Part-1

Sk Minhazul Islam

Tuesday, September 02, 2014

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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, I investigated the distribution of averages of 10, 20, and 40 exponential(0.2)s with a thousand simulated averages of 10, 20, and 40 exponentials.

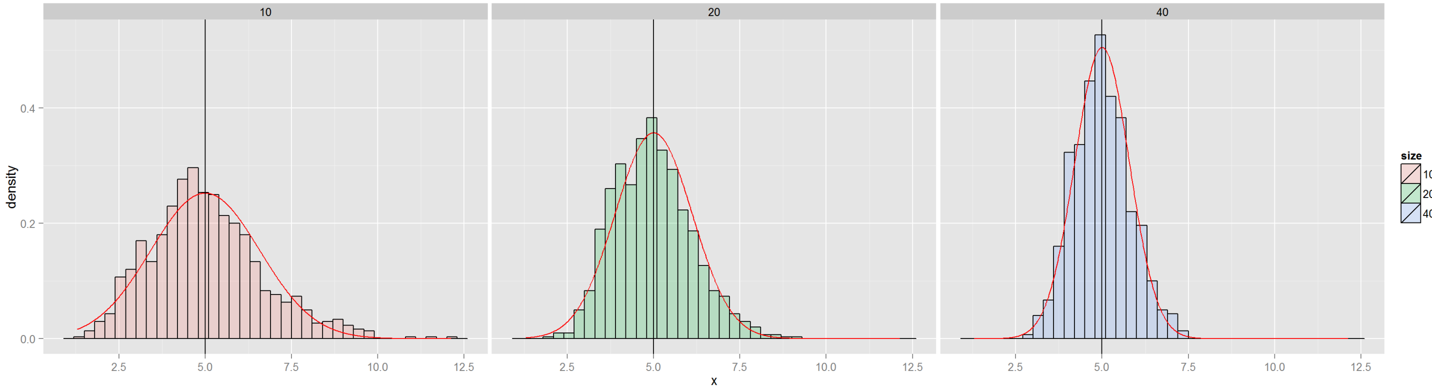
Result of our rexp(n=10,20,40, .2) experiment

```
nosim <- 1000
lambda <- 0.2
cfunc <- function(x, n) mean(x)
dat10<-rexp(nosim * 10, lambda)
dat20<-rexp(nosim * 20, lambda)
dat40<-rexp(nosim * 40, lambda)

dat <- data.frame(
  x = c(apply(matrix(dat10, nosim),1, cfunc, 10),
        apply(matrix(dat20, nosim),1, cfunc, 20),
        apply(matrix(dat40, nosim),1, cfunc, 40)
        ),size = factor(rep(c(10, 20, 40), rep(nosim, 3))))

grid <- with(dat, seq(min(x), max(x), length = 1000))
normaldens <- ddply(dat, "size", function(df) {
  data.frame(
    predicted = grid,
    density = dnorm(grid, mean=5, sd=5/sqrt(as.numeric(as.character((df$size)))))
  )})

g <- ggplot(dat, aes(x = x, fill = size)) + geom_histogram(alpha = .20, binwidth=.3, colour = "black"
, aes(y = ..density..))
g <- g + geom_line(data = normaldens, aes(x = predicted, y = density), colour = "red")
g <- g + geom_vline(xintercept = 5)
g <- g + facet_grid(. ~ size)
g
```



1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

The theoretical center of the distribution is 5. It can bee seen in the figure above that the distrubution centers around the vertical line at 5.

2. Show how variable it is and compare it to the theoretical variance of the distribution.

```
var10<-mean(apply(matrix(dat10, nosim),1, var))
var10
```

```
## [1] 25.97047
```

The variance for the 10 simulated exponentials is 25.9704674

```
var20<-mean(apply(matrix(dat20, nosim),1, var))
var20
```

```
## [1] 24.93207
```

The variance for the 20 simulated exponentials is 24.9320655

```
var40<-mean(apply(matrix(dat40, nosim),1, var))
var40
```

```
## [1] 25.39837
```

The variance for the 40 simulated exponentials is 25.3983703

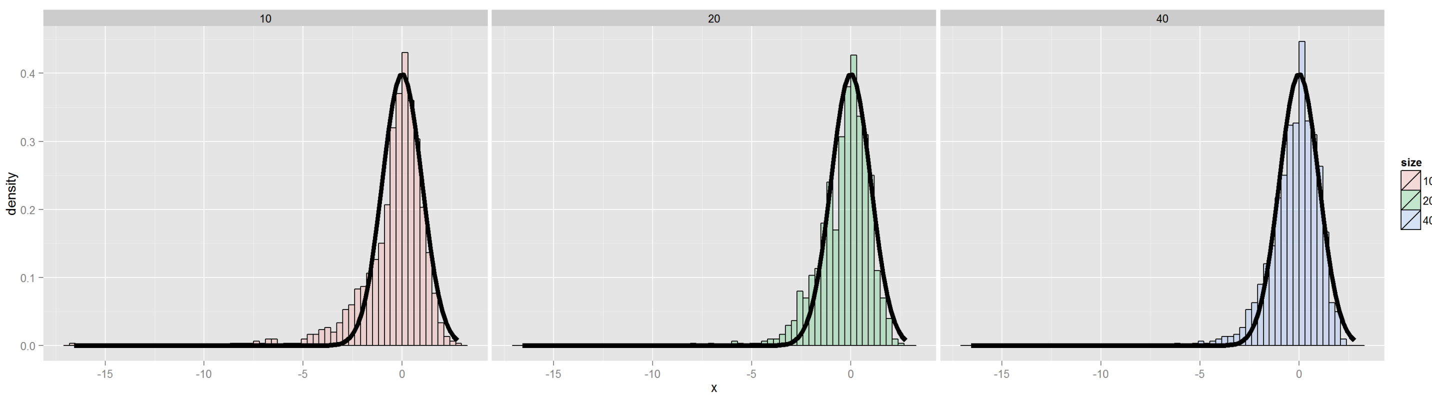
The theoretical variance is 25 or (1/.2)^2

3. Show that the distribution is approximately normal.

It can be seen in the figures above that the distribution is aproximately normal. The red line in each panel represents a normal distribution with mean 5 and standard deviation of 5/sqrt(number of simulated exponentials)

In the figure below, I test the Central Limit Theorem by taking the outcome of the simulation minus the mean 5 and dividing by the standard error which is 5 divided by the square root of the number of simulated exponentials. I overlayed the result with a normal equation. It can be seen that the results of the means of the simulation look like a standard Bell curve. Both the above figure and the figure below show that more simulated results approximate the standard Bell curve better.

```
cfunc2 <- function(x, n) sqrt(n) * (mean(x) - 5) / sd(x)
dat2 <- data.frame(
  x = c(apply(matrix(dat10, nosim),1, cfunc2, 10),
        apply(matrix(dat20, nosim),1, cfunc2, 20),
        apply(matrix(dat40, nosim),1, cfunc2, 40)
        ),  size = factor(rep(c(10, 20, 40), rep(nosim, 3))))
g <- ggplot(dat2, aes(x = x, fill = size)) + geom_histogram(alpha = .20, binwidth=.3, colour = "black
", aes(y = ..density..))
g <- g + stat_function(fun = dnorm, size = 2)
g + facet_grid(. ~ size)
```



4. Evaluate the coverage of the confidence interval for 1/lambda: \[\bar X \pm 1.96 \frac{S}{\sqrt{N}} \]

```
#mn<-aggregate(x ~ size, data=dat, FUN=mean)
#mn40<-mn[mn$size==40,]$x
#sd<-aggregate(x ~ size, data=dat, FUN=sd) # or
#sd40<-sd[sd$size==40,]$x
#mn + c(-1, 1) * pnorm(1.96, mean=mn, sd=s, lower.tail=FALSE) * s / sqrt(n)

mn40<-mean(apply(matrix(dat40, nosim), 1, mean))
sd40<-sd(apply(matrix(dat40, nosim),1, mean))
n <- 40
mn40 + c(-1, 1) * 1.96 * sd40 / sqrt(n)
```

```
## [1] 4.742561 5.249852
```