Simulating the Exponential Distribution

Dhruv Mishra

Synopsis

This report investigates the exponential distribution in R and compares 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 and the standard deviation is also . For this simulation, we set . In this simulation, we investigate the distribution of averages of 40 numbers sampled from exponential distribution with .

The number of samples taken is 40 which are simulated a thousand times.

Simulations

Using the numerical values stated above, the code below evaluates the `rexp(n, lambda)` function a thousand times and stores the mean of each iteration in the `means` data frame.

```
set.seed(31)

## set constants
nosim <- 1:1000
n <- 40
lambda <- 0.2

## simulate the means
means <- data.frame(m = sapply(nosim, function(x) {
    mean(rexp(n, lambda))
}))
```

Sample Mean versus Theoretical Mean

The mean of the means data frame is the sample mean while the theoretical mean is

```
#sample mean
sample.mean <- mean(means$m)
#theoretical mean
theoretical.mean <- 1/lambda
```

The sample mean is **4.9938666** and the theoretical mean is **5**. They are very close, as expected by the Central Limit Theorem(CLT).

Sample Variance versus Theoretical Variance

```
## sample variance
sample.var <- var(means$m)
## theoretical variance
theoretical.var <- ((1/lambda)/sqrt(40))^2
```

The sample mean is **0.6291041** and the theoretical mean is **0.625**. They are also very close, as expected by the CLT.

Distribution

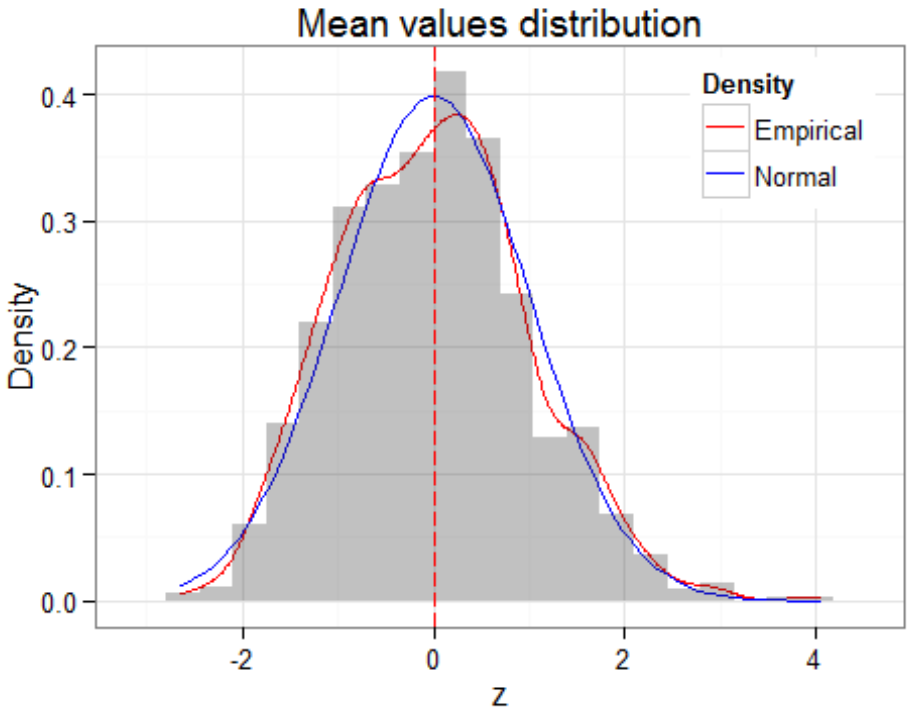
Also, according to the CLT, the distribution of the simulated means should be approximately normal. To illustrate this we will normalize the vectors and compare it to a distribution.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
zmean <- (means$m - sample.mean) / sqrt(sample.var)
qplot(zmean, geom = "blank") +
    geom_line(aes(y = ..density.., colour = 'Empirical'), stat =
'density') +
    stat_function(fun = dnorm, aes(colour = 'Normal')) +

    geom_histogram(aes(y = ..density..), , alpha = 0.3, binwidth=.35) +
    geom_vline(xintercept=0, colour="red", linetype="longdash") +
    scale_colour_manual(name = 'Density', values = c('red', 'blue')) +
    ylab("Density") + xlab("z") + ggtitle("Mean values distribution") +
    theme_bw() + theme(legend.position = c(0.85, 0.85))
```


Mean values distribution
Density
Empirical
Normal
0.4
0.3
0.2
0.1
0.0
Density
-2
0
2
4
z