Analysis of Simulation Distribution

Brady Chiu

May 2, 2016

Contents

1 Intro

This project investigates the exponential distribution in R and compare it with the Central Limit Theorem. We use the rexp(n, lambda) to generate our distributions, and observe 1000 simulations of 40 samples each with a lambda = 0.2.

We aim to show the following:

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

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

3. Show that the distribution is approximately normal

2 Simulation

- We define our simulation paramters

```
exponentials<-40
lambda<-0.2
simulation_count<-1000
```

- We run our simulation

```r
simulation_result<-simulator(exponentials,lambda,simulation_count)
head(simulation_result,10)
```

```
##        mean
## 1  3.383281
## 2  6.246429
## 3  4.308285
## 4  4.914561
## 5  5.190411
## 6  5.820374
## 7  5.375244
## 8  5.005151
## 9  4.195586
## 10 4.707139
```

3 Mean

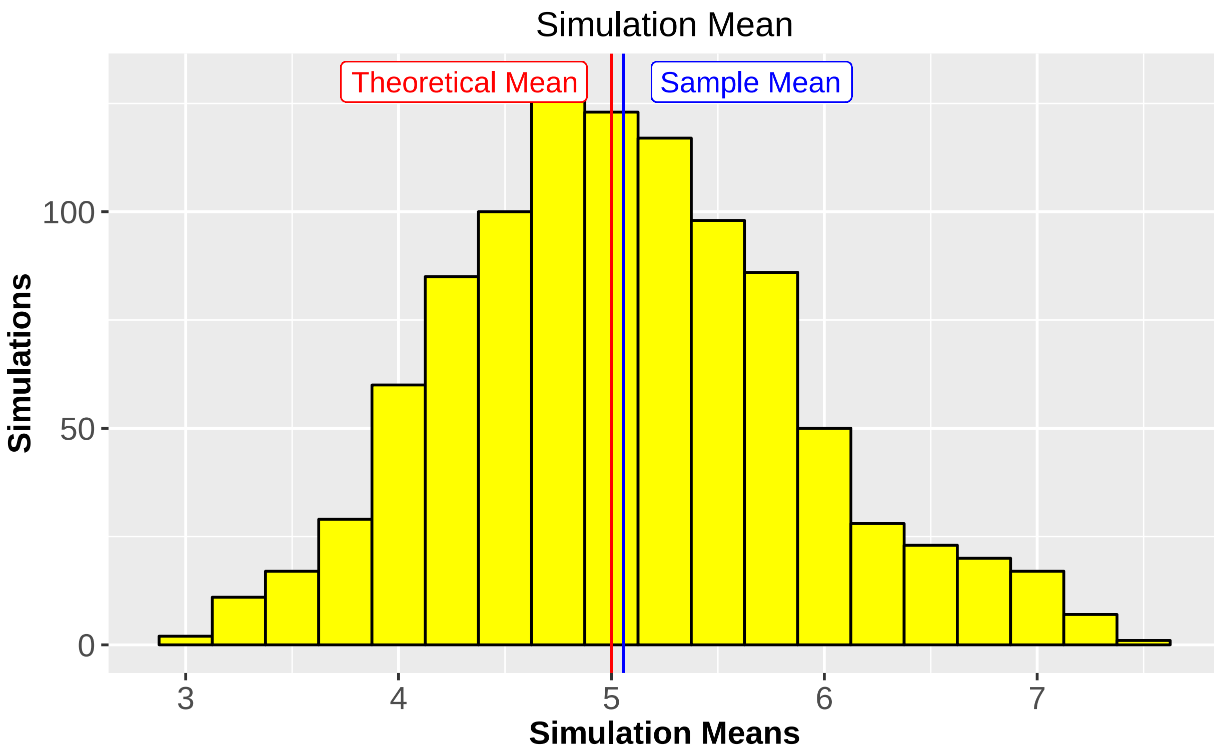
- We calculate the theoretical and sample means

```r
theoretical_mean<-1/lambda
sample_mean<-mean(simulation_result$mean)
paste("Theoretical Mean:",theoretical_mean)
```

```
## [1] "Theoretical Mean: 5"
```

```r
paste("Sample Mean:",sample_mean)
```

```
## [1] "Sample Mean: 5.05599525594939"
```



- Our theoretical and sample means are almost identical

4 Variance

- We calculate the theoretical and sample variances

```
theoretical_variance<-((1/lambda)^2)/simulation_count
sample_variance<-var(simulation_result$mean)
paste("Theoretical Variance:",theoretical_variance)
```

```
## [1] "Theoretical Variance: 0.025"
```

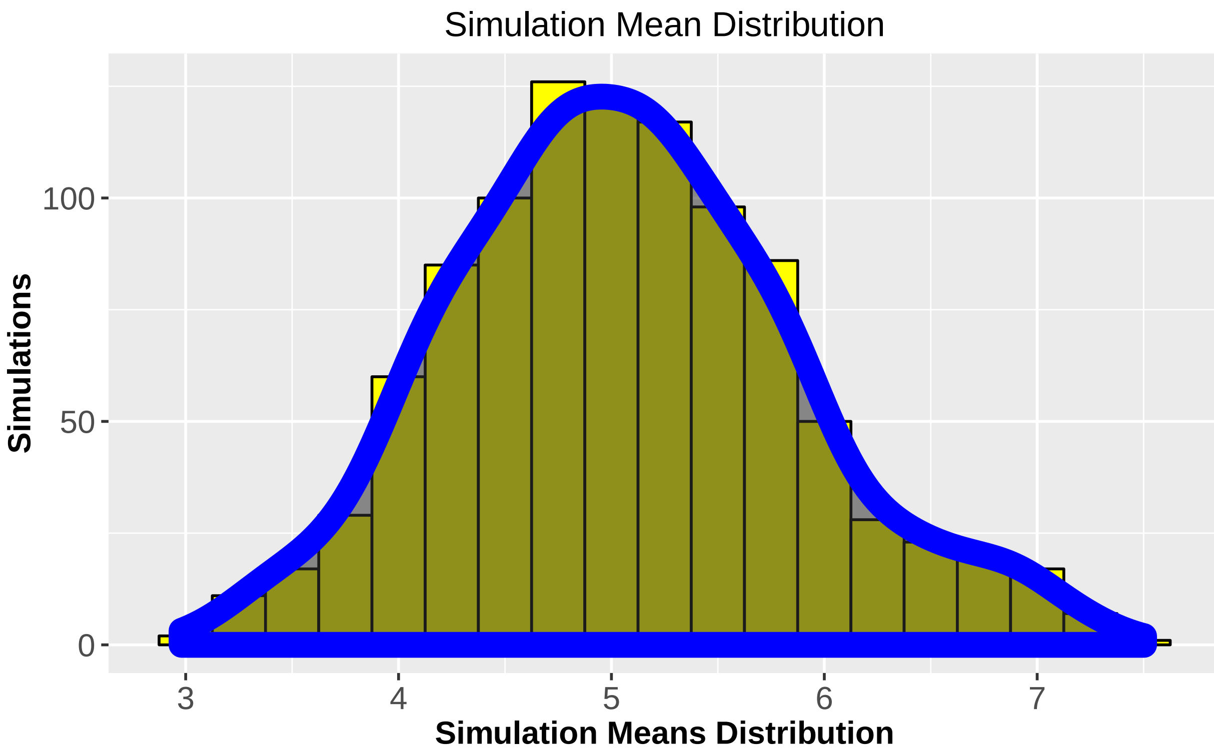
```
paste("Sample Variance:",sample_variance)
```

```
## [1] "Sample Variance: 0.654370303983734"
```

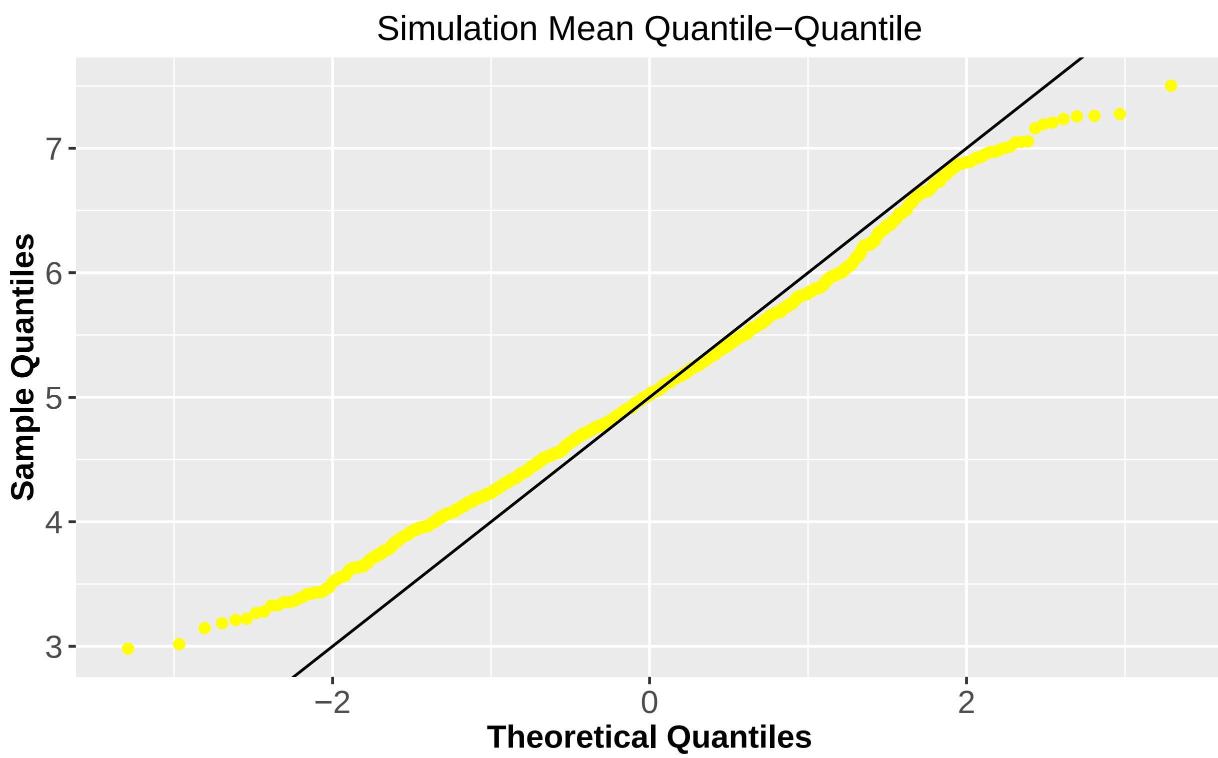
- Our theoretical and sample variances are almost identical

5 Distribution

- We show if our distribution is normal



- We see a shape very much like a gaussian normal distribution



- We see the Quantile-Quantile plot almost mirroring the y=x+b line, with a little skewage

6 Conclusion

Our analysis of 100 simulations of 40 samples each shows that the mean, variance and distrubtion of our sample is roughly equal to that of the population mean, variance, and distribution. This is consistent with the characteristic of the Central Limit Theorem.

7 Code

```
library(data.table, warn.conflicts=F)
library(dplyr, warn.conflicts=F)
library(ggplot2, warn.conflicts=F)
library(knitr, warn.conflicts=F)
library(lubridate, warn.conflicts=F)
library(tidyr, warn.conflicts=F)

# setwd("/Users/bradychiu/Dropbox (Uber Technologies)/R/Coursera/06_Statistical_Inference/stats_inferen
```

```
simulator<-function(n,lambda,simulation_count,seed=1337){
  set.seed(seed)
  simulation_means<-data.frame(mean=numeric(simulation_count))
  for(i in 1:simulation_count){
    simulation_means[i,]<-mean(rexp(n,lambda))
  }
  simulation_means
}
```

```
exponentials<-40
lambda<-0.2
simulation_count<-1000
```

```
simulation_result<-simulator(exponentials,lambda,simulation_count)
head(simulation_result,10)
```

```
theoretical_mean<-1/lambda
sample_mean<-mean(simulation_result$mean)
paste("Theoretical Mean:",theoretical_mean)
paste("Sample Mean:",sample_mean)
```

```
theoretical_variance<-((1/lambda)^2)/simulation_count
sample_variance<-var(simulation_result$mean)
paste("Theoretical Variance:",theoretical_variance)
paste("Sample Variance:",sample_variance)
```

```
ggplot(
  data=simulation_result
  ,aes(x=mean)
  )+
  geom_histogram(binwidth=0.25,color="black",fill="yellow")+
  geom_vline(aes(xintercept=theoretical_mean),color="red",show.legend=TRUE)+
```

```r
  geom_label(aes(label="Theoretical Mean",x=theoretical_mean,y=130,hjust=1.1),color="red")+
  geom_vline(aes(xintercept=sample_mean),color="blue",show.legend=TRUE)+
  geom_label(aes(label="Sample Mean",x=theoretical_mean,y=130,hjust=-0.2),color="blue")+
  ggtitle("Simulation Mean")+
  scale_x_continuous(name="Simulation Means")+
  scale_y_continuous(name="Simulations")+
  theme(
    axis.title=element_text(face="bold",size=12)
    ,axis.text=element_text(size=12)
    )
```

```r
ggplot(
  data=simulation_result
  ,aes(x=mean)
  )+
  geom_histogram(binwidth=0.25,color="black",fill="yellow")+
  stat_density(aes(y=0.25*..count..,alpha=0.9,size=0.5),color="blue")+
  ggtitle("Simulation Mean Distribution")+
  scale_x_continuous(name="Simulation Means Distribution")+
  scale_y_continuous(name="Simulations")+
  theme(
    axis.title=element_text(face="bold",size=12)
    ,axis.text=element_text(size=12)
    ,legend.position="none"
    )
```

```r
ggplot(
  data=simulation_result
  ,aes(sample=mean)
  )+
  geom_qq(color="yellow")+
  geom_abline(slope=1,intercept=5)+
  ggtitle("Simulation Mean Quantile-Quantile")+
  scale_x_continuous(name="Theoretical Quantiles")+
  scale_y_continuous(name="Sample Quantiles")+
  theme(
    axis.title=element_text(face="bold",size=12)
    ,axis.text=element_text(size=12)
    ,legend.position="none"
  )
```