Statistical Inference Course Project Part 1

Yan

5/8/2018

Part 1: Simulation Exercise Instructionsless

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. 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 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Set simulations

Run exponential distributions 1000 times, with lambda = 0.2 and n = 40

```
#The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate param
set.seed(123456)
# set lambda and n
lambda <- 0.2
n <- 40
# initialize
mns=NULL
vars=NULL
for (i in 1 : 1000) {
  edist <- rexp(n,lambda)
  mns = c(mns, mean(edist))
}
```

Sample mean vs Theretical mean

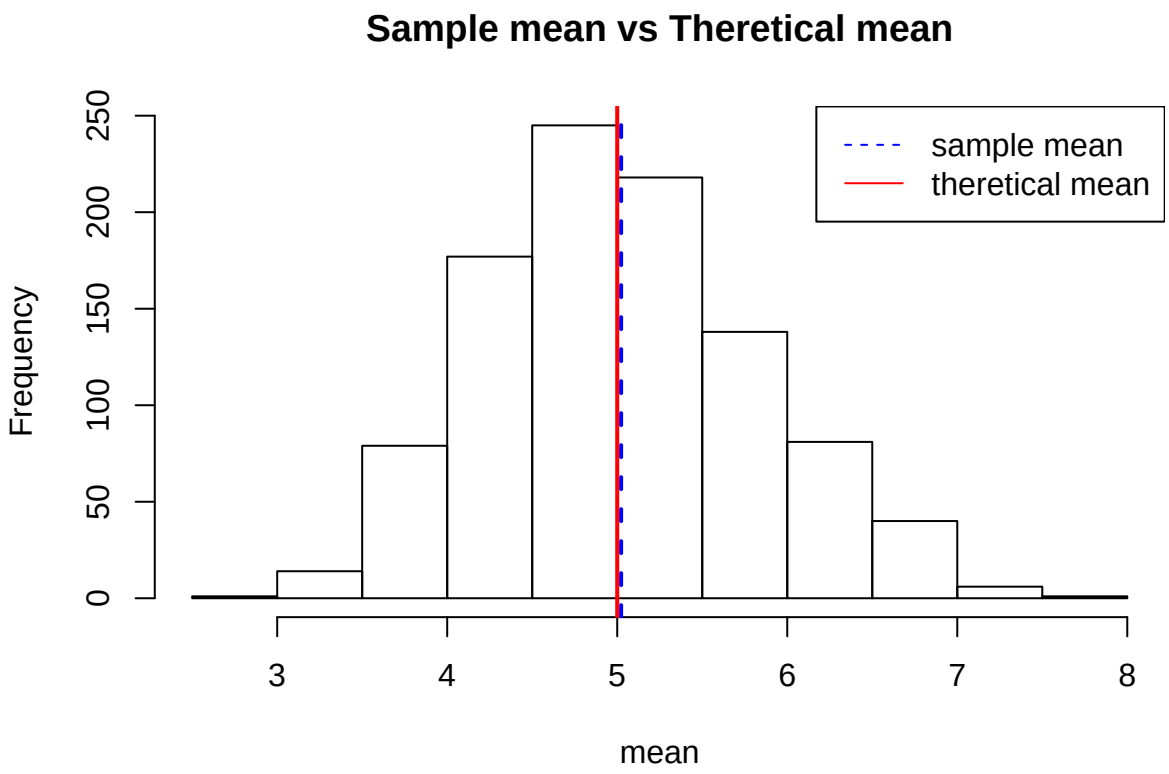
```
# theretical mean is 1/lambda
tmn <- 1/lambda
tmn
```

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

```
# sample mean
smn <- mean(mns)
smn
```

```
## [1] 5.022915
```

```
# make plot for the mean of 1000 averages of 40 random exponential distrubutions, and compare sample me
hist(mns, xlab = "mean",main = "Sample mean vs Theretical mean")
abline(v=smn,col="blue",lwd=2,lty=2)
abline(v=tmn,col="red",lwd=2,lty=1)
legend("topright",lty = c(2,1),col=c("blue","red"),legend=c("sample mean","theretical mean"))
```



Sample Variance vs Theoretical Variance

```
# theretical variance : standard deviation is 1/lambda
tvar <- (1/lambda)^2/n
tvar
```

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

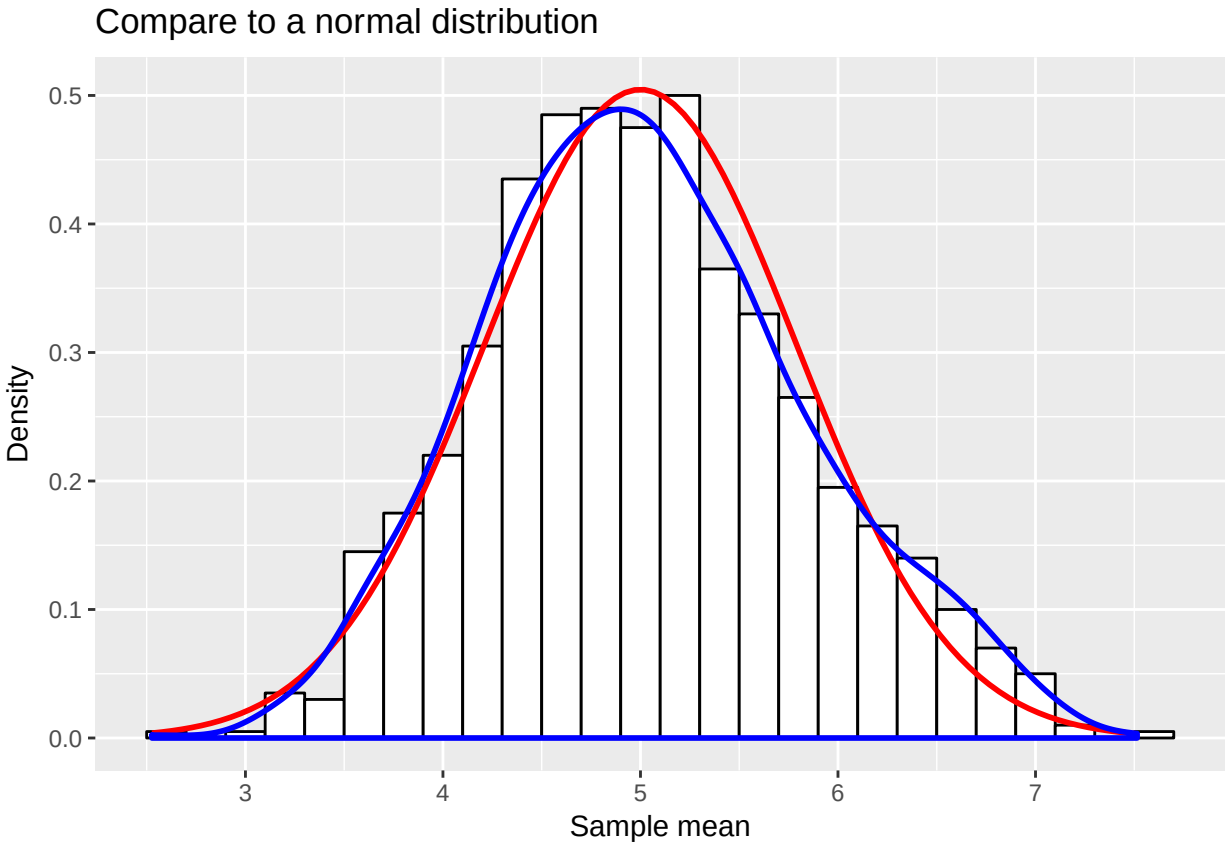
```
# sample variance
svar <- var(mns)
svar
```

```
## [1] 0.6570391
```

Distribution: normal?

Use ggplot2 to do this plot

```
mnsplot <- data.frame(mns)
g1 <- ggplot(data=mnsplot,mapping = aes(x=mns))  +
  geom_histogram(aes(y=..density..),binwidth = 0.2, color = "black", fill = "white") +
  stat_function(fun = dnorm, color = "red",size=1, args = list(mean=tmn,sd=sqrt(tvar))) +
  geom_density(color="blue", size=1) +
  ggtitle("Compare to a normal distribution") +
  xlab("Sample mean") + ylab("Density")
g1
```



We have checked that the mean and variance of the sample distributions are close to those for normal distribution as shown in the figure above. So this distribution is approximately normal.

Note: Normal Q-Q plot can also be used to check if the distribution is normal or not. Details can be refered to the links at the end.