The Exponential Distribution and The Averages Distribution

Guimiao Zhang

Overview

This project investigates the exponential distribution and the averages distribution in R. The mean of the exponential distribution and the standard deviation are all $\frac{1}{\lambda}$. The distribution of the averages of 40 exponentials would be investigated and compared with the original exponential distribution and normal distribution.

Simulations

For the illustration purpose, set $\lambda = 0.2$. 1000 simulations were obtained.

```
rep <- 1000; lambda <- 0.2; n <- 40
set.seed(4)
exps <- rexp(rep, lambda) # original exponential
expsim <- function(rep, n, lambda) {
    mat <- matrix(rexp(rep * n, lambda), nrow = rep, ncol = n, byrow = T)
    return (rowMeans(mat))
}
simed <- expsim(rep, n, lambda) # the ave. of 40 exponentials
```

Sample Mean v.s. Theoretical Mean

```
means <- c('theoretical' = mean(exps), 'sample' = mean(simed))
# set cols with transparency
cols <- c(rgb(0, 0, 1, 0.4), rgb(1, 0, 0, 0.4))
# hist exp
hist(exps, col = cols[1], breaks = 30, xlab = '',
     xlim = c(0, 38), ylim = c(0, 250), main = 'Histogram of 1000 simulations')
hist(simed, col = cols[2], breaks = 10, add = T)
# lines for mean
abline(v = means[1], col = cols[1], lwd = 1.8)
abline(v = means[2], col = cols[2], lwd = 1.8)
legend(14.5, 250, bty = 'n', lty = 1, lwd = 1.8, col = cols,
       legend = c(paste('mean of theoretical exponetial: ',
                        round(means[1], 3)),
                  paste('mean of averages of 40 exponetial samples: ',
                        round(means[2], 3))))
legend(15, 200, bty = 'n', fill = cols,
       legend = c('histogram of theoretical exponetial',
                  'histogram of averages of 40 exponetial samples'))
```

Theoretically, the mean of the exponential should be $\frac{1}{\lambda} = 5$. The simulations show both the exponential distribution and the averages of 40 exponentials have the mean very close to 5 (Figure 1).

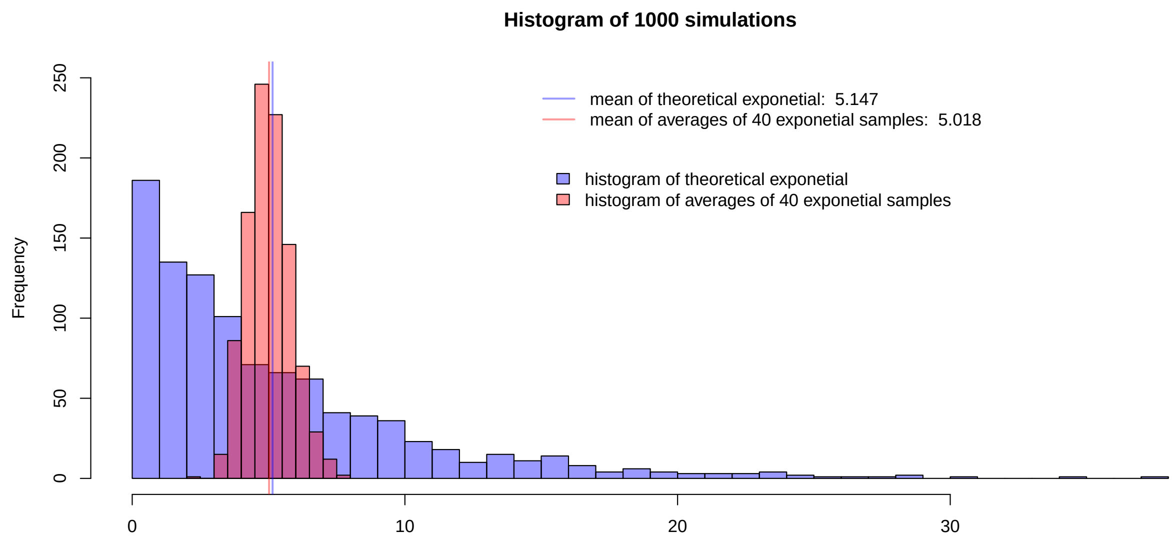

Figure 1: The Distributions of 1000 Simulations

Sample Variance v.s. Theoretical Variance

Theoretically, the variance (σ^2) of the exponential distribution is as follows: $X \sim Exp(\lambda) \Rightarrow \sigma^2 = \frac{1}{\lambda^2} = 25$. From figure 1 above, it is obvious that the exponential distribution has a wider spread than the distributions of averages of 40 exponentials, indicating a larger variance.

```
vars <- c('theoretical' = var(exps), 'sample' = var(simed))
relation <- vars[1]/n
names(relation) <- 'theoretical/n'
vars; relation
```

```
## theoretical      sample
##  27.0254576   0.6415888

## theoretical/n
##     0.6756364
```

From the simulation results, we find the simulated exponential distribution has variance close to the theoretical result, whereas the variance for averages of 40 exponentials is approximately one fortieth of the theoretical exponential variance.

Distributions

```
simed.CI <- quantile(simed, c(0.025, 0.975)) # 95% CI for simulation
norm.CI <- qnorm(c(0.025, 0.975), 1/lambda, 1/lambda/sqrt(n))  # 95% CI for normal
hist(simed, freq = F, col = NA, xlab = 'Simulated Averages of 40 Exponetials',
     main = expression(paste("The Distribution of Averages of 40 Exp(",
                             lambda, "= 0.2)")), breaks = seq(2, 9, by = 0.2))
lines(density(simed), col = 'blue', lwd = 1.8)
curve(dnorm(x, 1/lambda, 1/lambda/sqrt(n)), add = T, col = 'red', lwd = 1.8)
abline(v = simed.CI, col = 'blue', lwd = 1.8, lty = 2)
abline(v = norm.CI, col = 'red', lwd = 1.8, lty = 2)
legend('topright', lwd = 1.8, bty = 'n', lty = c(1, 1, 2, 2), col = c('red', 'blue'),
```

```r
      c('Theoretical Normal Density Curve', 'Simulated Empirical Density Curve',
        'Theoretical Normal C.I. limits', 'Simulated Empirical C.I. limits'))
```

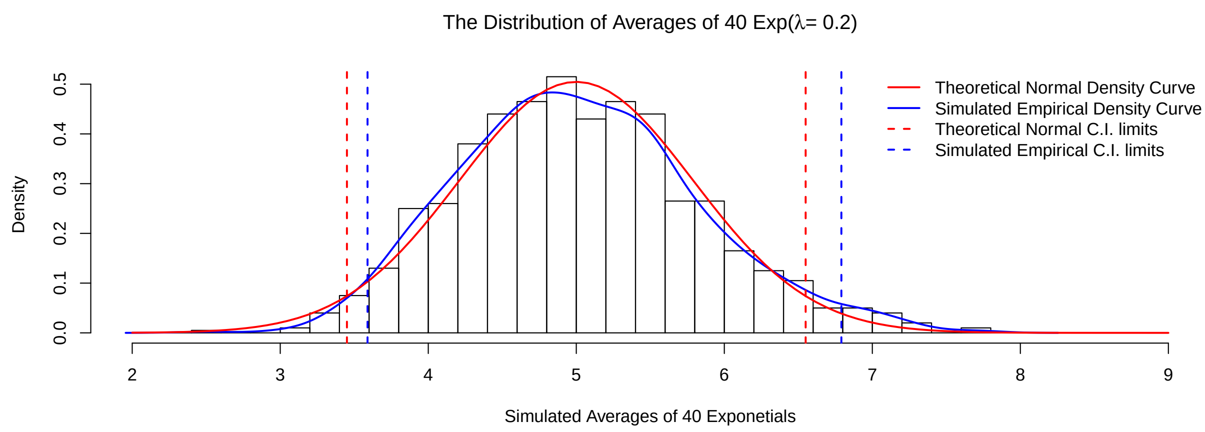

Figure 2: The Distribution Comparison for Normal & Simulation

```r
qqnorm(simed); qqline(simed, lty = 2, col = 'red')
```

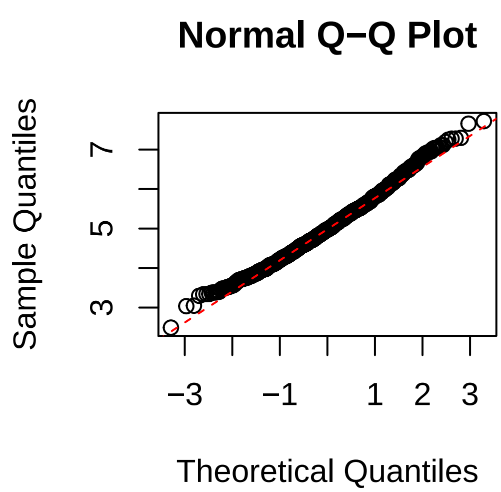

Figure 3: Normal Q-Q Plot

Figure 2 shows the distribution of the simulated averages is very similar to $N(mean = \frac{1}{\lambda}, variance = \frac{1}{n\lambda^2})$. The 95% Confidence Interval (C.I.) overlaps with that of $N(mean = \frac{1}{\lambda}, variance = \frac{1}{n\lambda^2})$. Figure 3 shows the sample quantiles roughly fall on the line of theoretical quantiles, indicating a roughly normal distribution. In addition, from the previous 2 sections, the mean and variance of the distribution of the averages are approximately equal to 5 and 0.625, respectively.

Theoretically, by taking n number of i.i.d samples from a distribution with $mean = \mu$ and $variance = \sigma^2$, the averages of the n samples will be asymptotically normally distributed with $mean = \mu$ and $variance = \sigma^2/n$.

Therefore, the observations are consistent with the theoretical expectation.