Exponential distribution simulation in R and comparison to the Central Limit Theorem

Wei Wei

May 15, 2016

Synopsis

The Central Limit Theorem states that the distribution of the mean of independent and identically distributed random variables becomes that of a normal distribution as the sample size increase, regardless of the underlying distribution. In this simulation analysis, the exponential distribution is investigated and is found to conform to the Central Limit Theorem.

Simulation of 1000 means of 40 exponentials

```
lambda <- 0.2
n <- 40
nosim <- 1000

## simulation: generate 1000 means of 40 exponentials
means <- NULL
for (i in 1:nosim) {
      means[i] <- mean(rexp(n, lambda))
}

str(means)
```

```
##  num [1:1000] 4.31 4.68 4.35 4.31 5.46 ...
```

A numeric vector called "means" containing 1000 means of 40 exponentials is generated in R by simulation.

Comparison of the sample mean and theoretical mean

```
## sample mean from the 1000 simulations
mean_sim <- mean(means)

## theoretical mean
mean_t <- 1/lambda

## comparison of the sample mean and theoretical mean
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.2.5
```

```
mean_df <- rbind(c("Sample mean", "Theoretical mean"), c(round(mean_sim, 2), mean_t))
kable(mean_df, caption="Comparison of Sample mean and Theoretical mean")
```

Table 1: Comparison of Sample mean and Theoretical mean

Sample mean	Theoretical mean
4.99	5

The sample mean of the 1000 means of 40 exponentials approximates the theoretical mean.

Comparison of the sample variance and theoretical variance

```
## sample variance from the 1000 simulations
var_sim <- var(means)

## theoretical variance
sd_t <- 1/lambda
var_t <- sd_t^2/n

## comparison of the sample variance and theoretical variance
var_df <- rbind(c("Sample variance", "Theoretical variance"), c(round(var_sim, 3), var_t))
kable(var_df, caption="Comparison of Sample variance and Theoretical variance")
```

Table 2: Comparison of Sample variance and Theoretical variance

Sample variance	Theoretical variance
0.589	0.625

The sample variance of the 1000 means of 40 exponentials approximates the theoretical variance that of a sample size 1000.

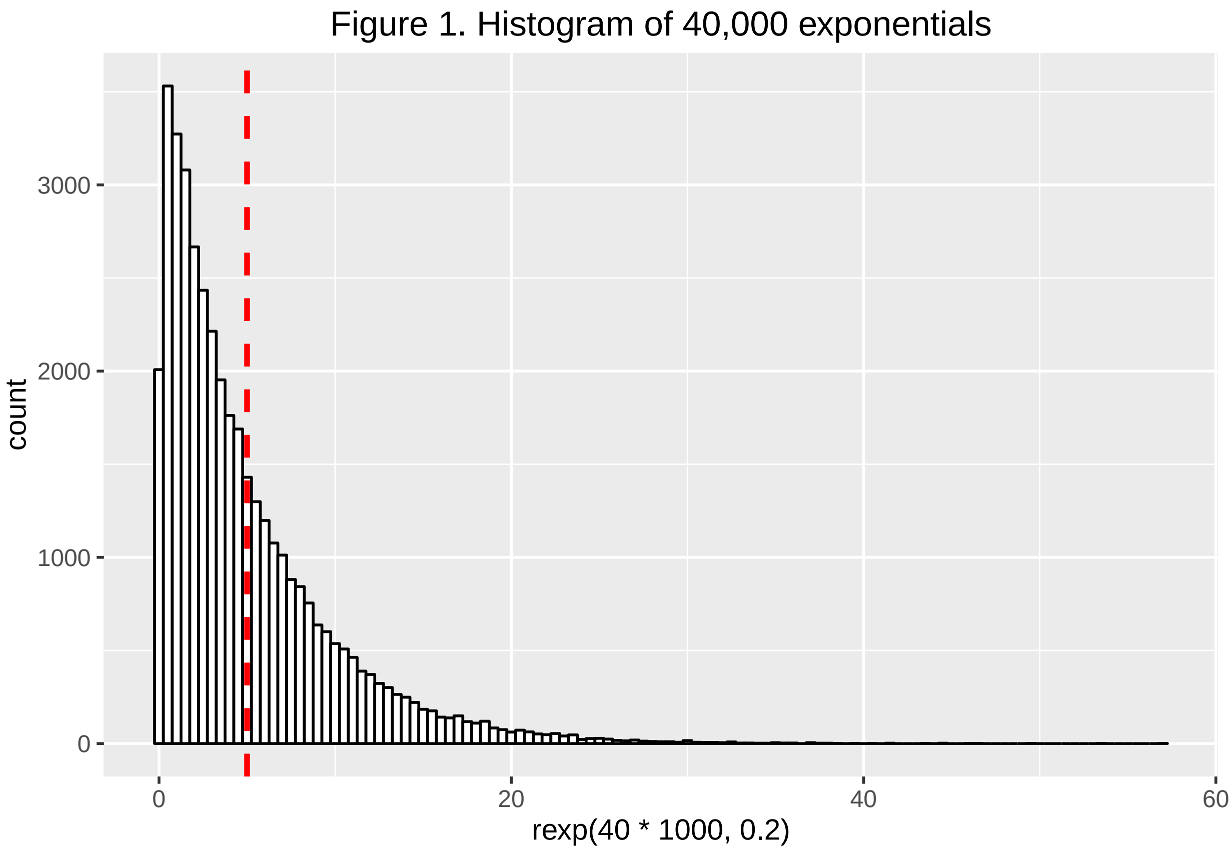
Test normal distribution of simulated sample

Compared to 40 * 1000 exponentials (see Appendix Figure 1, red dashed line = theoretical mean), the 1000 means of 40 exponentials have a apparently normal distribution (see Appendix Figure 2, red dashed line = theoretical mean) which is centered on the theoretical mean. The simulated 1000 means of 40 exponentials do not appear to substantially deviate from normality using the normal probability plot (see Appendix Figure 3, red line = normality).

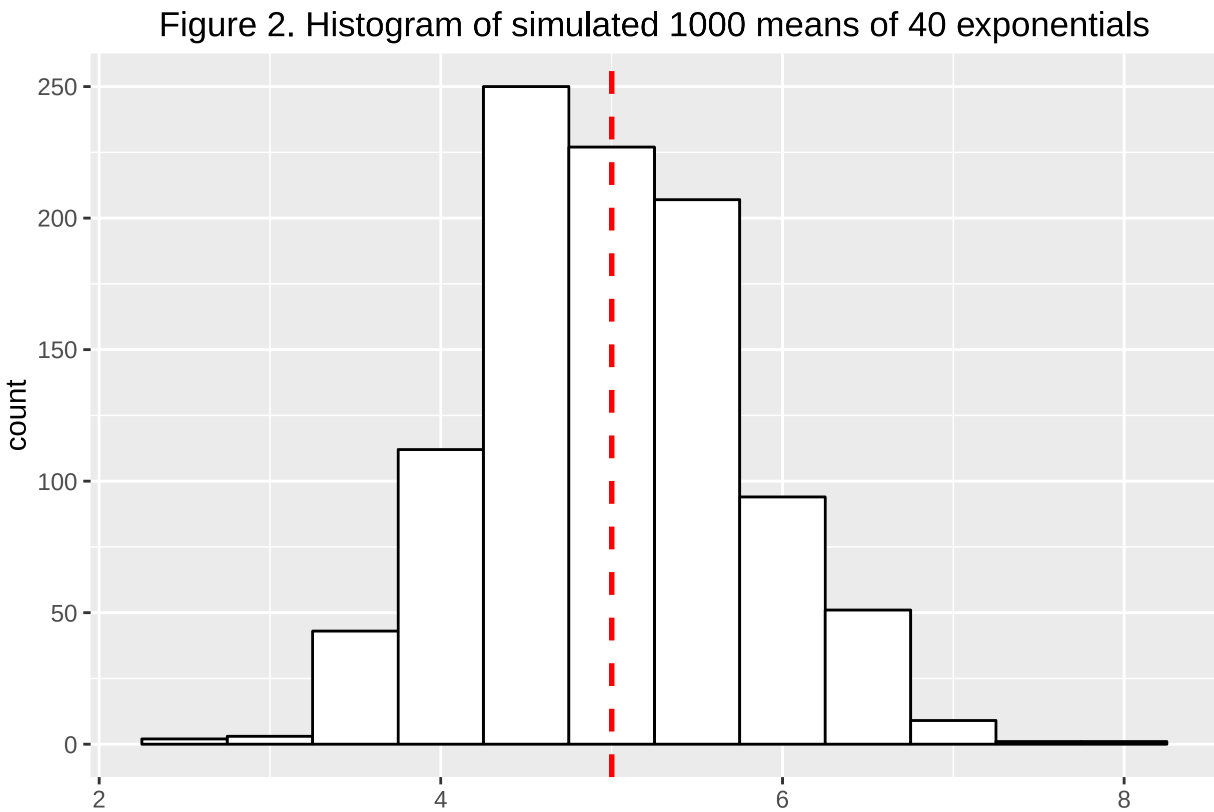
Appendix

```
## Appendix figure 1. histogram of 1000 * 40 exponential distribution with lambda = 0.2
library(ggplot2)
g <- ggplot()
g <- g + aes(rexp(40 * 1000, 0.2))
```

```r
g <- g + geom_histogram(binwidth=.5, colour="black", fill="white")
g <- g + geom_vline(aes(xintercept=mean_t), color="red", linetype="dashed", size=1)
g <- g + ggtitle("Figure 1. Histogram of 40,000 exponentials")
g
```

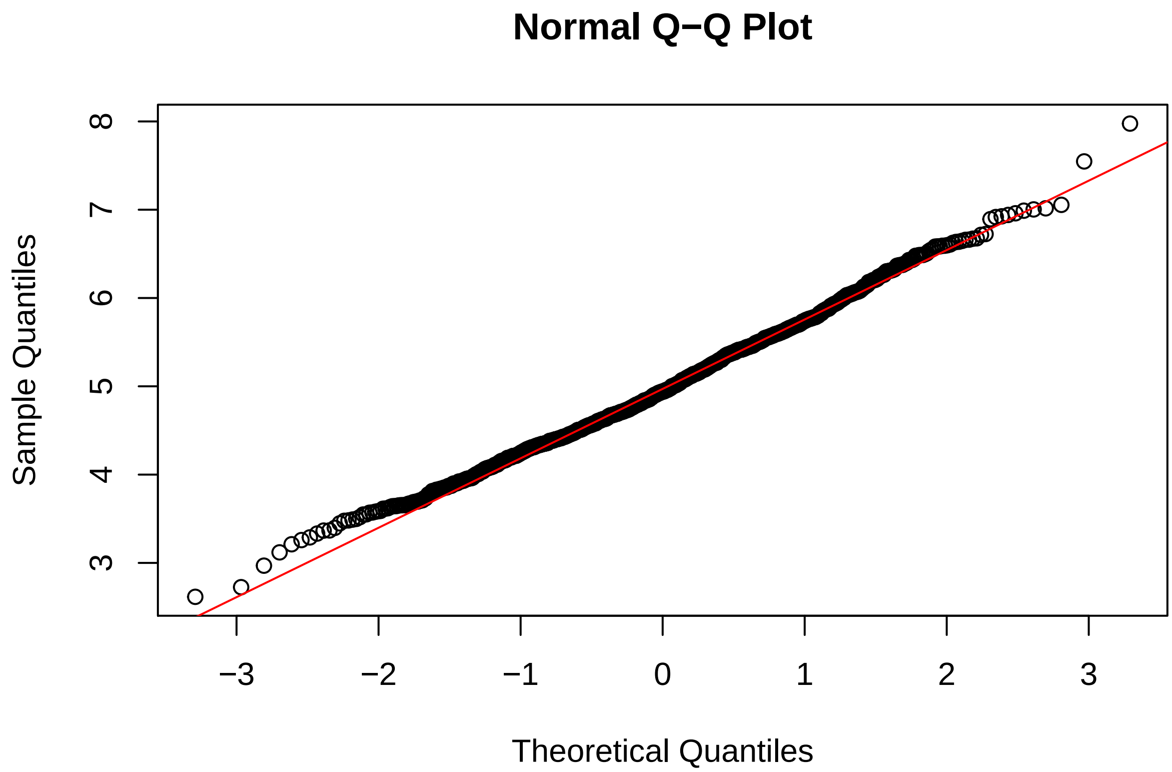



```r
## Appendix figure 2. histogram of the 1000 means of 40 exponentials by simulation
g <- ggplot()
g <- g + aes(means)
g <- g + geom_histogram(binwidth=.5, colour="black", fill="white") + labs(x='')
g <- g + geom_vline(aes(xintercept=mean_t), color="red", linetype="dashed", size=1)
g <- g + ggtitle("Figure 2. Histogram of simulated 1000 means of 40 exponentials")
g
```




```
## Appendix figure 3. use normal probability plot to check normality in the simulation sample
qqnorm(means)
qqline(means, col="red")
mtext(side=3, line=-1, text="Figure 3", font=2, adj=0, outer=TRUE)
```

Figure 3


Normal Q−Q Plot
Sample Quantiles
3
4
5
6
7
8
−3
−2
−1
0
1
2
3
Theoretical Quantiles