Statistical Inference : Course Project

Craig Garlick

17 July 2017

Part 1 : Simulation Exercise

As described in the exercise, this project compares the exponential distribution to the central limit theorum. For reference, an exponential distribution controlled by a rate of lambda has the following properties:

1. mean = 1/lambda
2. variance = 1/lambda^2 (and therefore standard deviation is equal to the mean)

The below figure illustrates the exponential function:

Histogram of rexp(1000, lambda)
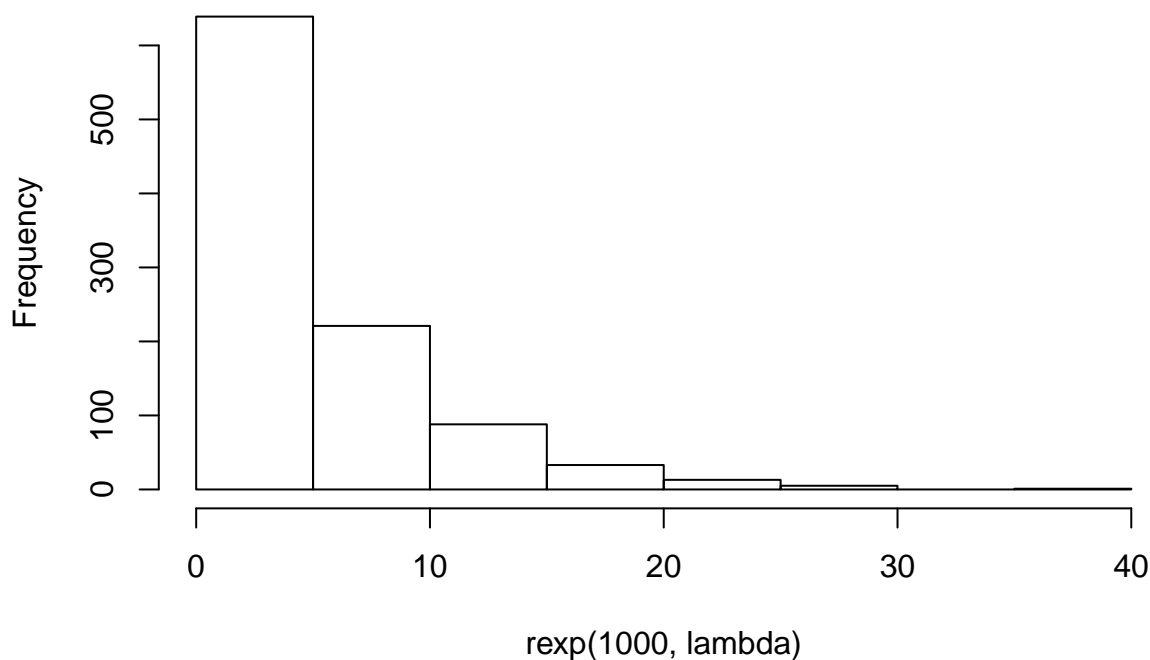
Simulations

For this exercise we'll be investigating the distribution of 40 exponentials across 1000 simulations, using a fixed rate of 0.2. The following code will generate all our data, by running the full number of simulations across all required distributions and arranging into a matrix, ie 1000 rows of 40..

```
lambda <- 0.2
no_of_simulations <- 1000
no_of_exponentials <- 40
distribution_matrix <- matrix(rexp(no_of_simulations * no_of_exponentials, lambda), ncol = no_of_exponen
```

Comparisons

```r
library(ggplot2)

means_of_distributions <- apply(distribution_matrix, 1, mean)

sample_mean <- mean(means_of_distributions)
theoretical_mean <- 1/lambda

sample_variance <- var(means_of_distributions)
theoretical_variance <- 1/(lambda^2 * no_of_exponentials)
```
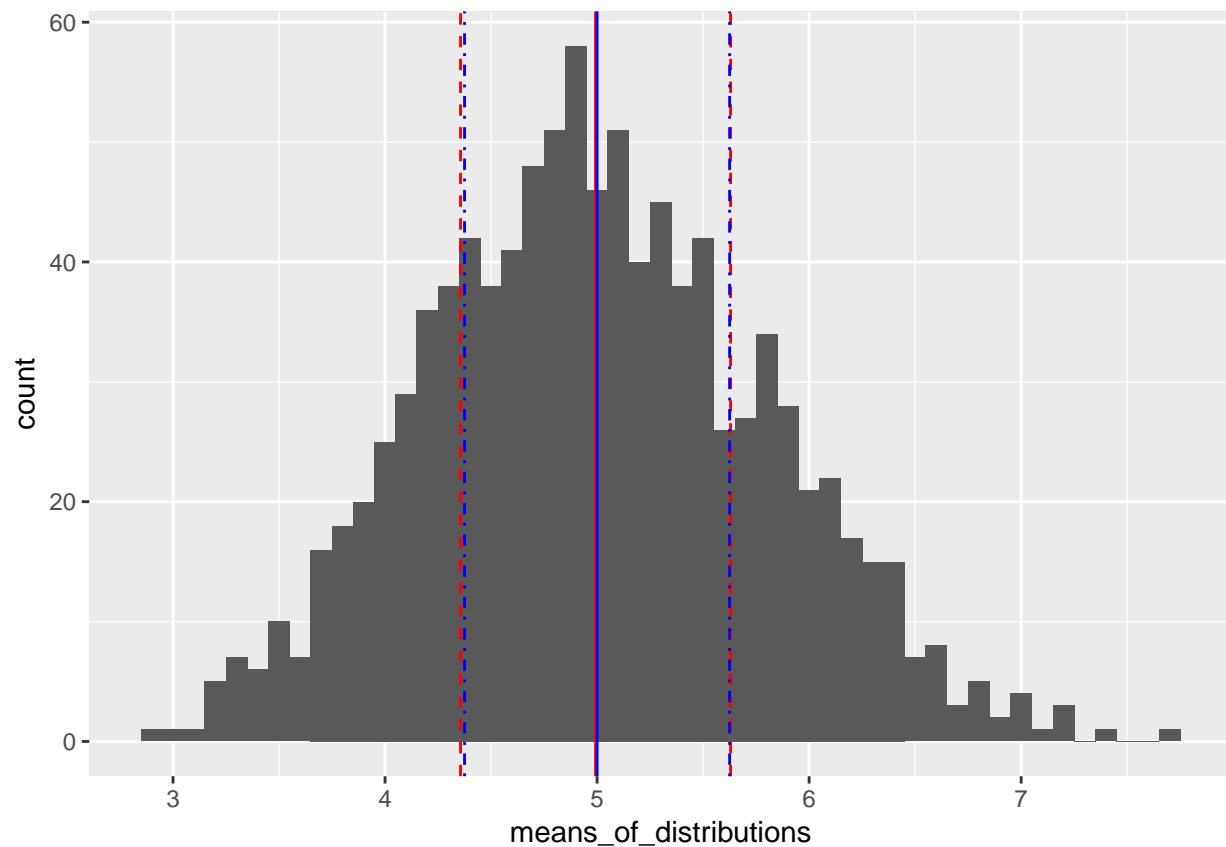
The sampled mean of the the simulations is 4.992765, very close to the theoretical mean of 5

The sampled variance of the above distribution is 0.6368553, very close to the theoretical variance of 0.625

The comparisons can be seen visually in the below plot, experimental values in red, theoretical values in blue

Finally we can see how the distribution of averages of exponentials is an approximately normal distribution (top), vs the exponential distribution (bottom)