Exploring Averaged Distributions

NU-PBA1

September 2015

Synopsis

This report is a description of an exploration of how the distribution of a random variable changes as averages of 40 draws are taken instead of individual draws. The expectation is that the while the individual draws will represent the random function, the averaged draws will behave closer to the behavior of the normal distribution.

The basis for the experiment will be simulations using drawings of a randomly distributed variable from the exponential distribution. The exponential distribution is governed by a probability density function of

$$f(x,\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0 \end{cases}$$

.

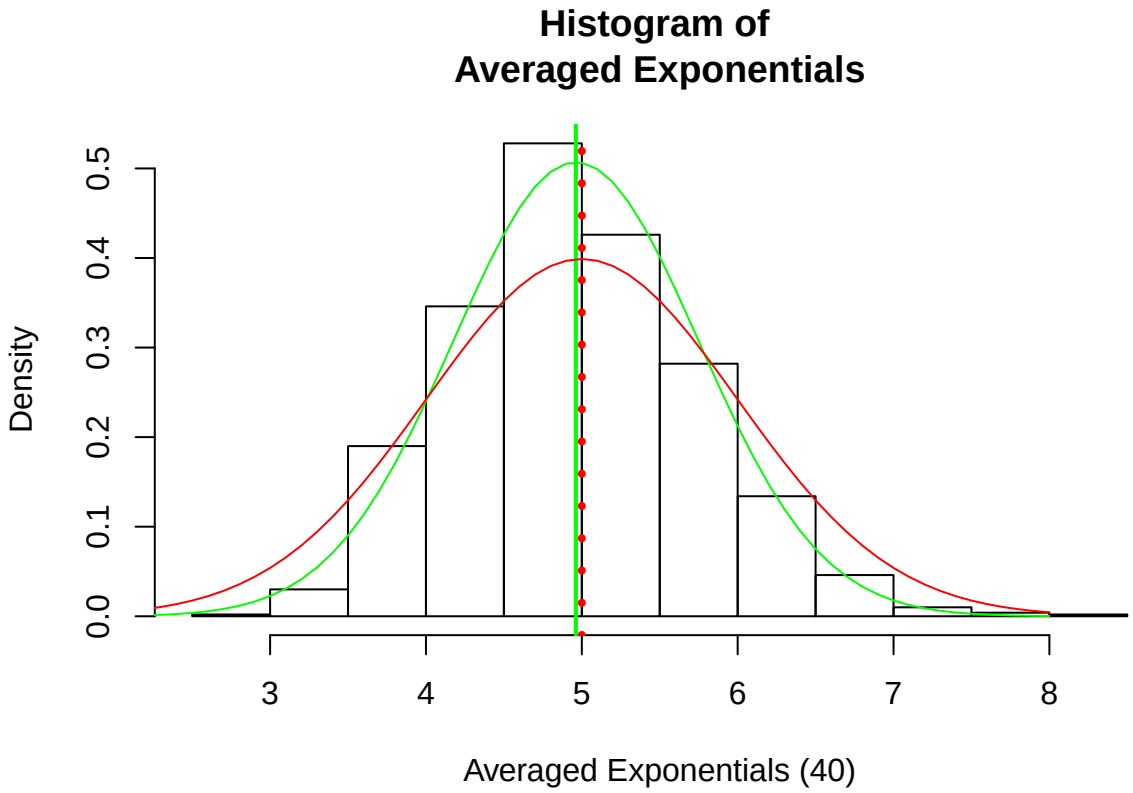
Procedure

The data will be created using the built in R exponential distribution with $\lambda = 0.2$.

```
## single draw exponentials
re<-rexp(1000,0.2)
## create averages of 40 draws of exponentials
mns = NULL
for (i in 1:1000) mns = c(mns,mean(rexp(40,0.2)))
```

After taking the draw, plot an histogram of the averages of the draws and overlay a normal distribution line with mean = $1/\lambda$ and sd = 1 to represent the normal with the theoretical mean of exponential distribution (seen in red) as well as the theoretical mean as a vertical line (dashed red). Also overlay a density plot of the averaged exponentials (green curve) and a vertical bar holding the mean of the averaged exponentials (solid green vertical).

```
x<-seq(2,8,0.1)
hist(mns,freq=FALSE,xlab="Averaged Exponentials (40)",main="Histogram of \nAveraged Exponentials")
lines(x,dnorm(x,mean=mean(mns),sd=sd(mns)),col="green")
m<-mean(mns)
abline(v=m,col="green",lwd=2)
abline(v=1/0.2,col="red",lty=3,lwd=4)
lines(x,dnorm(x,mean=1/0.2,sd=1),col="red")
```

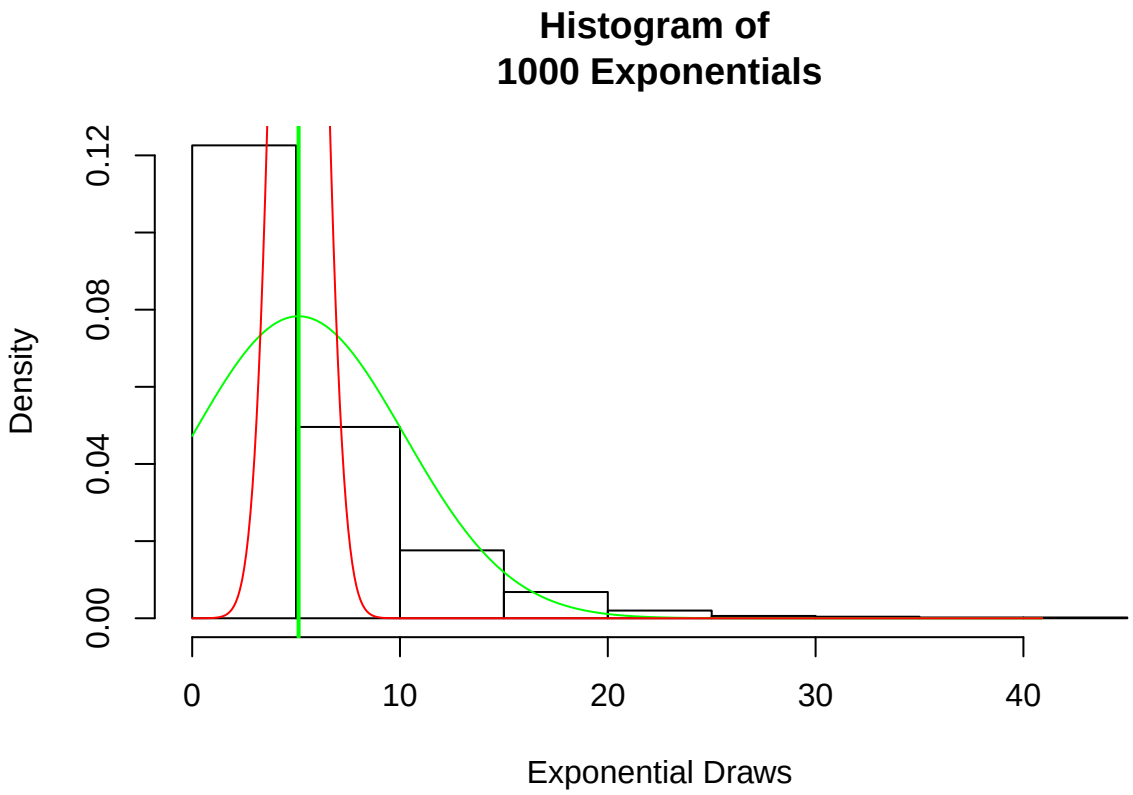



```
varm<-var(mns)
```

This plot shows that the averaged exponentials do approximate a normal distribution and the calculated mean (4.962) and the calculated variance (0.6192) approximate the the theoretical values of mean $= 1/\lambda = 5$ and variance $= n/\lambda^2 = 0.625$. This variance is smaller than the normal distribution, but approximates it.

Similarly, create the plot and perform the calculations on the individual draws data set. The resulting plot shows a skewed and broadened distribution not approximating the normal (red curve).

```
re<-rexp(1000,0.2)
x<-seq(min(re),max(re),0.1)
hist(re,freq=FALSE, xlab="Exponential Draws",main="Histogram of \n1000 Exponentials")
lines(x,dnorm(x,mean=mean(re),sd=sd(re)),col="green")
mre<-mean(re)
abline(v=mre,col="green",lwd=2)
lines(x,dnorm(x,mean=mean(re),sd=1),col="red")
```




```
varre<-var(re)
```

The calculated variance of the individual draw data is 25.9604 which approximates the theoretical variance of the exponential distribution, not the normal which is as expected.