Statistical Inference Project 1 - Part 1

Joao Clemencio

27 September 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.

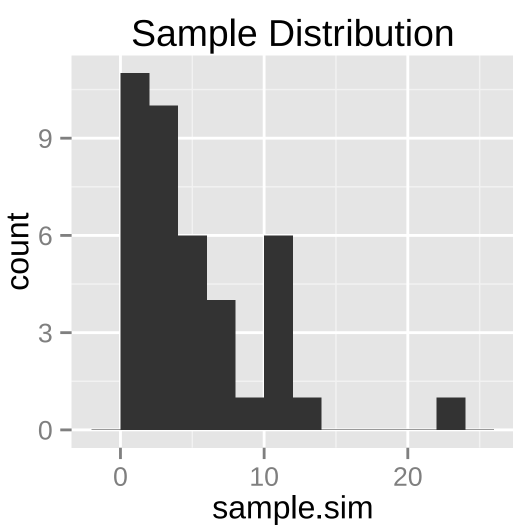
Simulations

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

```
set.seed(267)
n = 40
lambda = 0.2
sample.sim = rexp(n, lambda)
summary(sample.sim)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##  0.08953  1.87800  3.36700  5.05600  7.25100 22.74000
```

```
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

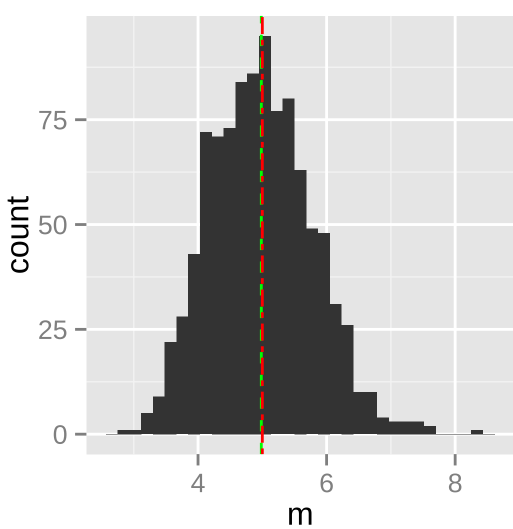
```
exp.means = apply(exp.sims, 1, mean)
theo.mean = 1/lambda

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

```
##        Expected Mean Actual Mean of Means
##             5.000000             4.984383
```

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

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




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

Sample Variance vs Theoretical Variance

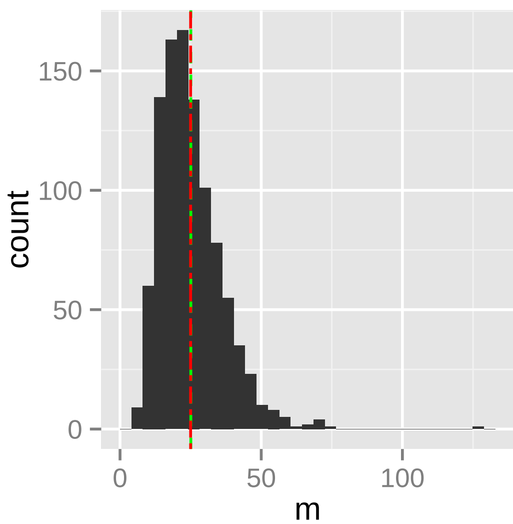
```
exp.vars = apply(exp.sims, 1, var)
theo.var = (1/lambda)^2

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

```
##        Expected Variance Actual Mean of Variances
##                 25.00000                 25.06968
```

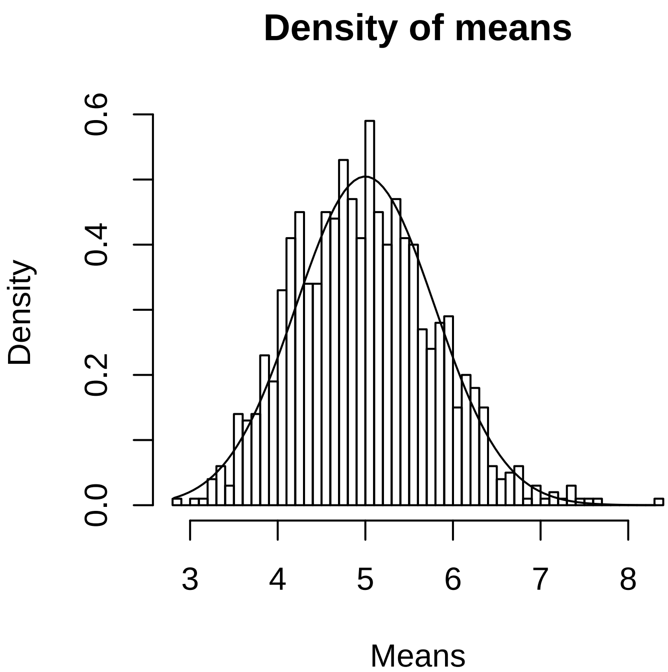
```
ggplot(data = data.frame(m = exp.vars) %>% tbl_df, aes(x=m)) + geom_histogram() + geom_vline(xintercept
```

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



Show that the distribution is apparently normal

```
# We first plot the means
hist(exp.means, breaks = n, prob = T, xlab = "Means", ylab="Density", main="Density of means")
# And then plot the distribution over it
x.fit = seq(min(exp.means), max(exp.means), length=100)
y.fit = dnorm(x.fit, mean = 1/lambda, sd = (1/lambda/sqrt(n)))
lines(x.fit, y.fit)
```



For a better view of the apparent normality, we can use a Q-Q plot.

```
qqnorm(exp.means)
qqline(exp.means)
```