Central Limit Theorem on an Exponential Distribution

Justin Pizzino

July 5, 2019

1. Summary:

Below I will prove the central limit theorem and show how sampling a population can create a normal distribution for data that may not be normal.

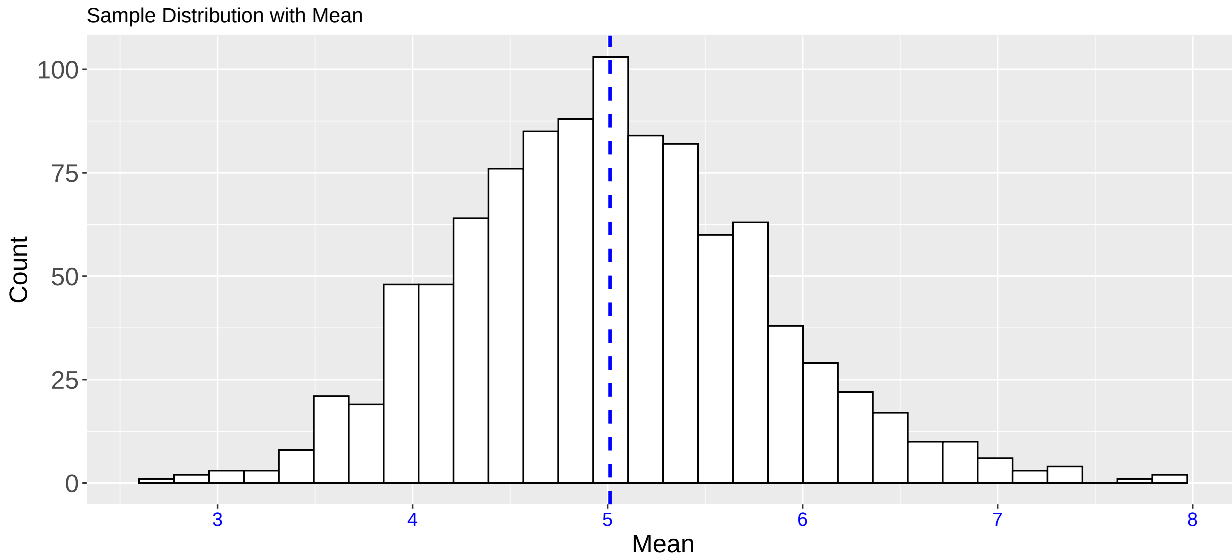
2. Create the simulation & plot the results

I used a for loop to run the rexp function to calculate 40 sample size simulation. I ran this 1000 times and added the means to a data table.

```
set.seed(8534)
rate<-0.2
n<-40
intervals<-1000
sim<-NULL
for (i in 1 : intervals) sim = c(sim, mean(rexp(n,rate)))


sim<-data.table(sim)
theoMean<-(1/rate)
theoVar<-theoMean^2/n
sampMean<-mean(sim$sim)
```

Below is a chart showing the mean of the 1000 simulations. The mean is the sample is near 5 which is close to our theoretical mean.



3. Now to compare sample variance & mean to the theoretical variance & mean, as well as review confidence levels

Confidence interval: c(4.96366993834471, 5.06164361678079)

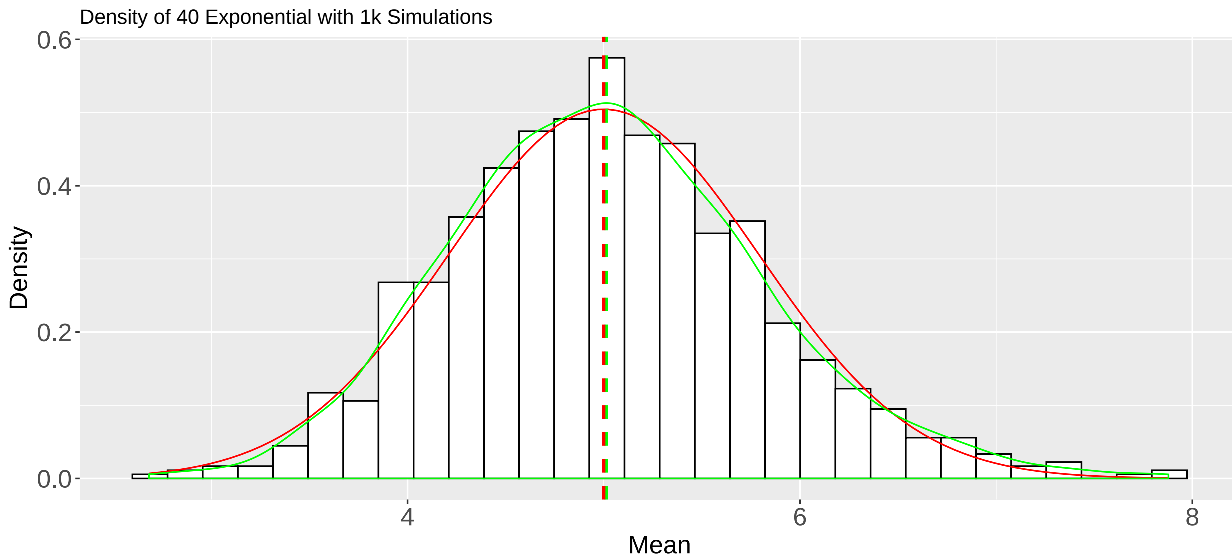
The sample mean is: 5.0126568 (also shown in blue below), which is very close to the theoretical mean of 1/lamda or 1/.2 = 5.

Theoretical variance is mean^2/n so in this case 0.04/40: 0.625

The sample variance using r function var() is: 0.6231745 These values are very close, and the graph below shows how close the two distributions are (red and green curves)

4. Distributions

Let's look at the distribution to show it is normal. As can be seen below the green curve appears symmetric bell shaped, with ~68% of the values with in 1 standard deviation.



The red shows sample mean and distribution, green shows theoretical. As Central Limit Theorem states that sample means of simple random elements from a large (infinite) population will take on a bell shape. It is acceptable to have 30 or more in a sample and we have 40 samples by 1000 simulations. As you can see here our distribution is bell shaped and is close to the theoretical.

5. Does the distribution means of 40 exponential behave as predicted by he Central Limit Theorem?

CLT states - the distribution of averages of iid variables (properly normalized) becomes that of a standard normal as the sample size increases.

Yes, based on the simulation run, the comparison of theoretical to sample, and charts displayed, the samples do behave as predicted by CLT. The sample means as charted appear normal. **Also the sample and theoretical means and variances are close to each other as predicted.**

6. Appendix

Plot 1 code

```
g<-ggplot(sim, aes(x=sim))
g<-g + geom_histogram(color="black",fill="white", bins=30)
g<-g + geom_vline(aes(xintercept=mean(sim)),
                  color="blue",linetype="dashed",size=1)
g<-g + theme(axis.text.x=element_text(size=12, color="#0000ff"))
g<-g+ scale_x_continuous(breaks=c(3:8))
g<-g+labs(title="Sample Distribution with Mean", x="Mean",y="Count")
print(g)
```

Plot 2 code

```
g2<-ggplot(sim, aes(x=sim))
g2<-g2 + geom_histogram(aes(y=..density..),color="black",fill="white", bins=30)
g2<- g2 + stat_function(fun=dnorm, args = list(mean=theoMean, sd=sqrt(theoVar)),
                        color="red")
g2<-g2 + labs(title="Density of 40 Exponential with 1k Simulations", x="Mean",
              y="Density")
g2<-g2+geom_density(color="green")
g2<-g2 + geom_vline(aes(xintercept=mean(sim)),
                  color="green",linetype="dashed",size=1)
g2<-g2 + geom_vline(aes(xintercept=theoMean),
                  color="red",linetype="dashed",size=1)
g2<-g2+theme(legend.position="right")
print(g2)
```