The Distribution of The Mean of 40 Exponentials

Hai Bacti

August 26, 2020

Synopsis

In this report, we will examine the distribution of averages of 40 exponentials and compare it with the Central Limit Theorem. We generate simulation from the rate parameter $\lambda = 0.2$, do calculation and use plots to show how sample distribution is approximately to the standard normal distribution.

Simulations

In probability theory and statistics, the exponential distribution is the probability distribution of the time between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate.

We simulate a sample of 1000 exponentials of rate $\lambda = 0.2$ and plot a histogram of the generated data in **Figure 1**. The yellow vertical line intercepts x-axis at mean $1/\lambda = 5$.

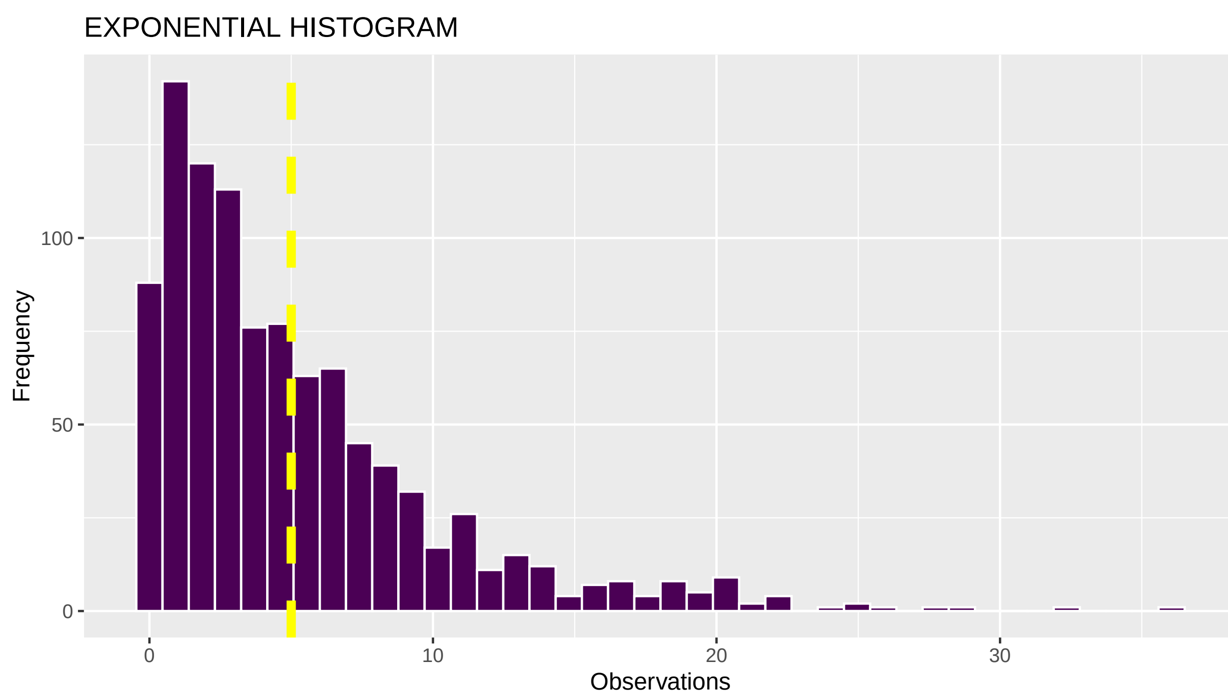

Figure 1: The Exponential Distribution

To examine the distribution of the mean of exponentials, we make 1000 simulations of the averages of 40 exponentials, randomly generated using R with rate $\lambda = 0.2$. So the theoretical mean is $1/\lambda = 1/0.2 = 5$ and the theoretical variance is $\frac{1}{\lambda^2 * n} = \frac{1}{0.2^2 * 40} = 0.625$.

```
# draw 1000 simulations of the averages of 40 exponentials
dat <- data.frame(X_bar = apply(matrix(data = rexp(nosim * 40, rate = lambda),
                                       ncol = 40),
                                MARGIN = 1, FUN = mean))
```

Now check mean of the sample

```
# sample mean
mean(dat$X_bar)
```

```
## [1] 5.012962
```

and the variance of the sample

```
# sample variance
var(dat$X_bar)
```

```
## [1] 0.6078168
```

The sample mean and variance of simulations are quite close to theoretical values.

In **Figure 2**, we show the histogram of the sample distribution, draw the probability density curve in red and the theoretical density curve of normal distribution $\mathcal{N}(\frac{1}{\lambda}, \frac{1}{\lambda^2 * 40})$ in green.

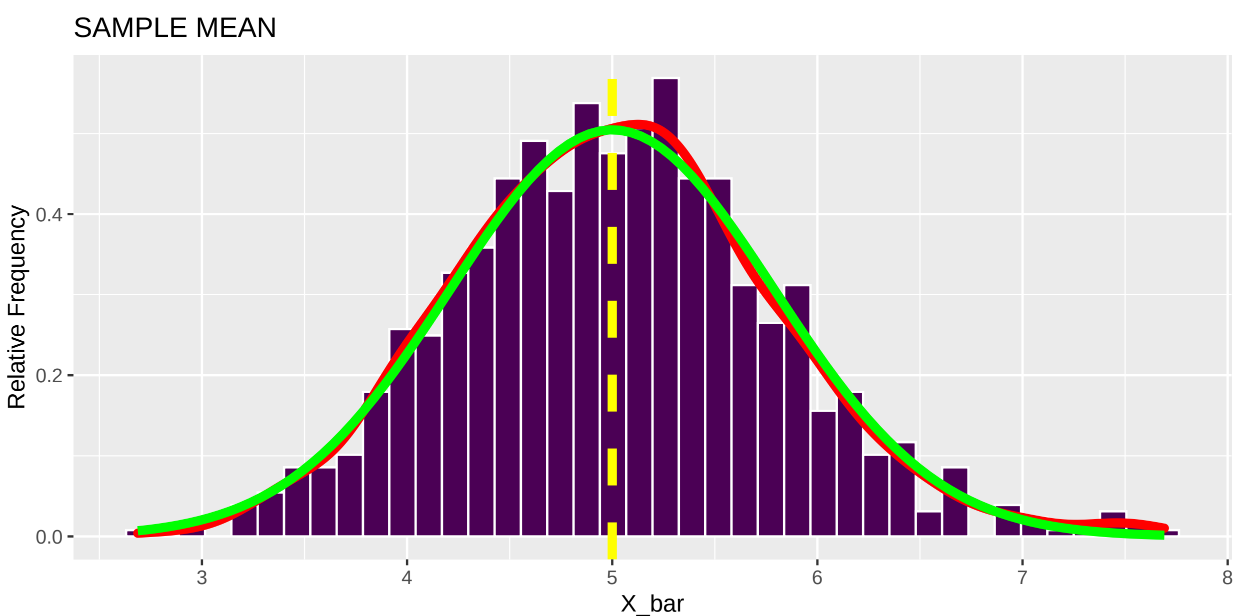

Figure 2: The Distribution of Means of 40 Exponentials

We also draw a yellow vertical line intercepts x-axis at mean.

Conclusion

We can easily see that contrast to the exponential distribution, our new sample draws the symmetric distribution and looks likely standard Normal distribution with a mean of $\frac{1}{\lambda}$. Moreover, both sample and theoretical density curves are almost matched.

To verify the sample curve is normally distributed, we deploy Q-Q plot in **Figure 3** over all simulations. In statistics, Q-Q (quantile-quantile) plots play a very vital role to graphically analyze and compare two probability distributions by plotting their quantiles against each other. If the two distributions which we are comparing are exactly equal then the points on the Q-Q plot will perfectly lie on a straight line $y = x$.

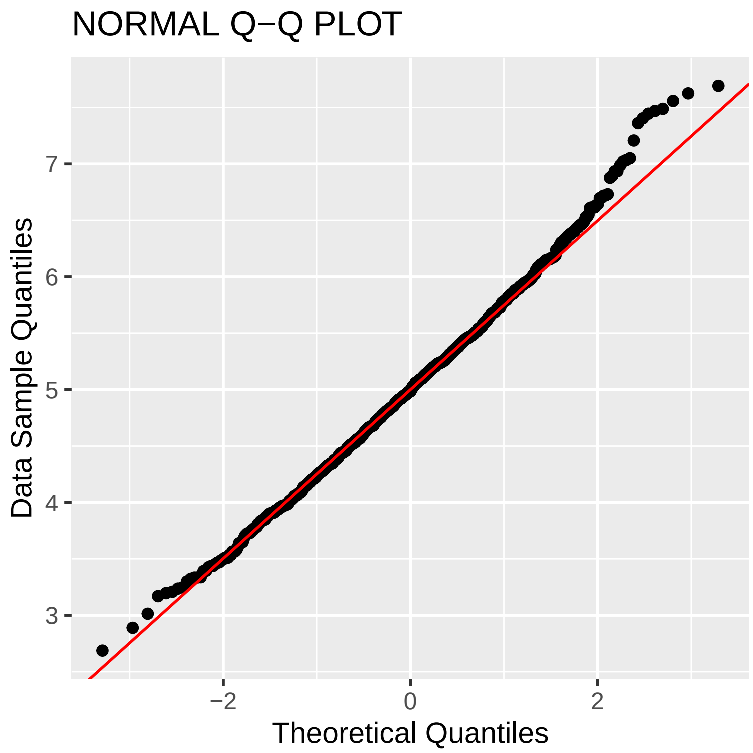

Figure 3: Q-Q plot

So it's clear that the sample observations fall about a straight line, especially in the middle. We confirm that the distribution of means of 40 exponentials is approximately normally distributed, and behave as predicted by the Central Limit Theorem.

Appendix

```r
library(ggplot2)
library(dplyr)

set.seed(123)
nosim <- 1000 # number of simulations
lambda <- 0.2 # the rate parameter of exponential distribution
T_mean = 1 / lambda # theoretical mean
T_variance = 1 / (lambda ** 2) / 40 # theoretical variance
x <- rexp(nosim, rate = lambda)
```

```r
# draw 1000 simulations of the averages of 40 exponentials
dat <- data.frame(X_bar = apply(matrix(data = rexp(nosim * 40, rate = lambda),
                                       ncol = 40),
                                MARGIN = 1, FUN = mean))
```

```r
# sample mean
mean(dat$X_bar)
```

```r
# sample variance
var(dat$X_bar)
```

```r
# the histogram of exponential distribution
qplot(x = x, geom = 'histogram',
      colour = I('white'), fill = I('#4B0055'), bins = 40) +
    geom_vline(xintercept = T_mean, colour = 'yellow', lty = 'dashed', size = 2) +
    xlab('Observations') +
    ylab('Frequency') +
    ggtitle('EXPONENTIAL HISTOGRAM')
```

```r
# show the distribution of Means of 40 Exponentials
# we also draw the sample and theoretical probability density curves
ggplot(dat, aes(x = X_bar, colour = I('white'))) +
    geom_histogram(aes(y = ..density..), bins = 40, fill = I('#4B0055')) +
    geom_density(aes(y = ..count.. / nosim), color = 'red', lwd = 2) +
    geom_function(fun = dnorm, args = list(mean = T_mean, sd = sqrt(T_variance)),
                  colour = 'green', size = 2) +
    geom_vline(xintercept = T_mean, colour = 'yellow', lty = 'dashed', size = 2) +
    ylab('Relative Frequency') +
    ggtitle('SAMPLE MEAN')
```

```r
# test quantiles of the sample simulations
segment <- c(0.25, 0.75)
slope <- diff(quantile(dat$X_bar, segment)) / diff(qnorm(segment))
ggplot(dat, aes(sample = X_bar)) +
    geom_qq() +
    geom_abline(slope = slope, intercept = T_mean, colour = 'red') +
    xlab('Theoretical Quantiles') +
    ylab('Data Sample Quantiles') +
    ggtitle('NORMAL Q-Q PLOT')
```