Statistical_Inference_Project

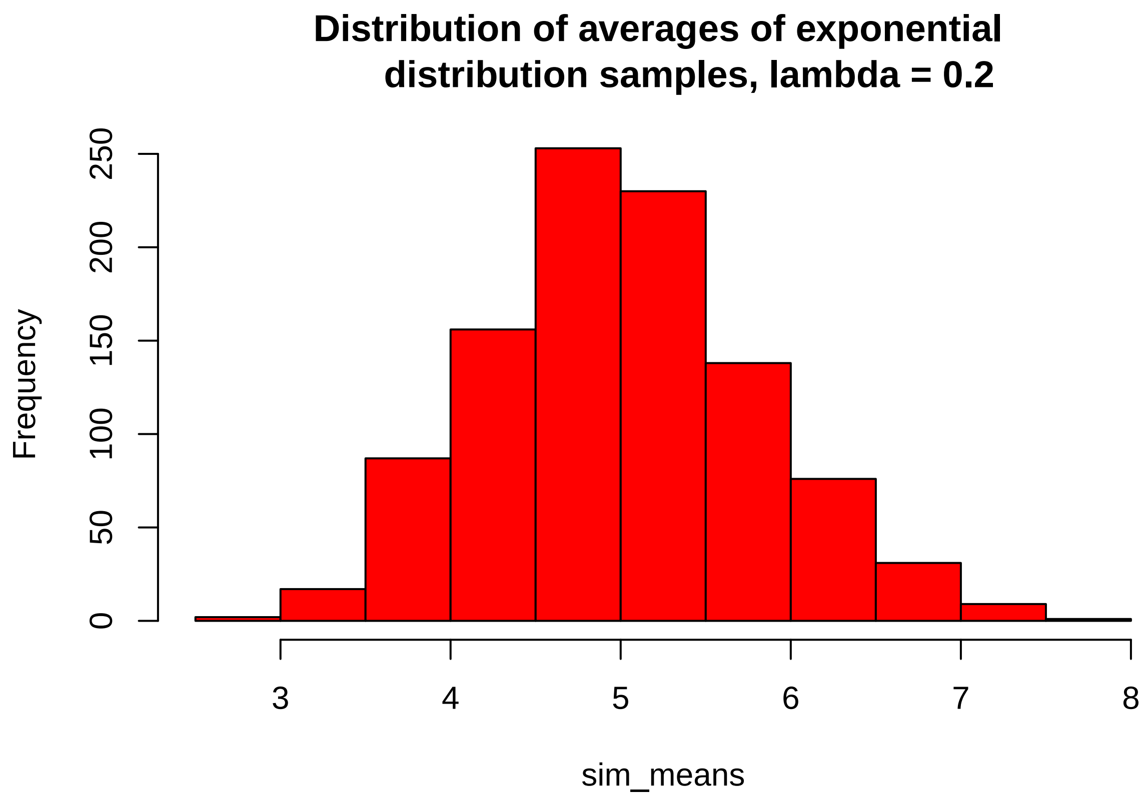
Patrick Close

21 November 2015

Summary: Investigate the exponential distribution by simulating 40 exponential distributions with lambda = 0.2 in a Monte Carlo simulation with 1000 runs. Use the resulting distribution to compare the mean and variance of the simulation to the theoretically expected mean and variance. Then observe the effect of the Central Limit Theorem by observing that the resulting distribution of simulated means is approximately normal (Gaussian).

Step1: simulate the distribution of means of 40 exponential distributions with mean = 1/lambda & stddev = 1/lambda

```r
nosim <- 1000
n <- 40
lambda <- 0.2
set.seed(8180)
#
#create nosim x n matrix to hold simulated random variables
simdata <- matrix(rexp(nosim * n, rate = lambda), nosim, n)
sim_means <- rowMeans(simdata)
#
#Plot distribution of simulation averages:
hist(sim_means, col = "red", main = "Distribution of averages of exponential
     distribution samples, lambda = 0.2")
```



Compare simulated and theoretical means

```r
#create matrix to store simulated and theoretical means
df_summary <- data.frame(mean(sim_means), 1/lambda)
colnames(df_summary) <- c("simulated", "theoretical")
print(round(df_summary,3))
```

```
##   simulated theoretical
## 1     5.012           5
```

As observed the mean of the simulation is close to the theoretically expected mean

Step2: compare variance

```r
df_summary[2,] <- c(var(sim_means), (1/lambda)^2/n)
rownames(df_summary) <- c("mean", "variance")
print(round(df_summary,3))
```

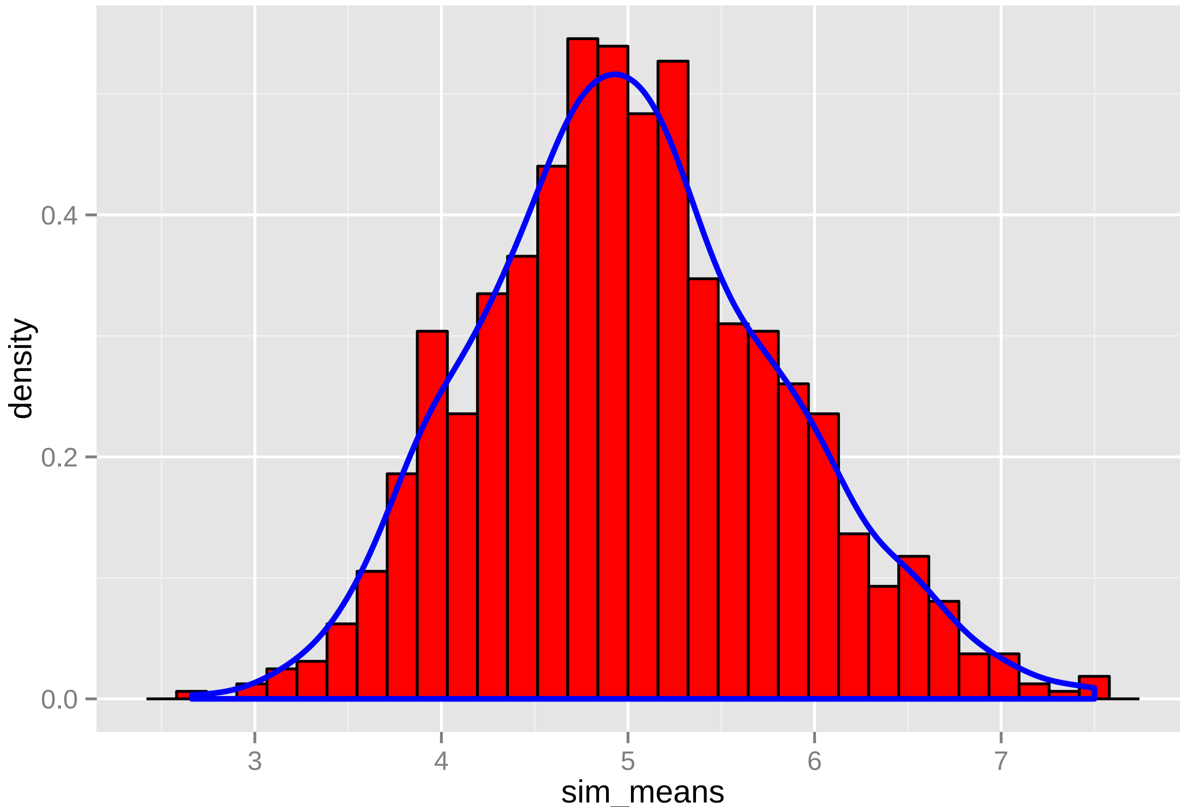
```
##          simulated theoretical
## mean         5.012       5.000
## variance     0.640       0.625
```

As observed the variance of the simulation is close to the theoretically expected variance

Step3: show the variable approximately follows a normal distribution

```r
library(ggplot2)
sim_means_g <- data.frame(sim_means)
g <- ggplot(sim_means_g, aes(x=sim_means))
g <- g + geom_histogram(aes(y=..density..), col = "black", fill = "red")
g <- g + geom_density(col = "blue", size = 1)
print (g)
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



The shape of the curve is approximately Gaussian

construct simulated confidence interval and compare to theoretical CI

```r
sim_CI <- mean(sim_means) + c(-1,1)*1.96*sd(sim_means)/sqrt(n)
theo_CI <- 1/lambda + c(-1,1)*1.96*((1/lambda^2)/n)/sqrt(n)
CI_summary <- data.frame(rbind(sim_CI, theo_CI), row.names = c("simulated", "theoretical"))
colnames(CI_summary) <- c("Lower Bound", "Upper Bound")
print(CI_summary)
```

```
##             Lower Bound Upper Bound
## simulated      4.763686    5.259412
## theoretical    4.806310    5.193690
```

```r
#
#normal quintile plot
qqnorm(sim_means); qqline(sim_means)
```