Investigating the Central Limit Theorem with Exponential Distribution

Rhyz C. Gomez

OVERVIEW OF THE REPORT

This paper investigates the exponential distribution and shows how the distribution of averages of 40 exponentials is approximately normal. Further, the paper reveals that the sample mean and sample variance are good approximates to the theoretical mean and theoretical variance of the exponential distribution using the concept of the Central Limit Theorem.

THE EXPONENTIAL DISTRIBUTION

The **Exponential distribution** is a continuous probability distribution arises naturally when events occur continuously and independently at a constant average rate.

The probability density function (pdf) of an exponential distribution has the form

$$f(x; lambda) = lambda * e^t, whenx >= 0,$$
$$f(x; lambda) = 0, whenx < 0$$

where lambda is the rate parameter and t is -lambda*r

The mean or expected value of an exponentially distributed random variable X is

$$E[X] = 1/lambda$$

and the variance of X is given by

$$Var[X] = 1/(lambda)^2$$

You can read more about Exponential Distribution at wikipedia.

Here, we will investigate the distribution of averages of 40 exponentials and compare it with the Central Limit Theorem. The lambda will be set at 0.2.

```
lambda <- 0.2
meanExp <- sdExp <- 1/lambda
varExp <- (sdExp)^2
```

Thus, the theoretical/population mean of an exponential distribution with lambda equal to 0.2 is 5 and the population variance is 25.

SIMULATION

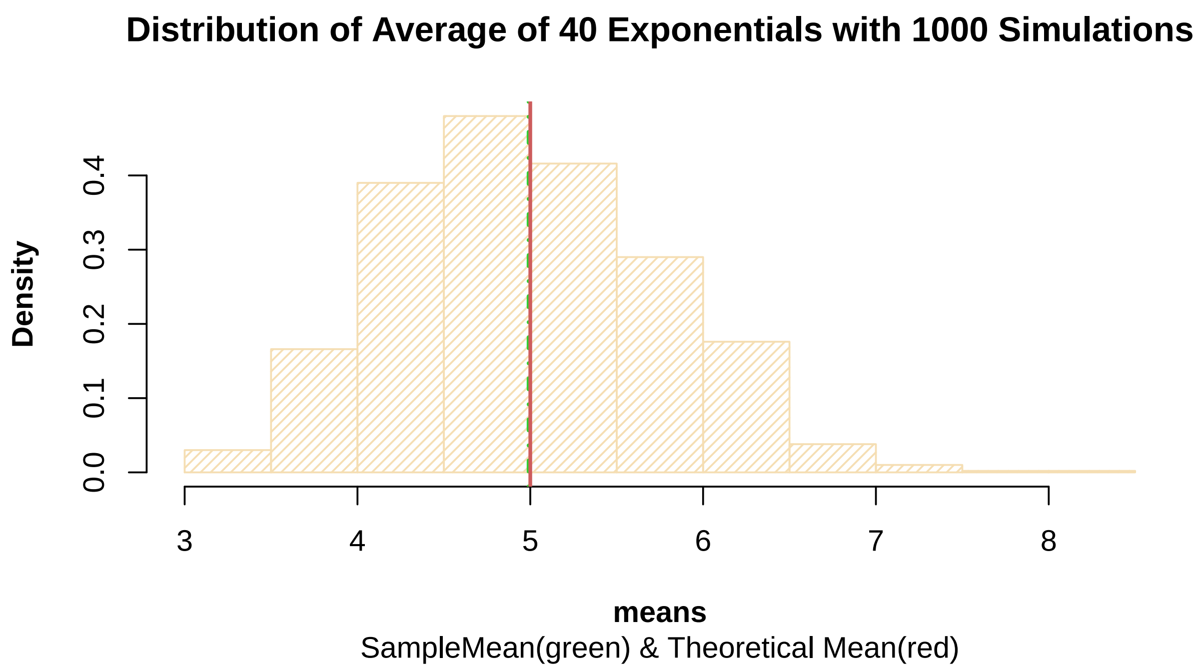
Based on the CLT, sample means are good estimates to the population/theoretical means when the sample size is large. To demonstrate this, 1000 simulations are made where each simulation consists of 40 exponentials. The result is saved on a matrix called `setofexp` (see Appendix, Code 1).

SAMPLE MEAN VS THEORETICAL MEAN

When the mean of every simulation is calculated using the apply() function (see Code 1), a numeric vector of averages is produced and it is stored in `means`. Thus, the mean of this vector is the sample mean.

```
sampleMean <- mean(means)
```

The mean of the averages of 40 exponentials is 4.9900252. Based on the figure, this is a good estimate to the theoretical mean of 5. See Appendix Code 2 for the complete code of the histogram.

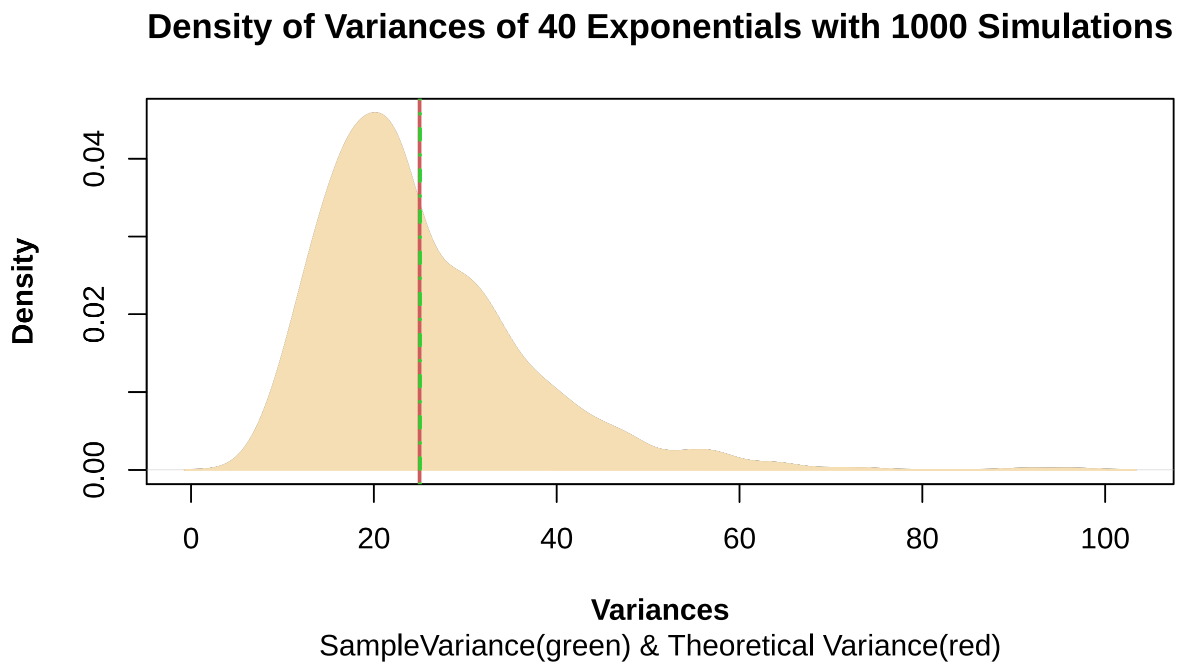


SAMPLE VARIANCE VS THEORETICAL VARIANCE

As mentioned, the theoretical variance is 25. In this study, the sample variance is the expected value of the 1000 variances of 40 exponentials. This expected value is a good estimate of the theoretical variance. Hence, calculate first the variances of 40 exponentials in 1000 simulations. Store the result in `variances`. The sample variance is then the mean of the variances.

```
variances <- apply(setofexp,1,var)
samVar <- mean(variances)
```

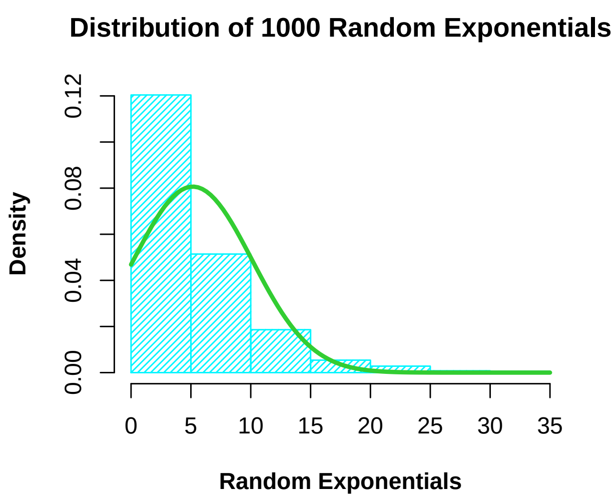
SampleVariance	TheoreticalVariance
25.05783	25

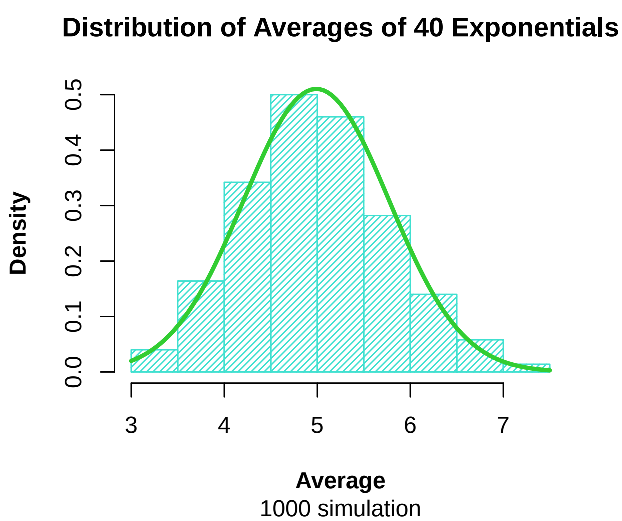


Based on the figure and the table, the sample variance and theoretical variance are approximately equal. Comparing the sample variance and theoretical variance, one can easily conclude that the sample variance is a good approximation to the theoretical variance.

DISTRIBUTION

The distribution of averages of exponentials is approximately normal. To illustrate this, we simulate 1000 averages of 40 exponentials using the for loop function and use the hist() function to plot the averages. We compare it to the distribution of 1000 exponentials. (see Appendix, Code 4)





In the figure above, the distribution of averages of exponentials is nearly a bell curved. Thus, it shows that the distribution of averages of 40 exponentials is approximately normal.

APPENDIX

Code 1

```
set.seed(1)
n <- 40
simulation <- 1000
setofexp <- matrix(rexp(n*simulation,lambda),nrow=simulation)
means <- apply(setofexp,1,mean)
```

Code 2

```
hist(means,density=25,prob=TRUE,col="wheat",
main="Distribution of Average of 40 Exponentials with 1000 Simulations",
font.lab=2, sub="SampleMean(green) & Theoretical Mean(red)")
abline(v=sampleMean,col="limegreen",lwd=2,lty=4)
abline(v=meanExp,col="indianred",lwd=2)
```

Code 3

```
dense <- density(variances)
plot(dense, main="Density of Variances of 40 Exponentials with 1000 Simulations",
xlab="Variances",font.lab=2,
sub="SampleVariance(green) & Theoretical Variance(red)")
polygon(dense,col="wheat",border="wheat")
abline(v=varExp,col="indianred",lwd=2)
abline(v=samVar,col="limegreen",lwd=2,lty=4)
```

Code 4

```
set.seed(1)
expn = NULL
for (i in 1 : 1000) expn = c(expn, mean(rexp(40,0.2)))
set.seed(1)
rexp <-rexp(1000,0.2)
par(mfrow=c(1,2))
hist(rexp,density = 30,prob=TRUE,col="turquoise1",
main="Distribution of 1000 Random Exponentials",xlab="Random Exponentials",
font.lab=2)
curve(dnorm(x,mean=mean(rexp),sd=sd(rexp)),lwd=3,col="limegreen",add=TRUE,yaxt="n")
hist(expn,density=30,prob=TRUE, main ="Distribution of Averages of 40 Exponentials",
xlab="Average",col="turquoise",sub="1000 simulation",font.lab=2)
curve(dnorm(x,mean=mean(expn),sd=sd(expn)),lwd=3,col="limegreen",add=TRUE,yaxt="n")
```