Statistical Inference Project Part1 Exponential Distribution

rachelxi

26 April 2015

Overview

For the 1000 simulations of the average of 40 random exponentials, the means seem to follow a normal distribution rather than exponential distribution. The population mean and population variance approxiamtes the theoretical ones under lambda=0.2. Detailed figures are demonstrated in following parts.

Simulation

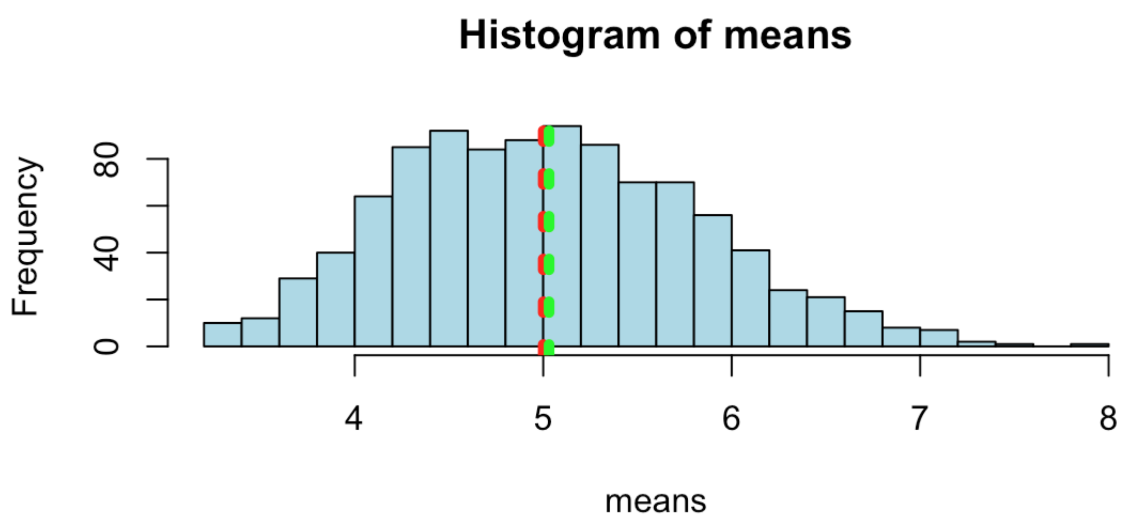
1. Sample Mean Versus Theoretical Mean

The first thing is to define the number of random exponentials generated each observation (simulation) and the value of lambda.

Then, create the random exponential distribution of the average 40 distributions per simulation and run 1000 simulations to create the original data

Then we need to explore the sample mean vector of the simulations

We can plot the histogram of the sample mean of simulations, where the red dashed line is the theoretical population mean (1/lambda=5),and the green dashed line is the mean of the sample mean in simulation.



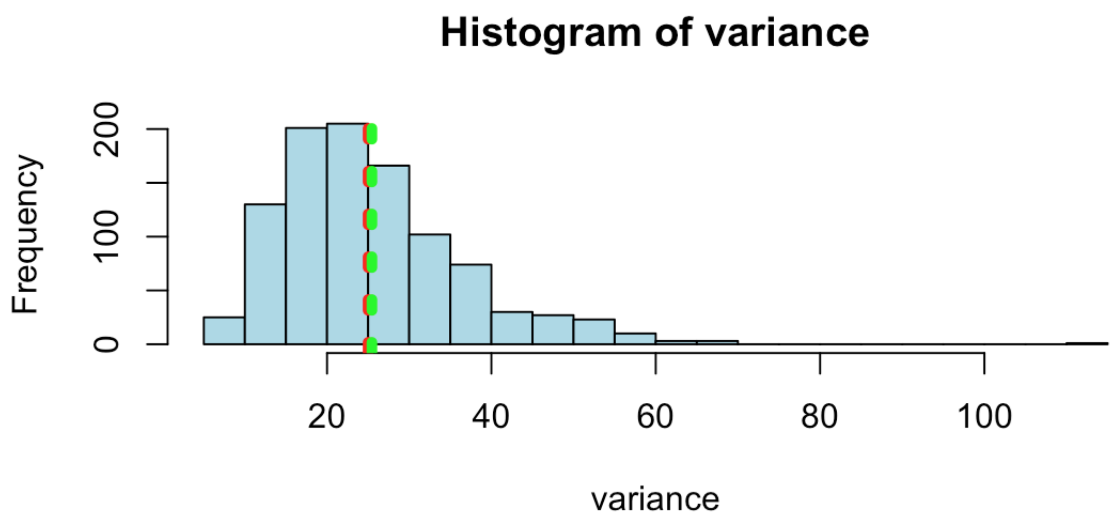
Then we can further find the value of the actual mean of sample mean in simulation:

```
## [1] 5.031
```

2. Sample Variance versus Theoretical Variance

First figure out the sample variance of the simulations

Then plot the histogram of the sample variance, where the red dashed line is the theoretical population variance (1/lambda=5),and the green dashed line is the mean of the sample variance in simulation.



Hence the mean of the variance in the simulation is (theoretical is (1/lambda)^2=25)

```
## [1] 25.45763
```

Besides, the population variance of the sample mean is

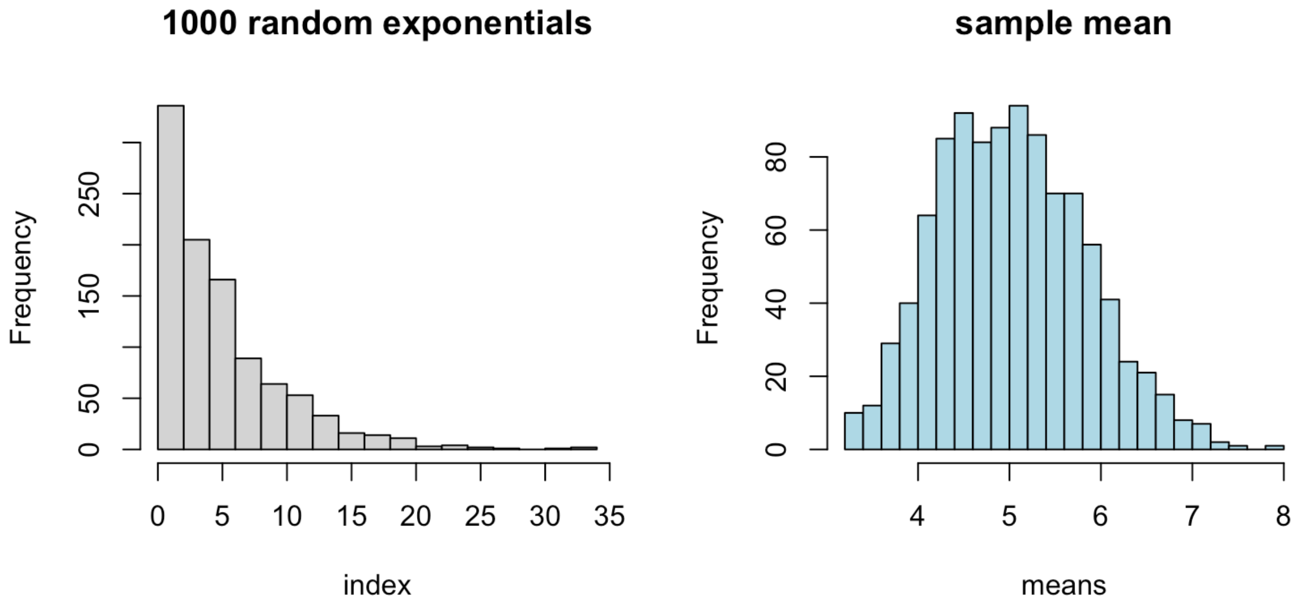
```
## [1] 0.6506353
```

And theoretical population mean is

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

3. Distribution

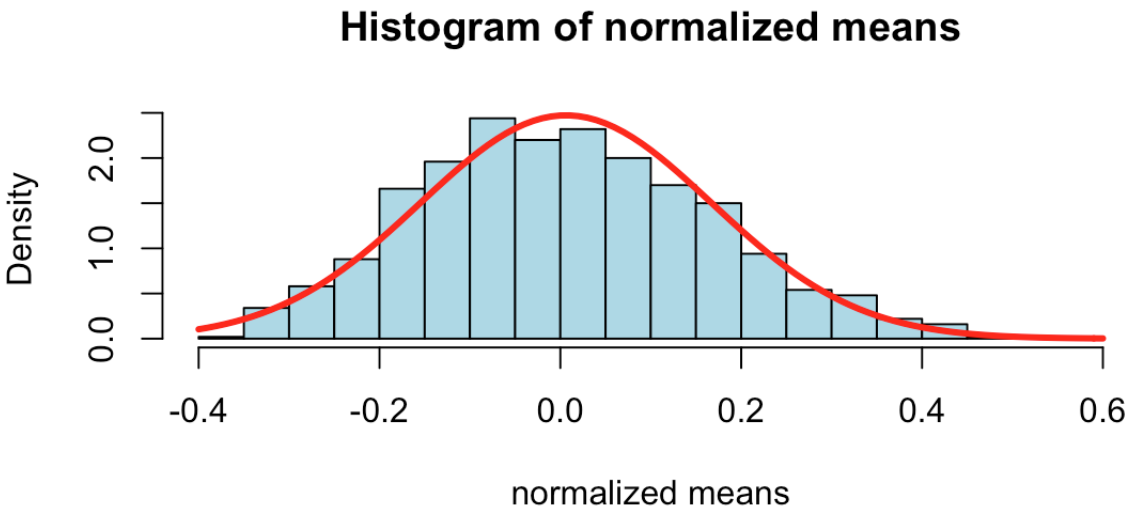
First we can compare the the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials



It is clear that 1000 random exponentials follow a exponential distribution while the sample mean of the 40 random exponentials follow an approxiamte normal distribution. Now we will look whether the judgement is true. First we normalize the data into a new vector.

Then we can plot the histogram of the adjusted data and compare it with a standard normal distibution

line



And we can conclude that it is approximately a normal distribution and the mean of new vector approximates 0

```
## [1] 0.006200092
```

Appendix–R codes

lambda <- 0.2

n <- 40

```
myexp <- NULL

for(i in 1:1000)myexp <- rbind(myexp,rexp(n,lambda))

means <- NULL

for(i in 1:1000)means <- c(means,mean(myexp[i,]))

hist(means,breaks=20, col=“lightblue”)

abline(v=1/lambda,col=“red”,lwd=5)

abline(v=mean(means),col=“green”,lwd=5)

mean(means)

variance <- NULL

for(i in 1:1000)variance <- c(variance,var(myexp[i,]))

hist(variance,breaks=20, col=“lightblue”)

abline(v=(1/lambda)^2,col=“red”,lwd=5)

abline(v=mean(variance),col=“green”,lwd=5)

var(means)

(1/lambda)^2/n

quartz()

par(mfrow=c(1,2))

hist(rexp(1000,0.2),breaks=20,col=“lightgray”, xlab=“index”,main=“Histogram of 1000 random exponentials”)

hist(means,breaks=20,col=“lightblue”, main=“Histogram of 1000 averages of 40 exponentials”)

dev.off()

new <- (means-(1/lambda))/(1/lambda)

hist(new,breaks=20,col=“lightblue”,xlab=“normalized means”, main=“Histogram of normalized means”,prob=TRUE)

curve(dnorm(x, mean=mean(new), sd=sd(new)), col=“red”, lwd=3, add=TRUE)
```