Exponential Distributions vs the Central Limit Theorem

Nils Sandøy

August 20, 2015

This is a course project for the Coursera class in Statistical Inference. The goal of this project is to investigate the Exponential Distribution and make a comparison with the Central Limit Theorem (CLM). Ref Appendix 1. for details.

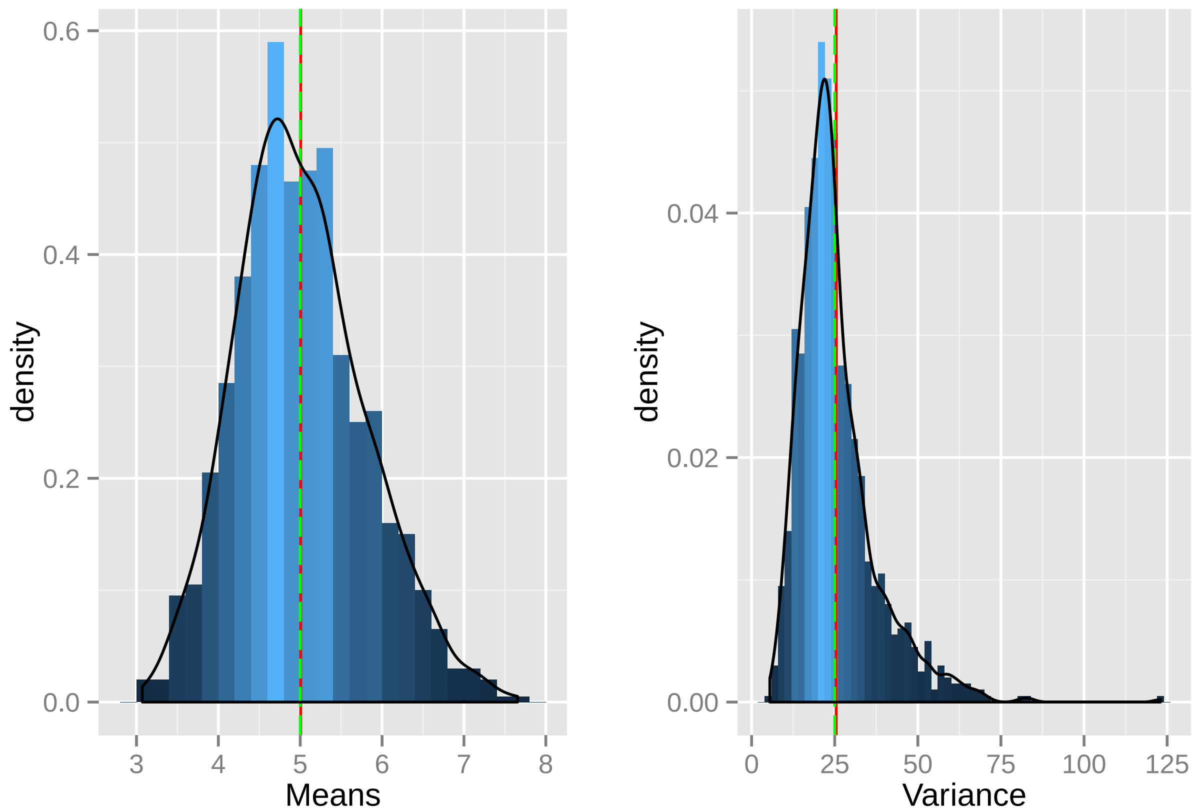
I will show that, when using the average of many simulations of the exponential distribution, the sample mean approximates the theoretical mean, the sample variance approximates the theoretical variance, and the distribution of the sample means is normal.

Illustration of the properties of the distribution of the mean of exponentials.

```
lambda <- 0.2
n <- 40
theoretical_mean <- 1/lambda
theoretical_variance <- 1/(lambda^2)
nosim <- 1000

dist <- matrix(nrow=nosim, ncol=n)
for (i in 1:nosim) dist[i,] <- rexp(n, lambda)

means <-apply(dist, MARGIN=1, FUN=mean)
variances <- apply(dist, MARGIN=1, FUN=var)

sample_mean <- mean(means)
sample_variance <- mean(variances)
```

Here I have run 1000 simulations each generating 40 random exponentials. λ is set to **0.2** for all simulations. I have also calculated the the mean and variance of each of these simulations. Finally I calculate the **sample mean**, 5.0061811, and **sample variance**, 25.4309141 accross all these 1000 simulations.

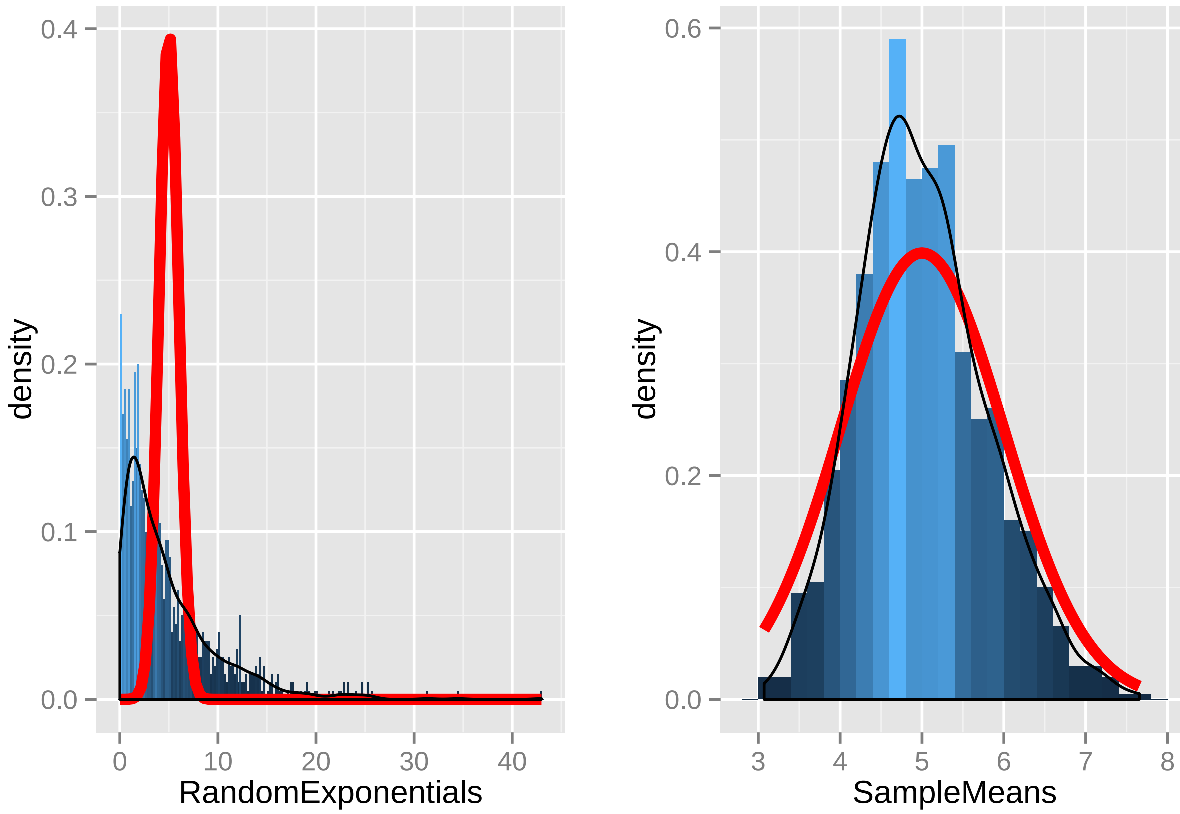
Comparison of sample mean & variance to the theoretical mean & variance of the distribution.



The left histogram above shows the means of each of the 1000 simulations with $n = 40$. The red sample mean, the average of the individual means of each of the 1000 samples, is very close to the theoretical mean ($\frac{1}{\lambda}$) of **5** shown in green. **This corresponds with the CLT which predicts that the sample mean will approximate the theoretical mean when we do many simulations.**

Similarly in the right histogram we have plotted the distribution of the variance of each of the 1000simulations. The average variance, shown as a red vertical line, is $Var(\bar{X}) = \frac{\sigma^2}{n} = \mathbf{25.4309141}$, which is close to the theoretical variance (green line) of $Var(X) = \frac{1}{\lambda^2} = \mathbf{25}$. **This confirms again the CLT by showing that, when running many simulations, the sample variance estimates the population variance.**

Show that the distribution of the means of the simulations is approximately normal.



This is a comparison of the normal distribution, drawn in red, with the distribution of 1000 random exponentials, contrasted with the distribution of the simulation/sample means (average of 1000 simulations of sice n=40). It is clear that the distribution of the simulation means is more "bell shaped" than the distribution of 1000 random exponentials. I conclude that **when running many simulations the distribution of the sample mean for these simulations is normal**

Appendix 1: Short definition of the CLT and the Exponential distribution

The **CLT** states that the distribution of averages of **independent** and **identically distributed** (**iid**) variables (properly normalized) becomes that of a standard normal as the sample size increases. The result is The result is that

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\text{Estimate} - \text{Mean of estimate}}{\text{Std. Err. of estimate}}$$

has a distribution like that of a standard normal for large n. A useful way to think about the CLT is that $\bar{X}_n$ is approximately $N(\mu, \sigma^2/n)$

The **Exponental Distribution** is the probability distribution that describes the time between events in a **Poisson process**, in which events occur continuosly and independently at a constant average rate.

The **probability density function** (PDF) of an exponential function is

$$f(x;\lambda) = \begin{Bmatrix} \lambda \mathrm{e}^{-\lambda x} & x >= 0. \\ 0 & x < 0. \end{Bmatrix}$$

The mean or expected value of an exponential distribution is

$$E[X] = \frac{1}{\lambda} = \beta$$

The standard deviation (σ) is equal to the mean ($\frac{1}{\lambda}$), as the variance of X is given by

$$Var[X] = \frac{1}{\lambda^2}$$

The exponential distribution with rate λ has density

$$f(x) = \lambda \mathrm{e}^{-\lambda x}$$