Statistical Inference - Peer Graded Assignment

Pete Petersen III

6/20/2020

Part 1: Simulation Exercise

Overview:

In this paper we demonstrate that the Central Limit Therom is in congruence with theoretical normal distributions. In order to accomplish this we will generate a 1000 columns and 40 rows, We will then take the means of each of these columns. This will show that the density of the means is very sumilar to the theoretic density. The siulation will use rexp(n,lambda) where lambda is the rate parameter and n is the number of rows. rexp uses Ahrens, J. H. and Dieter, U. (1972). Computer methods for sampling from the exponential and normal distributions. Communications of the ACM, 15, 873–882.

The Exponential Distribution Simulation:

The mean of distribution and the standard deviation are both 1/lambda are given as lambda = 0.2, we will perform 1000 averages.

```
Usage
   dexp(x, rate = 1, log = FALSE)
   pexp(q, rate = 1, lower.tail = TRUE, log.p = FALSE)
   qexp(p, rate = 1, lower.tail = TRUE, log.p = FALSE)
   rexp(n, rate = 1)
```

Arguments: x, q vector of quantiles. p vector of probabilities. n number of observations. If length(n) > 1, the length is taken to be the number required. rate vector of rates.
log, log.p logical; if TRUE, probabilities p are given as

$$log(p)$$

. lower.tail logical; if TRUE (default), probabilities are

$$P(X \leq x)$$

, otherwise,

$$P(X > x)$$

.

```
 # Set the initial simulations paremeters wher lamda is the the rate parameter in
 # exponential distrubution.
  lambda <- 0.2
# Set the seed for the random number generator for reproducabkility
  set.seed(11)
```

```r
  # Set the number of observations. If length(n) > 1, the length is taken to be the number
  # required
  n <- 40
  # Iterate  a thousand times of the exponential distribution where length = 40 and lamda = .2
  # This will result in  a 40 row by 1000 column array.
  simulation <- replicate(1000, rexp(n, lambda))

# we use the apply to take means over the simulation distribution columns and arrive ar 1 mean per colu
  mean_simulation <- apply(simulation, 2, mean)
```

Sample Mean versus Theoretical Mean: Include figures with titles. In the figures, highlight the means you are comparing. Include text that explains the figures and what is shown on them, and provides appropriate numbers.

```r
  sampleMean <- mean(mean_simulation)
  simulationmean <- mean(simulation)
  theoretical_mean <- (1/lambda)
```

The mean of the simulations is 4.9871567
The theoretical mean of:

$$\mu = \frac{1}{\lambda} = .5$$

Sample Variance versus Theoretical Variance: Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

```r
# sample deviation & variance
sample_sd <- sd(mean_simulation)
sample_var <- sample_sd^2
simulationmean <- mean(mean_simulation)
# theoretical deviation & variance
theoretical_sd <- (1/lambda)/sqrt(n)
theoretical_var <- ((1/lambda)*(1/sqrt(n)))^2
theoretical_mean <- 1/lambda
```

We now compare data from the simulation to the values the CTL. CTL predicts that the distribution of the simulation should be very close to normal distribution.

Defined As

$$N(\mu, \sigma)$$

where

$$\mu = \frac{1}{\lambda}$$

and

$$\sigma = \frac{1}{\lambda / n^2}$$

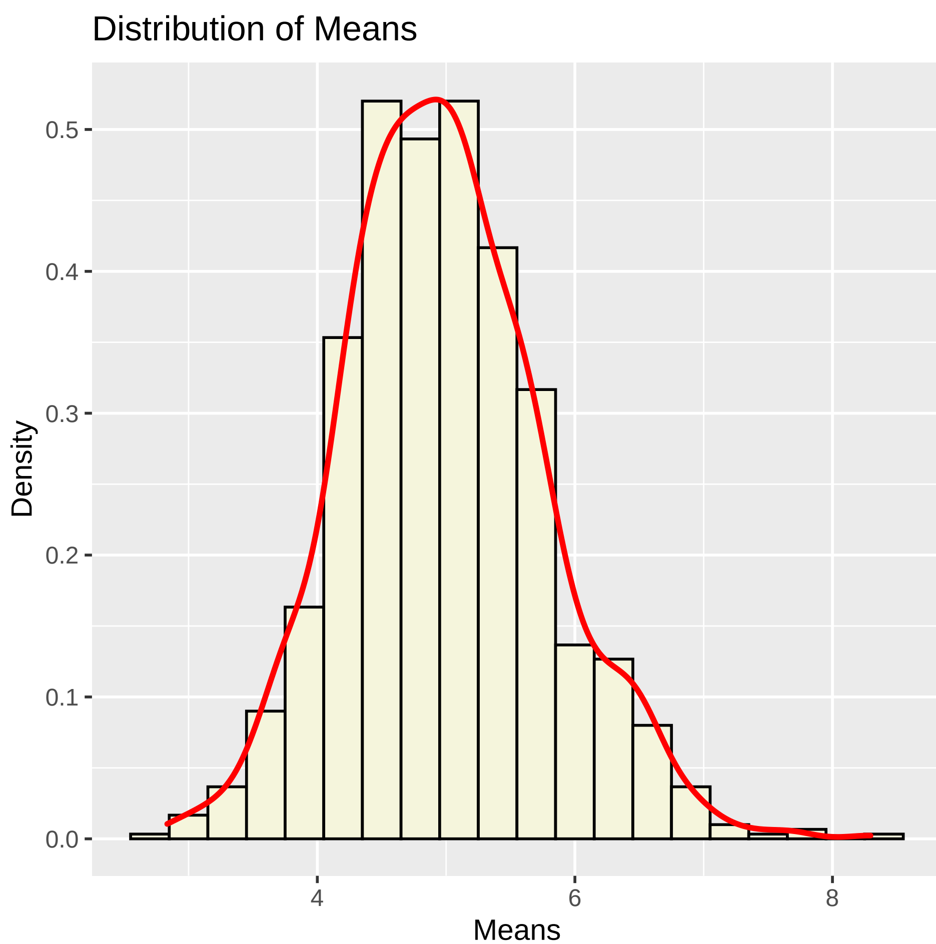

Figure 1: Fig. 1 Simulation Distribution snd Density

Table 1: Simulated vs Theoretical Values

	Mean	Std	Variance
Theoretical	5.000	0.791	0.625
Simulation	4.987	0.775	0.601

Distribution:

In the figure below, we plot the histogram of the means and the density of the means in order to note the appearant normalcy. We also note that the distribution appears centered around.5 and is symetric. Further, the similarity of the simulation density in blue to the theoretical density in red highlights the proximity to normalcy of the simulation means which exemplifies the CLT's correctneess.

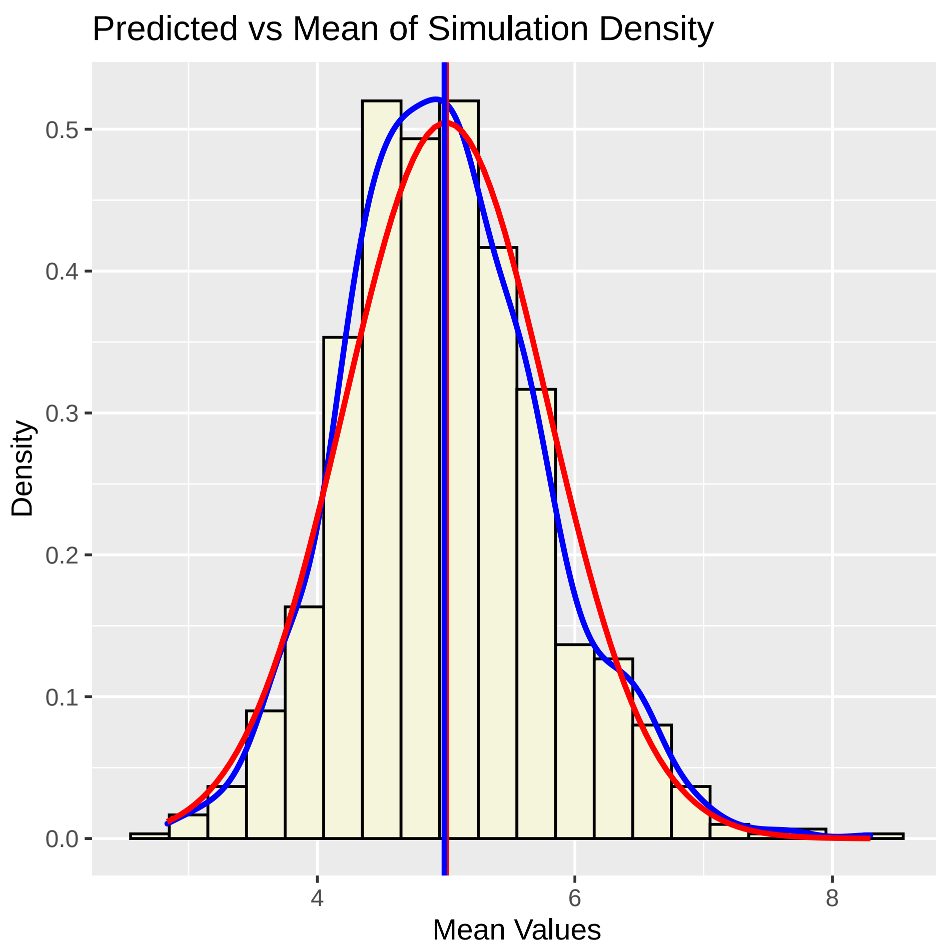

Figure 2: Fig. 2 Distribution Analysis

Q-Q Normal Plot also demonstrates a normal distribution. qqnorm is a generic function the default method of which produces a normal QQ plot of the values in y. qqline adds a line to a "theoretical", by default normal, quantile-quantile plot which passes through the probs quantiles, by default the first and third quartiles. The proximity to the xy plot of the simulation pairs demonstrated very little deviation from normalcy.

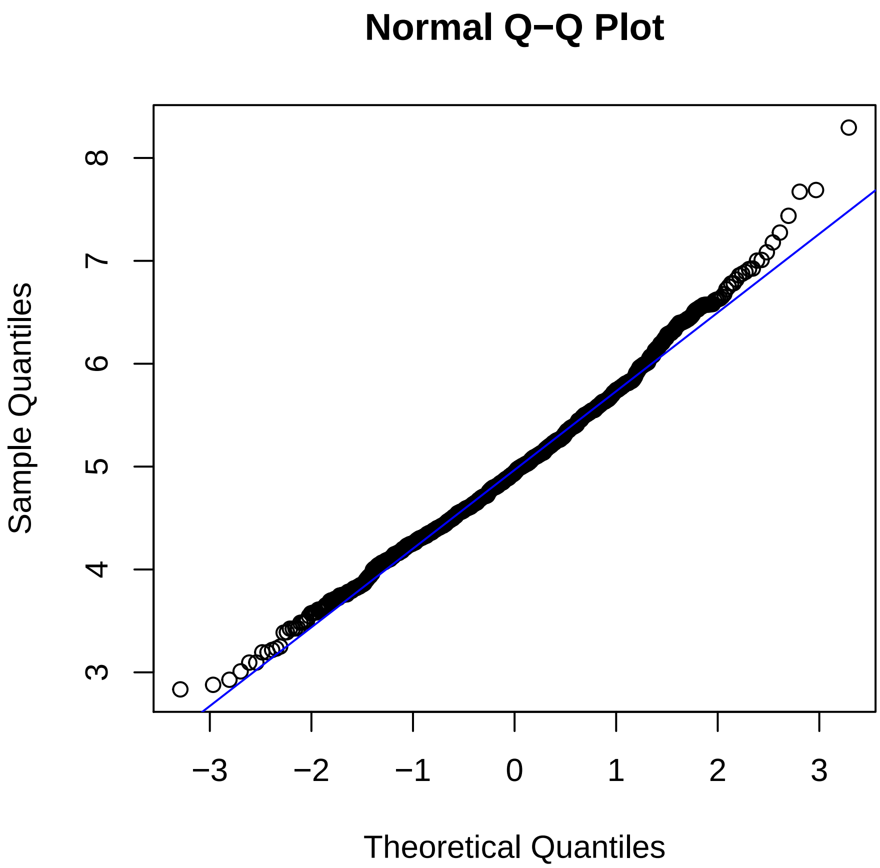

Figure 3: Fig. 3 Distribution Analysis