Statistical Inference: Course Project

Created by H.Wang on November 19, 2015

Basic Settings

```
echo= TRUE  # make scripts visible to others
```

Description of the Problem

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. Investigate the distribution of averages of 40 exponentials, need to do a thousand or so simulated averages of 40 exponentials. Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

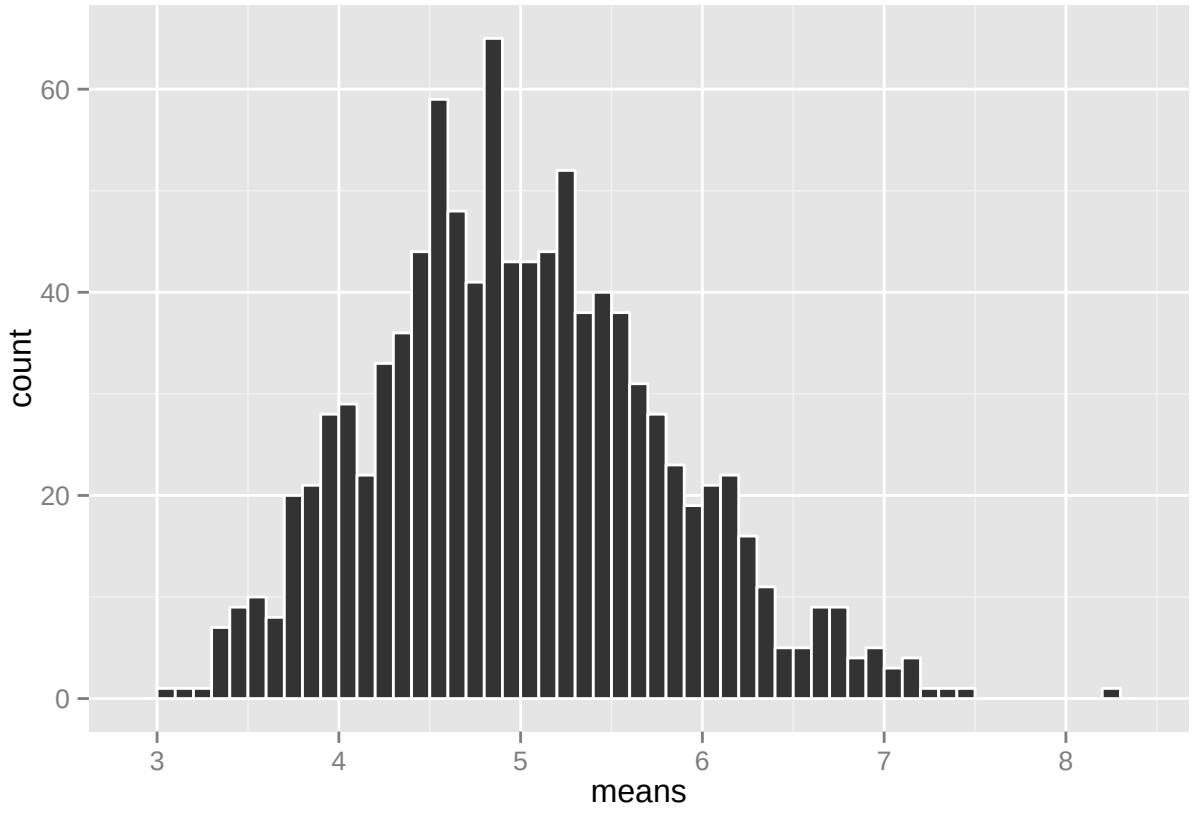
Simulations

```
library(ggplot2)
# simulation constants
lambda <- 0.2
n <- 40
sim_num<- 1000 #simulation number

# set the seed to reproduce simluation sample values
set.seed(5523028)

# run the simulation in 1000*40 matrix
exp_dist <- matrix(data= rexp(n*sim_num, lambda),nrow= sim_num)
exp_dist_mean <- data.frame(means = apply(exp_dist, 1, mean))

# histrogram of simulation results
ggplot(exp_dist_mean, aes(means)) +  stat_bin(binwidth=0.1, color = 'white')
```



Sample Mean vs. Theoretical Mean

```
t_mean <- 1/lambda #theoretical Mean
s_mean <- mean(exp_dist_mean$means) #simulation Mean

t_mean
```

```
## [1] 5
```

```
s_mean
```

```
## [1] 5.003705
```

As comparison, the average sample mean and expected mean are very close.

Sample Variance vs. Theoretical Variance

```
t_var <- (1/lambda/sqrt(n))^2 #theoretical variance
s_var <- var(exp_dist_mean$means) #simulation variance

t_var
```

```
## [1] 0.625
```

```
s_var
```

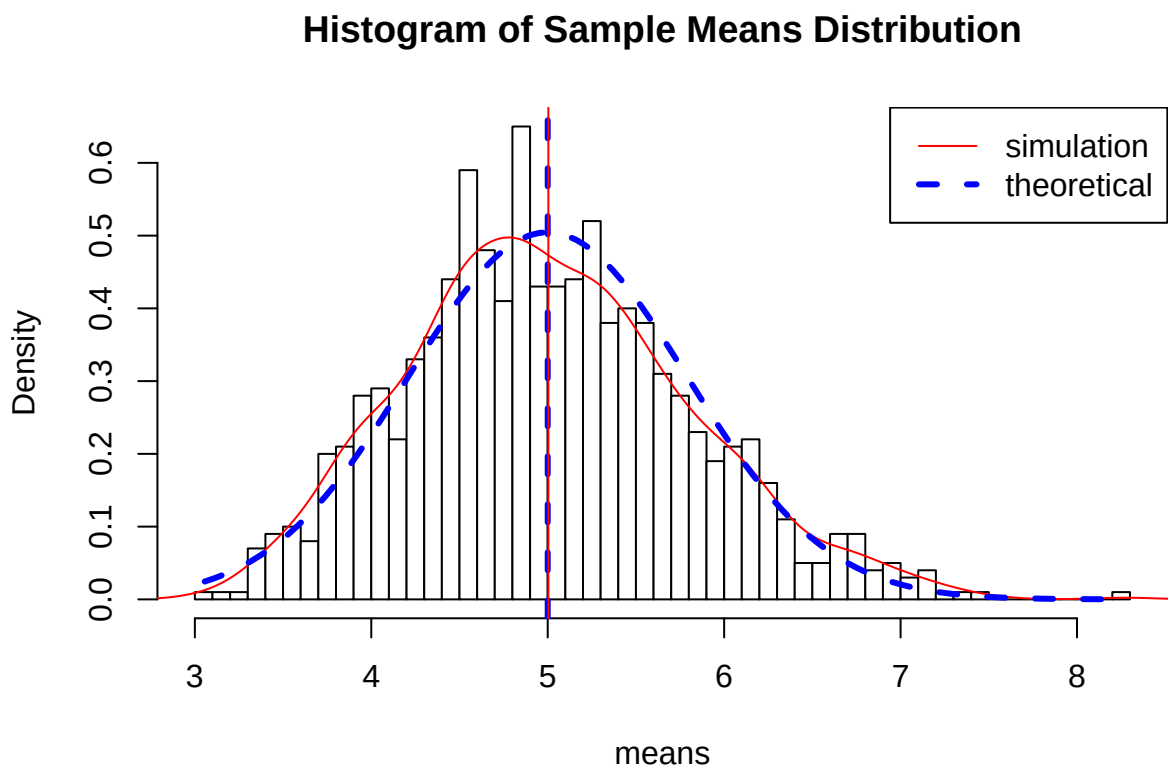
```
## [1] 0.6423574
```

As comparison, both variance are very close. Since variance is the square of the standard deviations, minor differences will be enhanced but are still pretty close.

Distribution

Comparing the population means & standard deviation with a normal distribution of the expected values. Lines for the simulated and expected means are added.

```
hist(exp_dist_mean$mean, breaks=50, prob=TRUE,
     main = "Histogram of Sample Means Distribution", xlab="means", col="white")
lines(density(exp_dist_mean$mean), col="red")
abline(v = c(1/lambda, mean(exp_dist_mean$means)), col = c("blue","red"), lty=c(2,1), lwd=c(3, 1))
xfit <- seq(min(exp_dist_mean$mean), max(exp_dist_mean$mean), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
lines(xfit, yfit, pch=22, col="blue", lty=2, lwd = 3)
legend('topright', c("simulation", "theoretical"), lty=c(1,2),lwd=c(1, 3), col=c("red", "blue"))
```



Therefore, the distribution of averages of 40 exponentials is centered at 5.0037045 and the same is close to the theoretical center of the distribution, which is $\lambda^{-1} = 5$.