Statistical Inference Project 1

Friday, December 19, 2014

Abstract

- In this analysis, the distribution of averages of 40 exponential-distributed random variables is investegated by simulation approach.

- The following questions are answered:

    - Show where the distribution is centered at and compare it to the theoretical center of the distribution.
    - Show how variable it is and compare it to the theoretical variance of the distribution.
    - Show that the distribution is approximately normal.

Data Analysis

Random Variables Generation

1. Setting the parameters of the exponential distribution.

```r
library(ggplot2)
library(grid)
# Setting parameters
lambda <- 0.2
# Average number
n <- 40
# Number of simulation
N.sim <- 1000
```

2. Random variables generation.

```r
# Set seed for ruproducible property
set.seed(12345)
# Simulate the average of 40 exp-dist random variables
Sample.sim <- sapply(1:N.sim, function(x) mean(rexp(n,lambda)))
```

Compare the Center of Theoretical Distrubution and Sample Distribution

```r
# Construct the data frame for Center of Theoretical and sample distribution
Med <- data.frame(Median=c(1/lambda, median(Sample.sim)),
                  Type=c("Theoretical","Simulation"))
# Visualize the result by ggplot
qplot(x=Sample.sim,geom="histogram") +
  geom_histogram(fill="skyblue3",colour="darkgrey") +
  geom_vline(aes(xintercept=Median,colour=Type),data=Med,show_guide = T,size=0.8) +
```

```
scale_colour_manual(values=c("red","black")) +
theme(legend.key.width=unit(0.5, "cm"),legend.key.size=unit(1, "cm"),
        legend.position=c(0.8,0.8),legend.background=NULL)
```
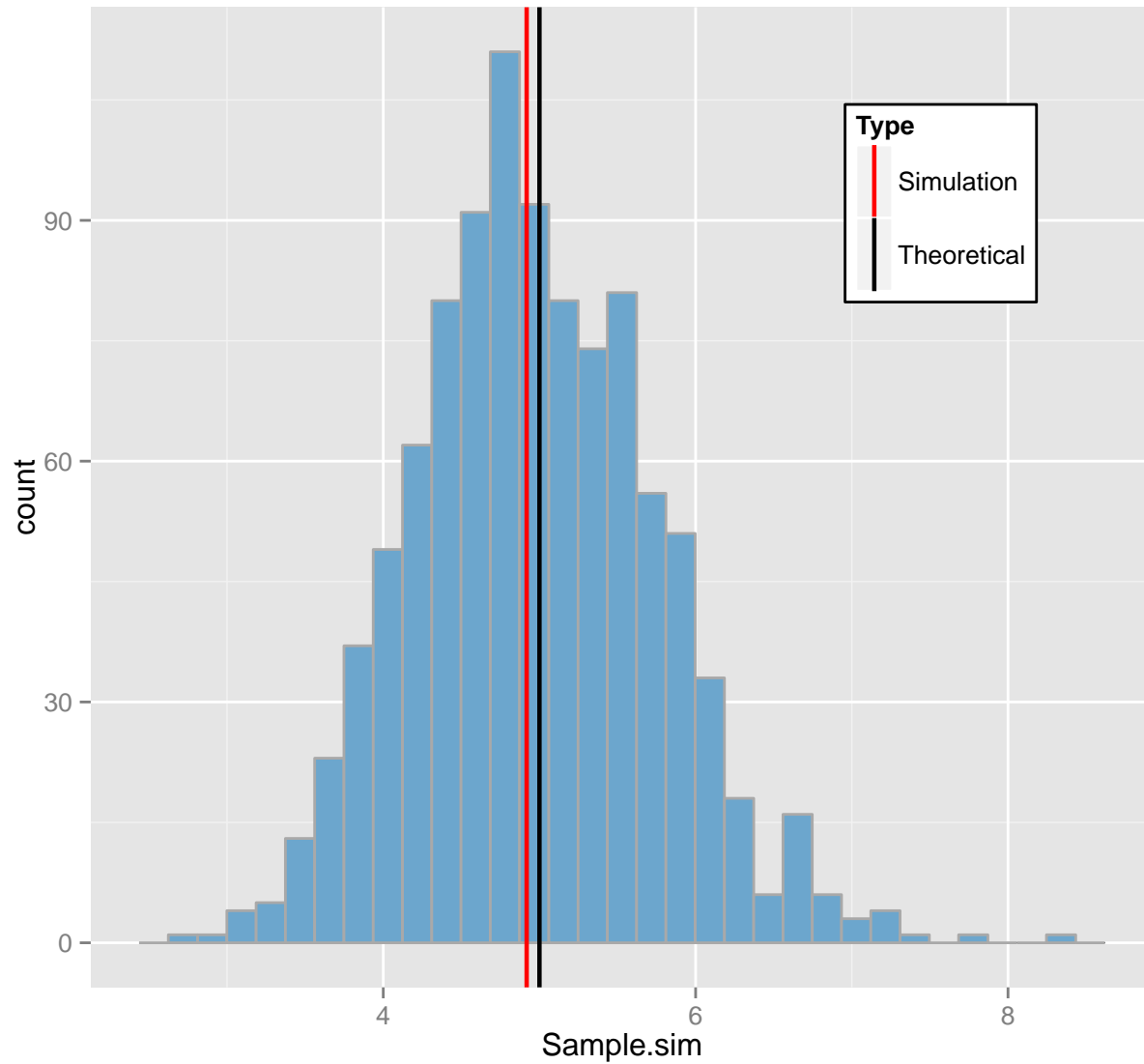

Variance Comparison

The theoretical variance is

$$var = \frac{1}{40 * lambda^2} = 0.625$$

.

Now we compute the sample variance

```
var(Sample.sim)
```

```
## [1] 0.5954369
```

The Sample Distribution is Approximately Normal

We can show that the distribution of averages of 40 exponential-distributed random variables is apporximately normal by QQ-plot.

```r
qplot(sample=Sample.sim,stat="qq",distribution = qnorm) +
  labs(title="Quantile-Quantile Plot: Sample v.s. Normal  Distribution",
       x= "Theoretical Normal Disribution",
       y= "Sample Distribution") +
  geom_abline(slope=sd(Sample.sim),intercept=median(Sample.sim),size=0.8,colour="red")
```