Exponential Distribution and Central Limit Theorem

Taedong Yun

Overview

In this article, we will investigate the exponential distributions and compare it with the Central Limit Theorem. We will analyze statistics(mean and variance) of a random sample from a distribution of averages of a set of exponentials and compare it with the Gaussian distribution, in an attempt to visually understand an application of the Central Limit Theorem.

Simulations

First we generate random samples we will use. The exponential distribution is a continuous probably distribution that depends on a single parameter λ. We are interested in the distribution of *averages* of 40 exponentials, so we will generate 1000 samples from that distribution. The following R code snippet generates the desired samples. Note that we set $\lambda = 0.2$ throughout this article.

```
set.seed(1)
samples = numeric()
for (i in 1 : 1000) samples = c(samples, mean(rexp(40, 0.2)))
```

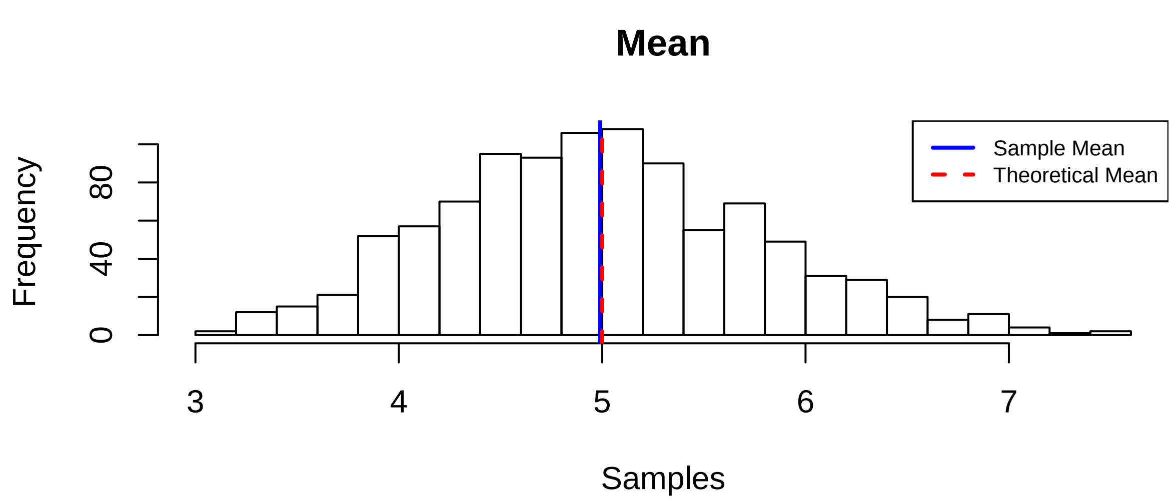
Sample Mean versus Theoretical Mean

The sample mean of the distribution is:

```
mean(samples)
```

```
## [1] 4.990025
```

Note that the theoretical mean of an exponential distribution is $1/\lambda = 5$ so the theoretical mean of the *averages* of 40 exponentials is also 5. The sample mean 4.99 is very close to the theoretical mean 5 as we can see in the following figure. (See **Code2** in Appendix for R code.)



Sample Variance versus Theoretical Variance

Let us first compute the theoretical variance of the distribution. The variance of a single exponential distribution is $\sigma^2 = 1/\lambda^2$, so by the Central Limit Theorem, the variance of the mean of 40 exponentials can be approximated by $\sigma^2/40 = 1/(40\lambda^2)$. By setting $\lambda = 0.2$, the theoretical variance is:

```
1 / (40 * 0.2^2)
```

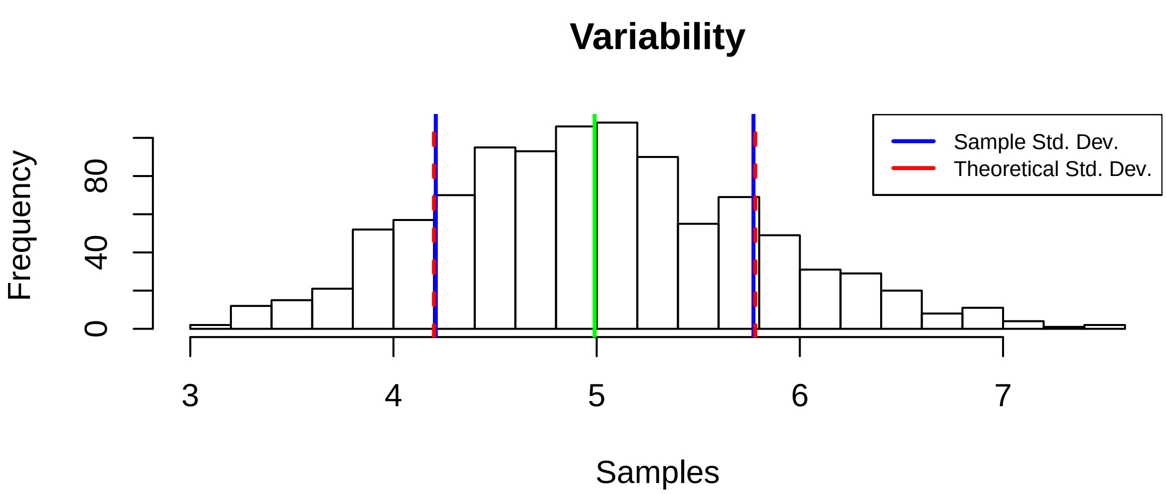
```
## [1] 0.625
```

On the other hand, the sample variance is:

```
var(samples)
```

```
## [1] 0.6111165
```

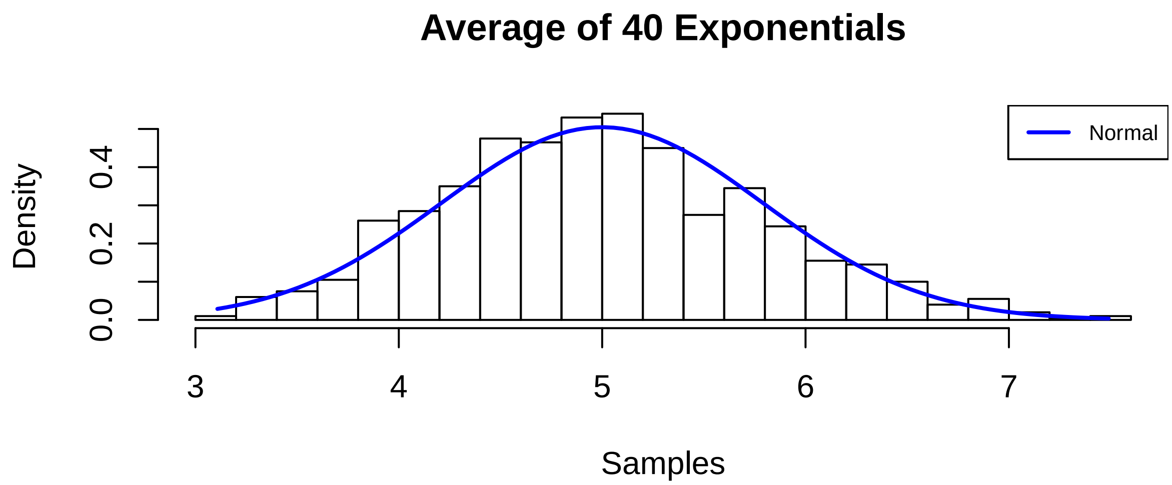
The following figure shows the variability of the samples compared to the theoretical value. Note that the **standard deviations**, which are the square roots of variances, are compared instead of variances, since the standard deviation has the same unit as the mean. The blue and red lines indicates the one-standard-deviation intervals centered at the sample mean. The green line represents the sample mean. (See **Code3** in Appendix for R code.)



Again, we have verified that the sample variance (equivalently, standard deviation) is very close to the theoretical variance (equivalently, standard deviation).

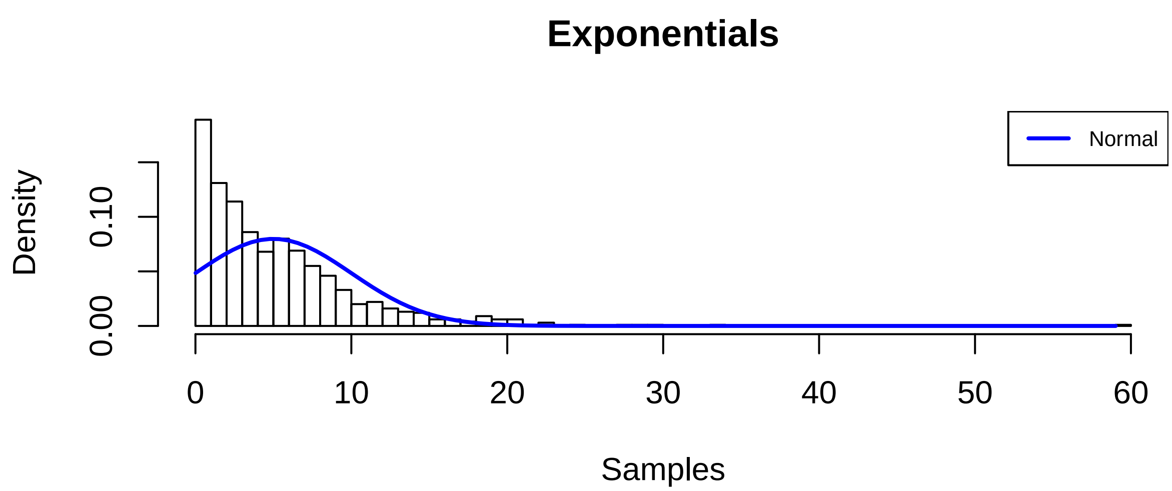
Distribution

The distribution of the averages of 40 exponentials can be closely approximated by the normal distribution with the same mean and variance. The following figure shows how the histogram of samples we had above is close to a normal distribution. (See **Code4** in Appendix for R code.)



The blue Gaussian curve closely matches the histogram of our samples. This is a consequence of the **Central Limit Theorem**.

On the other hand, the distribution of a large collection of random exponentials *cannot* be approximated by a normal distribution. Here we plot a histogram of 1000 random exponentials (with $\lambda = 0.2$), compared with a normal distribution with the same mean and variance. (See **Code5** in Appendix for R code.)



Appendix

Code1.

```
set.seed(1)
samples = numeric()
for (i in 1 : 1000) samples = c(samples, mean(rexp(40, 0.2)))
```

Code2.

```
hist(samples, main="Mean", xlab="Samples", ylab="Frequency", breaks = 20)
abline(v = mean(samples), col = "blue", lty = 1, lwd = 2)
abline(v = 5, col = "red", lty = 2, lwd = 2)
legend("topright", legend=c("Sample Mean", "Theoretical Mean"),
       col = c("blue", "red"), lty = c(1, 2), lwd = 2, cex = 0.7)
```

Code3.

```
hist(samples, main="Variability", xlab="Samples", ylab="Frequency", breaks = 20)
abline(v = mean(samples), col = "green", lty = 1, lwd = 2)
abline(v = mean(samples) - sd(samples), col = "blue", lty = 1, lwd = 2)
abline(v = mean(samples) + sd(samples), col = "blue", lty = 1, lwd = 2)
abline(v = mean(samples) - sqrt(0.625), col = "red", lty = 2, lwd = 2)
abline(v = mean(samples) + sqrt(0.625), col = "red", lty = 2, lwd = 2)
legend("topright", legend=c("Sample Std. Dev.", "Theoretical Std. Dev."),
       col = c("blue", "red"), lty = c(1, 1, 2), lwd = 2, cex = 0.7)
```

Code4.

```
hist(samples, main="Average of 40 Exponentials", xlab="Samples", ylab="Density",
     breaks = 20, freq=FALSE)
min = min(samples)
max = max(samples)
x <- seq(min, max, length=100)
gx <- dnorm(x, mean = 5, sd = sqrt(0.625))
lines(x, gx, type="l", lty=1, lwd = 2, col="blue")
legend("topright", legend=c("Normal"),
       col = "blue", lty = 1, lwd = 2, cex = 0.7)
```

Code5.

```
samples2 = rexp(1000, 0.2)
hist(samples2, main="Exponentials", xlab="Samples", ylab="Density", breaks = 50,
     freq=FALSE)
min2 = min(samples2)
max2 = max(samples2)
x <- seq(min2, max2, length=100)
gx <- dnorm(x, mean = 5, sd = 5)
lines(x, gx, type="l", lty=1, lwd = 2, col="blue")
legend("topright", legend=c("Normal"),
       col = "blue", lty = 1, lwd = 2, cex = 0.7)
```