Simulation of Exponential Distribution using R

Shalini Ruppa Subramanian

2015-09-24

Overview

In this project, the exponential distribution using R will be investigated. The distribution of averages of 40 exponentials will be investigated with a large number of simulations to understand the Central Limit Theorem.

Simulations

The code for generating random exponential distribution in R is `rexp(n,lamda)` where n refers to the sample size and lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. In our exercise, lambda is set to 0.2 for all the simulations. Refer to the Appendix that explains the basic exponential distribution. The assumptions involved in this exercise are that the samples are random.

Distribution of averages of 40 exponentials

We will generate the distribution of the averages of 40 exponentials with lambda (rate parameter) = 0.2.

```
set.seed(1)
mns=NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(40,0.2)))
data <- data.frame(mns,size=40)
```

The distribution of the averages of 40 exponentials is shown in Figure 1 below with the sample mean indicated.

Sample Mean versus Theoretical Mean

```
sampleMean <- mean(mns)
```

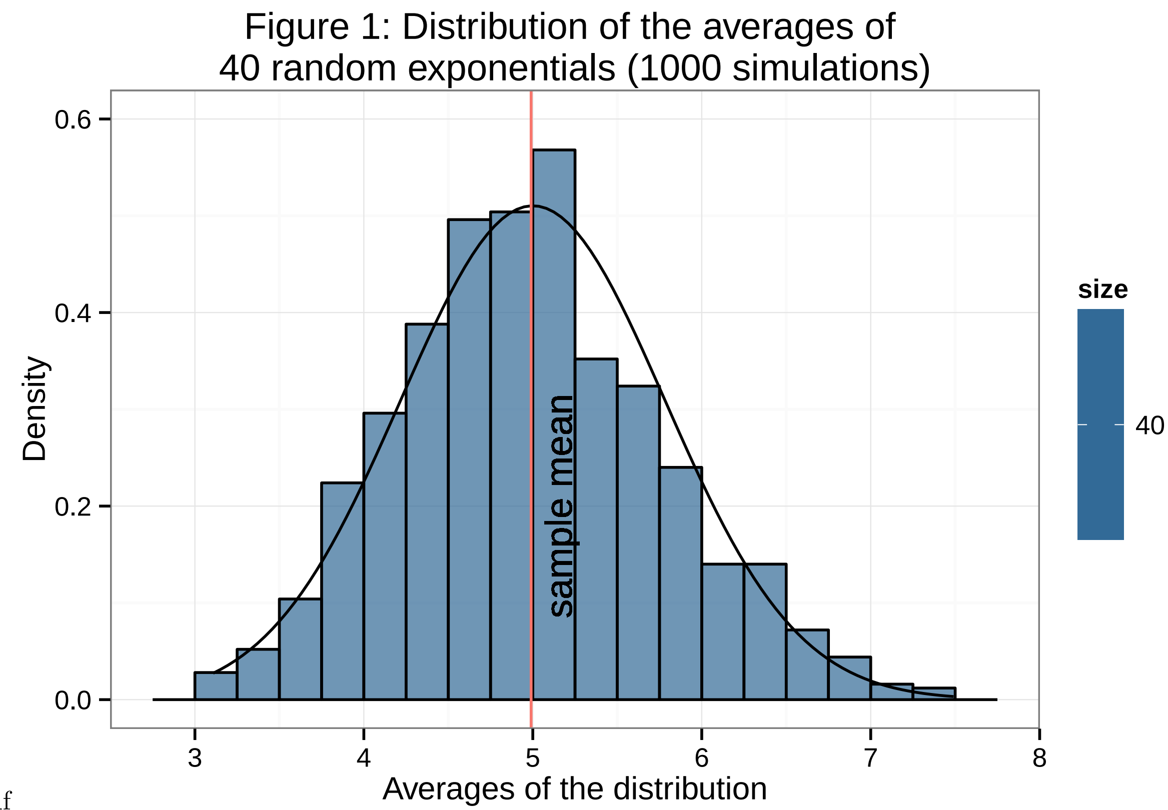
The sample mean is 4.9900252 and is also shown in Figure 1.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
ggplot(data,aes(x=mns,fill=size))+
  theme_bw()+
  geom_histogram(aes(y=..density..),alpha=0.7,binwidth=.25,col="black") +
  ylim(c(0,0.6))+
  stat_function(fun=dnorm,arg=list(mean=5,sd=sd(mns)))+
```

```
  geom_vline(aes(xintercept=mean(mns),colour="red")) +
  geom_text(aes(x=mean(mns),label="\nsample mean",y=0.2),colour="black",angle=90, text = element_text(si
              xlab("Averages of the distribution") + ylab("Density")+
            ggtitle("Figure 1: Distribution of the averages of \n40 random exponentials (1000 simulatio
```



1-1.pdf

```
theoreticalMean <- 1/0.2
```

The mean of the exponential distribution is 1/lambda. Hence, the theoretical mean is 5.

The sample mean of 4.9900252 is close to the theoretical mean of 5.

Sample Variance versus Theoretical Variance

```
sampleVariance <- var(mns)
```

The sample mean is 0.6111165.

```
theoreticalVariance <- ((1/0.2)/sqrt(40))^2
```

The standard deviation of the exponential distribution is 1/lambda. The variance is calculated using the standard deviation and sample size. Hence, the theoretical variance is 0.625.

The sample variance of 0.6111165 is close to the theoretcal variance of 0.625.

Distribution

The averages distribution is centered around the theoretical mean. By comparing an exponential distribution with Figure 1 earlier, the distribution of the averages shown in Figure 1 appears to be a normal distribution.

The normal distribution curve is also superimposed using the `stat_function` and it can be seen that the averages of exponential distribution aligns with the normal distribution.

In addition, I have shown the distribution of averages of 40 random eponentials with 10,000 simulations presented in the appendix. This is to illustrate that with more simulations, the sample averages appear to be normally distributed.

Conclusion

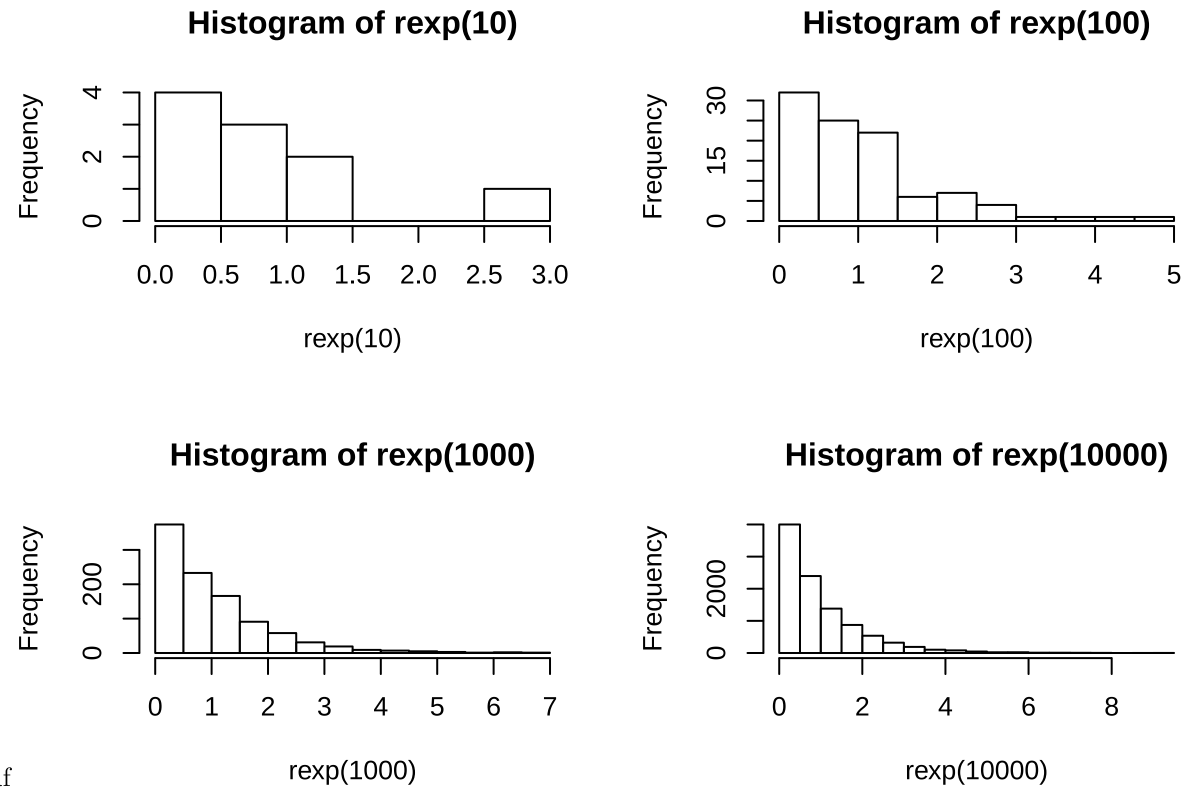
This exercise illustrates the Central Limit Theorem - states that the distribution of averages of independent and identically distributed (IID) variables becomes that of a standard normal as the sample size increases.

Appendix

Basic Exponential Distribution

Let's see how an exponential distribution looks like. The seed is set so that the results are reproducible when someone verifies the code. The random samples are drawn from exponential distribution generated using `rexp(n)` with the sample size (n) ranging from 10,100,1000,10000. The default value of 1 is used as the rate parameter. In the below histogram plot Figure A1, it can be seen that as sample size increases, the sample distribution approaches to be an exponential distribution.

```
set.seed(1)
par(mfrow=c(2,2))
set.seed(1)
hist(rexp(10))
hist(rexp(100))
hist(rexp(1000))
hist(rexp(10000))
```



exp dist-1.pdf

Distribution of averages with 10,000 simulations

The below figure shows the distribution of averages of 40 random eponentials with 10,000 simulations.

```r
set.seed(1)
mns2=NULL
for (i in 1 : 10000) mns2 = c(mns2, mean(rexp(40,0.2)))
data2 <- data.frame(mns2,size=40)
ggplot(data2,aes(x=mns2,fill=size))+
  geom_histogram(aes(y=..density..),binwidth=.25,col="black") +
  ylim(c(0,0.6))+
  stat_function(fun=dnorm,arg=list(mean=5,sd=sd(mns2)))+
  geom_vline(aes(xintercept=mean(mns2),colour="red"))+
  geom_text(aes(x=mean(mns2),label="\nsample mean",y=0.2),
          colour="black",angle=90, text = element_text(size=11))+
  xlab("Averages of the distribution") + ylab("Frequency")+
  ggtitle("Figure A2: Distribution of the averages of
        40 random exponentials (10000 simulations)")
```

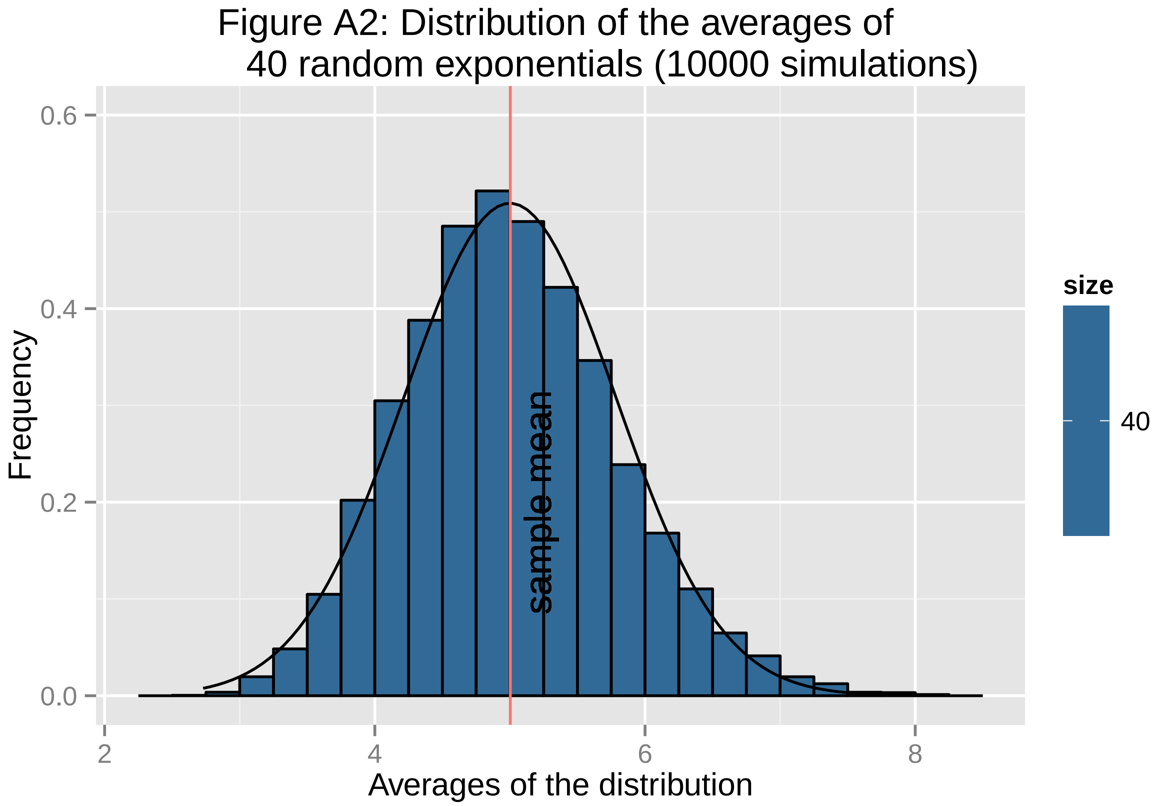


Figure A2: Distribution of the averages of
40 random exponentials (10000 simulations)
Frequency
0.6
0.4
0.2
0.0
sample mean
2
4
6
8
Averages of the distribution
size
40