Comparison of the Exponential Distribution with the Central Limit Theoreom

Gene Kaufman

February 7, 2016

Overview

The Exponential Distribution can be simulated with the R function *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. Using lambda = 0.2, I will investigate the distribution of 1000 simulations of size 40 exponentials and compare to the Central Limit Theorem.

Simulations

First, load some libraries and set some options

```
require(knitr)
opts_chunk$set(echo=TRUE, results="asis", warning=FALSE, message=FALSE)
```

Initialize Variables

```
lambda <- 0.2
n <- 40
num_sims <- 1000
```

Create Simulations

```
set.seed(42) # Reproducibility!
```

Build matrix of 1000 simulations of 40 Exponentials

```
exp_dist <- matrix(rexp(n * num_sims,lambda),num_sims)
```

Calculate means of the simulations

```
exp_means <- apply(exp_dist, MARGIN=1,FUN=mean)
```

Sample Mean versus Theoretical Mean:

The Theoretical Mean of the Exponential Distribution is *1/lambda*

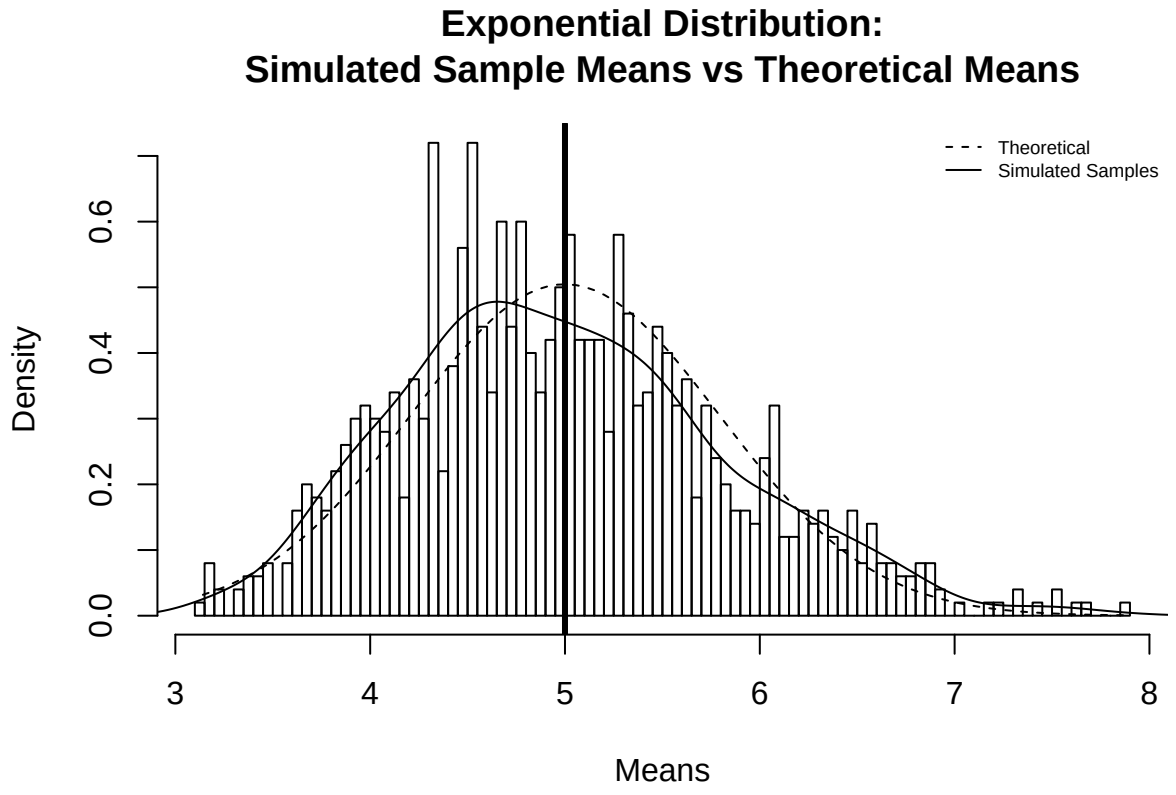
```
theory_mean <- 1/lambda
```

The average Sample Mean:

```
avg_sample_mean <- mean(exp_means)
```

Theoretical mean: 5, Sample mean: 4.9865083

Plotting everything onto one chart makes it easy to compare that the Sample Mean and Theoretical Mean are pretty close:



Sample Variance versus Theoretical Variance:

The Theoretical Variance of the Exponential Distribution is the square of the Theoretical SD: *(1/lambda)^2* or *theory_sd^2*

```
theory_sd <- 1/lambda
theory_var <- theory_sd ^ 2
```

However, we are comparing to a sample, so we have to account for the sample size:

```
theory_sd_samp <- theory_sd / sqrt(n)
theory_var_samp <- theory_sd_samp ^ 2
```

The average Sample Variance:

```
avg_sample_var_means <- var(exp_means)
```

Theoretical Variance: 0.625,

Average of Sample Variance Means: 0.6793521

Distribution:

The distribution is approximately normal because the Central Limit Theoreom states that the means of large number of iterations of a distribution will be approximately normally distributed. Here we can see that the histogram of means has a very Gaussian look, centered around the distribution mean