Statistical Inference-Simulation Exercise

Overview

This project investigates the averages of 40 exponentials. The investigation involves determining the sample mean, sample variance and comparing them to the theoritical mean and variance and finally determining if the distribution is normal. The objective of this analysis to determine if the exponential distribution is normal bu comparing the mean and variances. I will be conducting 1000 simulations to achieve this goal.

Environment

1. OS: Windows 10
2. Tool: R version 3.2.3; R Studio version 0.99.491; R packages used: knitr, ggplot2; Publishing tool: RPubs, PDF
3. Data: Simulated exponential distribution
4. Analyst: Uma Venkataramani; Date of Analysis: Feb 2016
   ##Simulations ### Simulating the data

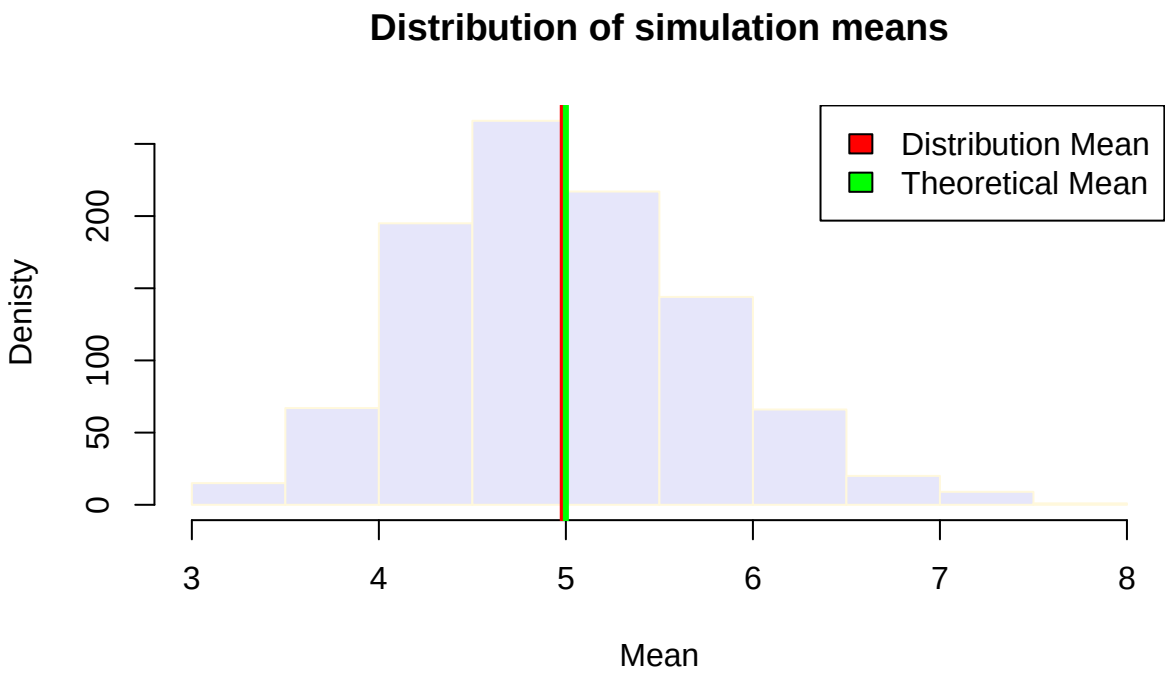
```
#libraries
    library(knitr)
    library(ggplot2)
#Set variables for simulations> sim; number of exponentials>n; lambda>lambda
    sim=1000
    n=40
    lambda=0.2
#Simulate and clculate the mean for 1000 simulations of 40 exponentials
    set.seed(7)
    simexp=replicate(sim,rexp(n,lambda))
    simmean=apply(simexp,2,mean)
```

Determine the center of the distribution and compare it to the theoretical mean

```
#Distribution mean
    distmean=mean(simmean)
#Theoetical mean
    theomean=1/lambda
#Comparing distribution mean and theoretical mean
    meandiff=theomean-distmean
    cat("Distribution mean=",round(distmean,3),", Theoretical mean=",round(theomean,3),", Difference= "
```

```
## Distribution mean= 4.983 , Theoretical mean= 5 , Difference=  0.017
```

```
#Plotting the means of the simulations
    hist(simmean, col="lavender",ylab="Denisty",main="Distribution of simulation means",xlab="Mean",bor
    abline(v=distmean, col="red",lwd=3)
    abline(v=theomean, col="green",lwd=3)
    legend("topright",legend=c("Distribution Mean","Theoretical Mean"),fill=c("red","green"))
```



Analysis

Comparing the two means it is clear the distribution mean and theoretical mean are similarly centered with a marginal difference
Determine the variance of the distribution and compare it to the theoretical variance

```
#Distribution mean
    distsd=sd(simmean)
    distvar=distsd^2
#Theoetical mean
    theosd=theomean/sqrt(n)
    theovar=theosd^2
    vardiff=theovar-distvar
#Comparing distribution mean and theoretical mean
    cat("Distribution var=",round(distvar,3),", Theoretical var=",round(theovar,3),", Difference = ", r
```

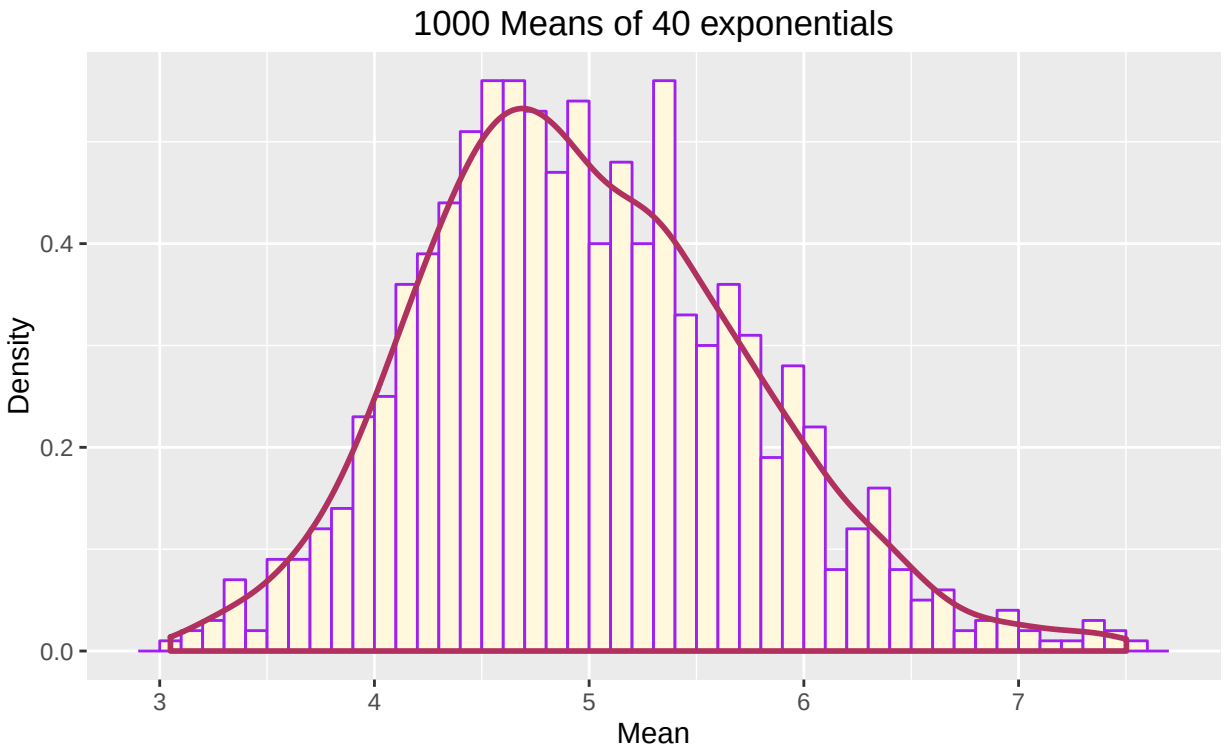
```
## Distribution var= 0.579 , Theoretical var= 0.625 , Difference =  0.046
```

Analysis

Comparing the two variances it is clear the distribution variance and theoretical variance are more or less similar in distribution with a small difference
Determining if the simulated distribution is normal

```
#Plotting the simulated means
    x=ggplot(data.frame(simmean), aes(x=simmean))
    x=x+geom_histogram(aes(y=..density..),color="purple",fill="cornsilk",binwidth=0.1)
```

```
x<-x+scale_fill_hue(c=45, l=80)
x<-x+xlab("Mean")
x<-x+ylab("Density")
x<-x+ggtitle("1000 Means of 40 exponentials")
x= x+geom_density(colour="maroon", size=1)
x
```



Analysis

As we can see from plot, the distribution is close to a bell curve with left and right side of the mean more or less equally distributed. We may be able to conclude that the distribution is normal.
Method applied for this analysis 1. Generated 1000 simulations of a random sample of 40 exponentials. Each simulation will be referred to as a set and determined the mean of each set of simulation and stored it in a matrix. 3. Determined the distribution mean by calculting the mean of the means of all the 1000 sets and determined the theoretical mean given by 1/lambda 5. Compared the distribution mean and theoretical mean and determined that they lie quite close together 6. Determined the distribution standard deviation by calculting the standard deviation of the means of all the 1000 sets and determined the theoretical variance by squaring the distribution standard deviation 8. Determined the theoretical standard deviation given by (theoretical mean)/sqrt(n) and determined the theoretical variance by squaring the theoretical standard deviation 10.Compared the distribution variance and theoretical variance and determined that they are more or less similar 11.Finally plotted the means of the all the simulations and determined that they may be considered to be normally distributed
###Summary Comparing the distribution mean and theoretical mean, it is clear the distribution mean and theoretical mean are similarly centered with a marginal difference. Comparing the distribution variance and theoretical variance, t is clear the distribution variance and theoretical variance are more or less similar in distribution. The distribution is close to a bell curve with left and right side of the mean more or less equally distributed. We may be able to conclude that the distribution is normal.