Exponential Distribution in R Compared to Central Limit Theorem

By Aiman D.

Overview

Part 1 of the project 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, λ) where λ is the rate parameter. The mean of exponential distribution is 1/λ and the standard deviation is 1/λ. We will set λ = 0.2 for all of the simulations and investigate the distribution of averages of 40 exponentials.

Simulations

To compare, we will execute 1000 simulations that has 40 exponentials with λ = 0.2 and and apply the function rexp(n, λ).

The following code was executed:

```
NumSims = 1000;
n = 40;
Lambda = 0.2

Means <- vector("numeric")
Sum_of_Means <- vector("numeric")
Cum_of_Means <- vector("numeric")

for (i in 1:NumSims)
{
  Means[i] <- mean(rexp(n, Lambda))
}
Sum_of_Means <- Means[1]

for (i in 2:NumSims)
{
  Sum_of_Means[i] <- Sum_of_Means[i-1] + Means[i]
}

for (i in 1:NumSims)
{
  Cum_of_Means[i] <- Sum_of_Means[i]/i
}

print(sprintf("The Means of the sample equal: %f", Cum_of_Means[NumSims]))
print(sprintf("The theoretical mean is equal to: %f", 1/Lambda))
```

This was the output:

```
[1] "The Means of the sample equal: 5.003542"
```

```
[1] "The theoretical mean is equal to: 5.000000"
```

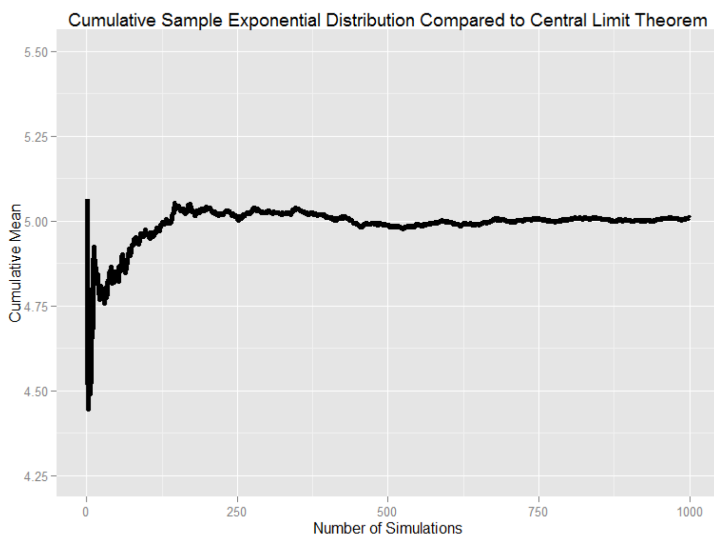
The end of the simulations was very close to the theoretical mean as the plot will also show after running the following code:

```r
library(ggplot2)
NumSims = 1000;
n = 40;
Lambda = 0.2

Means <- vector("numeric")
Sum_of_Means <- vector("numeric")
Cum_of_Means <- vector("numeric")

g<-ggplot(data.frame(x = 1:NumSims, y = Cum_of_Means), aes(x = x, y = y))
g<-g+geom_hline(yintercept = 0) + geom_line(size = 1)
g<-g+scale_y_continuous(breaks=c(4.25, 4.50, 4.75, 5.00, 5.25, 5.50), limits =c(4.25, 5.5))
g<-g+theme(plot.title=element_text(size=12, face="bold", vjust=2, hjust=0. 5))
g<-g+labs(title="Cumulative Sample Exponential Distribution Compared to Central Limit Theorem")
g<-g+labs(x="Number of Simulations", y="Cumulative Mean")
print(g)
```

The code produced the following plot:



Sample variance compared to the theoretical variance of the distribution

```r
NumSims = 1000;
n = 40;
Lambda = 0.2

Means <- vector("numeric")

for (i in 1:NumSims)
{
  Means[i] <- mean(rexp(n, Lambda))
}

print(sprintf("The Sample Variance Means: %f", var(Means)*n))
print(sprintf("The theoretical variance is: %f", (1/Lambda)^ 2))
```

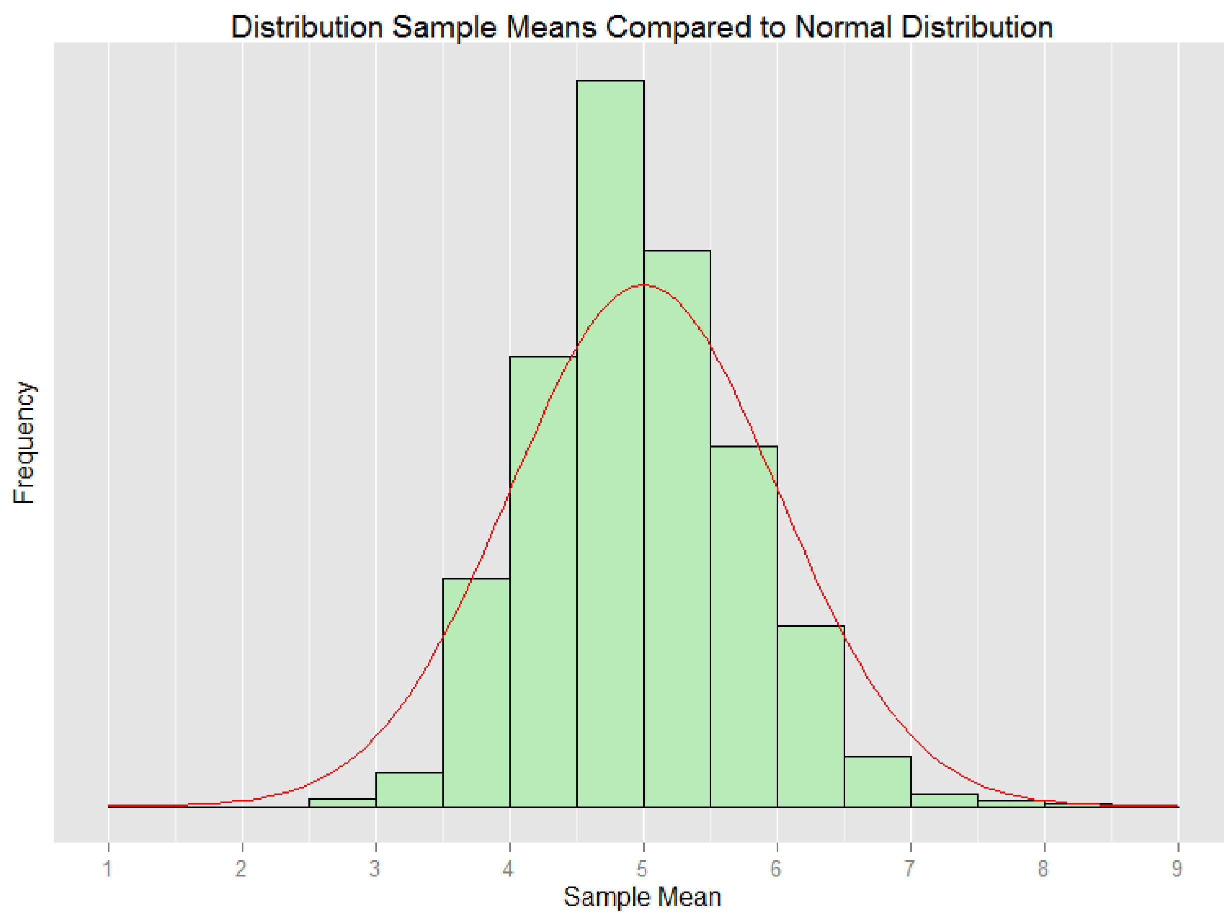
The output was also very close:

```
> print(sprintf("The Sample Variance Means: %f", var(Means)*n))
[1] "The Sample Variance Means: 23.834909"
> print(sprintf("The theoretical variance is: %f", (1/Lambda)^ 2))
[1] "The theoretical variance is: 25.000000"
```

Sample Distribution Means Compared to Normal Distribution

```
library(ggplot2)
Means <- vector("numeric")
for (i in 1:NumSims)
{
  Means[i] <- mean(rexp(n, Lambda))
}
g<-ggplot(data.frame(x = Means), aes(x = x ))
g<-g+geom_histogram(position="identity", fill="green", color="black", alpha= 0.2, binwidth=0.5, aes(y=..density..))
g<-g+stat_function(fun=dnorm, colour="red", args=list(mean=5))
g<-g+scale_x_continuous(breaks=c(1, 2, 3, 4, 5, 6, 7, 8, 9), limits=c(1, 9))
g<-g+scale_y_continuous(breaks=c())
g<-g+theme(plot.title=element_text(size=14, face="bold", vjust=2, hjust=0. 5))
g<-g+labs(title="Distribution Sample Means Compared to Normal Distribution")
g<-g+labs(x="Sample Mean", y="Frequency")
print(g)
```

The code produced the following plot:



The sample shows normal distribution around the mean of 5.

Statistical Inference Project Part 2

By Aiman D.

Overview

Part 2 of the project will analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
4. State your conclusions and the assumptions needed for your conclusions.

Description

The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

Usage

ToothGrowth

Format

A data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length

[,2] supp factor Supplement type (VC or OJ).

[,3] dose numeric Dose in milligrams.

Source

C. I. Bliss (1952) The Statistics of Bioassay. Academic Press.

References

McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley.

Data Loading and Analysis

```
> library(datasets)
> data(ToothGrowth)
> head(ToothGrowth)
   len supp dose
1  4.2   VC  0.5
2 11.5   VC  0.5
3  7.3   VC  0.5
4  5.8   VC  0.5
5  6.4   VC  0.5
6 10.0   VC  0.5
```

```
> str(ToothGrowth)
'data.frame':	60 obs. of  3 variables:
 $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
 $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
 $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Basic Summary of Data

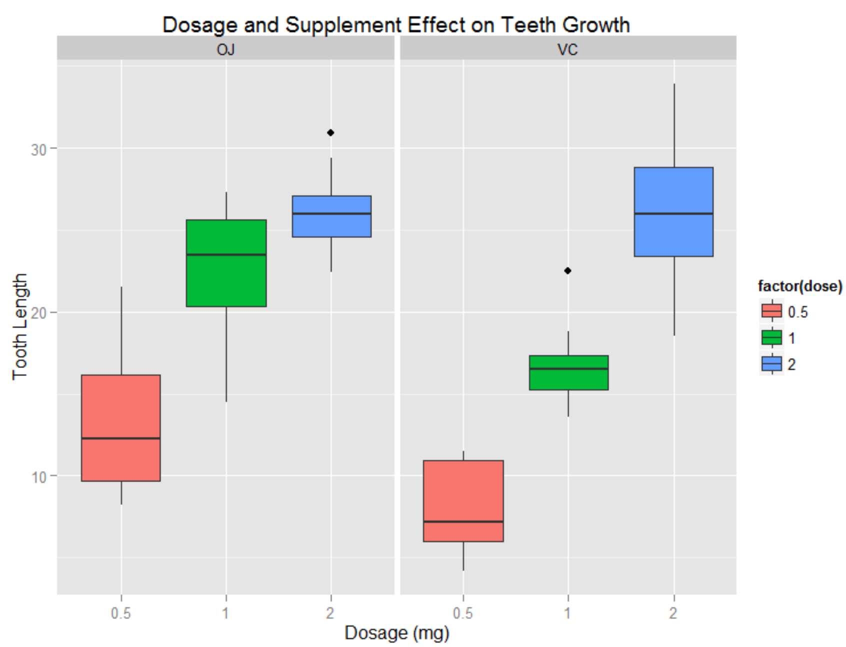
```
> ToothGrowth$dose <- as.factor(ToothGrowth$dose)
> table(ToothGrowth$supp, ToothGrowth$dose)

     0.5  1  2
  OJ  10 10 10
  VC  10 10 10
```

```
> summary(ToothGrowth)
      len        supp     dose   
 Min.   : 4.20   OJ:30   0.5:20  
 1st Qu.:13.07   VC:30   1  :20  
 Median :19.25           2  :20  
 Mean   :18.81                   
 3rd Qu.:25.27                   
 Max.   :33.90                   
> mean(ToothGrowth$len)
[1] 18.81333
> sd(ToothGrowth$len)
[1] 7.649315
```

Plotting Data

```
require(ggplot2)
plot <- ggplot(ToothGrowth, aes(x=factor(dose),y=len,fill=factor(dose)))
plot + geom_boxplot(notch=F) + facet_grid(.~supp) +
  scale_x_discrete("Dosage (mg)") +
  scale_y_continuous("Tooth Length") +
  ggtitle("Dosage and Supplement Effect on Teeth Growth")
```



Confidence Intervals and/or hypothesis tests to compare tooth growth by supp and dose

```
supp.t1 <- t.test(len~supp, paired=F, var.equal=T, data=ToothGrowth)
supp.t2 <- t.test(len~supp, paired=F, var.equal=F, data=ToothGrowth)
supp.result <- data.frame("p-value"=c(supp.t1$p.value, supp.t2$p.value),
                 "Low Conf"=c(supp.t1$conf[1],supp.t2$conf[1]),
                 "High Conf"=c(supp.t1$conf[2],supp.t2$conf[2]),
                  row.names=c("Equal Var","UnEqual Var"))
supp.result
```

```
               p.value   Low.Conf High.Conf
Equal Var   0.06039337 -0.1670064  7.567006
UnEqual Var 0.06063451 -0.1710156  7.571016
```

Conclusions

Analyzing the data and the plots:

1. The dosage seems to have the biggest effect on the growth if teeth.
2. OJ is better for teeth growth than VC at lower dosages (0.5 mg - 1 mg).
3. OJ and VC at the higher dosage (2 mg) seem to be statistically indifferent but with a slight edge for VC.