Course 6 Projects

Rohit Jain

9/11/2018

Part1:Simulation Exercise

In this part of the project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem.

1. Show the sample mean and compare it to the theoretical mean of the distribution.

We are capturing the 1000 values of mean of the "random exponentials distribution for 40 observations and lambda = 0.2". we have set the seed to 123478 in R to get a reproducible result every time we knit the Rmd file.

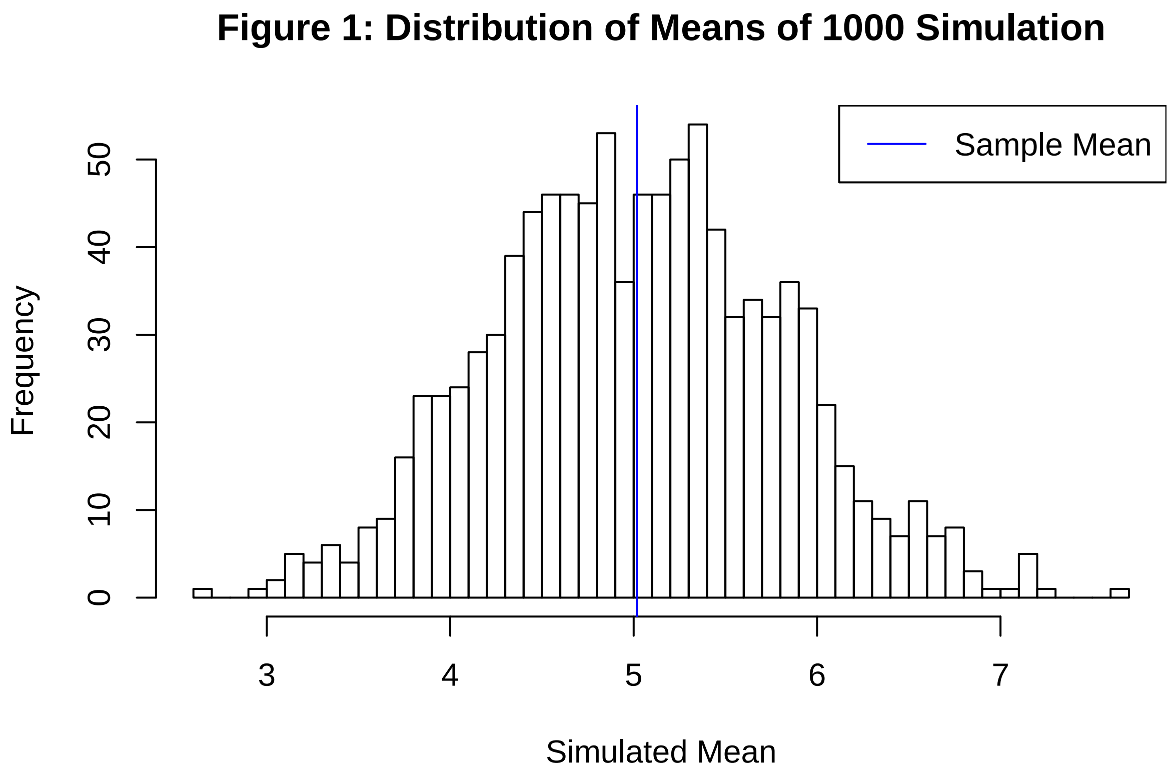
Initializing simulation data:

```
sampleSize <- 40
lambda <- 0.2
```

Simulating:

```
set.seed(123478)
simP1 <- vector()
for (i in 1:1000) {
  simP1 <- c(simP1, mean(rexp(sampleSize, 0.2)))
}
```

Below is the distribution of mean values captured from the simulation and the Blue line is showing the mean of



the simulated data.

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
library(knitr)
theoreticalValues <- c(1/lambda, (1/0.2)^2/sampleSize)
sampleValues <- c(mean(simP1),var(simP1))
comData <- rbind(theoreticalValues,sampleValues)
colnames(comData) <- c("Mean","Variance")
```

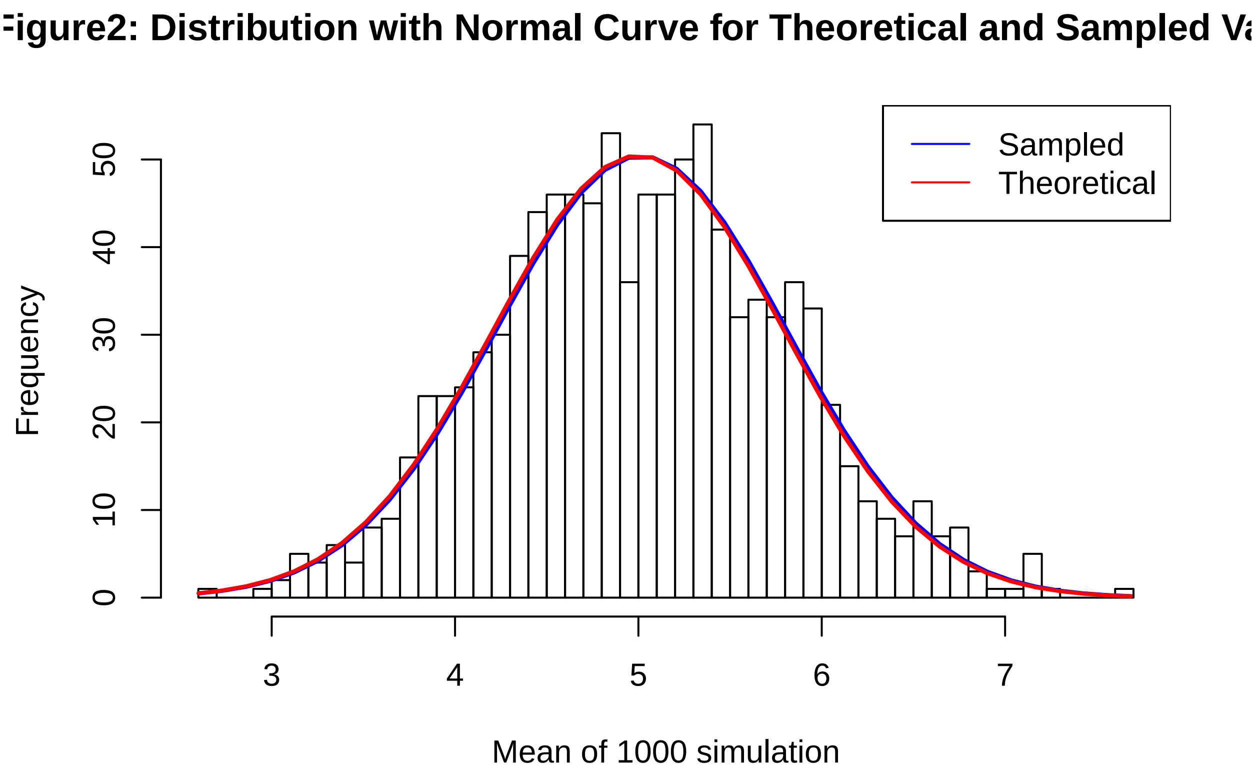
Below is the table in which we have compared the theoretical and sample, mean and variance.

Table 1: Table1 : Comparision for the mean and variance of sampled and theoretical values

	Mean	Variance
theoreticalValues	5.000000	0.6250000
sampleValues	5.017829	0.6267851

3. Show that the distribution is approximately normal.

The Blue and Red curves are for the sample and theoretical data repectively is fitted on the Simulated data distribution which is showing that the normal curve is perfectly fitted on the distribution.



Appendix 1: Code for Tables and Figures:

```
# Code of Figure1

hist(simP1, breaks=40, main = "Figure 1: Distribution of Means of 1000 Simulation", xlab = "Simulated Me
abline(v=mean(simP1), b = 90 , col = "blue")
legend('topright', c("Sample Mean"),
       col=c("blue"), lty=c(1))
```

```
# Code of Table1

kable(comData, caption = "Table1 : Comparision for the mean and variance of sampled and theoretical valu
```

```
#Code for Figure2

x <- simP1
h<-hist(x, breaks=40, main="Distribution with Normal Curve for Theoretical and Sampled Values", xlab =
xfit<-seq(min(x),max(x),length=40)
yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)

yfit1 <-dnorm(xfit,mean=1/0.2,sd=sqrt((1/0.2)^2/40))
yfit1 <- yfit1*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)
lines(xfit, yfit1, col= "red", lwd = 2)
legend('topright', c("Sampled", "Theoretical"),col=c("blue", "red"), lty=c(1,1))
```