Comparison of Exponential Distribution with CLT

Nima Taghidoost

1/18/2020

In this project we will have 1000 simulations which each one has 40 exponentials.

Samples

The lambda is equal to 0.2 , the size of each simulation is 40 and we are going to simulate 1000 times:

```
lambda <- 0.2
Size <- 40
Sim <- 1000
```

Now we can simulate 1000 times and calculate the mean of each simulation

```
AllSamples <- matrix(data=rexp(Size*Sim,lambda),Sim)
MeansOfSamples <- apply(AllSamples,1,mean)

MeansOfSamples <- data.frame(MeansOfSamples)
names(MeansOfSamples) <- "Mean"
```

Comparing means

We calculate the mean from the means that we have and from CLT

```
SampleMean <- mean(MeansOfSamples$Mean)

RealMean <- 1/lambda
```

```
SampleMean
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
## [1] 4.990025
```

```
RealMean
```

```
## [1] 5
```

They are too close.

Comparing variances

And here is the variance from the means that we have and from CLT

```
SampleVar <- var(MeansOfSamples$Mean)

RealVar <- ((1/lambda)^2)/Size
```

```
SampleVar
```

```
## [1] 0.6177072
```

```
RealVar
```

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

They are also too close.

Confidence Intervals

Now we can calculate the intervals from samples that we have and from CLT

```
SampleMean + (c(-1,1)*qnorm(0.975)*sqrt(SampleVar/Size))
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
## [1] 4.746463 5.233587
```

```
RealMean + (c(-1,1)*qnorm(0.975)*sqrt(RealVar/Size))
```

```
## [1] 4.755005 5.244995
```

They are approximately the same.

Plotting the distribution of samples and CLT

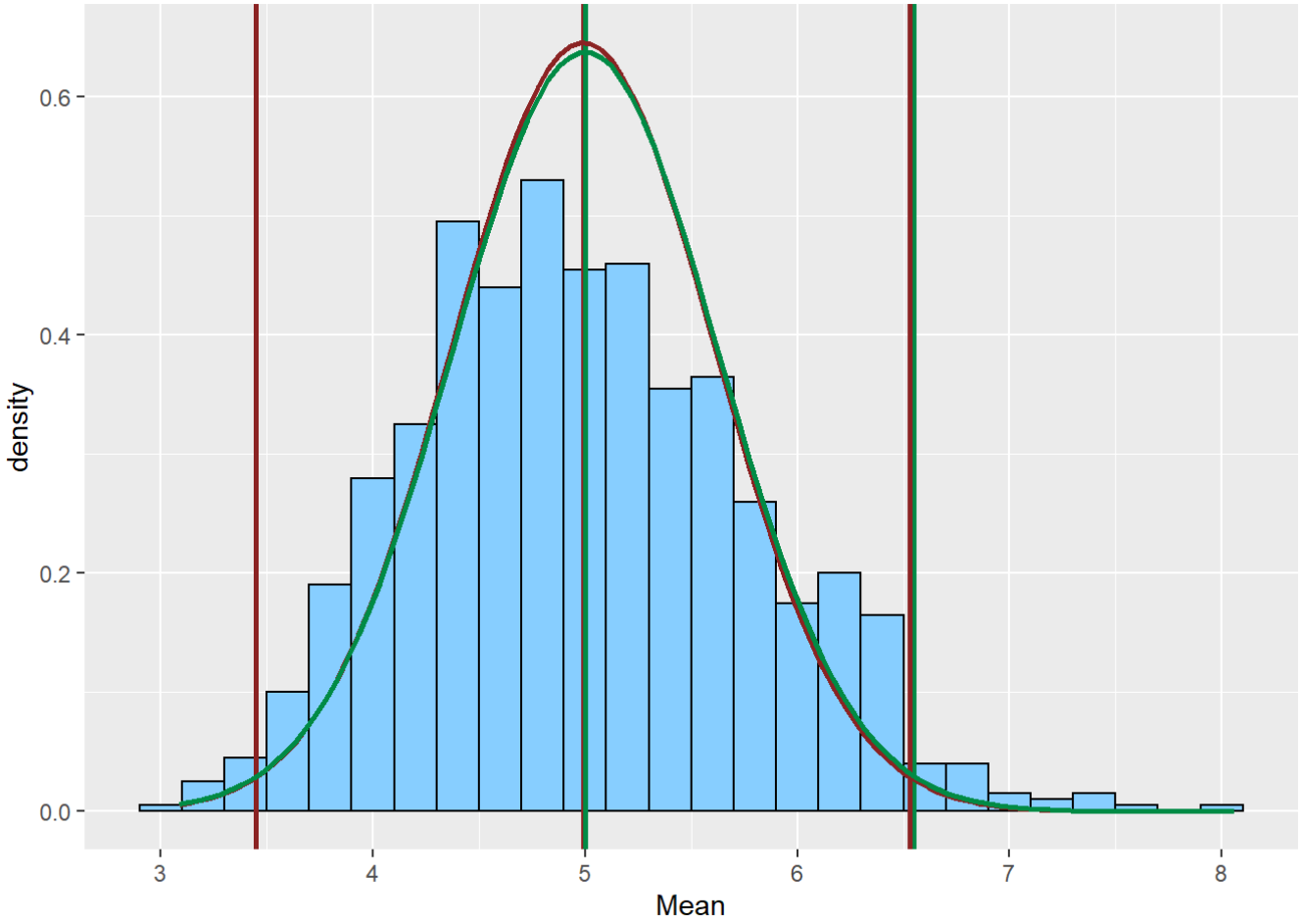
Now we can plot both to see the similarity:

```
RealLI <- RealMean + (-1*qnorm(0.975)*sqrt(RealVar))
RealHI <- RealMean + (1*qnorm(0.975)*sqrt(RealVar))


SampleLI <- SampleMean + (-1*qnorm(0.975)*sqrt(SampleVar))
SampleHI <- SampleMean + (1*qnorm(0.975)*sqrt(SampleVar))


ggplot(MeansOfSamples,aes(x=Mean))+
  geom_histogram(aes(y=..density..),binwidth = 0.2,color="black",fill="skyblue1" )+
  geom_vline(aes(xintercept= mean(MeansOfSamples$Mean)),color="brown4",size=1)+
  geom_vline(aes(xintercept= mean(RealMean)),color="springgreen4",size=1)+
  geom_vline(aes(xintercept= RealLI),color="springgreen4",size=1)+
  geom_vline(aes(xintercept= RealHI),color="springgreen4",size=1)+
  geom_vline(aes(xintercept= SampleLI),color="brown4",size=1)+
  geom_vline(aes(xintercept= SampleHI),color="brown4",size=1)+
  stat_function(fun = dnorm, args = list(mean = SampleMean,
                sd = SampleVar), colour = "brown4",size=1)+
  stat_function(fun = dnorm, args =list(mean = RealMean,
                sd = RealVar), colour = "springgreen4",size=1)
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD



The red lines are the data from samples and the green lines are the data from CLT. We can clearly see that their mean,confidence intervals and distributions are very very close to each other and conclude that the distribution demonstrates CLT.

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD