Exponential Distribution and the Central Limit Theorem: A Simulation

Ajay Ghanti

11/6/2016

Overview

The exponential distribution, a.k.a. negative exponential distribution is a 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 (called lambda, the event rate). The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

The Central Limit Theorem (CLT) states that the distribution of averages of IID variables becomes that of a standard normal as the sample size increases. In other words, the sampling distribution of the sample mean is approximately normal, when the sample size is sufficiently large.

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem.

1. We will show the sample mean and compare it to the theoretical mean of the distribution
2. We will show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution
3. We will show that the distribution is approximately normal

Simulations

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

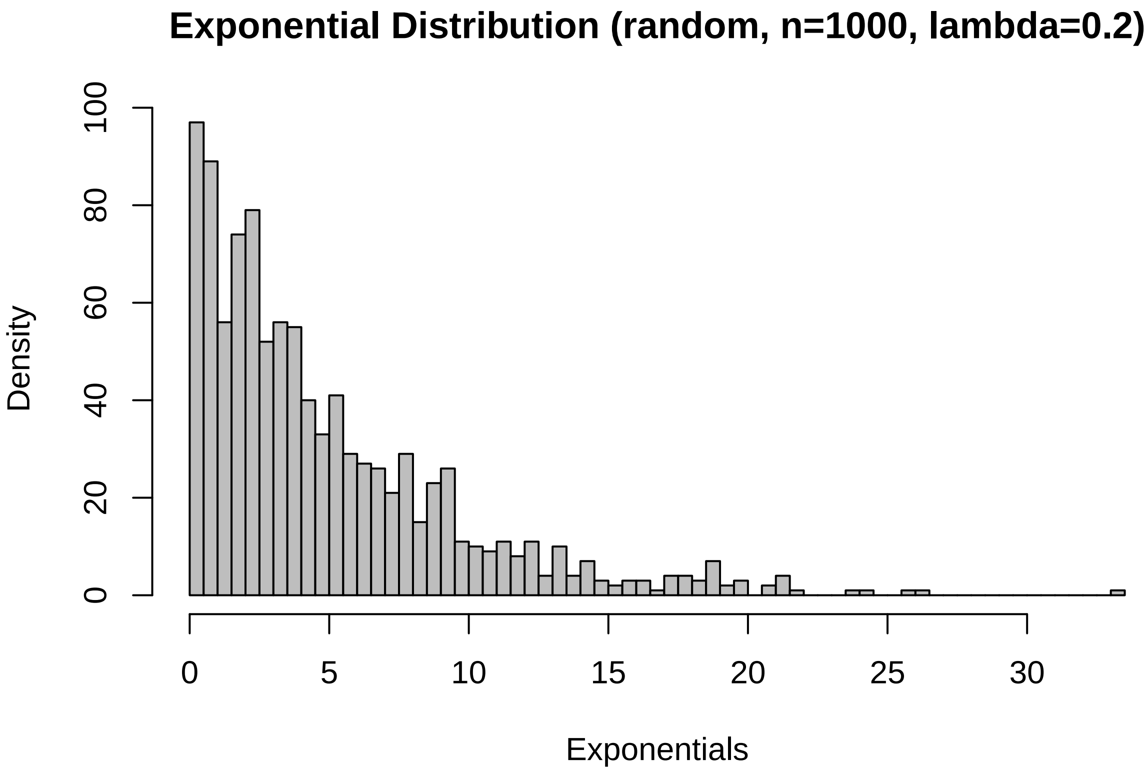
For our simulations, we will set lambda = 0.2. First we will consider the distribution of 1000 random exponentials.

```r
# Set a seed for reproducibility of data
set.seed(8888)

# Rate parameter
lambda <- 0.2

# Draw 1000 random exponentials
exp <- rexp(1000, rate = lambda)

# Plot the data
hist(exp, breaks=50, col="gray", xlab="Exponentials", ylab="Density",
     main="Exponential Distribution (random, n=1000, lambda=0.2)")
```



From the above plot, you can see that the graph of an exponential distribution starts on the y-axis at a positive value and decreases to the right.

Next, let's generate the distribution of averages of 40 random exponentials, over 1000 simulations.

```
avgs = NULL
for (i in 1 : 1000) {
    avgs = c(avgs, mean(rexp(40, rate = lambda)))
}
```

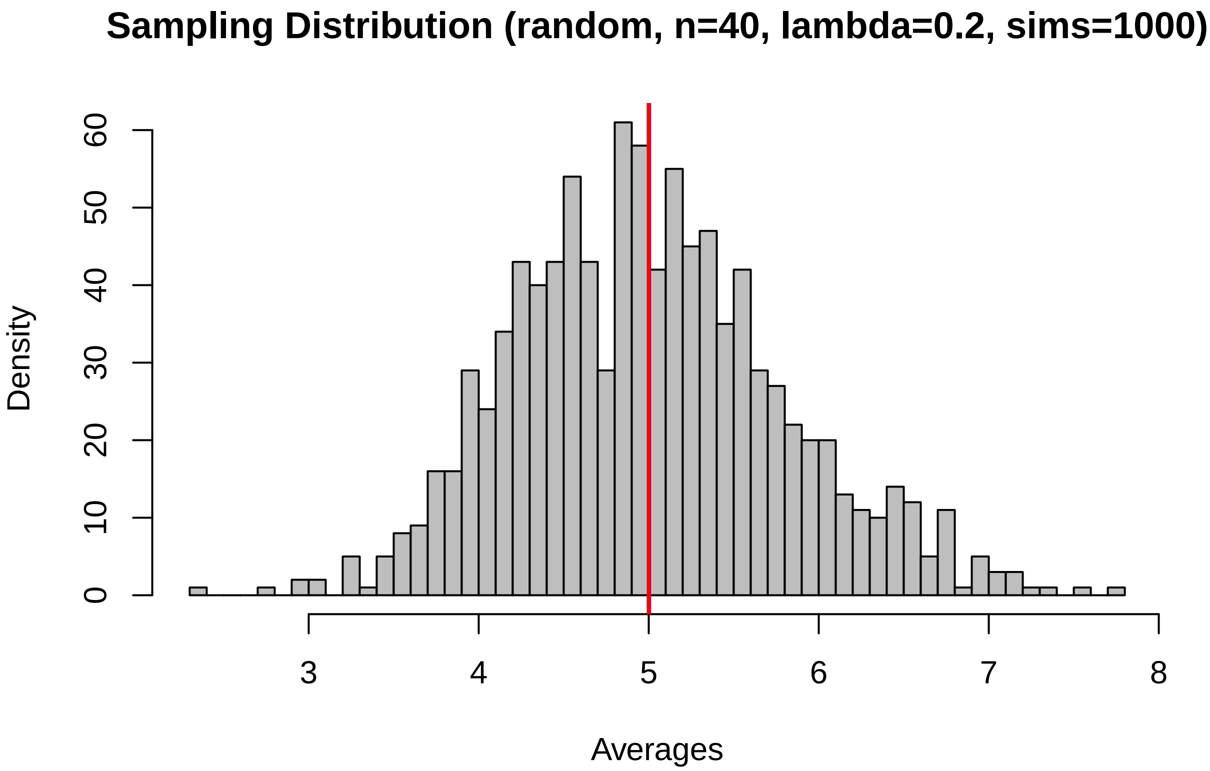
Sample Mean versus Theoretical Mean

The mean for an exponential distribution is 1/lambda. Let us compute the theoretical and sample means.

```
sampleMean <- round(mean(avgs), 3)
theoreticalMean <- 1/lambda
```

Type	Mean
Sample	5.002
Theoretical	5

```
# Plot the sampling distribution, sample and theoretical means
hist(avgs, freq=T, breaks=50, col="gray", xlab="Averages", ylab="Density",
     main="Sampling Distribution (random, n=40, lambda=0.2, sims=1000)")
abline(v=sampleMean, col='blue', lwd=2)
abline(v=theoreticalMean, col='red', lwd=2)
```



As you can see, the sample mean is very close to the theoretical mean. You cannot see both the lines on the graph because the values are very close

Sample Variance versus Theoretical Variance

The variance for an exponential distribution is (1/lambda)^2/n. Let us compute the theoretical and sample variances.

```
sampleVar <- round(var(avgs), 3)
theoreticalVar <- (1/lambda)^2/40
```

Type	Variance
Sample	0.632
Theoretical	0.625

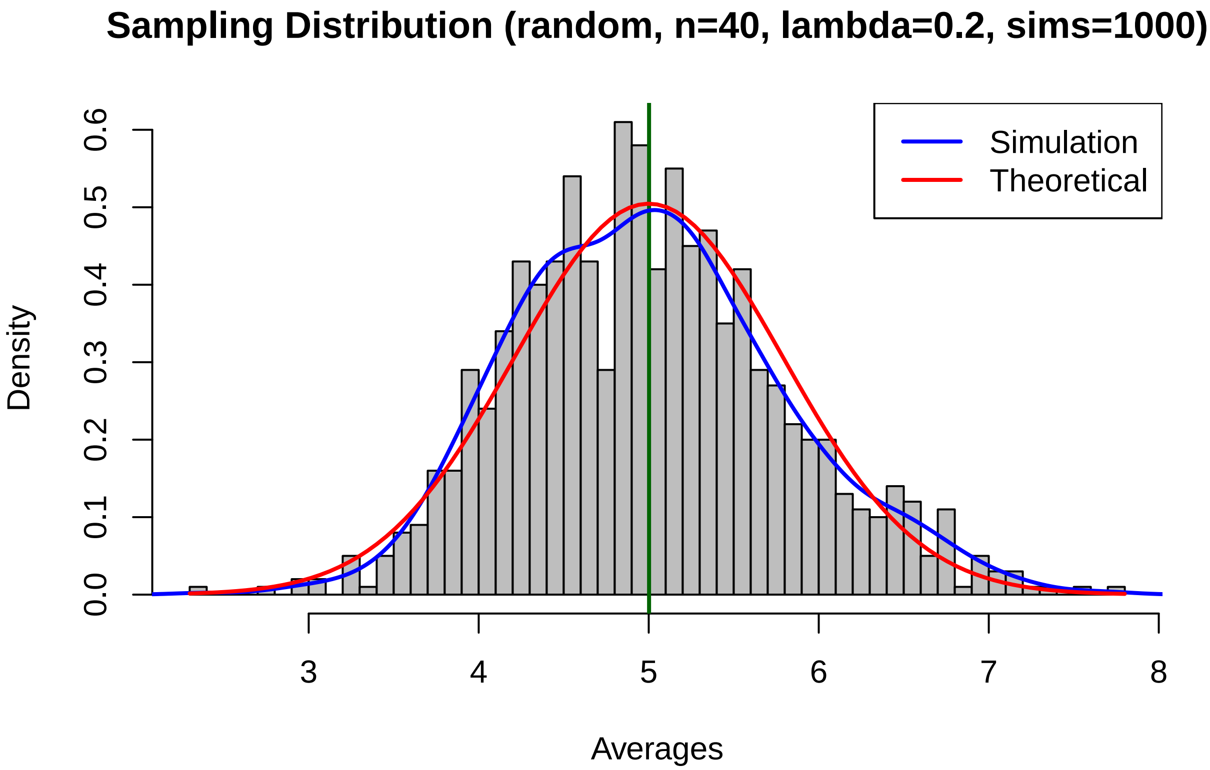
As you can see, the sample variance is very close to the theoretical variance.

Distribution

From the first plot in this report, it is seen that the graph of an exponential distribution starts on the y-axis at a positive value and decreases to the right. However, when we generate the distribution of averages of 40 random exponentials, over 1000 simulations, the graph looks different.

```
# Plot the data
hist(avgs, prob=T, breaks=50, col="gray", xlab="Averages", ylab="Density",
     main="Sampling Distribution (random, n=40, lambda=0.2, sims=1000)")
abline(v=sampleMean, col='dark green', lwd=2)
```

```r
# Curve for sampling distribution
lines(density(avgs), col="blue", lwd=2)
# Curve for normal distribution with mean & sd 1/lambda
curve(dnorm(x, mean=theoreticalMean, sd=sqrt(theoreticalVar)), add=TRUE, col="red", lwd=2)
# Legend
legend('topright', c("Simulation", "Theoretical"),
       col=c("blue", "red"), lty=c(1,1), lwd=c(2,2))
```



In the above graph, we have plotted the sampling distribution and the curve for the distribution (in blue). For comparison, we have also plotted the curve for a normal distribution (in red). You can see that the blue curve (sampling distribution) is very close to the red curve (normal distribution), and hence is **approximately normal**. This proves that the distribution of means of exponentials behave as predicted by the Central Limit Theorem.