Simulation Exercise

yangru1q1

2019-06-27

```r
library(tidyverse)
```

Synopsis

The main goal for this project is to explore the exponential distributin and compare it with the Central Limit Theorem. We can ovserve from the output that the sample mean estimate the theoretical mean very well; the sample variance is closed to the theoretical variance; the kernal density for the distribution of the average of 40 exponential with rate 0.2 is approximately a normal distribution.

Create the dataset

Let's create a matrix with 10000 rows, with 40 random generated number from $exp(0.2)$ in each column. To make the result from this report reproduciable, let's set a seed before the simulation.

```r
set.seed(1)
# number of simulations
m <- 1000
# number of observations in each group
n <- 40
# rate
lambda <- 0.2
data <- matrix(data = sample(rexp(n = n * m, rate = lambda)),
               nrow = m, ncol = n)
```

Show the sample mean and compare to the theoretical mean of the distribution

$X_i \sim exp(\lambda)$, we will have the theoretical mean for average of 40 exponential distribution as:

$$E(\bar{X}) = E(\frac{1}{40} \sum_{i=1}^{40} X_i)$$

$$= \frac{1}{40} \sum_{i=1}^{40} E(X_i)$$

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

```r
means <- apply(data, 1, mean)
meanCompare <- data.frame(mean(means), 1 / lambda)
names(meanCompare) <- c("Sample Mean", "Theoretical Variance")
kable(meanCompare)
```

Sample Mean	Theoretical Variance
4.990025	5

Show how vraiable the sample is (via variance) and compare it to the theoretical variance of the distribution

$X_i \sim exp(\lambda)$, we will have the theoretical variance for average of 40 exponential distribution as:

$$Var(\bar{X}) = Var(\frac{1}{40}\sum_{i=1}^{40}X_i)$$

$$= \frac{1}{40^2}\sum_{i=1}^{40}Var(X_i)$$

$$= \frac{1}{40}\frac{1}{\lambda^2}$$

```
varCompare <- data.frame(sd(means), (1 / lambda^2) / 40)
names(varCompare) <- c("Sample Variance", "Theoretical Variance")
kable(varCompare)
```

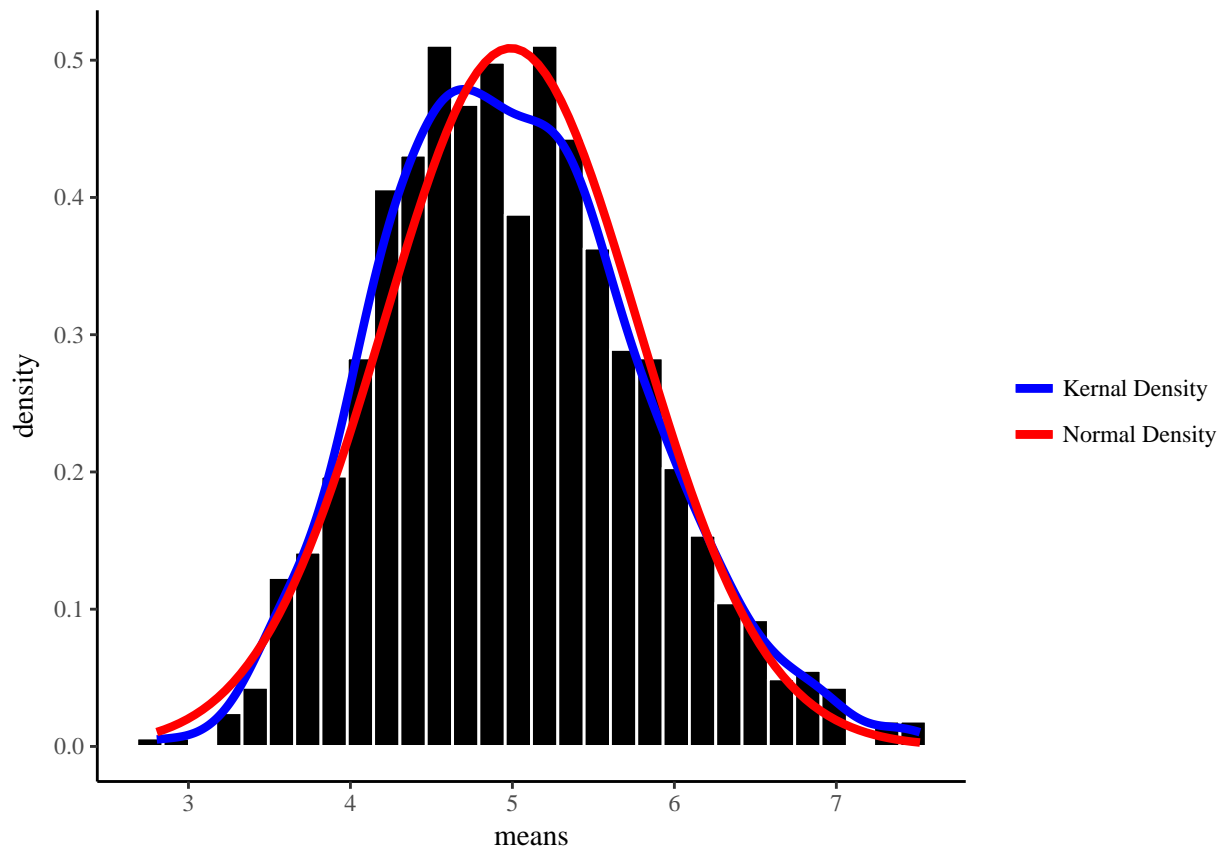| Sample Variance	Theoretical Variance
0.7840859	0.625

Show that the distribution is approximately normal

By Central limit theorem: if $X_1, X_2, ...$ are i.i.d. with $E(X_i) = \mu < \inf$, and $Var(X_i) = \sigma^2$ then

$$\sqrt{n}(\bar{X}_n - \mu) \to \sigma N(0, 1)$$

So, we expected to see $\bar{X}_{40} \sim N(\frac{1}{\lambda}, \frac{1}{40}\frac{1}{\lambda^2})$.

```
data.frame(means = means) %>%
        ggplot(aes(means)) +
        geom_histogram(aes(y = ..density..),
                        fill = "black", color = "white") +
        stat_density(aes(color = "Kernal Density"),
                lwd = 1.5,
                geom = "line") +
        stat_function(fun = dnorm,
                aes(color = "Normal Density"),
                lwd = 1.5,
                args = list(mean = mean(means),
                        sd = sd(means)),
                show.legend = TRUE) +
        scale_colour_manual("",values = c("blue", "red")) +
        theme_classic(base_family = "Times")
```

Conclusion

Observed from above:

1. The sample mean estimated the theoretical mean very well.

2. The sample variance is closed to the theoretical variance.

3. The kernal density is approximately normal with the theoretical mean and variance.

From the Central Limit Theorem, we know that if we can have a $\bar{X}_n$ for n large enough, we will have a better performance for above three properties.