Coursera - Statistical Inference - Course Project

Sina Eetezadi

September 24, 2015

Evaluation of the Central Limit Theorem for the exponential distribution

Overview

The goal of this project is to compute simulated values of the exponential distribution and assess the validity of the Central Limit Theorem. Specifically, the following three scenarios are illustrated.

1. Comparison of the simulated sample mean to the theoretical mean of the distribution.
2. Comparison of the simulated variance to the theoretical variance of the distribution.
3. Illustration that the distribution of means is approximately normal.

Initial setup

According to the instructions the following parameters were defined initially.

```
# Rate parameter as given in instructions
lambda <- 0.2

# Theoretical mean and sd
t_mean <- 1/lambda
t_sd <- t_mean

# Number of simulations and number of averages
nosim <- 1000
n <- 40

# Set seed for reproducability of simulations
set.seed(7)
```

Task 1: Simulated Mean vs Theoretical Mean

The following computes the mean sample mean of n means of nosim simulation of rexp().

```
# Simulation
means <- NULL
for (i in 1:n) means = c(means, mean(rexp(nosim, lambda)))
mean(means)
```

```
## [1] 4.983294
```

```
# Theoretical mean
print(t_mean)
```

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

It was shown that the simulated mean approximates the theoretically known mean.

Task 2: Simulated Variance vs Theoretical Variance

The following computes the mean sample variance of n variances of nosim simulations of rexp().

```
# Simulation
vars <- NULL
for (i in 1:n) vars = c(vars, var(rexp(nosim, lambda)))
mean(vars)
```

```
## [1] 24.91688
```

```
# Theoretical variance
print(t_sd ^ 2)
```

```
## [1] 25
```

It was shown that the simulated variance approximates the theoretically known variance.

Task 3: Distribution of Means follows Normal Distribution

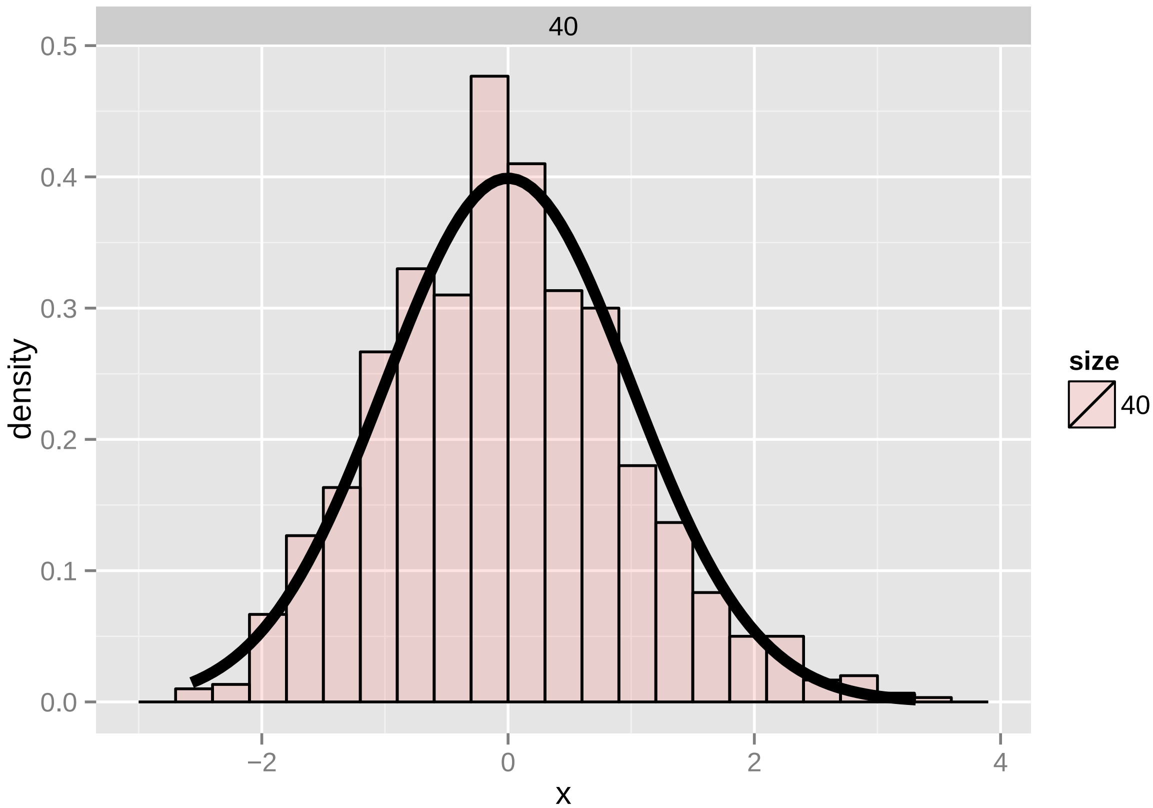
The following creates a matrix of nosim * n simulated means of rexp() in nosim rows. Therefore each row has n values (columns). Each value is normalized using cfunc to compare to the standard normal distribution. In the plot the histogram of the normalized, simulated data is shown and overlayed with the standard normal density function.

```
  # Correction function to calculate back to standard normal
  cfunc <- function(x, n) sqrt(n) * (mean(x) - t_mean) / t_sd

  # Matrix with n columns and nosim rows filled with exponentials
  mat <- matrix(rexp(n * nosim, lambda), nosim)

  # Apply cfunc for every row with n values
  # Results in distribution of means normalized to standard normal
  dist <- apply(mat, 1, cfunc, n)

  # Plotting
  library(ggplot2)
  dat <- data.frame(x = dist, size = factor(rep(n, nosim)))
  g <- ggplot(dat,
              aes(x = x, fill = size)) + geom_histogram(alpha = .20,
                                                        binwidth=.3,
                                                        colour = "black",
                                                        aes(y = ..density..))
  g <- g + stat_function(fun = dnorm, size = 2) # Normal distribution density function
  g + facet_grid(. ~ size)
```



It was shown that the distribution of means follows the normal distribution and therefore the central limit theorem is valid.

Conclusion

All three simulations illustrate the validity of the Central Limit Theorem for the exponential distribution.