Distribution of the Mean of Exponentials

Shuai Wang

Saturday, April 25, 2015

Overview

In this report, I am going to investigate the exponential distribution and the distribution of averages of those exponentials, compare them, and explain the result with the Central Limit Theorem.

Simulations

Exponential distribution is one kind of standard distribution. It has only one parameterlambda, and both its theoretical mean and theoretical variance equal to lambda. Now we are going to investigate the distribution of averages of exponentials, and understand the Central Limit Theorem in this case.

For this purpose, we will generate samples from exponential distribution by simulation. We need fix lambda first, say 0.2. Then let us get the average of 40 exponentials, and do it for a thousand times. Those are the 1000 samples of averages of exponentials we get.

```
lambda<-0.2
simdata<-NULL
for (i in 1:1000) simdata<-c(simdata,mean(rexp(40,lambda)))
```

Sample Mean versus Theoretical Mean

First we want to compare the sample mean of the averages of exponentials to theoretical mean of exponential distribution, which is 1/0.2=5 in this case.

```
mean<-mean(simdata)
mean
```

```
## [1] 5.007779
```

The sample mean of the averages of exponentials is **5.0077791**, very close to **5**, the theoretical mean of exponential distribution. We can also see it from the histogram of averages of exponentials (see Appendix A), on which two vertical lines are overlayed. The green vertical line represents the sample mean of the averages of exponentials, and the red vertical line represents the theoretical mean of exponential distribution. Those two lines are very close to each other, not only in absolute values, but also relative to the distribution of the averages of exponentials.

Sample Variance versus Theoretical Variance

Next we will compare the sample variance of the averages of exponentials to theoretical variance of exponential distribution, which is also 1/0.2=5 in this case.

```
var<-var(simdata)
var
```

```
## [1] 0.6636646
```

The sample variance of the averages of exponentials is **0.6636646**, much smaller than **5**, the theoretical variance of exponential distribution. We can also see it from the histogram of averages of exponentials (see Appendix B), on which four vertical lines are overlayed. The two dashed green vertical line enclose the range of one sample standard deviation from the sample mean of the averages of exponentials, and the two dotted red vertical line enclose the range of one theoretical standard deviation from the theoretical mean of exponential distribution (Note that standard deviation is the square root of variance). The range enclosed by the two dashed green lines are much narrower than that enclosed by the two dotted red lines, not only in absolute values, but also relative to the distribution of the averages of exponentials.

Distribution

The Central Limit Theorem says that the distribution of averages of independently identical distributed variables (properly normalized) becomes that of standard normal as the sample sizes increases. If not normalized, the distribution becomes normal instead of standard normal. In our case, the sample size is 40, and let's see if the averages of 40 exponentials is approximately normal.

First let us take a look at the distribution of a large collection of random exponentials.
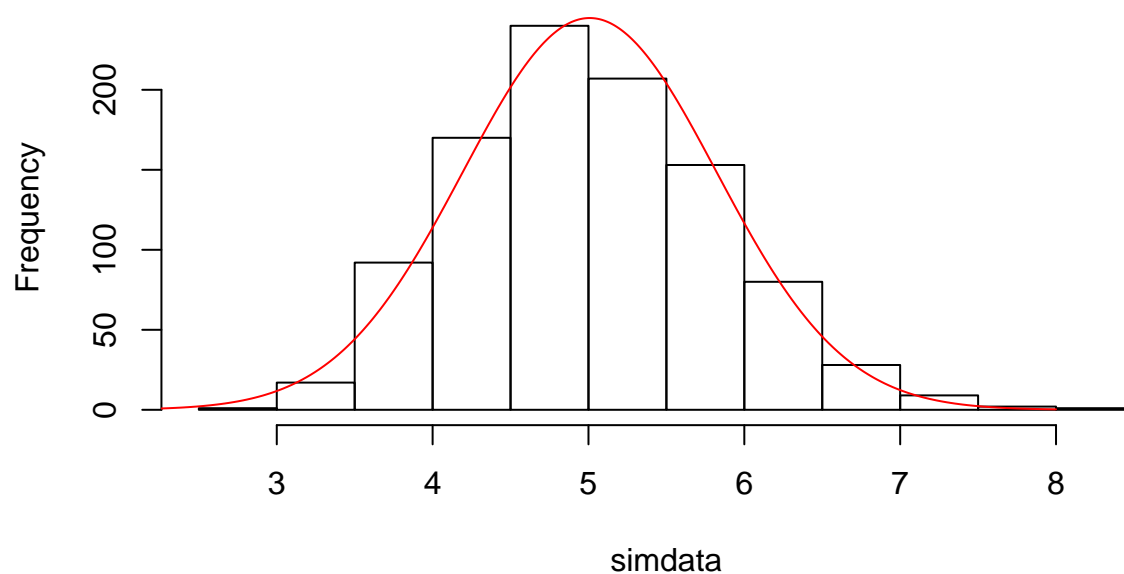
```
expdistr<-rexp(1000,lambda)
hist(expdistr,main="Histogram of Exponentials")
```
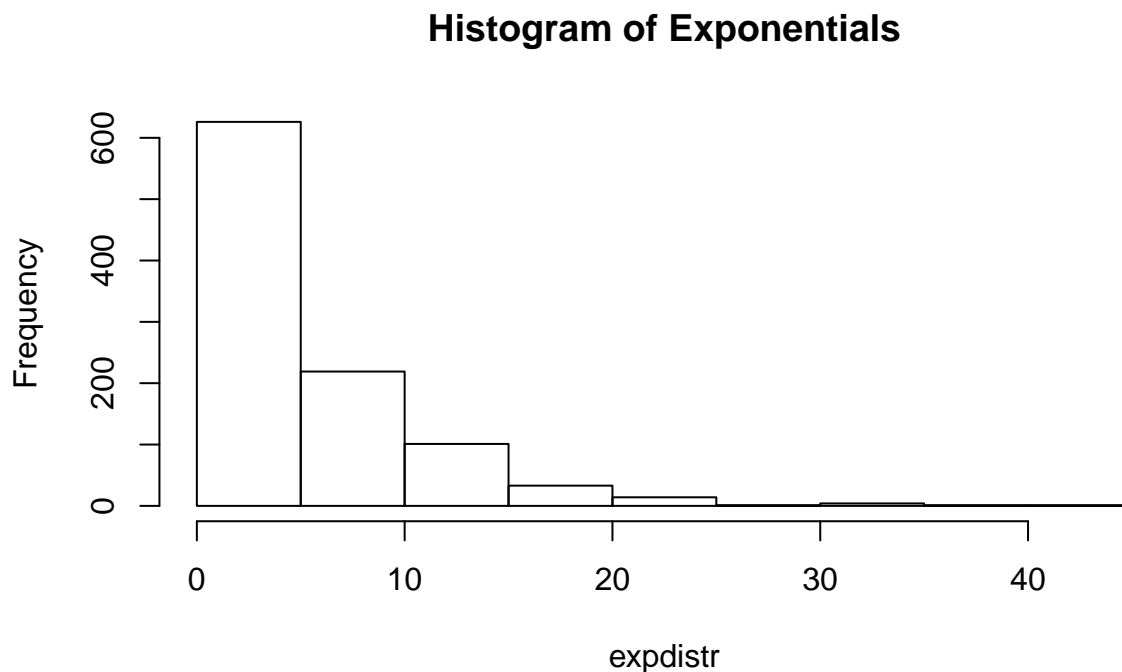

Then we look at the distribution of the averages of 40 exponentials.

```
xvals<-seq(2,8,length=1000)
hist(simdata, main="Histogram of the Averages of Exponentials")
lines(xvals,500*dnorm(xvals,mean(simdata),sd(simdata)),col=2)
```

Histogram of the Averages of Exponentials

One can easily tell the two distributions are totally different. Among the two, the latter distribution looks far more Gaussian than the former exponential distribution. What's more, on the histogram of the averages of 40 exponentials I overlay the density function from normal distribution with the same mean and variance as those of the the averages of 40 exponentials, multipled by 500 (500 is chosen to approximate the expected counts in each bar from 1000 random normal variables), and you can see that they match very well. Thus it helps us to understand the the Central Limit Theorem, i.e., the distribution of averages of independently identical distributed variables becomes normal as the sample sizes increases, no matter what kind of distribution is for those independently identical distributed variables.

Appendix A: Histogram of Averages of Exponentials

Averages of Exponentials
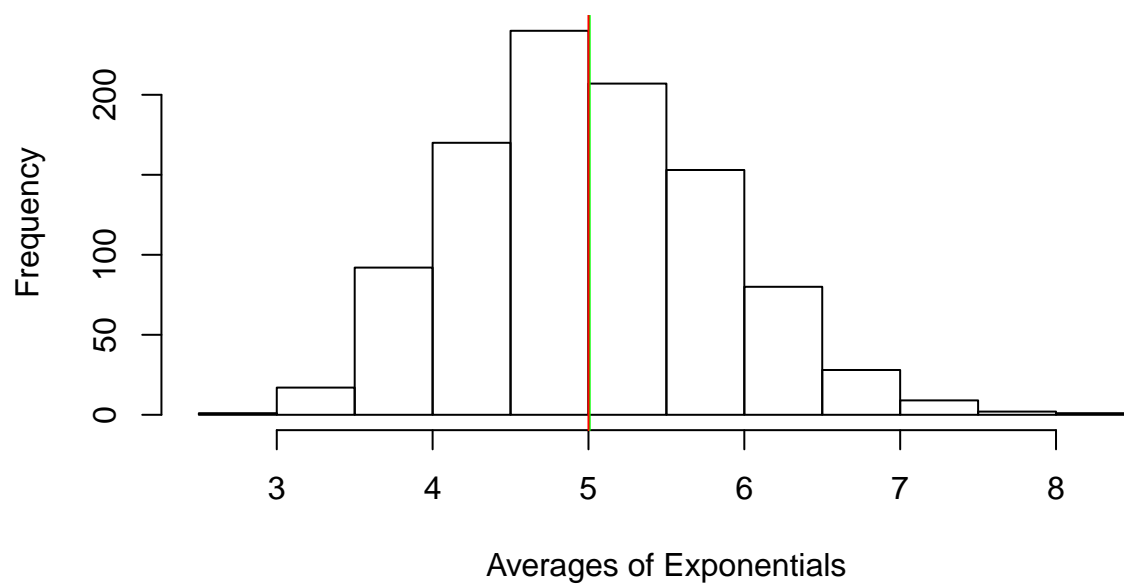
Appendix B: Histogram of Averages of Exponentials

Averages of Exponentials