Illustrating the Central Limit theorem Using the Exponential Distribution

NobumasaHayashi

8/20/2020

Overview

In this project, the exponential distribution is investigated and compared with the **Central Limit Theorem**(CLT). The analysis reveals that the distribution of averages of exponentials gets close to the normal distribution as the number of simulation increases.

Simulation

The **CLT** states that the distribution of averages of iid variables becomes that of a standard normal as the sample size increases even if the original variables themselves are not normally distributed. Here, **the exponential distribution** is investigated as an example. The probability density function of an exponential distribution is expressed as

$$f(x;\lambda) = \lambda e^{-\lambda x}$$

λ is the rate paramter, and both of the mean and standard deviation of exponential distribution are $\frac{1}{\lambda}$. In this project, for all of the simulation, $\lambda = 0.2$. Here, **40 random exponential variables** are investigated. We start with the distribution of 100 simulations of 40 random exponential variables, that is, the exponential distribution of 4,000 values and move to the distribution of averages of 100 simulations of 40 random exponentials, where 40 random exponentials are averaged 100 times. Finally, the distribution of averages of 1000 simulations is compared with them.

Sample Mean versus Theoretical Mean

Fig 1.(a) illustrates **the distribution of 100 simulations of 40 random exponential** whose probabilty decreases with the increaseing number. On the other hand, **the distribution of averages of 100 simulations of 40 random exponentials**, which is illustrated in **Fig 1.(b)**, has a similar shape of the normal distribution. Similarly, **the distribution of averages of 1000 simulations (Fig 1.(c))** has a similar shape of the normal distribution with smaller standard deviation. **Fig 2.(a)** illustrates **the cumulateve mean of 40 exponentials**.As the figure shows, the mean gets close to the theoretical mean ($\frac{1}{\lambda} = \frac{1}{0.2} = 5.0$) as the number of observations increases.

Sample Mean Variance versus Theoretical Mean Variance

Similarly, **Fig 2.(b)** illustrates **the cumulateve variance of averages of 0 exponentials**. As the figure shows, the variance gets close to the theoretical variance ($\frac{(\frac{1}{\lambda})^2}{\sqrt{n}} = \frac{5^2}{\sqrt{40}} = 0.625$) as the number of observations increases.

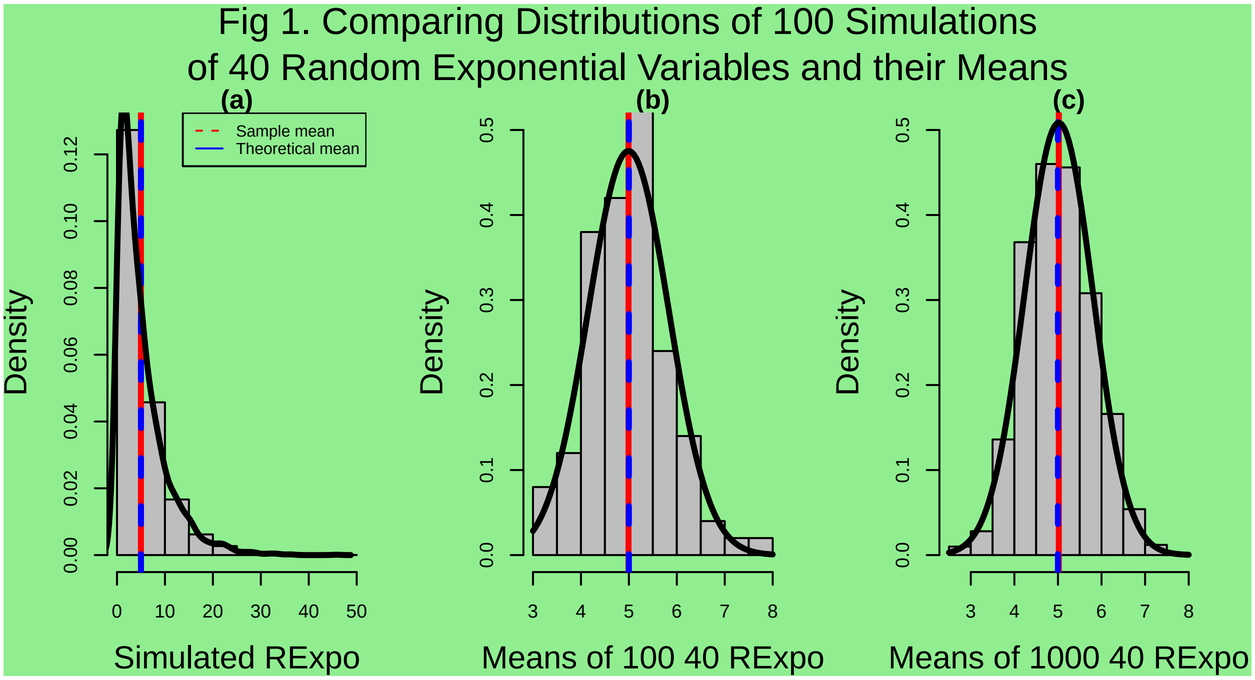


Fig 1. Comparing Distributions of 100 Simulations
of 40 Random Exponential Variables and their Means
(a)
(b)
(c)
Sample mean
Theoretical mean
Density
Simulated RExpo
Means of 100 40 RExpo
Means of 1000 40 RExpo

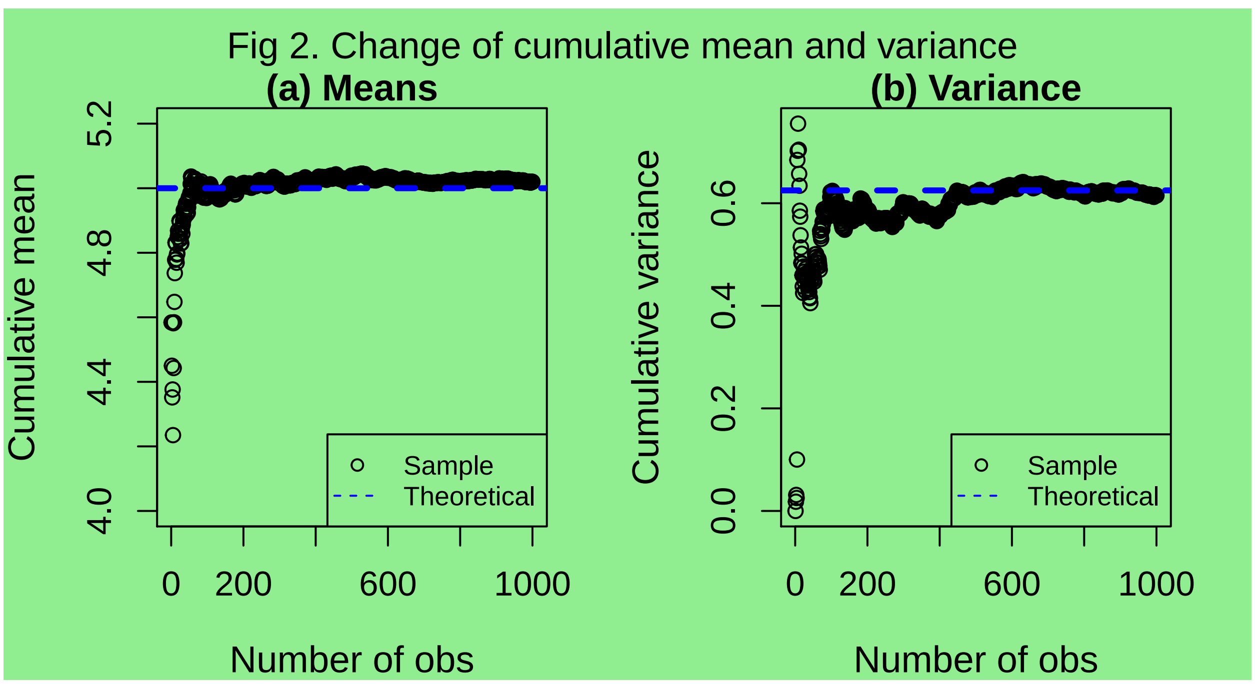
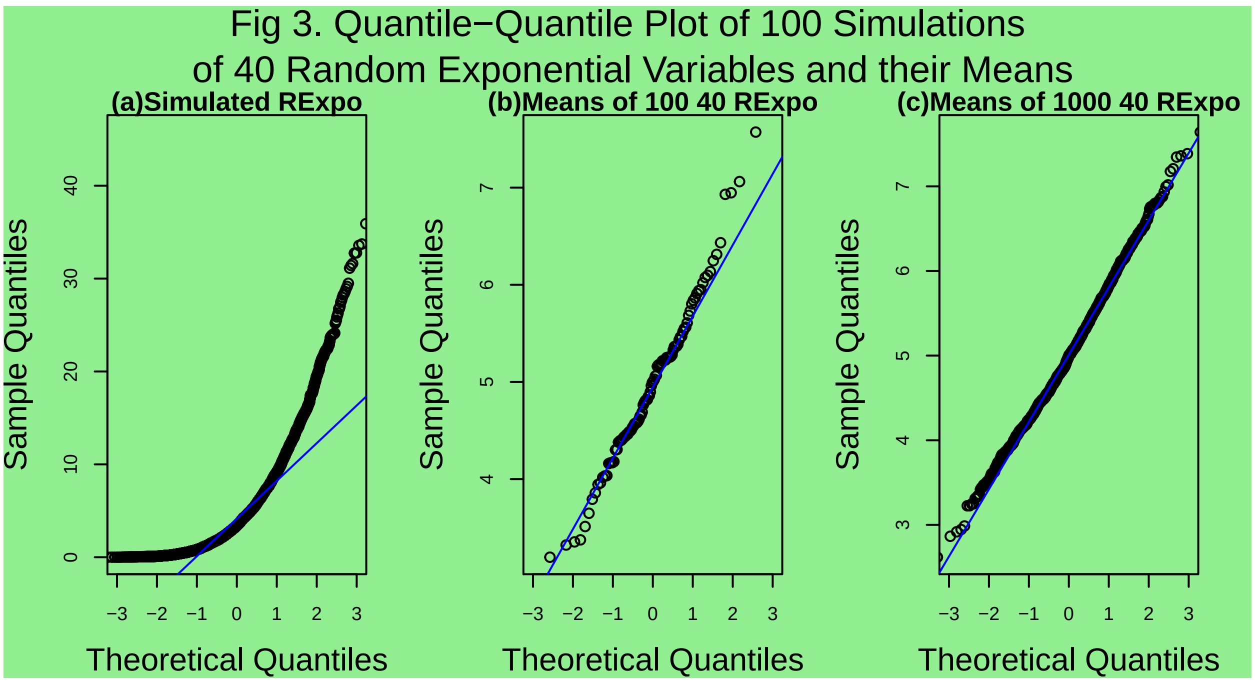


Fig 2. Change of cumulative mean and variance
(a) Means
(b) Variance
Cumulative mean
Cumulative variance
Sample
Theoretical
Number of obs

Distribution

Fig 1. indicates that the distribution of averages of 100 and 1,000 simulations have similar shapes of the normal distribution while the original exponential distribution does not. Here, **quantile-quantile plots** are used to investigate the similarity between the sample distribution and the normal distribution by plotting their quantiles against the normal theoretical quantiles. If the sample distributions have similarities to the normal distribution, the plots will follow the 45° lines (blue lines). **Fig 3.(a)** indicate that the original distribution has no similarity to the normal distribution. However, **Fig 3.(b) and (c)** indicate that the distribution of averages of 100 and 1,000 simulations have an obvious similarity to the normal distribution. Therefore, the distributions of averages of a large collection of exponentials represent normal even if the original variables are not normally distributed. This corresponds to the statement of **CLT**.



Appendix: All code for this report

```
#General Setting
knitr::opts_chunk$set(echo = FALSE)
#Fig 1. Comparing Distribution
Rexpo1<-NULL
n1<-100
set.seed(6784)
for (i in 1 : n1)
{Rexpo1 <-c(Rexpo1, rexp(40, 0.2))}
m1<-mean(Rexpo1)
s1<-sd(Rexpo1)

n2 = 100
Rexpo2 = NULL
set.seed(6784)
for (i in 1 : n2)
{Rexpo2 = c(Rexpo2, rexp(40, 0.2))}
means2<-apply(matrix(data=Rexpo2,nrow = n2),1, mean)
m2<-mean(means2)
s2<-sd(means2)

n3 = 1000
Rexpo3 = NULL
set.seed(6784)
for (i in 1 : n3)
{Rexpo3 = c(Rexpo3, rexp(40, 0.2))}
means3<-apply(matrix(data=Rexpo3,nrow = n3),1, mean)
m3<-mean(means3)
s3<-sd(means3)

par(mfrow = c(1,3), bg = "lightgreen", cex.axis = .9, cex.lab = 1.5)
par(mar=c(4.1,4.1,1.1,2.1))
par(oma = c( 0, 0, 3.2, 0 ))

hist(Rexpo1,col="grey", prob = TRUE, xlab = "Simulated RExpo", main = "(a)",
     cex = 1.2)
abline(v = mean(Rexpo1), col = "red", lwd = 3, lty = 1)
abline(v = 5, col = "blue", lwd = 3, lty = 2)
lines(density(Rexpo1,bw=1), col='black', lwd=3)
legend("topright",legend=c("Sample mean", "Theoretical mean"),
       col=c("red", "blue"), lty=c(2,1), cex=0.7)

hist(means2, breaks = 10, prob = TRUE, col="grey",
     xlab = "Means of 100 40 RExpo", main = "(b)",
     cex = 1.2, ylim = c(0,0.5))
abline(v = mean(means2), col = "red", lwd = 3, lty = 1)
abline(v = 5, col = "blue", lwd = 3, lty = 2)
curve(dnorm(x, mean=m2, sd=s2), col="black", add=T, lwd = 3)

hist(means3, breaks = 10, prob = TRUE, col="grey",
     xlab = "Means of 1000 40 RExpo", main = "(c)",
     cex = 1.2, ylim = c(0,0.5))
```

```
abline(v = mean(Rexpo3), col = "red", lwd = 3, lty = 1)
abline(v = 5, col = "blue", lwd = 3, lty = 2)
curve(dnorm(x, mean=m3, sd=s3), col="black", add=T, lwd = 3)
mtext("Fig 1. Comparing Distributions of 100 Simulations\nof 40 Random Exponential Variables and their M
      outer = TRUE, col = "Black", cex.axis = 1, cex = 1.2)

#Fig 2. Cumulative mean and variance
par(mfrow = c(1,2), bg = "lightgreen", cex.axis = 1.1, cex.lab = 1.2)
par(mar=c(4,4,1.1,2.1))
par(oma = c( 0, 0, 1.5, 0 ))

mn3<-cumsum(means3)/(1:n3)
plot(x=1:n3, y=mn3, ylim=c(4.0, 5.2),xlab = "Number of obs",
     main ="(a) Means", ylab="Cumulative mean")
abline(h=5,col = "blue", lwd = 3, lty = 2)
legend("bottomright",legend=c("Sample", "Theoretical"),
       col=c("black", "blue"), pch=c(1,NA), lty=c(0,2), cex=0.8)

mn33<-cumsum((means3^2))/(1:n3)
v3<-mn33-mn3^2
plot(x=1:n3, y=v3, xlab = "Number of obs",
     ylab = "Cumulative variance",main ="(b) Variance")
abline(h= (1/0.2)^2/40,col = "blue", lwd = 3, lty = 2)
legend("bottomright",legend=c("Sample", "Theoretical"),
       col=c("black", "blue"), pch=c(1,NA), lty=c(0,2), cex=0.8)
mtext("Fig 2. Change of cumulative mean and variance ",
      outer = TRUE, col = "Black", cex.axis = 1,cex = 1.2)

#Fig 3. Quantile-Quantile plot
par(mfrow = c(1,3), bg = "lightgreen", cex.axis = .9, cex.lab = 1.5)
par(mar=c(4.1,4.1,1.1,2.1))
par(oma = c( 0, 0, 3.2, 0 ))

qqnorm(Rexpo1,xlim = c(-3,3), main = "(a)Simulated RExpo")
qqline(Rexpo1, col = "blue")

qqnorm(means2,xlim = c(-3,3), main = "(b)Means of 100 40 RExpo")
qqline(means2, col = "blue")

qqnorm(means3,xlim = c(-3,3), main = "(c)Means of 1000 40 RExpo")
qqline(means3, col = "blue")

mtext("Fig 3. Quantile-Quantile Plot of 100 Simulations\n of 40 Random Exponential Variables and their M
      outer = TRUE, col = "Black", cex.axis = 1, cex = 1.2)
```