illustrate Central Limit Theorem

Yinyan Guo

October 9, 2015

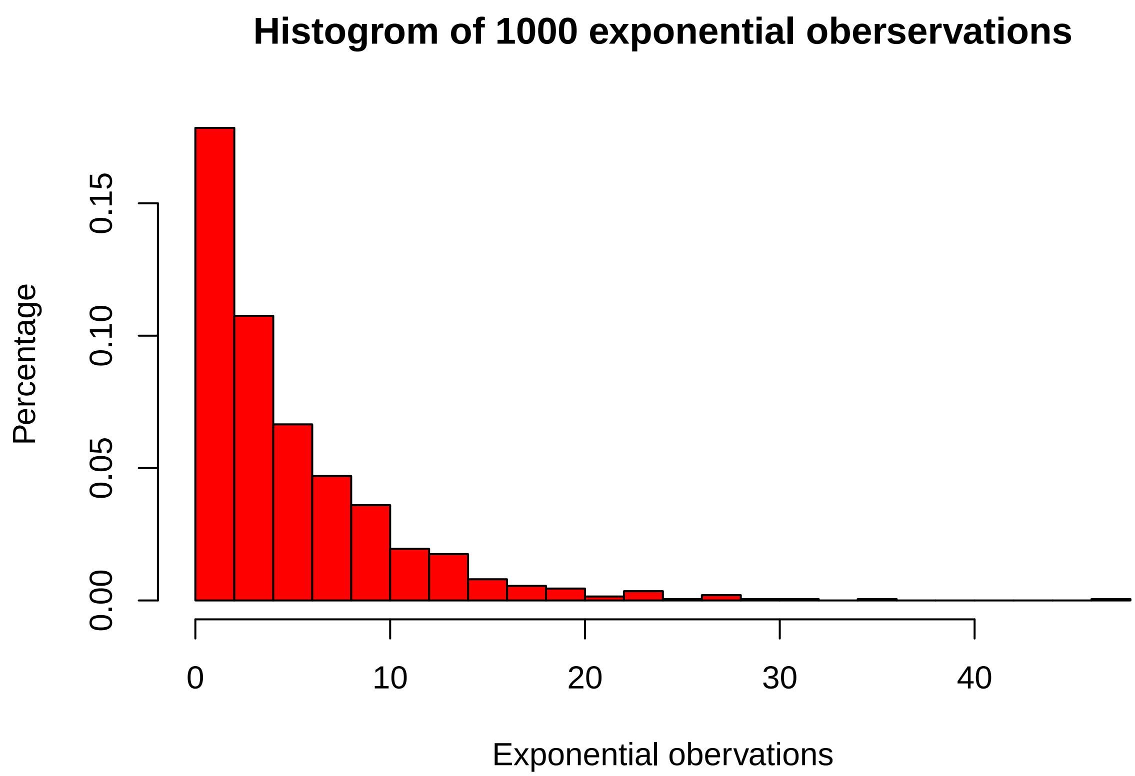
Overview

This is a report for "Statistical Inference Course Project. Simulated random data was generated by rexp() fuction to illustrate the Central Limit Theorem.

A sample of 1000 random exponential number

- A given population ~ Exponential distribution (the rate of the distribution = 0.2, thus population mean = 1/0.2 = 5, variance = (1/0.2)^2 = 25)
- Extract 1000 random number from this population and show distribtion of this single sample
- Calcuate the mean and variance of this single sample

```
set.seed(100000)
x <- rexp(1000,0.2)
hist(x, breaks=30, xlab="Exponential obervations", ylab="Percentage", main="Histogrom of 1000 exponenti
```



```r
print(paste("Mean=", format(mean(x), digits=8), "      ", "Variance=", format(var(x), digits=8)))
```

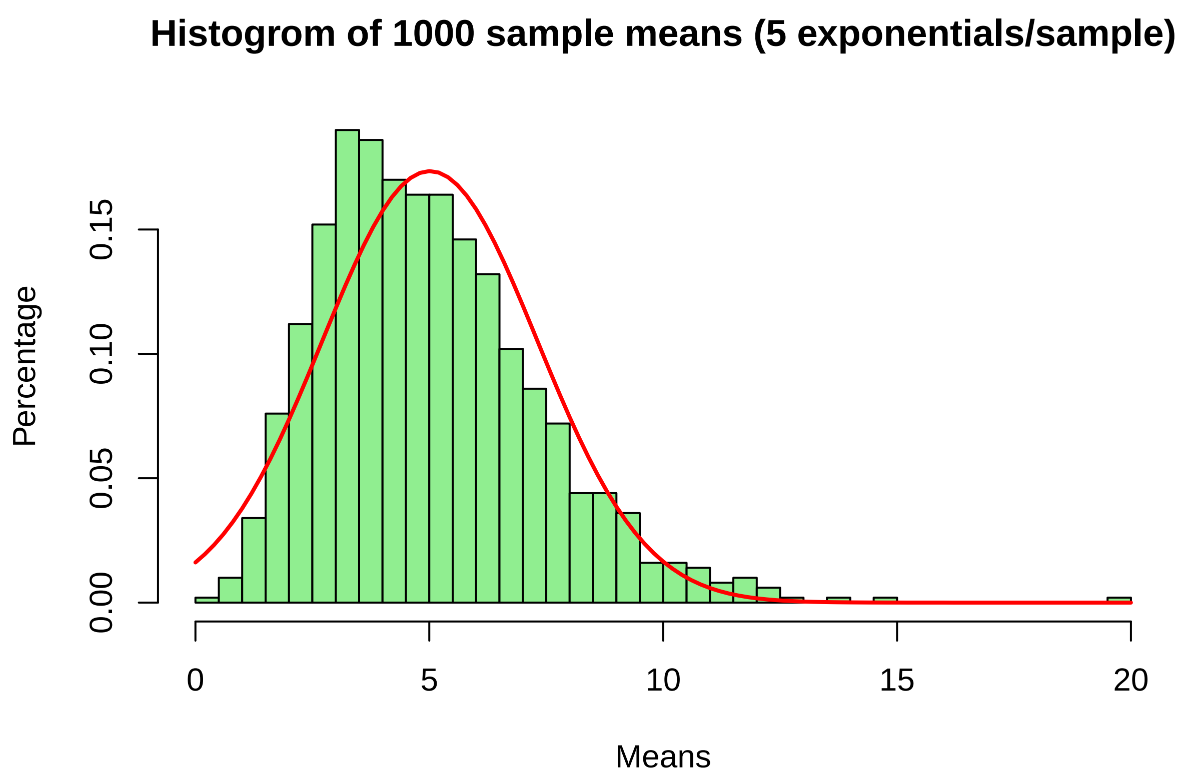
```
## [1] "Mean= 4.8526482       Variance= 26.125292"
```

The results showed:
- This single sample (1000 number) is exponential distibution

1000 samples of 5 exponential obervations/sample

- Show distribtion of samples means
- Calcuate the mean and variance of sample means

```r
##Mean distribution of 1000 samples (5 exponential obervations/sample)
mns1 = NULL
set.seed(100000)
for (i in 1 : 1000) mns1 = c(mns1, mean(rexp(5, 0.2)))
hist(mns1, xlab="Means", breaks=30,ylab="Percentage", main="Histogrom of 1000 sample means (5 exponenti
curve(dnorm(x, mean=mean(mns1), sd=sd(mns1)), add=TRUE, col="red", lwd=2)
```




```r
print(paste("Mean=", format(mean(mns1), digits=8), "      ", "Variance=", format(var(mns1), digits=8)))
```

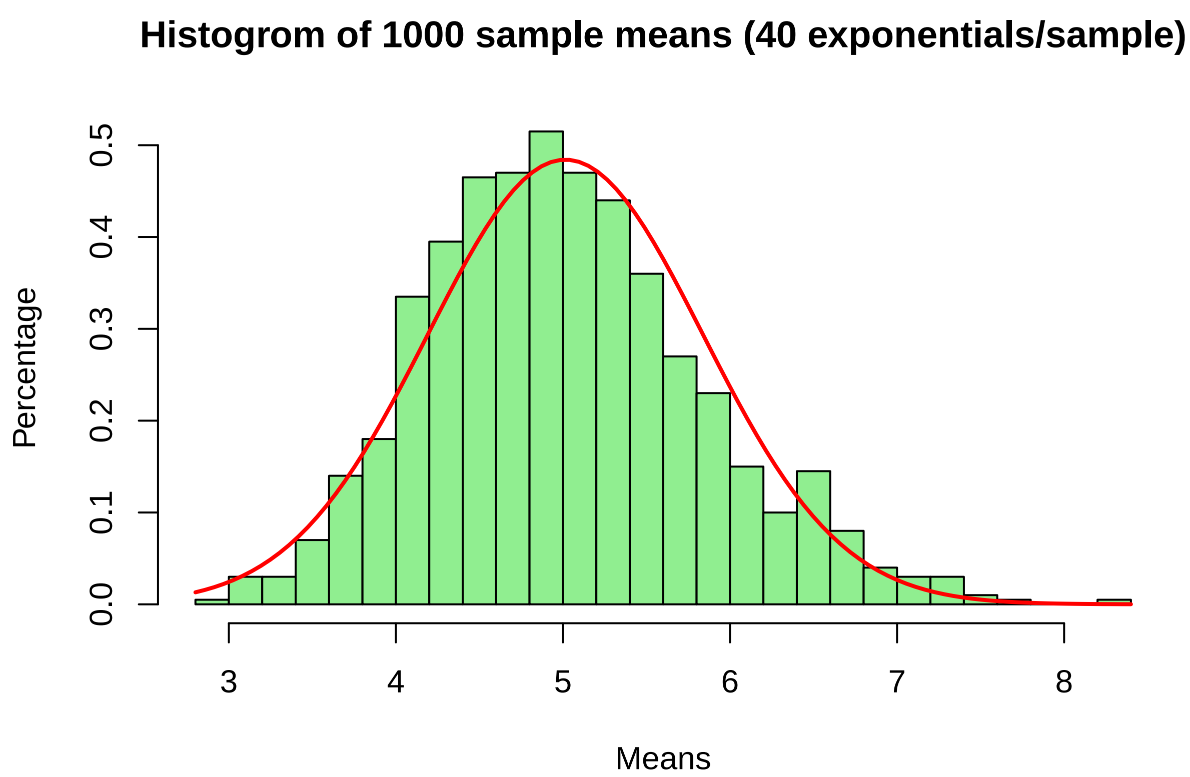
```
## [1] "Mean= 5.0116088       Variance= 5.2878955"
```

The results showed:
- The mean of sample mean distribution is close to 1/0.2=5 (original exponential distribution mean)
- the variance of sample mean distibution is close to the original population variance /5, i.e., (1/0.2)^2/5= 5
- The distribution of left skewed

1000 samples of 40 exponential obervations / sample

- Show distribtion of samples means
- Calcuate the mean and variance of sample means

```
##Mean distribution of 100,000 samples (40 exponential obervations/sample)
mns2 = NULL
set.seed(100000)
for (i in 1 : 1000) mns2 = c(mns2, mean(rexp(40, 0.2)))
hist(mns2, xlab="Means", breaks=30,ylab="Percentage", main="Histogrom of 1000 sample means (40 exponenti
curve(dnorm(x, mean=mean(mns2), sd=sd(mns2)), add=TRUE, col="red", lwd=2)
```




```
print(paste("Mean=", format(mean(mns2), digits=8), "     ", "Variance=", format(var(mns2), digits=8)))
```

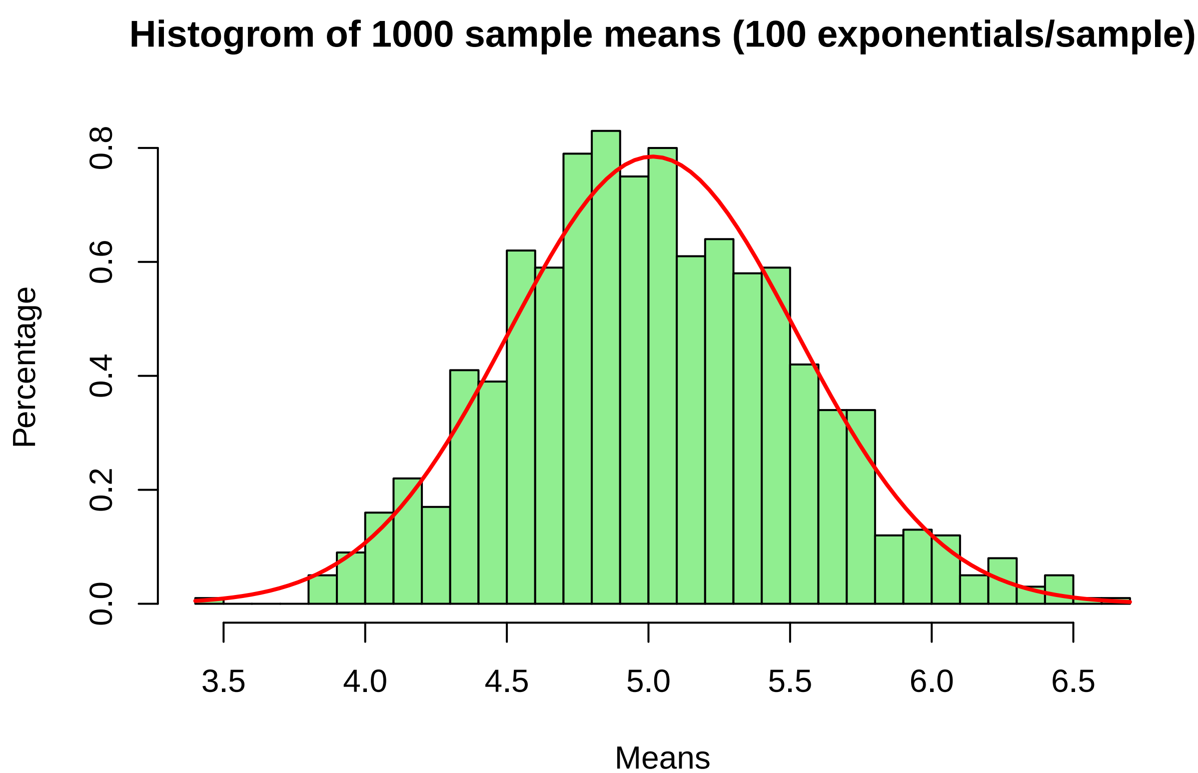
```
## [1] "Mean= 5.0146137       Variance= 0.6787186"
```

The results showed:
- The mean of sample mean distribution is close 1/0.2=5 (original exponential distribution mean)
- the variance of sample mean distibution is close (1/0.2)^2/40= 0.625

1000 samples of 100 exponential obervations / sample

- Show distribtion of samples means
- Calcuate the mean and variance of sample means

```
##Mean distribution of 1000 samples (100 exponential obervations/sample)
mns3 = NULL
set.seed(100000)
for (i in 1 : 1000) mns3 = c(mns3, mean(rexp(100, 0.2)))
hist(mns3, xlab="Means", breaks=30,ylab="Percentage", main="Histogrom of 1000 sample means (100 exponen
curve(dnorm(x, mean=mean(mns3), sd=sd(mns3)), add=TRUE, col="red", lwd=2)
```




```
print(paste("Mean=", format(mean(mns3), digits=8), "     ", "Variance=", format(var(mns3), digits=8)))
```

```
## [1] "Mean= 5.0150349       Variance= 0.25834567"
```

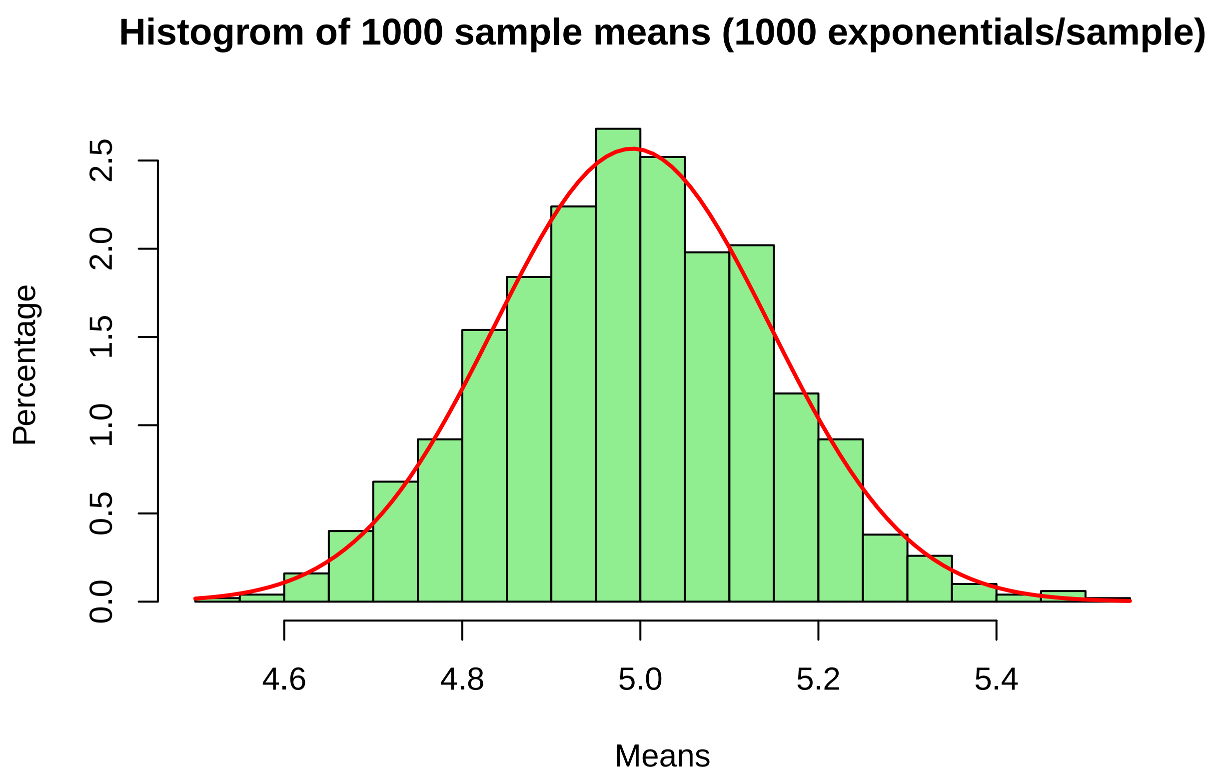
The results showed:
- The mean of sample mean distribution is ~ 1/0.2=5 (original exponential distribution mean)
- the variance of sample mean distibution is ~ (1/0.2)^2/100= 0.25

1000 samples of 1000 exponential obervations / sample

- Show distribtion of samples means

- Calcuate the mean and variance of sample means

```
##Mean distribution of 1000 samples (1000 exponential obervations/sample)
mns4 = NULL
set.seed(100000)
for (i in 1 : 1000) mns4 = c(mns4, mean(rexp(1000, 0.2)))
hist(mns4, xlab="Means", breaks=30,ylab="Percentage", main="Histogrom of 1000 sample means (1000 expone
curve(dnorm(x, mean=mean(mns4), sd=sd(mns4)), add=TRUE, col="red", lwd=2)
```




```
print(paste("Mean=", format(mean(mns4), digits=8), "     ", "Variance=", format(var(mns4), digits=8)))
```

```
## [1] "Mean= 4.9909701       Variance= 0.024150154"
```

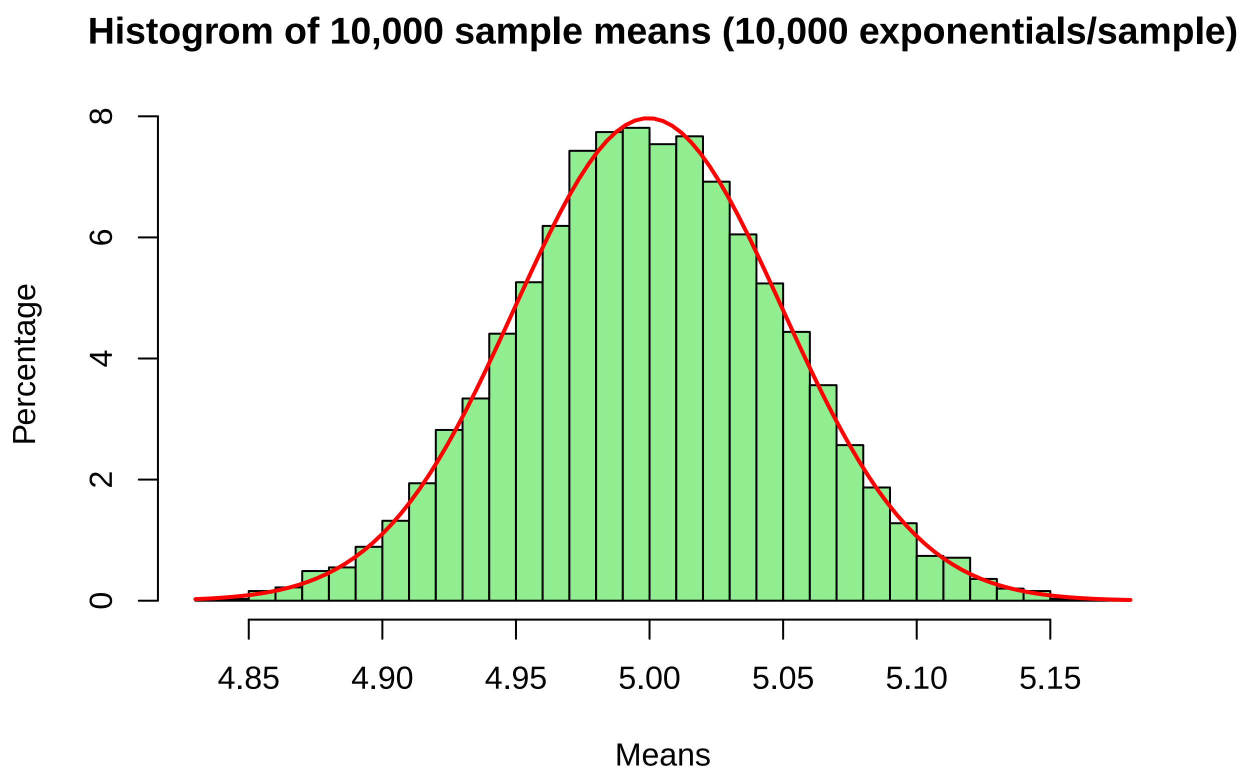
The results showed:
- The mean of sample mean distribution is ~ 1/0.2=5 (original exponential distribution mean)
- the variance of sample mean distibution is ~ (1/0.2)^2/1000= 0.025

10,000 samples of 10,000 exponential obervations / sample

- Show distribtion of samples means

- Calcuate the mean and variance of sample means

```
##Mean distribution of 10,000 samples (10,000 exponential obervations/sample)
mns5 = NULL
set.seed(100000)
for (i in 1 : 10000) mns5 = c(mns5, mean(rexp(10000, 0.2)))
hist(mns5, xlab="Means", breaks=30,ylab="Percentage", main="Histogrom of 10,000 sample means (10,000 ex
curve(dnorm(x, mean=mean(mns5), sd=sd(mns5)), add=TRUE, col="red", lwd=2)
```




```
print(paste("Mean=", format(mean(mns5), digits=8), "     ", "Variance=", format(var(mns5), digits=8)))
```

```
## [1] "Mean= 4.9995505       Variance= 0.0025058074"
```

The results showed:
- The mean of sample mean distribution is ~ 1/0.2=5 (original exponential distribution mean)
- the variance of sample mean distibution is ~ (1/0.2)^2/10000= 0.0025

Conlusion

The Central Limit Theorem states that the sampling distribution of the sampling means approaches a normal distribution as the sample size gets larger, no matter what the shape of the population distribution. The mean of the sample mean distribution is the same as mother population while variance will be variance of the mother population divided by sample size n.