Part1: Simulation Exercise

Haonan

October 26, 2016

Overview

This is a simulation practice which we would like to compare the simulation means and variances with the theoratical means and variances of a certain distribution. The distribution is an exponential distribution with lamda equals 0.2, 40 simulations will make up a sample and totally there are 1000 samples. This simulation will produce mean and variance plots and verify the assumption that the mean and the variance distribution is approximately normal.

Set Up the Simulation

We use the code as follows to set up the simulation results:

```
setwd("C:/Users/Haonan/Desktop/Study/R/statistical inferring project")

library(ggplot2)

random <- data.frame()
for (i in 1:1000){
  a <- rexp(40,.2)
  meanr <- mean(a)
  stdr <- sd(a)
  random[i,1] <- meanr
  random[i,2] <- stdr
}
colnames(random) <- c("Mean","Std_Dev")

head(random)
```

```
##       Mean  Std_Dev
## 1 4.445450 4.079561
## 2 4.407662 3.950253
## 3 5.236816 6.142477
## 4 5.113871 5.427024
## 5 5.481151 6.559435
## 6 5.129172 4.418907
```

Comparison

Sample Means

The theoretical mean of exponential distributions is 1/lamba, which, in this case, was 1/0.2 = 5 Firstly, it is reasonable to conduct a t test to see whether there's a significant difference between the sample mean and the theoretical mean

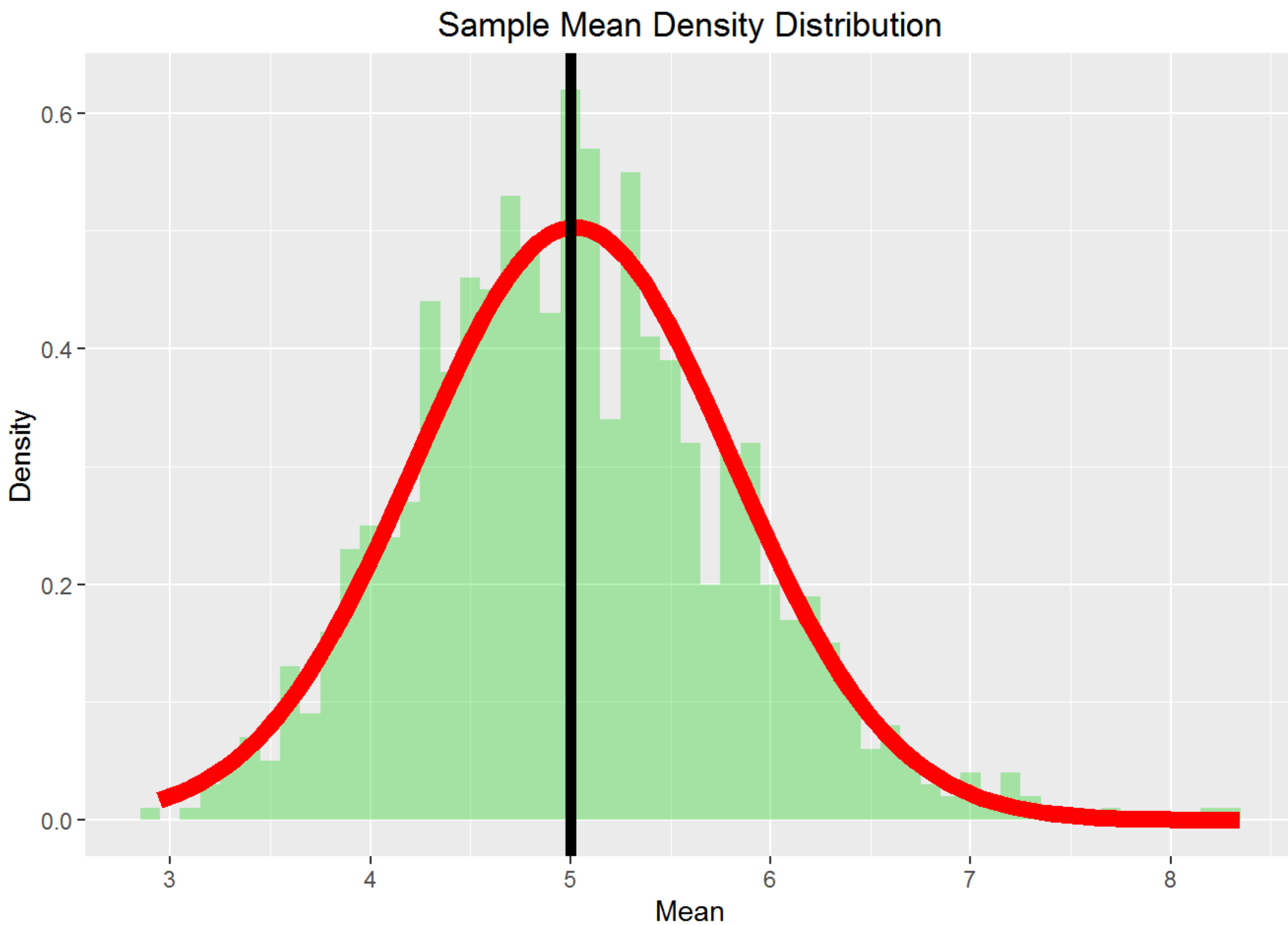
```
t.test(random$Mean, mu = 5)
```

```
##
##  One Sample t-test
##
## data:  random$Mean
## t = 0.74452, df = 999, p-value = 0.4567
## alternative hypothesis: true mean is not equal to 5
## 95 percent confidence interval:
##  4.969461 5.067879
## sample estimates:
## mean of x
##   5.01867
```

From the result we could clearly identify that the p-value is very large. This means there's no significance between the sample mean and the theoretical mean, they're equal.

Then we graph it, add a vertical line which indicates the theoretical mean = 5, and finally add a normal line as follows:

```
gg1 <- ggplot(random, aes(Mean)) + geom_histogram(alpha = .3, fill = 3, binwidth = .1, aes(y =
..density..))+ stat_function(fun=dnorm, size = 3, color = "red", args=list(mean=mean(random$Mean),
 sd=sd(random$Mean))) + geom_vline(xintercept = 5, size = 2)+labs(x = "Mean", y = "Density", tit
le = "Sample Mean Density Distribution")
gg1
```



From the plot, the sample mean is almost normally distributed with mean = 5. Compared with the red line, the sample mean is approximately normal

Sample Variance

The theoretical variance of exponential distributions is 1/lamba^2, which, in this case, was 1/0.2^2 = 25 Firstly, it is reasonable to conduct a t test to see whether there’s a significant difference between the sample mean and the theoretical mean

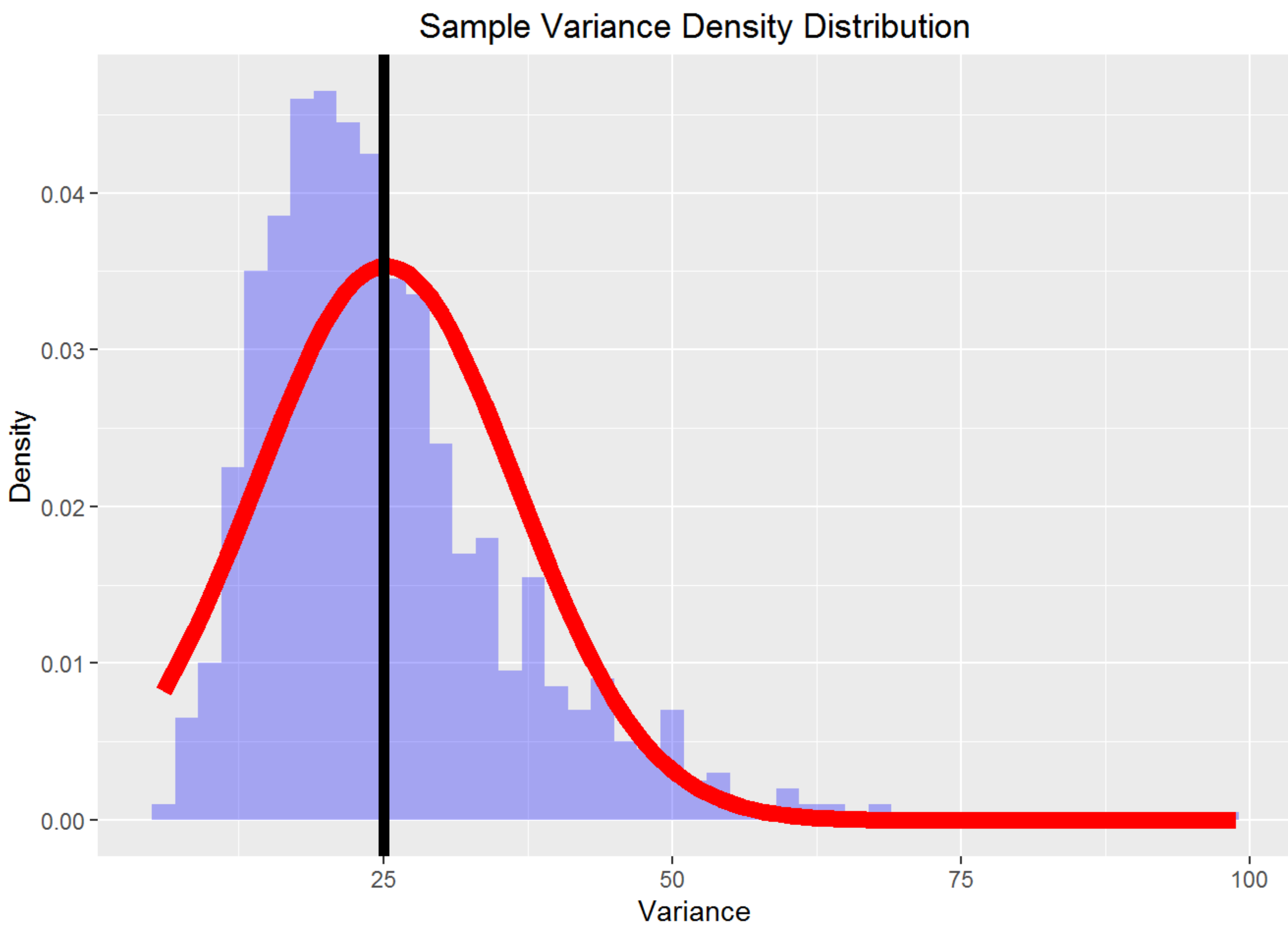
```
with(random, t.test(Std_Dev^2, mu = 1/.04))
```

```
##
##  One Sample t-test
##
## data:  Std_Dev^2
## t = 0.6666, df = 999, p-value = 0.5052
## alternative hypothesis: true mean is not equal to 25
## 95 percent confidence interval:
##  24.53738 25.93860
## sample estimates:
## mean of x
##  25.23799
```

From the result we could clearly identify that the p-value is quite large. This means there's no significance between the sample variance and the theoretical variance, they're equal

Then we graph it, add a vertical line which indicates the theoretical variance = 25, and finally add a normal line as follows:

```
gg2 <- ggplot(random, aes(Std_Dev^2)) + geom_histogram(alpha = .3, fill = 4, binwidth = 2, aes(y
 = ..density..))+ stat_function(fun=dnorm, size = 3, color = "red", args=list(mean=mean(random$S
td_Dev^2), sd=sd(random$Std_Dev^2)))+ geom_vline(xintercept = 25, size = 2) + labs(x = "Varianc
e", y = "Density", title = "Sample Variance Density Distribution")
gg2
```



From the plot, the sample variance seems like normally distributed with mean = 25, but still, a little skewed to the right. Compared with the red line, the sample variance distribution is approximately normal

Conclustsion

The simulation exercise has produced the desired results that the sample mean and the sample variance is nearly normal