Statistical Inference: Programming Assignment 1

LarionovaAnna

Sunday, May 24, 2015

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

### switch on libraries

library('ggplot2')

### terms of the problem

simulations\_number <- 1000  
n <- 40  
lambda <- 0.2

### simulations

mn = NULL  
variance = NULL  
for (i in 1 : simulations\_number) {  
 expd <- rexp(n, lambda) #Exponential Distribution  
 mn <- c(mn, mean(expd)) #mean  
 variance <- c(variance, var(expd)) #Variance   
}  
sample\_mean <- mean(mn) # Sample Mean  
mean\_theoretical <- 1/lambda # Theoritical Mean  
sample\_mean

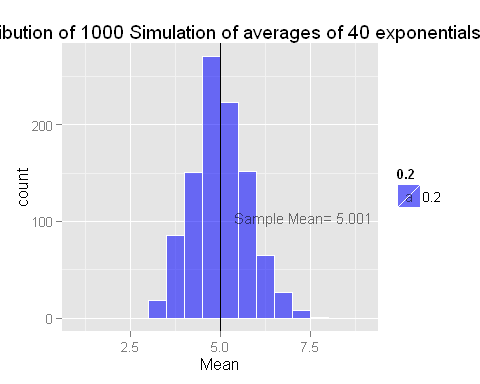
## [1] 5.001223

mean\_theoretical

## [1] 5

### create graphics

plot1 <- qplot(mn, fill = I("blue"), color = I("white"), geom = "histogram",   
 xlab = "Mean", binwidth = 0.5, xlim = c(1,9), alpha = 0.2,  
 main = "Distribution of 1000 Simulation of averages of 40 exponentials")   
plot1 <- plot1 + geom\_vline(xintercept = sample\_mean, color = "black")   
plot1 <- plot1 + geom\_text(mapping = aes(x = sample\_mean, y = 110,   
 label = paste("Sample Mean=",   
 round(sample\_mean, 3))),   
 size = 4, vjust = 1, hjust = -0.1)  
plot1



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

sample\_variance <- mean(variance) # Sample Variance  
variance\_theoretical <- (1/lambda)^2 # Theoritical Variance  
sample\_variance

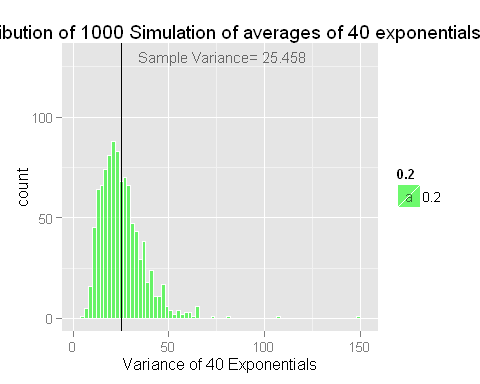
## [1] 25.45849

variance\_theoretical

## [1] 25

### create graphics

plot2 <- qplot(variance, fill = I("green"), color = I("white"),  
 geom = "histogram", binwidth = 2,   
 xlab = "Variance of 40 Exponentials",  
 alpha = 0.2,  
 main = "Distribution of 1000 Simulation of averages of 40 exponentials")   
plot2 <- plot2 + geom\_vline(xintercept = sample\_variance, color = "black")   
plot2 <- plot2 + geom\_text(mapping = aes(x = sample\_variance, y = 130,   
 label = paste("Sample Variance=",   
 round(sample\_variance, 3))),   
 size=4, hjust= -0.1)  
plot2



## 3. Show that the distribution is approximately normal.

expdistrib <- rexp(simulations\_number, lambda)  
expdistrib\_mean = mean(expdistrib) #Mean Exponential Distribution   
expdistrib\_variance = var(expdistrib) #Variance Exponential Distribution   
expdistrib\_mean

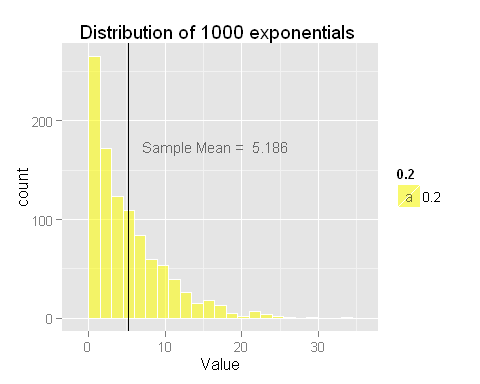
## [1] 5.185809

expdistrib\_variance

## [1] 24.40464

### create graphics

plot3 <- qplot(expdistrib, fill = I("yellow"), color = I("white"),  
 alpha = 0.2,  
 geom = "histogram", xlab = "Value", binwidth = 1.5,  
 main = "Distribution of 1000 exponentials ")   
plot3 <- plot3 + geom\_vline(xintercept = expdistrib\_mean, color = "black")   
plot3 <- plot3 + geom\_text(mapping = aes(x = expdistrib\_mean, y = 180,   
 label = paste("Sample Mean = ",   
 round(expdistrib\_mean, 3))),   
 size = 4, hjust = -0.1, vjust = 1)  
plot3



## Conclusions:

* On figure #1: the sample mean is very close to the theoretical mean
* On figure #2: the sample variance is very close to the theoretical variance.
* On figure #3: the exponentional distribution is approximately close to normal.
* Figure 1-3: the sample mean and distribution mean is very close to theoretical values. This is illustrated in Central Limit Theorem, which states that the distribution of averages of IID variables becomes that of a standard normal as the sample size increases.