Exponential Distribution Exploration

Daniel Hertenstein

December 29, 2016

Overview

In this analysis, we see that averages of samples random variables provide accurate insight ino the characteristics of the distribution of those random variables. In particular, we will show that averages of samples of an exponential random variable can tell us the mean and variance of that random variable without prior knowledge of those characteristics. We also see that distributions of collections of variables will be approximately normal as stated in the Central Limit Theorem.

Simulations

For this project we want to look at the distributions of 1000 simulations of 40 exponentials. To get the raw data for the analysis, I generate 1000 * 40 exponentials, and reshape the data into a matrix that has 1000 rows and 40 columns. This will allow me to easily compute different statistics describing the data.

```r
lambda = 0.2
num_samples <- 1000
sample_size <- 40
samples <- matrix(rexp(sample_size * num_samples, lambda), nrow=num_samples, ncol=sample_size)
```

Sample Mean versus Theoretical Mean

Let's compute the sample mean. We do this by computing the mean of each simulation and then take the mean of those 1000 means.

```r
means <- apply(samples, 1, mean)
mean(means)
```

```
## [1] 5.015726
```

According to the theory, the mean of this distribution should be 1 / lambda. In our case, lambda is 0.2, so the mean should be 1 / 0.2 = 5. Our sample mean is very close.

Sample Variance versus Theoretical Variance

Let's do the same thing for the sample variance. We compute the variance for each simulation and then take the mean of those 1000 variances.

```r
variances <- apply(samples, 1, var)
mean(variances)
```

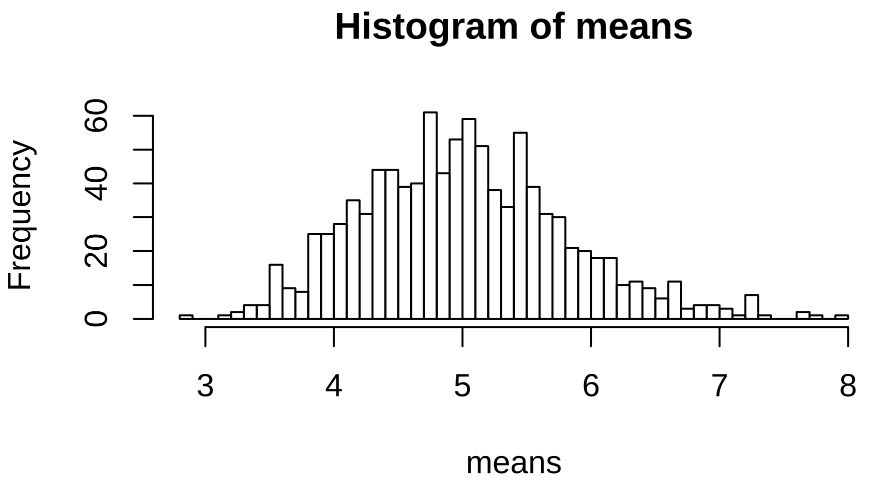
```
## [1] 25.24465
```

According to the theory, the variance of this distribution should be (1 / lambda)^2. In our case, lambda is 0.2, so the variance should be (1 / 0.2)^2 = 25. Our sample variance is very close.

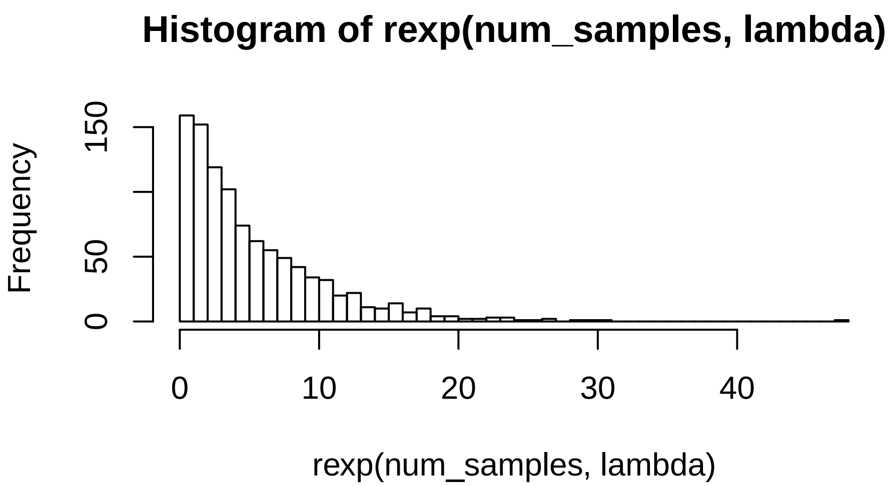
Distribution

Let's compare the distribution of our simulations to the distribution of 1000 simulations of collections of 1 random exponentials.

```r
hist(means, 50)
```




```r
hist(rexp(num_samples, lambda), 50)
```



The distribution of our simulations of 40 exponentials is vastly more normal looking than the 1000 simulations of 1 exponential. And, we can use the Central Limit Theorem to further show that the distribution of our simulations of 40 exponentials is approximately normal. This can be seen by the symmetric range around the mean.

```r
normal <- (means - mean(means)) / (sqrt(mean(variances)) / sqrt(num_samples))
hist(normal, 50)
```