The Exponential Distribution with the Central Limit Theorem*

Aliakbar Safilian†

December 12, 2018

1 Overview

We show that the distribution of averages of exponentially distributed[1] variables becomes that of a standard normal as the sample size increases (according to the Central Limit Theorem (CLT)[2]).

2 Simulations

Let us first fix some parameters, including the sample size, the number of simulations, the rate of our exponential distribution.

```
n <- 40    # sample size
lambda <- 0.2 # the rate of the exponential distribution
t <- 1000 # number of simulations
mu <- 1/lambda # the mean of the distribution
sigma <- 1/lambda # the standard deviation of the distribution
```

The following script simulates 1000 (**t**) sample means with size 40 (**n**). The result is stored into a variable named **mns**.

```
mns <- NULL
for(i in 1:t) mns = c(mns, mean(rexp(n, lambda)))
```

3 Sample Mean versus Theoretical Mean

The sample mean, i.e., **mean(mns)** $\approx$ **4.98**, and the theoretical mean, i.e., **mu** $=$ **5**, are almost equal. The following plot shows that as number of simulations increases, the sample mean (in *blue*) converges to the theoretical mean (in *red*).

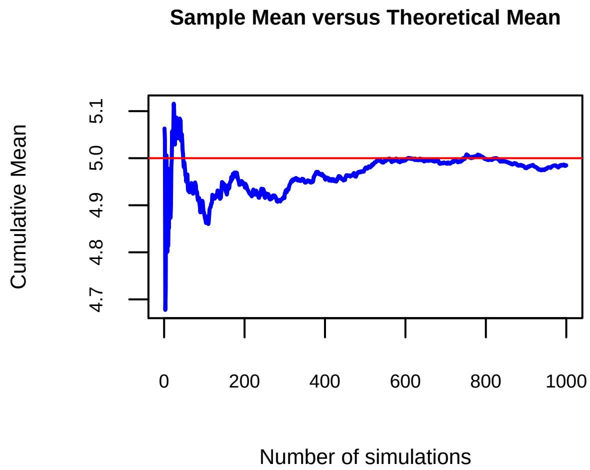
```
cummns <- cumsum(mns)/(1:t)
plot(cummns, type="l", lwd=2, col = "blue",
     cex=1, cex.lab = .7, cex.axis = .6, cex.main = .7, cex.sub=.5,
     main = "Sample Mean versus Theoretical Mean",
     xlab = "Number of simulations", ylab = "Cumulative Mean")
abline(h=sigma, col="red", lwd=1)
```

*This analysis report is the final project-Part1 of the Statistical Inference (Coursera) course at Johns Hopkins University

†Email: a.a.safilian@gmail.com

[1] https://en.wikipedia.org/wiki/Exponential_distribution

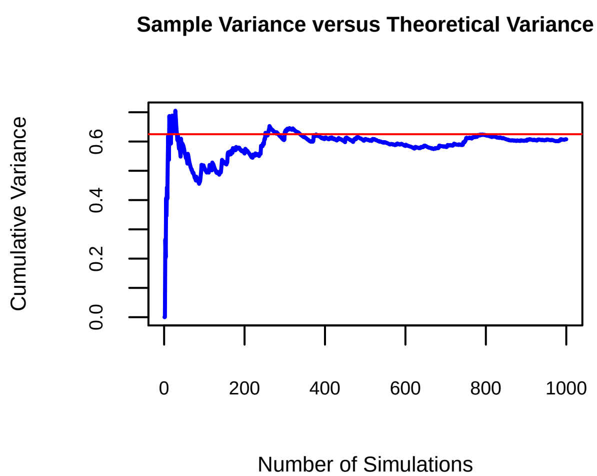
[2] https://en.wikipedia.org/wiki/Central_limit_theorem



4 Sample Variance versus Theoretical Variance

The sample variance, i.e, **var(mns)** $\approx$ **0.61**, and the theoretical variance, **sigma^2/n = 0.625**, are almost the same. The following plot shows that as number of simulations increases, the sample variance converges to the theoretical variance:

```
vars <- cumsum(mns^2)/(1:t)-cummns^2
plot(vars, type="l", lwd=2, col = "blue",
     cex=1, cex.lab = .7, cex.axis = .6, cex.main = .7, cex.sub=.5,
     main = "Sample Variance versus Theoretical Variance",
     xlab = "Number of Simulations", ylab = "Cumulative Variance")
abline(h=sigma^2/n, col="red", lwd=1)
```



5 Distribution

Let us denote the distribution of sample means by $\bar{X}_n$, where n denotes the sample size. According to the CLT, $\bar{X}_n \sim N(\mu, \sigma^2/n)$, where μ and σ are the mean and standard deviation of the distribution, respectively. We are going to show that this holds in our case, where $\mu = 5$, $\sigma = 5$, and $n = 40$.

The following script visualizes the density of our sample mean (denoted by a *blue* line), and the normal distibution $N(\mu, \sigma^2/n) = N(5, 0.625)$ (denoted by a *red* line). The densities are shown over a histogram of our simulation, **mns**. The blue and red straight lines indicate where the mean of the distributions are, respectively (they almost overlap). As we see in the figure, our simulation makes a good approxmiation of the corresponding normal distribution.

```
hist(mns, density = 20, breaks = 20, prob = TRUE,
     cex.lab = .8, cex.axis = .8, cex = 1, cex.main = .7,
     main = "The Distribution of Averages of 40 Exponentials", xlab = "means")
lines(density(mns), lwd = 2, col = "blue")
curve(dnorm(x, mean = mu, sd = sigma/sqrt(n)), col = "red", lwd = 3, add = TRUE)
abline(v = mean(mns), col = "blue")
abline(v = mu, col = "red")
```