Exponential Distribution and Central Limit Theory

Ronald Hensbergen

24 november 2016

Overview

In this project I investigate the exponential distribution (https://en.wikipedia.org/wiki/Exponential_distribution) and compare it to the Central Limit Theorem (CLT). As you may know, CLT shows that the distribution of averages of indipendent and identically distributed (iid) variables which are properly normalized follows the standard normal distribution as long as the sample size n is large enough.

The Exponential distribution I will simulate in R with `rexp(1/lambda)` where lambda (λ) is the rate parameter and I will set this at 0.2 for all simulations.

The last paragraph in the appendix contains information about the session of this document, such as the setting values and all the libraries loaded with their versions.

Simulations

The following I executed to come to the correct simulations:

- Set a seed, so the randomized numbers can be reproduced and the code in total is reproducible.
- Create an exponential distribution with 40 exponentials and 1000 simulations.
- Get the means of the 1000 simulations via `apply`.x

```
set.seed(1234)
n <- 40
lambda <- 0.2
simuls <- 1000
exp.dist <- data.frame(x = rexp(simuls * n, lambda))
exp.means <- data.frame(x = apply(matrix(exp.dist$x,simuls),1,mean))
exp.variances <- data.frame(x = apply(matrix(exp.dist$x,simuls),1,var))
```

Figure 1 in the appendix shows the exponential distribution compared to the theory.

- In the next paragraphs I get the mean and variance of the means to compare this to the theoretical mean and variance.

Sample Mean vs Theoretical Mean

Below code is executed to calculate the sample mean:

```
exp.s.mean <- mean(exp.means$x)
exp.s.mean
```

```
## [1] 4.974239
```

This mean is indeed quite close to the theoretical mean: $\mu = \frac{1}{\lambda} = \frac{1}{0.2} = 5$.

Figure 2 in the appendix shows the distribution of the means together with the mean of the means and also compared to a normal distribution. As the plot shows the distribution matches the normal distribution quite well, as CLT tells us.

Also, now compare the figures 1 and 2 and see the difference of the distributions. The first is the original distribution, and the second is the distribution of the means of the first distribution. One can wonder how those 2 are so different, since the origin is the same after all. This comes directly from the CLT: with a sample distribution the mean always approaches a normal distribution as the sample size increases.

Sample Variance vs Theoretical Variance

```
exp.s.variance <- var(exp.means$x)
exp.s.variance
```

```
## [1] 0.5949702
```

This variance is indeed quite close to the theoretical variance $var(\bar{X}) = \frac{\sigma^2}{n} = \frac{1}{40*0.2^2} = 0.625$.

For showing the distribution of the variances, I need to take another step

QQ Plot

Appendix

Figure 1

```r
library(ggplot2)
g <- ggplot(exp.dist, aes(x = x))
g <- g + geom_histogram(colour = "brown", fill = "brown", alpha = 0.1, binwidth = 1,
                        aes(y = ..density..))
g <- g + stat_function(fun = dexp, args = list(rate = lambda),
                       size = 0.5, colour = "red")
g <- g + ggtitle("Exponential distribution of\n1000 simulations with n=40")
g
```

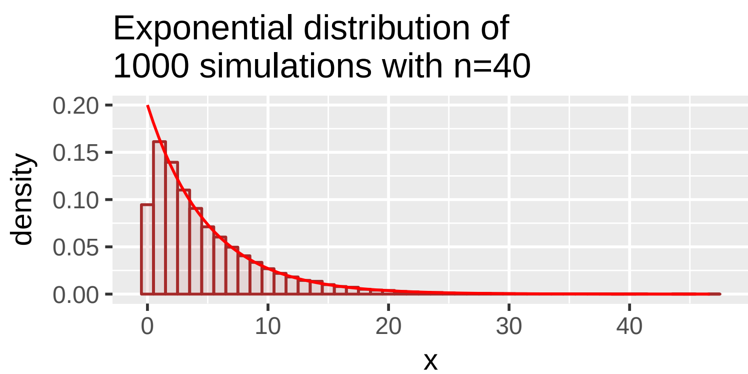


Figure 1: In above figure the histogram represents the exponential distribution I created at the beginning of this document, the red line shows the exponential distribution from the mathematical formula

Figure 2

```r
g <- ggplot(exp.means, aes(x = x))
g <- g + geom_histogram(alpha = .10, binwidth=0.3, colour = "brown", aes(y = ..density..))
g <- g + stat_function(geom = "line", fun = dnorm,
                       args = list(mean = 1/lambda, sd = 1/(sqrt(n)*lambda)),
              size = 1, colour = "red")
g <- g + geom_vline(xintercept = exp.s.mean, size = 2, color = "blue")
g <- g + ggtitle("Means of an Exponential distribution\ncompared to a normal distribution")
g
```

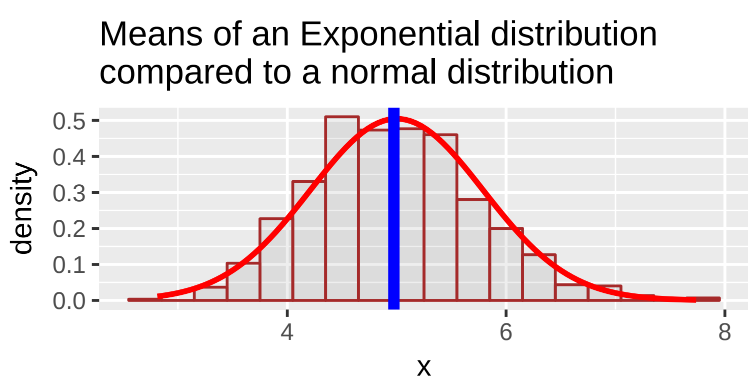
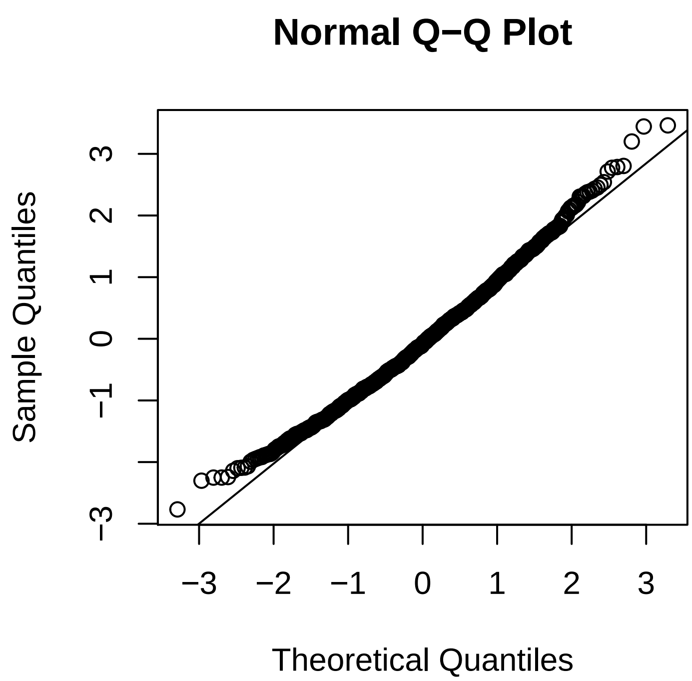


Figure 2 shows the distribution of means of 1000 simulations of an exponential distribution with n=40 (blue area) the blue vertical line the (sample) mean of means of that distribution. Finally the red line is a normal distribution with $\mu = 1/\lambda$ and $\sigma = 1/\lambda\sqrt{n}$.

Figure 3

```
x <- sqrt(n)*(lambda * exp.means$x -1)
qqnorm(x)
qqline(x)
```



Session information

```
library(devtools)
devtools::session_info()
```

```
## Session info -------------------------------------------------------------

##  setting  value
##  version  R version 3.3.2 (2016-10-31)
##  system   x86_64, mingw32
##  ui       RTerm
##  language (EN)
##  collate  Dutch_Netherlands.1252
##  tz       Europe/Berlin
##  date     2016-11-28

## Packages ------------------------------------------------------------------
```

```
##  package    * version date       source        
##  assertthat   0.1     2013-12-06 CRAN (R 3.3.1)
##  backports    1.0.4   2016-10-24 CRAN (R 3.3.2)
##  colorspace   1.3-1   2016-11-18 CRAN (R 3.3.2)
##  devtools   * 1.12.0  2016-06-24 CRAN (R 3.3.2)
##  digest       0.6.10  2016-08-02 CRAN (R 3.3.1)
##  evaluate     0.10    2016-10-11 CRAN (R 3.3.1)
##  ggplot2    * 2.2.0   2016-11-11 CRAN (R 3.3.1)
##  gtable       0.2.0   2016-02-26 CRAN (R 3.3.1)
##  htmltools    0.3.5   2016-03-21 CRAN (R 3.3.1)
##  knitr        1.15.1  2016-11-22 CRAN (R 3.3.2)
##  labeling     0.3     2014-08-23 CRAN (R 3.3.1)
##  lazyeval     0.2.0   2016-06-12 CRAN (R 3.3.1)
##  magrittr     1.5     2014-11-22 CRAN (R 3.3.1)
##  memoise      1.0.0   2016-01-29 CRAN (R 3.3.1)
##  munsell      0.4.3   2016-02-13 CRAN (R 3.3.1)
##  plyr         1.8.4   2016-06-08 CRAN (R 3.3.1)
##  Rcpp         0.12.8  2016-11-17 CRAN (R 3.3.2)
##  rmarkdown    1.2     2016-11-21 CRAN (R 3.3.2)
##  rprojroot    1.1     2016-10-29 CRAN (R 3.3.2)
##  scales       0.4.1   2016-11-09 CRAN (R 3.3.2)
##  stringi      1.1.2   2016-10-01 CRAN (R 3.3.1)
##  stringr      1.1.0   2016-08-19 CRAN (R 3.3.1)
##  tibble       1.2     2016-08-26 CRAN (R 3.3.1)
##  withr        1.0.2   2016-06-20 CRAN (R 3.3.1)
##  yaml         2.1.14  2016-11-12 CRAN (R 3.3.2)
```