Statistical Inference Project 1

Joao Clemencio

22 July 2015

Overview

In this project we will attempt to show that the averages of many exponential distributions follow the Central Limit Theorem, having an approximately Gaussian distibution.

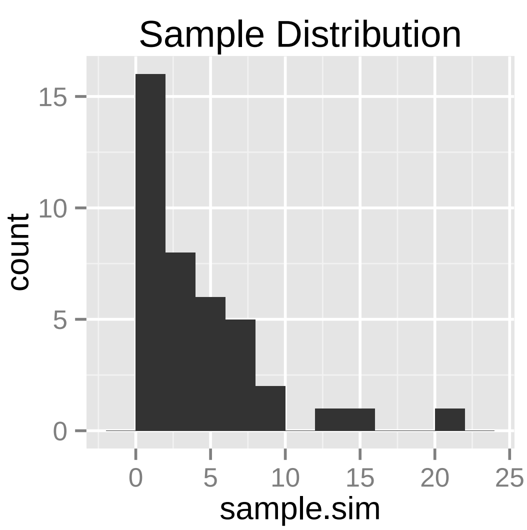
Simulations

To run the simulations the `rexp` function was used, with 40 simulations and lambda = 0.2. What follows is an example of this simulation.

```
n = 40
lambda = 0.2

sample.sim = rexp(n, lambda)

summary(sample.sim)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   0.146   1.395   3.046   4.332   5.984  20.270
```

```
qplot(sample.sim, binwidth = 2, main = "Sample Distribution")
```



We will now run 1000 of these simulations. The `exp.sims` variable is a matrix with 1000 rows (for each simulation group) and 40 columns (for each simulation in each group).

```
sims = 1000
n = 40
lambda = 0.2

exp.sims = matrix(
  data = rexp(n*sims, lambda),
  nrow = 1000,
  ncol = 40,
  byrow = TRUE)
```

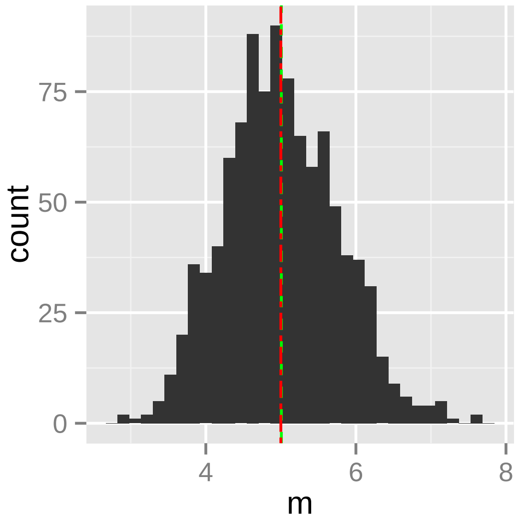
Sample Mean vs Theoretical Mean

Calculating the means for each of the simulations

```
exp.means = apply(exp.sims, 1, mean)
theo.mean = 1/lambda

c("Expected Mean" = theo.mean, "Actual Mean of Means" = mean(exp.means))
```

```
##        Expected Mean Actual Mean of Means
##             5.000000             5.006563
```

```
exp.means.plot = ggplot(
  data = data.frame(m = exp.means) %>% tbl_df,
  aes(x=m)) +

  geom_histogram() +
  geom_vline(
    xintercept = mean(exp.means),
    colour = "green",
    linetype = "dashed") +
  geom_vline(
    xintercept = theo.mean,
    colour = "red",
    linetype = "twodash")

exp.means.plot
```

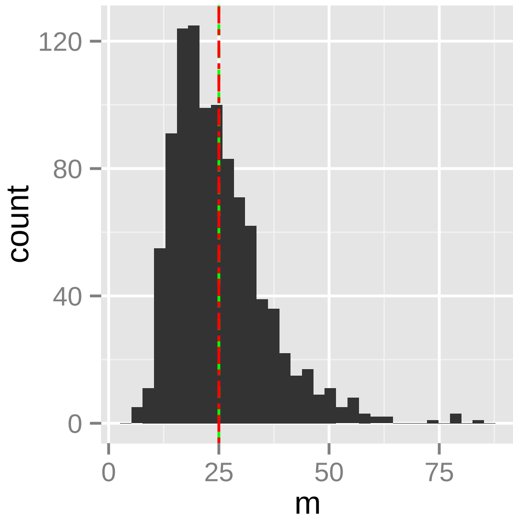



```r
theo.sd = 1/lambda
```

```r
exp.vars = apply(exp.sims, 1, var)
theo.var = (1/lambda)^2

c("Expected Variance" = theo.var, "Actual Mean of Variances" = mean(exp.vars))
```

```
##        Expected Variance Actual Mean of Variances
##                 25.00000                 24.99911
```

```r
exp.vars.plot = ggplot(
  data = data.frame(m = exp.vars) %>% tbl_df,
  aes(x=m)) +

  geom_histogram() +
  geom_vline(
    xintercept = mean(exp.vars),
    colour = "green",
    linetype = "dashed") +
  geom_vline(
    xintercept = theo.var,
    colour = "red",
    linetype = "twodash")

exp.vars.plot
```



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. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.