Exponential Distributions Simulations Analysis in R

JBrand

Overview

In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. I will investigate the averages of 40 exponentials with lambda = 0.2 over 1000 simulations.

Simulations

To start this simulation we will set our constants and create our empty variable `mns`. Then I will run the 1000 simulations for the means of the sets of 40 exponentials.

```
lambda = .2
n = 40
mns = NULL

for (i in 1:1000) {
    mns = c(mns, mean(rexp(n,lambda)))
}
```

Let's just take a quick peak at what we have created.

```
str(mns)
```

```
##  num [1:1000] 5.57 5.57 6.52 5.51 4.87 ...
```

Looks good - a vector of length 1000 with each element containing the mean of 40 exponentials.

Sample Mean vs. Theoretical Mean

Now let's compare the sample mean and the theoretical mean. Theoretically, the mean is calculated as 1/lambda. So for our simulations, the theoretical mean is equal to 5.

```
mean(mns)
```

```
## [1] 5.021678
```

And as you can see, our mean of the calculated means of the exponentials is almost exactly there as well. This is very much thanks to the Law of Large numbers (LLN), stating that as you collect more and more data (or in our case, virtually running larger numbers of experiments), you will eventually estimate the population mean exactly. Thanks statistics!

Sample Variance vs. Theoretical Variance

Now for the variances, it gets slightly trickier for the theoretical variance. We are given that the standard deviation for the exponential distribution is 1/lambda. In this case, since it's being applied to a large sample, we need to convert it first to the standard error (sigma/sqrt(n)), and then square it to get our variance. So the theoretical variance becomes ((1/lambda)/sqrt(n))^2.

```
var(mns)
```

```
## [1] 0.6124554
```

```r
((1/lambda)/sqrt(n))^2
```

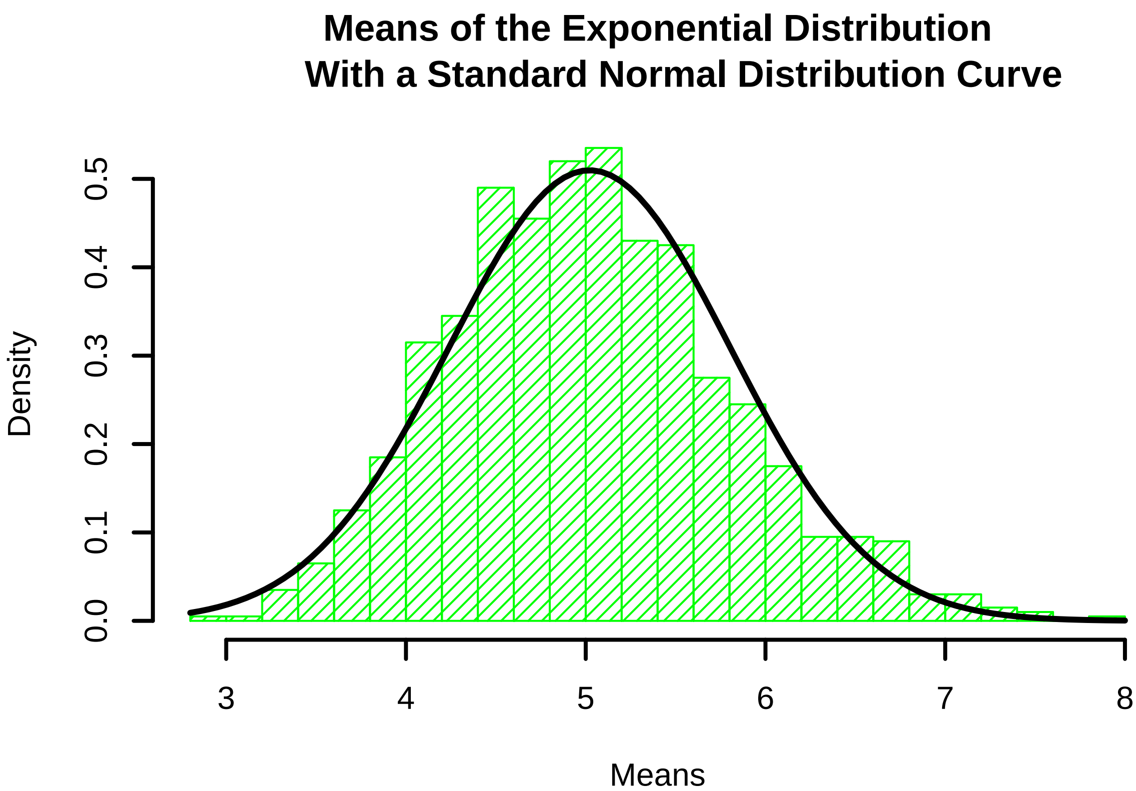
```
## [1] 0.625
```

You'll see that our theortical value of 0.625 is very closely approximated with the sample variance of .612~.

Distribution

Finally, let's look at the distribution.

```r
hist(mns, breaks = 20, probability = TRUE, density = 20, lwd = 2, col = "green",
     xlab = "Means", main = "Means of the Exponential Distribution
     With a Standard Normal Distribution Curve")
curve(dnorm(x, mean = mean(mns), sd = sd(mns)), add = TRUE, lwd =3)
```



With a Standard Normal Distribution Curve overlaid on the histogram, it's pretty easy to see that the Central Limit Theorem is at work. The CLT states that the distribution of averages of iid variables becomes that of a standard normal as the sample size increases. And I think it's safe to say that we have covered that base.