Statistical Inference Course Project Part 1

domesc

June 16, 2015

Assignment

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. 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. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

Analysis

First we need to set up the variables and simulations to be used for the assignment

```
library(ggplot2)
set.seed(42)
lambda <- 0.2
nosim <- 1000
n <- 40

sim <- data.frame(mean=numeric(nosim))
for (i in 1 : nosim) {
  sim[i,1] = mean(rexp(n, lambda))
}
```

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

We compute the theoretical mean through the formula and we use the R function "mean" to compute the sample mean of our simulations vector:

```
theo_mean <- 1/lambda
sample_mean <- mean(sim$mean)

theo_mean
```

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

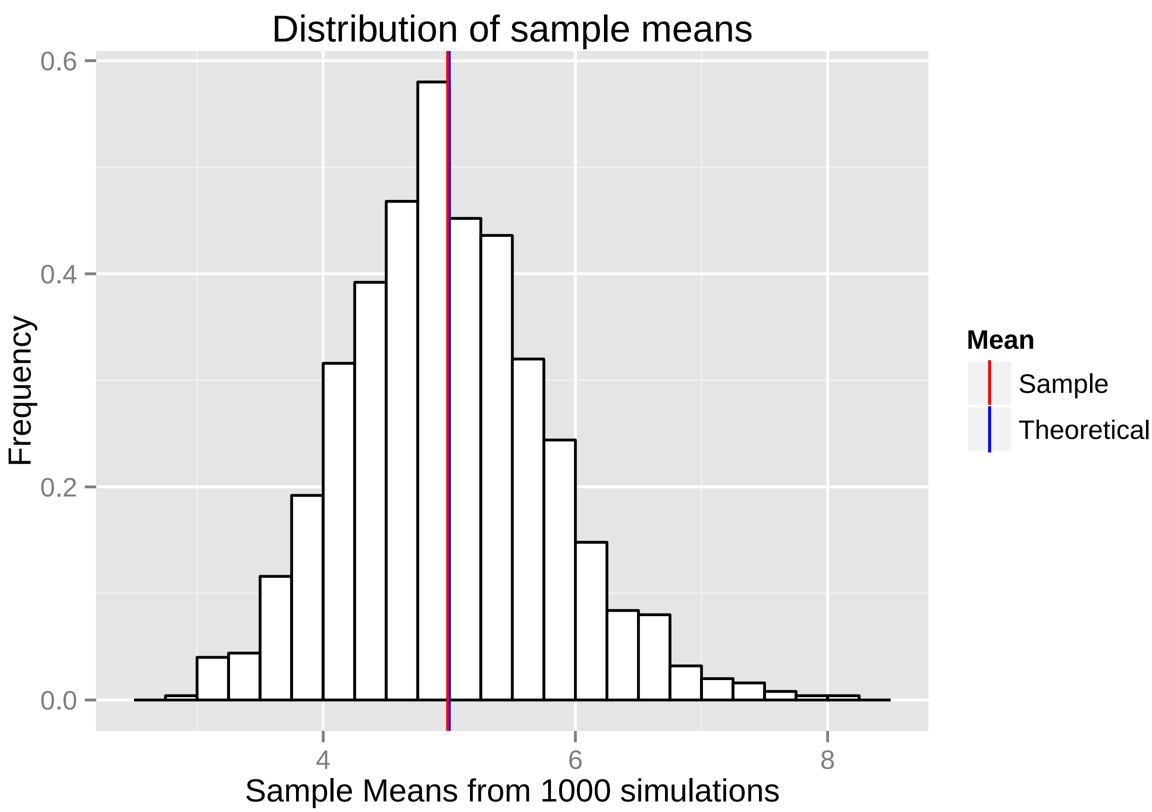
```
sample_mean
```

```
## [1] 4.986508
```

We can see also with a plot that the two means are quite close:

```r
plot <- ggplot(data = sim, aes(x=sim$mean))+
  ggtitle("Distribution of sample means") +
  xlab("Sample Means from 1000 simulations") +
  ylab("Frequency") +
  geom_histogram(aes(y=..density..), color = 'black', fill = 'white', binwidth = 0.25)

plot +
  geom_vline(aes(xintercept=theo_mean, color='Theoretical'), show_guide=T) +
  geom_vline(aes(xintercept=sample_mean, color='Sample'), show_guide=T) +
  scale_color_manual("Mean",
                     values=c("Theoretical"="blue","Sample"="red"))
```



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

We do the same for the variance:

```r
theo_var <- (1/(lambda^2))/n
sample_var <- var(sim$mean)

theo_var
```

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

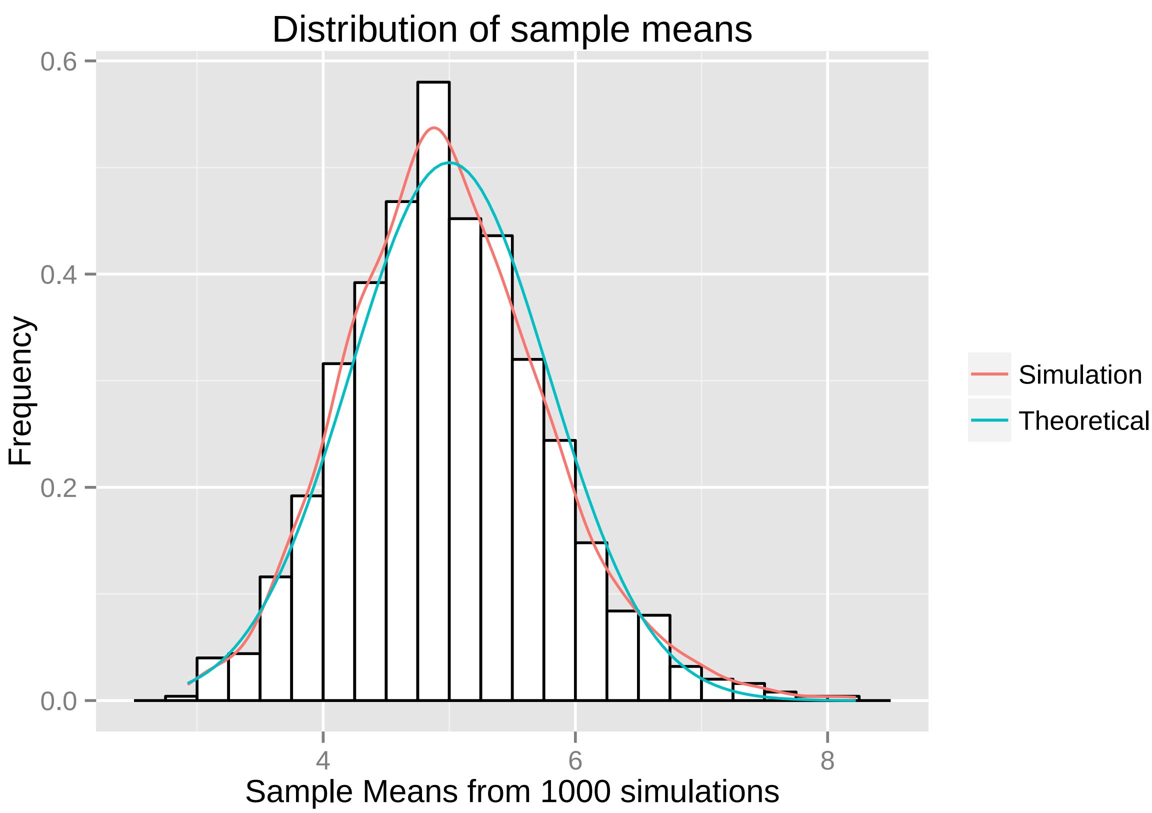
```
sample_var
```

```
## [1] 0.6344405
```

3. Show that the distribution is approximately normal.

We plot the histogram of our distribution and compare it with the normal distribution associated to the theoretical mean:

```
plot +
  geom_line(aes(x = sim$mean,y = ..density.., colour = 'Simulation'), stat = 'density') +
  stat_function(
    fun = dnorm, args = list(mean = theo_mean, sd = sqrt(theo_var)),
    aes(colour = 'Theoretical')
    ) +
  theme(legend.title=element_blank())
```



The plot shows well that our distribution is close to a normal distribution.