Exponential Distribution Simulations

Nicholas Nappy

September 2014

This is a basic data analysis completed for the Johns Hopkins University Statistical Inference course project on Coursera (https://www.coursera.org/course/statinference) in September 2014.

The R script used in this analysis can be found at https://github.com/nnappy/JHU-Statistical-Inference/blob/master/Exponential.R

The mean of forty random numbers from an exponential distribution where lambda = 0.2 were computed one thousand times. The thousand sample means were then compared to the theoretical mean of the distribution (1/lambda) and the theoretical variance of the distribution (1/lambda^2)

```
nosim <- 1000
n <- 40
lambda = 0.2

data <- matrix(rexp(nosim * n, lambda), nosim)
data <- apply(data, 1, mean)

sample_mean <- mean(data)
sample_sd <- sd(data)
```

The sample mean obtained from the simulation approximates the theoretical mean of the exponential distribution of 1/lambda which in this case is 5.

```
print(sample_mean)
```

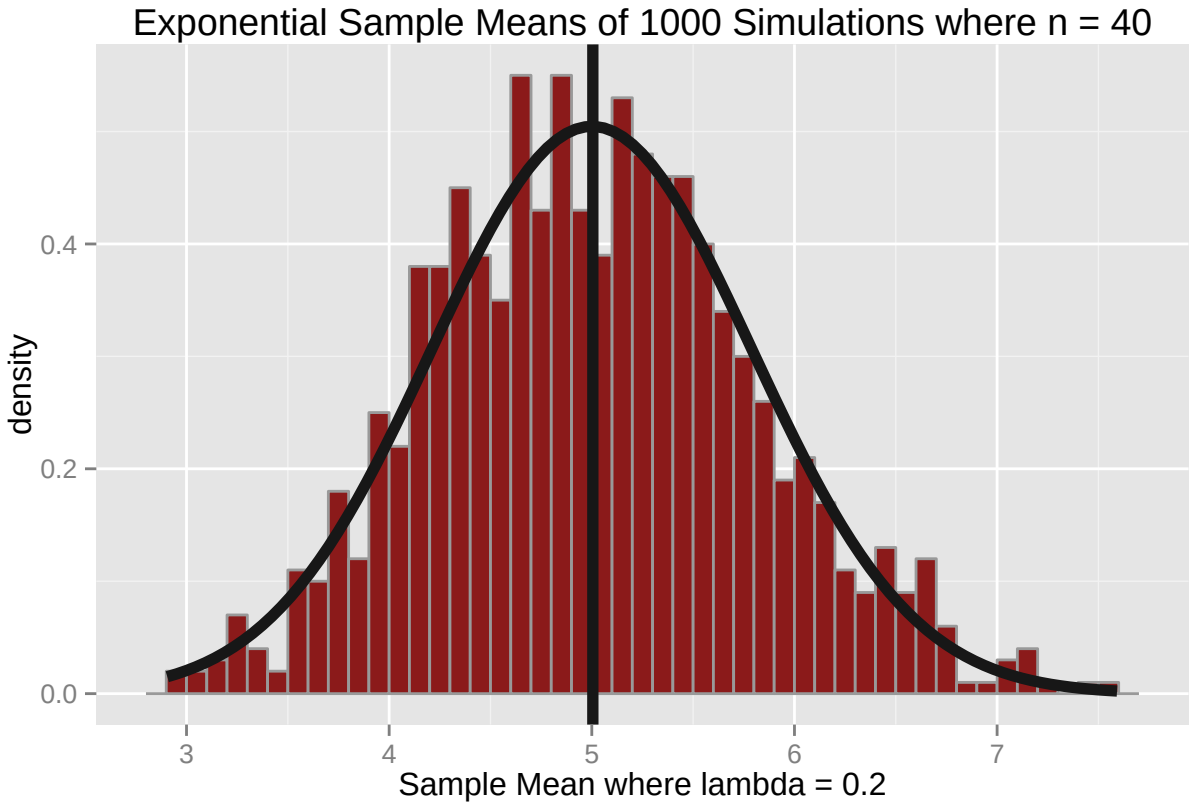
```
## [1] 5.005
```

The sample variance obtained from the simulation approximates the theoretical variance of the exponential distribution of 1/lambda^2 (25) so long as we divide the population variance by n, the number of random samples drawn in the simulation (25/40 = .625)

```
print(sample_sd^2)
```

```
## [1] 0.6262
```

To illustrate that the sample distribution of means is approximately normal we can plot a histogram of the sample means. A straight black vertical line indicates the mean of the sample distribution. The curved black line plots a normal distribution with the sample means and variance obtained from the simulation as parameters.



The coverage of the sample mean with a 95% confidence interval (ie 1.96 standard errors from the mean) is

```
## [1] 4.76 5.25
```

If we did not know the population mean beforehand, we could say we are 95% confident that this interval contains the true population mean, which in this case, knowing the population mean is 5, it obviously does.