Statistical Inference - Part 1: A simulation exercise

Khoo Jia Jun

14 Sep 2016

Synopsis

In this project you will investigate exponential distribution in R and compare it with Central Limit Theorem. Exponential distribution can be simulated in R with rexp(n, lambda) where lambda is rate parameter. Mean of exponential distribution is 1/lambda and standard deviation is also 1/lambda. Set lambda = 0.2 for all of simulations. You will investigate distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text properties of distribution of mean of 40 exponentials. You should

1. Show sample mean and compare it to theoretical mean of distribution.
2. Show how variable sample is (via variance) and compare it to theoretical variance of distribution.
3. Show that distribution is approximately normal.

Simulation Section

```
#load necessary library packages
library(ggplot2);
# install.packages("survey")
library(survey)
```

```
## Loading required package: grid
```

```
## Loading required package: Matrix
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survey'
```

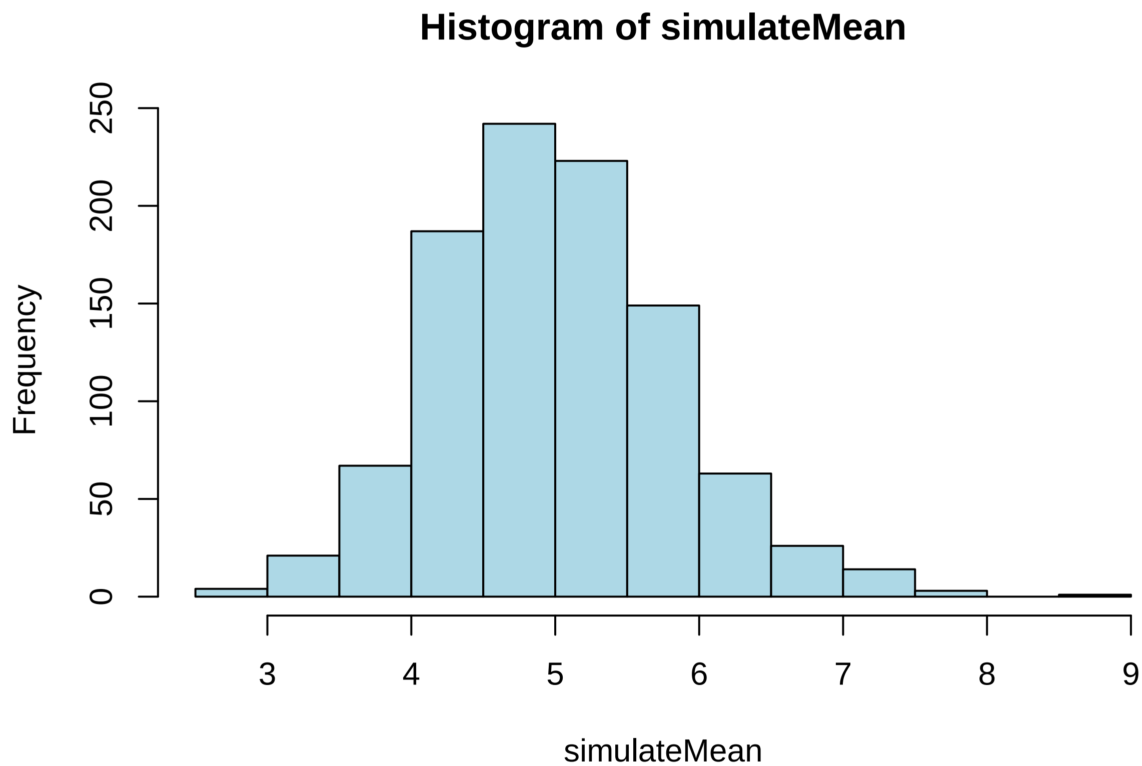
```
## The following object is masked from 'package:graphics':
##
##     dotchart
```

```
#variables that control simulation
numSimulations <- 1000;
lambda <- 0.2;
numExp <- 40;

set.seed(234);

#Create a matrix of 1000 rows with columns corresponding to random simulation 40 times
```

```
simulateMatrix <- matrix(rexp(numSimulations * numExp, rate=lambda), numSimulations,
numExp);
simulateMean <- rowMeans(simulateMatrix);
hist(simulateMean, col = "light blue");
```



Results Section

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

```
sampleMean <- mean(simulateMean);
print (paste("Sample mean of distribution = ", sampleMean));
```

```
## [1] "Sample mean of distribution =  5.0015728501858"
```

```
theoryMean <- 1/lambda;
print (paste("Theoretical mean of distribution = ", theoryMean));
```

```
## [1] "Theoretical mean of distribution =  5"
```

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

```
actualVariance <- var(simulateMean);
print (paste("Sample variance of distribution = ", actualVariance));
```

```
## [1] "Sample variance of distribution =  0.66315043736661"
```

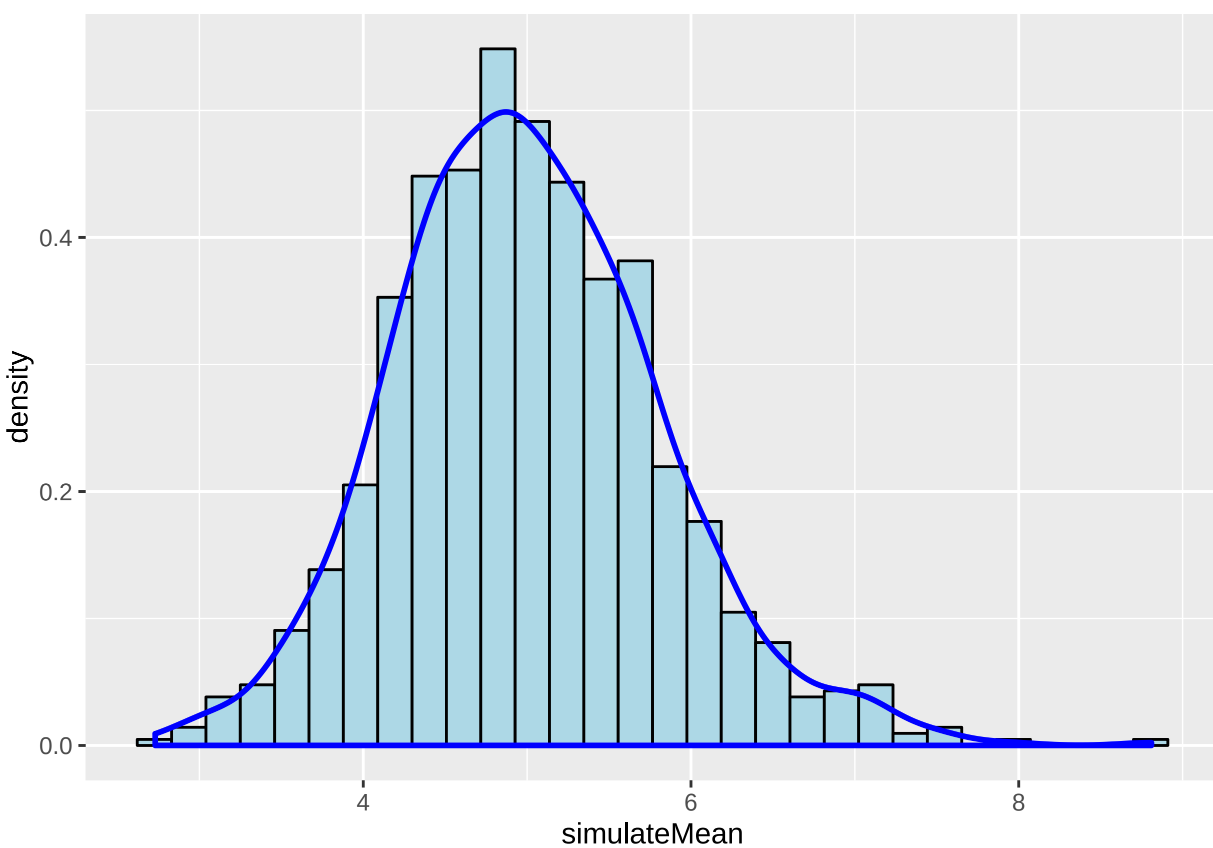
```
theoryVariance <- (1/lambda)^2/numExp;
print (paste("Theoretical variance of distribution = ", theoryVariance));
```

```
## [1] "Theoretical variance of distribution =  0.625"
```

3. Show that distribution is approximately normal.

```
plotdata <- data.frame(simulateMean);
m <- ggplot(plotdata, aes(x=simulateMean));
m <- m + geom_histogram(aes(y=..density..), colour="black", fill = "light blue");
m + geom_density(colour="blue", size=1);
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
actualConfidenceInterval <- round (mean(simulateMean)
                                   + c(-1,1)*1.96*sd(simulateMean)/sqrt(numExp),3);
print (paste("Sample confidence interval of distribution = ", actualConfidenceInterval));
```

```
## [1] "Sample confidence interval of distribution =  4.749"
## [2] "Sample confidence interval of distribution =  5.254"
```

```
theoryConfidenceInterval <- theoryMean + c(-1,1)*1.96*sqrt(theoryVariance)/sqrt(numExp);
print (paste("Theoretical confidence interval of distribution = ", theoryConfidenceInterval));
```

```
## [1] "Theoretical confidence interval of distribution =  4.755"
## [2] "Theoretical confidence interval of distribution =  5.245"
```