- mean(row.Means); observed.mean;
```

```
## [1] 4.98662
```

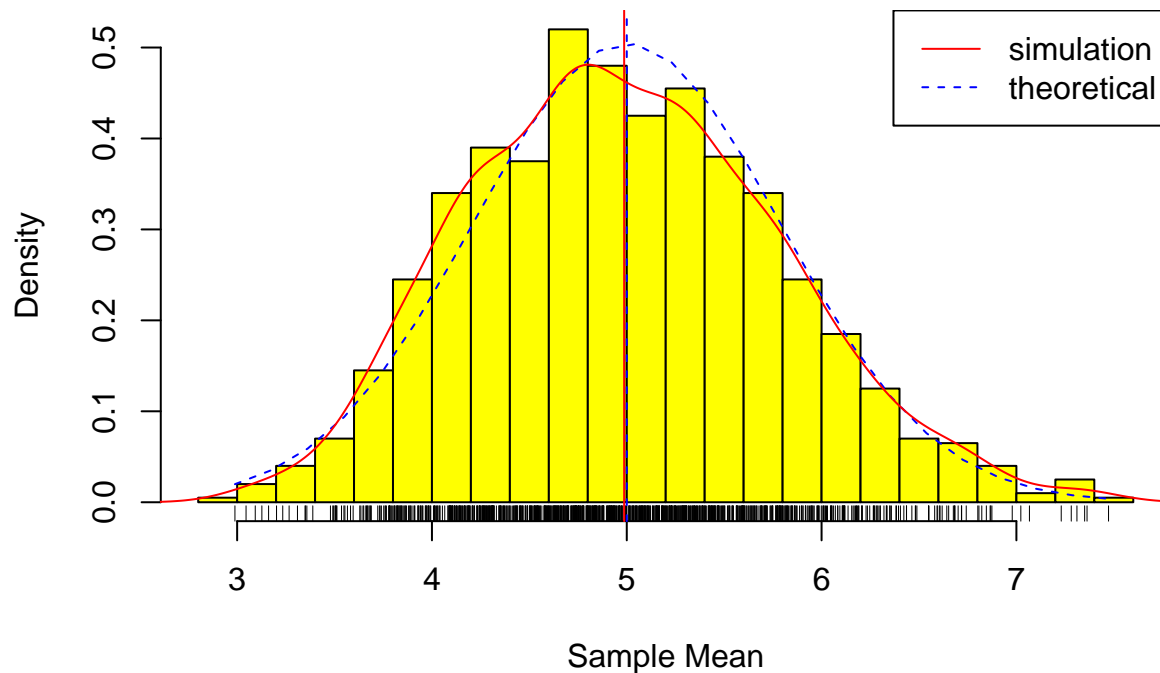
This can be seen below in the figure-1 below showing the histogram of distribution of averages of 1000 samples of size 40 drawn from population with $\lambda = 0.2$.

```
hist(row.Means, prob=T, col="yellow", xlab="Sample Mean",
      ylab="Density", main="Figure-1: Distribution of Averages of 1000 samples \nof size 40 drawn f
      breaks=25)
rug(row.Means)
#Plot the theoratical mean and the distribution
abline(v=mu, col="blue", lty=2)
x.Fit <- seq(min(row.Means), max(row.Means), length=25)
y.Fit <- dnorm(x.Fit, mean=mu, sd=(1/lambda/sqrt(sample.size)))
lines(x.Fit, y.Fit, col="blue", lty=2, pch=21)

#Plot the observed
abline(v=observed.mean, col="red", lty=1)
lines(density(row.Means), col="red", lty=1)

legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("red", "blue"))
```

Figure-1: Distribution of Averages of 1000 samples of size 40 drawn from exponentially distributed Population (Rate = 2)



3.2 Sample Variance versus Theoretical Variance

The theoretical variance of the sampling distribution of means of sample size n is given by $\frac{\sigma^2}{n}$. Here σ^2 is $\frac{1}{\lambda^2}$.

```
theoretical.variance=1/(lambda^2)/sample.size
theoretical.variance
```

```
## [1] 0.625
```

```
observed.variance=var(row.Means)
observed.variance
```

```
## [1] 0.6257575
```

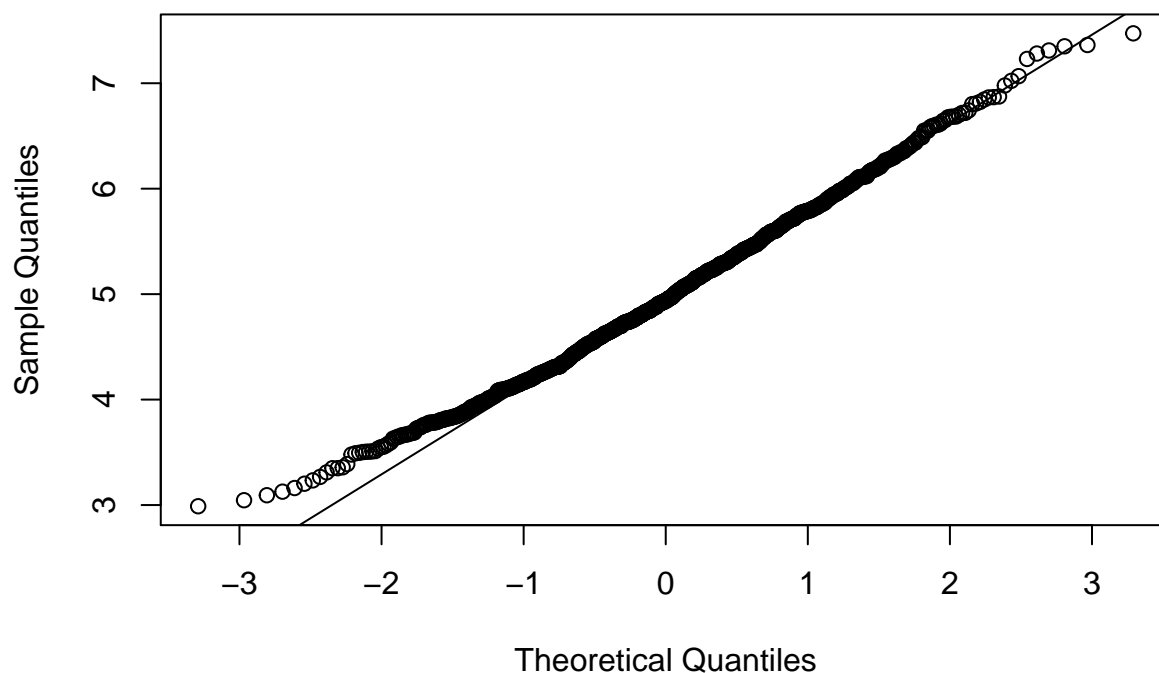
3.3 Is the sampling Distribution of Means Normal?

To show that the sampling distribution of the means is normal, we can have a look at the distribution of observed (Simulated) data in Figure-1 shown earlier. It looks somewhat normal

Apart from that, we can also demonstrate the same using normal probability plots which compare the observed versus theoretical quantiles.

```
qqnorm(row.Means, main="Normal Q-Q Plot of Sample Averages"); qqline(row.Means);
```

Normal Q-Q Plot of Sample Averages



Again the output seems to suggest some ambiguity in lower and upper quartiles.

We also confirm the normality using Shapiro-Wilk Normality Test

```
shapiro.test(row.Means)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  row.Means  
## W = 0.9943, p-value = 0.0007348
```

The shapiro-Wilk normality test assume a null hypothesis of data coming from normal and the low p-value seems to negate the Null Hypothesis that data is normal.

We can run some quick tests to see if increasing the sample size will make shapiro-wilk report a statistically significant p-value which can allow us to conclude that sampling distribution of means is coming from normally distributed data.

Following are the numbers reported for some larger sample size.

```
set.seed(3)  
sample.sizes = c(40,50,60,70,80,90,100)  
for (i in sample.sizes) {  
  simulated <- matrix(rexp(no.of.simulations*i, rate=lambda), no.of.simulations, i)  
  row.Means <- rowMeans(simulated)  
  t<- shapiro.test(row.Means)  
  cat("Sample Size = ", i, "p-Value for Normality Shapiro Wilks= ", t$p.value, "\n" )  
}
```

```
## Sample Size = 40 p-Value for Normality Shapiro Wilks= 0.0007348222
## Sample Size = 50 p-Value for Normality Shapiro Wilks= 0.000322532
## Sample Size = 60 p-Value for Normality Shapiro Wilks= 0.06177817
## Sample Size = 70 p-Value for Normality Shapiro Wilks= 0.0001424101
## Sample Size = 80 p-Value for Normality Shapiro Wilks= 0.0004966392
## Sample Size = 90 p-Value for Normality Shapiro Wilks= 0.6924742
## Sample Size = 100 p-Value for Normality Shapiro Wilks= 0.213554
```

4 Conclusions

The means of samples of size 40 extracted from an exponential distribution over 1000 such simulated draws showed

- The observed mean of averages is close to the theoretical mean (which is the mean of original population)
- The observed variance of averages is close to the theoretical variance of the distribution.
- However we were not able to conclusively prove that the distribution of sampling means is normal. It requires a sample size higher than 40.