

Simulation of Exponential Distribution

Yanfei Wu

June 15, 2016

Overview

This project investigates the exponential distribution in R and compares it with the Central Limit Theorem. Specifically, the sample mean and the sample variance are shown by simulations and are compared to the theoretical values of the distribution. Also, the distribution is shown to be approximately normal.

Simulations

In probability theory and statistics, the exponential distribution is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate.

The exponential distribution can be simulated in R with `rexp(n , λ)` in the “stats” package. n is number of observations and λ is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. For all the simulations in this project, $\lambda = 0.2$. The distribution of averages of 40 exponentials is investigated with 1,000 simulations.

```
require(stats)
nosim <- 1000
n <- 40
lambda <- 0.2
set.seed(3250)
mn_expdist <- NULL
for (i in 1 : nosim) mn_expdist[i] <- mean(rexp(n, lambda))
## Simulation of 1000 collections of averages of 40 exponentials
radm_expdist <- rexp(n*nosim, lambda)
## simulation of 40,000 collections of random exponentials
```

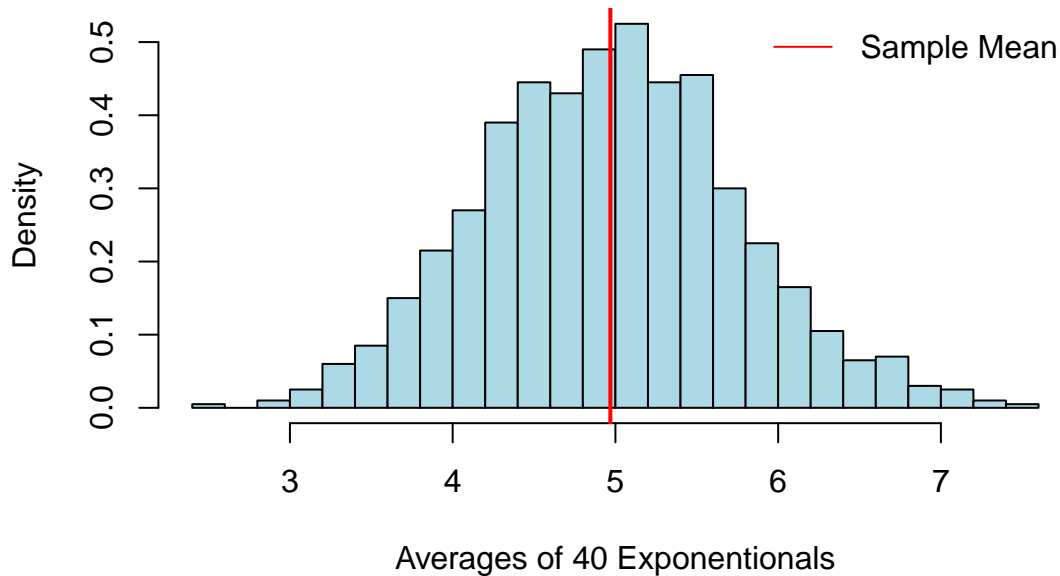
Results

Sample Mean versus Theoretical Mean

First, let's compare the sample mean to the theoretical mean of the distribution.

```
mean_sample <- round(mean(mn_expdist), 3)
mean_theory <- 1/lambda
hist(mn_expdist, breaks = 30, probability = T,
     col = "lightblue",
     main = "Distribution of Averages of 40 Exponentials",
     xlab = "Averages of 40 Exponentials")
abline(v = mean_sample, col = "red", lwd = 2)
legend("topright", lty = 1, col = "red", legend = "Sample Mean", bty = "n")
```

Distribution of Averages of 40 Exponentials



The distribution of 1000 collections of averages of 40 exponentials is shown above. The sample mean is 4.969, as indicated by the red vertical line. The theoretical mean of exponential distribution is $1/\lambda$, and is 5 in this case ($\lambda = 0.2$).

Clearly, the sample mean of 1000 collections of averages of 40 exponentials is very close to the theoretical mean. This is consistent with the law of large numbers (LLN).

Sample Variation versus Theoretical Variation

Next, let's look at the sample variation versus the theoretical variation. The variance of the sample is calculated as:

```
var_sample <- round(var(mn_expdist),3)
```

And the theoretical variance is calculated as the square of standard error of the estimated mean (σ^2/n):

```
var_theory <- round((1/lambda)^2/n,3)
```

The sample variance is calculated to be 0.619, and the theoretical variance is 0.625. Again, they are very close to each other, consistent with LLN.

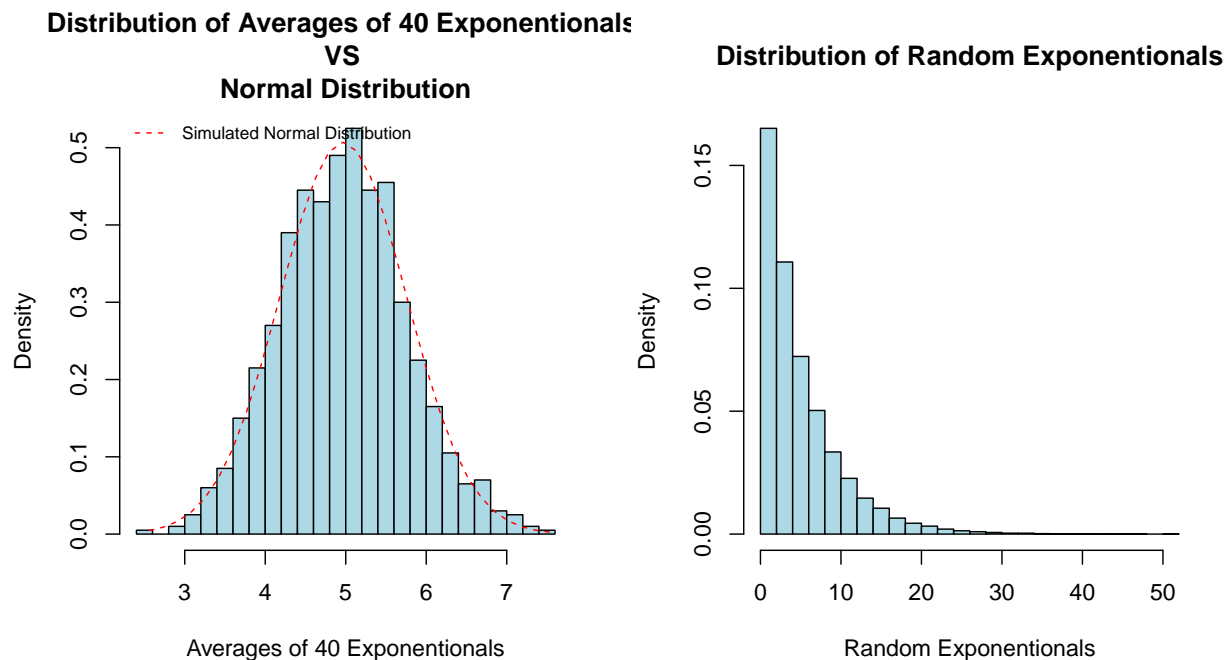
Distribution

Finally, let's compare the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials, and examine if they are approximately normal.

```

par(mfrow = c(1, 2))
hist(mn_expdist, breaks = 30, probability = T,
     col = "lightblue",
     main = "Distribution of Averages of 40 Exponentials \nVS \nNormal Distribution",
     xlab = "Averages of 40 Exponentials")
x_n <- seq(min(mn_expdist), max(mn_expdist), length = 30)
y_n <- dnorm(x_n, mean = mean(mean_sample), sd = sqrt(var_sample))
lines(x_n, y_n, lty = 2, col = "red")
legend("topleft", lty = 2, col = "red",
      legend = "Simulated Normal Distribution", bty = "n", cex = 0.75)
hist(radn_expdist, breaks = 30, probability = T,
     col = "lightblue",
     main = "Distribution of Random Exponentials",
     xlab = "Random Exponentials")

```



According to the Central Limit Theorem, the distribution of averages of iid variables becomes that of a standard normal if the sample size increases. This is exactly the case here. The dashed line in the top figure represents a normal distribution simulated with the sample mean and sample standard deviation. Clearly, it matches well with the actual distribution of 1000 collections of *averages of 40 exponentials*. However, the distribution of 40,000 *random exponentials* in the bottom figure does not look normal.