Report 1

Report 1

Intro

This document is a report to submit analysis asked for in the Coursera Statistical Inference Project, Question 1.

Question

The exponential distribution can be simulated in R with *rexp(n, lambda)* where lambda is the rate parameter. The mean of exponential distribution is *1/lambda* and the standard deviation is also also *1/lambda*. Set ***lambda = 0.2*** for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should:
1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Note that for point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

Answer

Show where the distribution is centered at and compare it to the theoretical center of the distribution.

By the central limit theorum, we know that the distribution should be centered around the mean, should be approximately Gaussian and the standard deviation is also known. Mean is $1/lambda = 1/0.2 = 5$ Variance is $(1/lambda)^{2/n.obs} = 1/(0.04$\(\times\)$40) = 0.625$ So the distribution should be centered around 5.

```
set.seed(123)
# Run Simulation:
sim<- matrix(
      data = rexp(n=1000*40,
                  rate=0.2),
      ncol = 1000)

# Get the sample mean from each set of 40 independent observations
meanv<-colMeans(sim)

# Compare the center of the distribution with the theoretical value.
# Remember that the theoretical value in this case is 1/lambda = 1/0.2 = 5
mean(meanv)
```

```
## [1] 5.011911
```

Here we ran the simulation for 40 variables and the number of simulations was 1000. We can see that the final value i.e, mean(sim) is very close to 5.

Show how variable it is and compare it to the theoretical variance of the distribution

This result can be shown taking the variance of columns :

```
# Compare the variability of the distribution with the theoretical value.
# theoretical value in this case is (1/(0.2))^2/40 = 0.625
var(meanv)
```

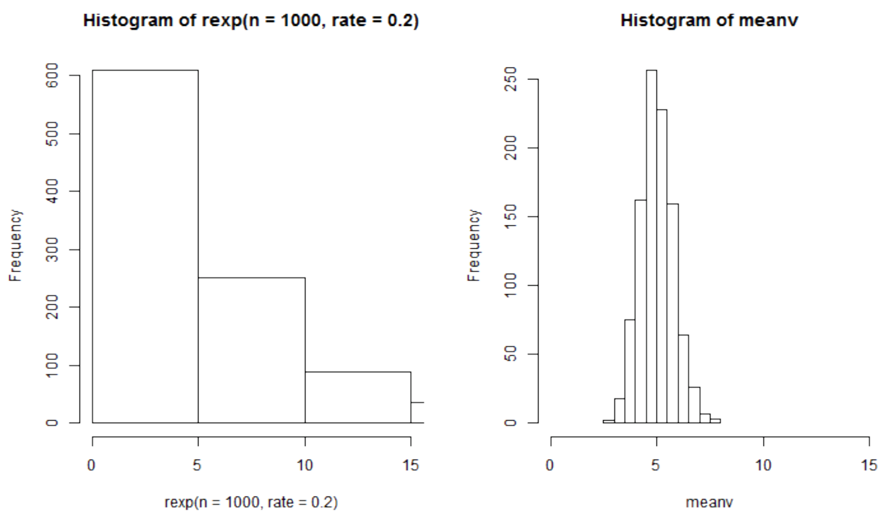
```
## [1] 0.6004928
```

We can see that the variability is close to the theoretical value.

Show that the distribution is approximately normal.

We can show that the distribution is approximately normal by comparing the distribution of means with the distribution of random exponentials.

```
par(mfrow=c(1,2))
q1<-hist(rexp(n=1000,rate=0.2),xlim=c(0,15))
q2<-hist(meanv,xlim=c(0,15))
```



We can see that the graph on the left is clearly not Gaussian while the exponential variables are much more so.