Exponential Distribution Simulation

Oliver Stroemer

2019-07-16

Synopsis

In this project we'll investigate the exponential distribution and compare it to the Central Limit Theorem (CLT).

We conclude that the means of observations from the exponential distribution are normally distributed.

Simulation

For making the simulations in this project reproducible, we need to set the seed of the RNG. We'll use an integer representation of the date, when this report was written.

Then we'll compute the mean of each 40 observations from an exponential distribution with $\lambda = .2$ a 1000 times; they'll be stored in `means`.

```
set.seed(20190716)

means <- sapply(1:1000, function(i) mean(rexp(40, 0.2)))
```

Sample Mean vs Theoretical Mean

In an exponential distribution, we expect the sample mean to be $\frac{1}{\lambda}$. Therefore in our simulation the expected mean of the sample means is $\frac{1}{0.2} = 5$.

The real mean of our sample means is:

```
mean(means)
```

```
## [1] 4.985932
```

This is fairly close to our expected value.

Sample Variance vs Theoretical Variance

The variance of our sample means is expected to be $s^2 = \frac{\frac{1}{\lambda}}{n}$. In our simulation the expected variance is $s^2 = \frac{(\frac{1}{.2})^2}{40} = .625$.

The real variance of our sample is:

```
var(means)
```

```
## [1] 0.6209562
```

Again this is fairly close to our expectations. The variance is only a little bit smaller; this means, that our sample is a little bit less spread out from the mean as we would have theoretically expected.

Comparsion of the means distribution to the normal distribution

The central limit theorem states that the distribution of averages of iid random variables approaches a normal with growing sample size. In the plot below, we plot a density histogram of our sample means and overlay it with a normal distribution with the same mean and the same standard deviation as our sample (red curve).

As we see the curve approximates the histogram quite closely. Therefore we can conclude that the CLT applies also to the means from the exponential distribution.

```
hist(means, freq=F, ylim=c(0,.51),
     main = "Histogram of rexp sample means\n in comparison to the normal distribution",
     xlab = "Sample means")
curve(dnorm(x, mean(means), sd(means)), col = "red", add = T)
```