Asymptotics for Exponential Distribution

by Pravin Kumar Choudhary

Overview

Objective of this project is to demonstrate Asymptotics for Exponential Distribution.

- Law of Large Numbers: if random variables are iid from a population with mean and variance then sample mean/variance converges in population mean / variance
- Central Limit Theorem: distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases

Simulations

For this project, we create random Exponentially Distributed sample data of size 40. To test the Asymptotic behaviour, we have consider a large number of simulation (1000).

Exponential Distribution with lambda / Rate = 0.2 has been considered for the data sample data generation and analysis. i.e. Mean and Std. Deviation as 5.

ggplot2 library has been used for plotting adn getting inference.

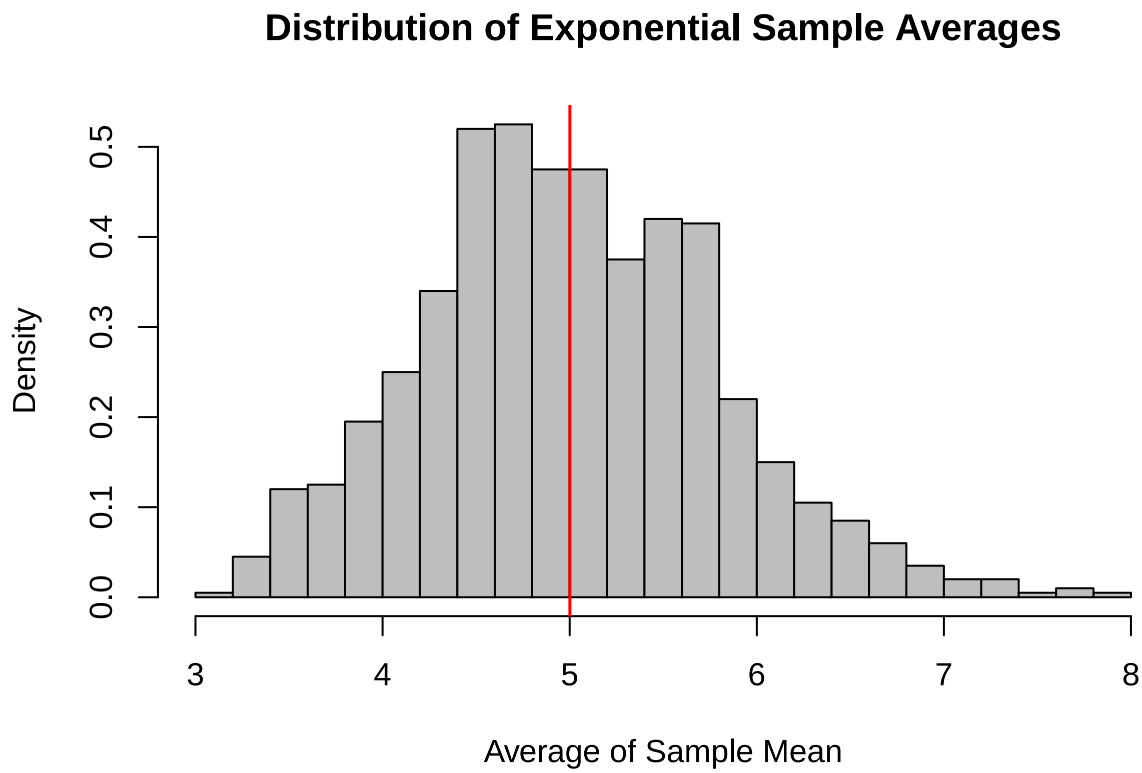
```r
# Import libraries
library(ggplot2)

# population mean, std deviation initialization
poplMean <- 1/0.2
poplSD <- 1/0.2
n <- 1000
set.seed(117)

# Create a sample data of 1000 simulations of exponential distributed data size of 40 each
# Convert the date in matrix format of 1000 records each with 40 size
sampleExpData <- matrix(rexp(n*40, rate = 0.2), ncol = 40)

# Get the average / mean of each record / rows
sampleExpMean <- apply(sampleExpData, 1, mean)
```

on drawing the histogram of the averages of the sample mean for the simulation size, we can see that sample mean converges to the population mean



Sample Mean versus Theoretical Mean

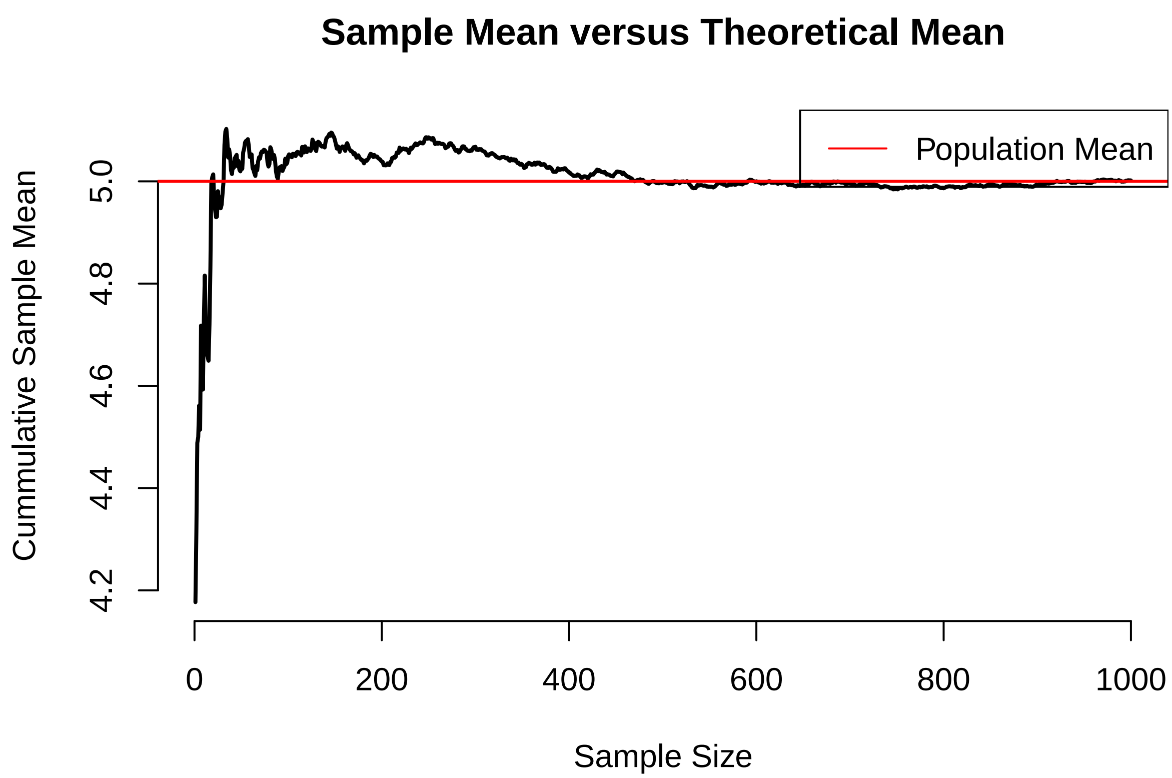
In the plot below, we can see that as the number of simulation go large, sample mean converges to population mean, which is 5 (1/lambda)

```r
set.seed(117)

# Create a sample data of 1000 simulations of exponential distributed data size of 40 each
# Convert the date in matrix format of 1000 records each with 40 size
sampleExpData <- matrix(rexp(n*40, rate = 0.2), ncol = 40)

# Get the average / mean of each record / rows
sampleExpMean <- apply(sampleExpData, 1, mean)

cumSampleMean <- cumsum(sampleExpMean) / (1:n)
plot(1:n, cumSampleMean, frame = FALSE, type = "l", lwd = 2,
     main = "Sample Mean versus Theoretical Mean",
     ylab = "Cummulative Sample Mean", xlab = "Sample Size")
abline(h = poplMean, col = "red", lwd = 1.5)
legend("topright", legend = c("Population Mean"), col = c("red"), lty = 1)
```

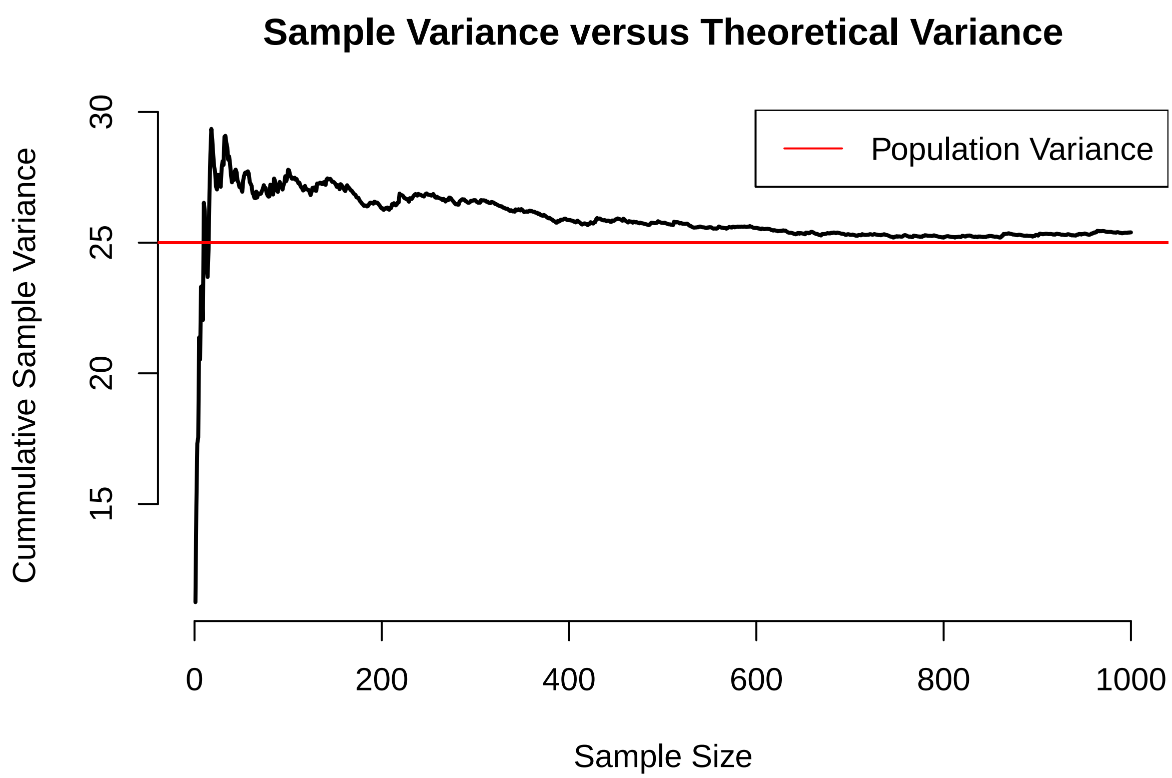


Sample Variance versus Theoretical Variance

In the plot below, we can see that as the number of simulation go large, sample variance converges to population variance, which is 25 (sqr(1/lambda))

```r
set.seed(117)
sampleExpData <- matrix(rexp(n*40, rate = 0.2), ncol = 40)
sampleExpVar <- apply(sampleExpData, 1, var)
cumSampleExpVar <- cumsum(sampleExpVar) / (1:n)

plot(1:n, cumSampleExpVar, frame = FALSE, type = "l", lwd = 2,
     main = "Sample Variance versus Theoretical Variance",
     ylab = "Cummulative Sample Variance", xlab = "Sample Size")
abline(h = poplSD^2, col = "red", lwd = 1.5)
legend("topright", legend = c("Population Variance"), col = c("red"), lty = 1)
```



Distribution

Using Central Limit Theorem, we will demonstrate here that, distribution of the average for a exponentially distribution sample data for a large number of simulation is approximately normal / Gaussian.

To demonstrate this, we will negate population mean from the sample data and divide it std error and plot the histogram with normal curve.

We can see that distribution is centered at 0 and with std deviation as 1, due to the data manimulation above, but we can also see that distribution has a proper bell shape which infers that distribution is very close to Normal one.

```
set.seed(117)
sampleExpData <- matrix(rexp(n*40, rate = 0.2), ncol = 40)
sampleExpMean <- apply(sampleExpData, 1, mean)

# Now subtract the population mean from sample mean and divide by std error of estimate
centralizedSampleMean <- (sampleExpMean - poplMean) / (poplSD/sqrt(40))

# Convert matrix to data frame
#sampleExpMeanDF <- as.data.frame(sampleExpMean)

# plot the histogram with line fitting to demonstrate the distribution of the average of 1000
# simulationss of exponential distributed sample data,each of size 40
hist(centralizedSampleMean, freq = FALSE, breaks = 20, xlim = c(-3,3), ylim = c(0.0,0.5),
     main = "Distribution of Average of iid Exponential Variables",
```

```r
      xlab = "Average of Sample Mean", ylab = "Density", col = "gray")
curve(dnorm(x, mean=mean(centralizedSampleMean), sd=sd(centralizedSampleMean)),
       add = TRUE, col = "red", lwd = 2)
```