Large Numbers and the Central Limit Theorem

Martí Cuquet

Abstract

This report simulates the exponential distribution. We first show that, as expected by the Law of Large Numbers, the sample mean and variance approach the theoretical mean and variance, respectively. Then, we compare the normalised distribution to that of a standard normal distibution, and observe that it is indeed approximately normal, as predicted by the Central Limit Theorem.

Introduction

In this project we investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is the probability distribution that describes the time between events in a Poisson process. Its probability density function is

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

and both its mean and standard deviation are equal to $1/\lambda$.

Simulations

Throughout this project we will set $\lambda = 0.2$. We will investigate the distribution of averages of $m = 40$ exponentials with a thousand ($n = 10^3$) simulations:

```
n <- 10^3
m <- 40
lambda <- 0.2
```

Let us also set the seed for reproducibility. We perform n simulations of m random variables each, and store them in a matrix whose n rows are the simulations:

```
set.seed(1)
simulations <- matrix(rexp(m * n, lambda), nrow = n, byrow = TRUE)
```

The sample means and variances for each simulation can be easily computed:

```
means <- rowMeans(simulations)
variances <- apply(simulations, 1, var)
```

Results

The aim of this project is to investigate the exponential distribution and compare it with the Law of Large Numbers and the Central Limit Theorem.

1

Sample Mean versus Theoretical Mean

We first compare the sample mean versus the theoretical mean, which is $\mu = 1/\lambda = 5$.

```
cumMeans <- cumsum(means) / (1:length(means))

g1 <- ggplot(data.frame(x = 1:length(cumMeans), y = cumMeans),
             aes(x = x, y = y))
g1 <- g1 + geom_hline(yintercept = 1 / lambda, color = "green") + geom_line()
g1 <- g1 + labs(title = "Sample mean versus theoretical mean",
                x = "Number of simulations", y = "Cumulative mean")
g1 <- g1 + theme(text = element_text(size=8))
```

As expected, the sample mean $\bar{X} = 4.9900252$ approaches the theoretical mean $\mu = 5$ after 1000 simulations of 40 random numbers each. This can be more clearly seen in Figure 1, which shows the cumulative mean after a certain number of simulations, which converges to the theoretical mean (in green).
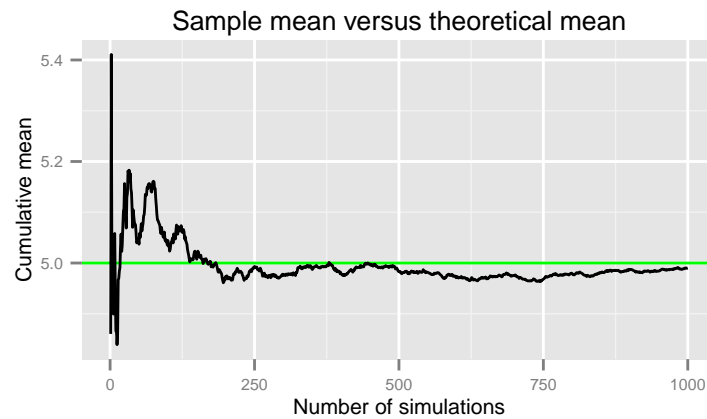

Figure 1: Cumulative mean versus number of simulations. The green, horizontal line depicts the theoretical mean. With increasing number of simulations, the sample mean approaches the theoretical mean.

Sample Variance versus Theoretical Variance

The same analysis can be done for the variance, which is $S^2 = 1/\lambda^2 = 25$.

```
cumVariances <- cumsum(variances) / (1:length(variances))

g2 <- ggplot(data.frame(x = 1:length(cumVariances), y = cumVariances),
             aes(x = x, y = y))
g2 <- g2 + geom_hline(yintercept = 1 / lambda^2, color = "green") + geom_line()
g2 <- g2 + labs(title = "Sample variance versus theoretical variance",
                x = "Number of simulations", y = "Cumulative variance")
g2 <- g2 + theme(text = element_text(size=8))
```

Again, the sample variance $\bar{S^2} = 25.064586$ approaches the theoretical variance $S^2 = 25$ after 1000 simulations of 40 random numbers each. Figure 2 shows the cumulative variance after a certain number of simulations, which converges to the theoretical variance (in green).
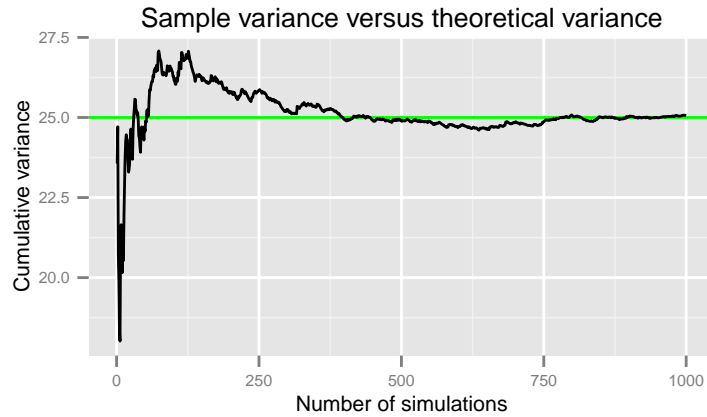
Figure 2: Cumulative variance versus number of simulations. The green, horizontal line depicts the theoretical variance. With increasing number of simulations, the sample variance approaches the theoretical variance.

Distribution

Finally, we check that the normalized mean, $\hat{X} := \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, has a distribution like that of a standard normal.

```
meansNorm <- sqrt(m) * (means - 1/lambda) / (1/lambda)
meanMeansNorm <- mean(meansNorm)
varMeansNorm <- var(meansNorm)

g3 <- ggplot(data.frame(x = meansNorm), aes(x = x))
g3 <- g3 + geom_histogram(binwidth= .1, aes(y=..density..),
                          colour = "black", fill = "#FF9999")
g3 <- g3 + stat_function(fun = dnorm, colour = "black")
g3 <- g3 + labs(title = "Distribution compared to the standard normal",
                x = "Normalized mean", y = "Density")
g3 <- g3 + theme(text = element_text(size=8))
```

Indeed, $\hat{X}$ has a mean of -0.013 and a variance of 0.978. Its distribution, compared to that of a standard normal, can be seen in Figure 3.
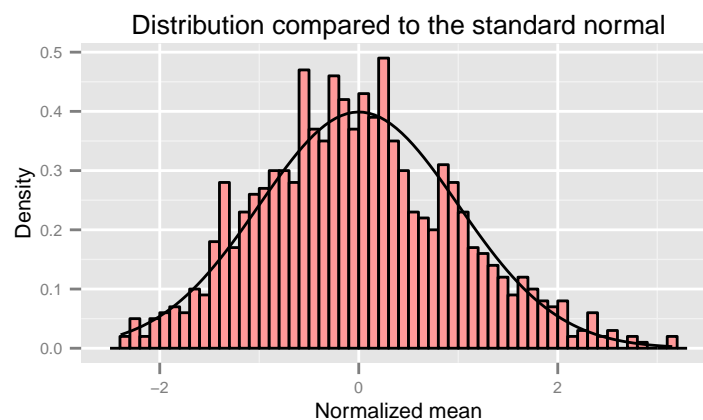

Figure 3: Density of the normalized mean. The superimposed line corresponds to the standard normal distribution.