The Means of 1000 Exponential Distribution Follows The Central Limit Theory

CF Hainey

August 11, 2016

Part 1 Simulation

The central limit theory (clt) is centered around the idea that with sufficient sampling and regardless of distribution type (normal, exponential, etc.), if one iterates the distribution (with given parameters) and takes the mean of each one, these means will follow a normal distribution.

Example

The occurrences of events in a Poisson process are depicted by the exponential distribution. These events are independent and identical with a constant average rate. For exponential events, both the mean and standard deviation is defined by the equation 1/lambda. For our example we will set lambda = 0.2, this will result in a mean of 5 and a standard deviation of 5 (1/0.2 = 5).

From figure 1 (supplemental data) you can see with 40 iterations, the mean of those iteration shows a Gaussian distribution, with the mean around 5, the same as the theoretical mean.

More iteration, with means, standard deviation and variation.

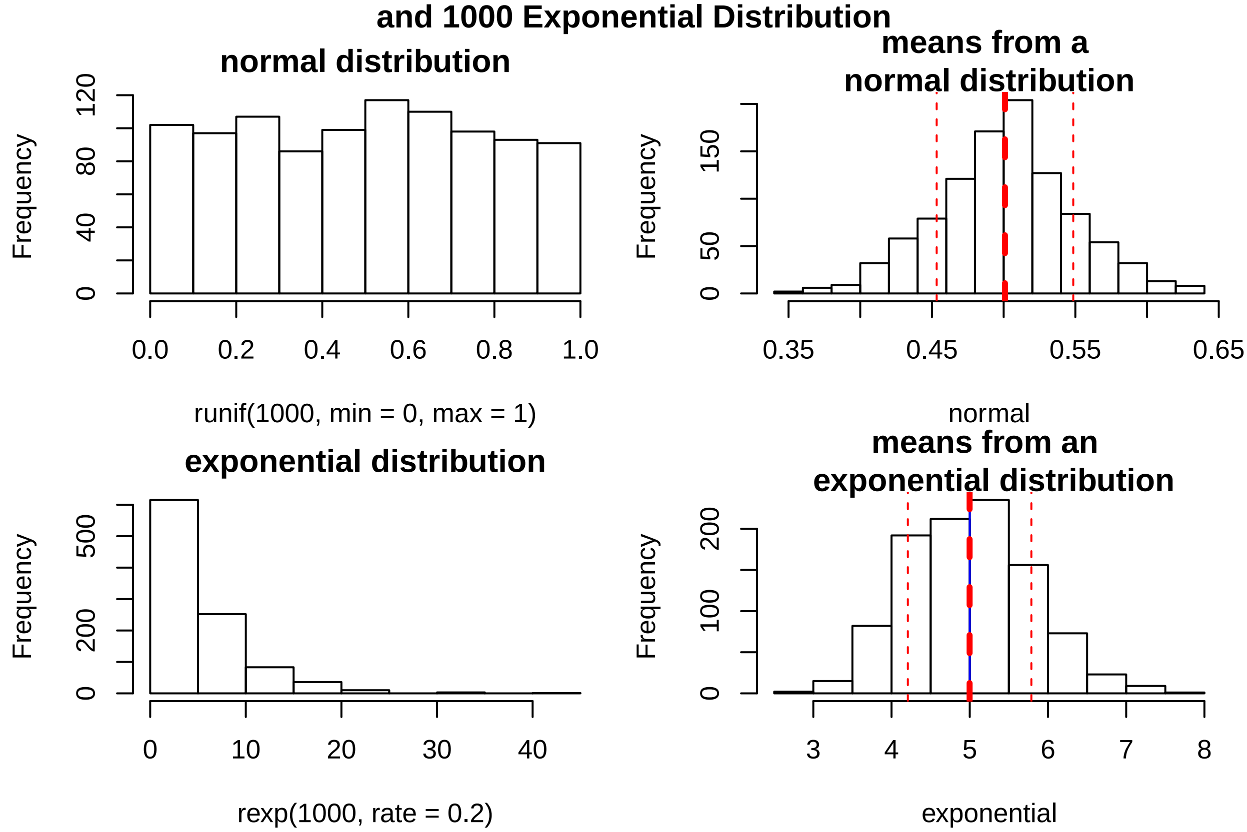
Sampling more iterations creates a more narrow window of where the mean of the means lie. This is depicted in Fig 2 (supplemental data). These histograms were created with iteration where mns### creates 1000 iteration. The ### indicates the number of iterations extracted for analysis. Test is 1000 exponential distribution (not their mean). When applying the CLT, the mean of the mean of 1000 exponentially distributions is similar to its theoretical 5. When taking the standard deviation or variance of the means of 1000 iterations of a exponentially distributed sample, they become smaller and diverge from the theoretical 5 to become as small as 0.1526 for our example. If one wanted to investigate the how the standard deviation changes, it also follows the central limit theory and the average standard deviation is similar to theoretical 5 (df2).

The variance and standard deviation will approach 0 as the number of samples considered. #Figures and Table of means, standard deviation and variance of exponential distributions

```r
#produce data
set.seed(456)
test <- (rexp(1000, rate = 0.2))
mns40 = NULL
for (i in 1 : 1000) mns40 = c(mns40, mean(rexp(40, rate = 0.2)))
sd40 = NULL
for (i in 1 : 1000) sd40 = c(sd40, sd(rexp(40, rate = 0.2)))
var40 = NULL
for (i in 1 : 1000) var40 = c(var40, sd(rexp(40, rate = 0.2)))
mns100 = NULL
for (i in 1 : 1000) mns100 = c(mns100, mean(rexp(100, rate = 0.2)))
sd100 = NULL
```

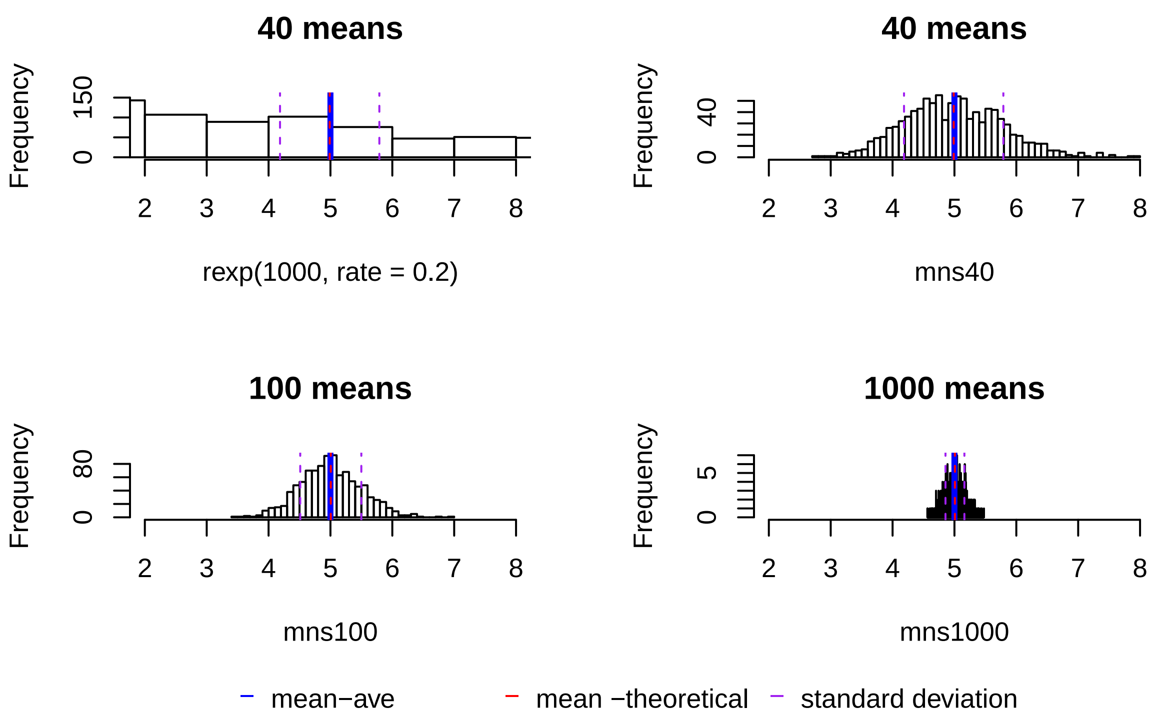
```
for (i in 1 : 1000) sd100 = c(sd100, sd(rexp(40, rate = 0.2)))
var100 = NULL
for (i in 1 : 1000) var100 = c(var100, sd(rexp(40, rate = 0.2)))
mns1000 = NULL
for (i in 1 : 1000) mns1000 = c(mns1000, mean(rexp(1000, rate = 0.2)))
sd1000 = NULL
for (i in 1 : 1000) sd1000 = c(sd1000, sd(rexp(40, rate = 0.2)))
var1000 = NULL
for (i in 1 : 1000) var1000 = c(var1000, sd(rexp(40, rate = 0.2)))

testname <- c("test","mns40","mns100", "mns1000", "theoretical")
test_mean <- c(mean(test), mean(mns40),mean(mns100),mean(mns1000),5)
test_mean_sd <- c(sd(test), sd(mns40),sd(mns100),sd(mns1000),0)
test_mean_var <- c(var(test), var(mns40), var(mns100), var(mns1000),0)
test_sd_mean <-c(mean(test),mean(sd40),mean(sd100),mean(sd1000),5)
test_sd_sd <- c(sd(test),sd(sd40),sd(sd100),sd(sd1000),0)
test_sd_var <-c(var(test), var(sd40),var(sd100),var(sd1000),0)
test_var_mean <-c(mean(test), mean(var40),mean(var100),mean(var1000),25)
test_var_sd <-c(sd(test), sd(var40),sd(var100),sd(var1000),0)
test_var_var <-c(var(test),var(var40),var(var100),var(var1000),0)
set.seed(456)
normal = NULL
for (i in 1 : 1000) normal = c(normal, mean(runif(40)))
exponential = NULL
for (i in 1 : 1000) exponential = c(exponential, mean(rexp(40, rate = 0.2)))

#plot data
mainmain <- "FIG 1. Comparison of 1000 Normal Distribution \n and 1000 Exponential Distribution"
par(mfrow = c(2, 2), mar = c(4, 4, 2, 1), oma = c(0, 0, 2, 0))
hist(runif(1000, min = 0, max = 1), main = "normal distribution")
hist(normal, main = "means from a \nnormal distribution")
abline(v=c(5, mean(normal)),  col=c("blue", "red"), lty=c(1,2), lwd=c(1, 3))
abline(v=c(mean(normal)-sd(normal),mean(normal)+sd(normal)),  col=c("red", "red"), lty=c(2,2), lwd=c(1,
hist(rexp(1000, rate = 0.2), main = "exponential distribution")
hist(exponential, main = "means from an \n exponential distribution")
abline(v=c(5, mean(exponential)),  col=c("blue", "red"), lty=c(1,2), lwd=c(1, 3))
abline(v=c(mean(exponential)-sd(exponential),mean(exponential)+sd(exponential)),  col=c("red", "red"),
title(mainmain, outer = TRUE, cex = 1.0)
```

FIG 1. Comparison of 1000 Normal Distribution and 1000 Exponential Distribution




```
mainmain2 <- "Fig 2. Comparison of Sampling 40, 100, and 1000 \n Exponential Distribution And How Their
par(mfrow = c(2, 2), mar = c(5.1,4., 4.1, 3), oma = c(1.5, 0, 3, 0))
hist(rexp(1000, rate = 0.2), breaks = 40, xlim = c(2,8), main = "40 means")
abline(v=c(5, mean(mns40)),  col=c("blue", "red"), lty=c(1,2), lwd=c(3, 1))
abline(v=c(mean(mns40)-sd(mns40),mean(mns40)+sd(mns40)),  col=c("purple", "purple"), lty=c(2,2), lwd=c(
hist(mns40,breaks = 40, xlim = c(2,8), main = "40 means")
abline(v=c(5, mean(mns40)),  col=c("blue", "red"), lty=c(1,2), lwd=c(3, 1))
abline(v=c(mean(mns40)-sd(mns40),mean(mns40)+sd(mns40)),  col=c("purple", "purple"), lty=c(2,2), lwd=c(
hist(mns100,breaks = 40, xlim = c(2,8), main = "100 means")
abline(v=c(5, mean(mns100)), col=c("blue", "red"), lty=c(1,2), lwd=c(3, 1))
abline(v=c(mean(mns100)-sd(mns100),mean(mns100)+sd(mns100)),  col=c("purple", "purple"), lty=c(2,2), lw
hist(mns1000,breaks = 1000, xlim = c(2,8), main = "1000 means")
abline(v=c(5, mean(mns1000)), col=c("blue", "red"), lty=c(1,2), lwd=c(3, 1))
abline(v=c(mean(mns1000)-sd(mns1000),mean(mns1000)+sd(mns1000)),  col=c("purple", "purple"), lty=c(2,2)
title(mainmain2, outer = TRUE, cex = 1.0)
par(fig = c(0, 1, 0, 1), oma = c(0, 0, 0, 0), mar = c(0, 0, 0, 0), new = TRUE)
plot(0, 0, type = "n", bty = "n", xaxt = "n", yaxt = "n")
legend("bottom", c("mean-ave","mean -theoretical","standard deviation"), xpd = TRUE,
       horiz = TRUE, inset = c(0,0), bty = "n", pch = c("-","-","-"),col = c("blue","red","purple"), ce
```

Fig 2. Comparison of Sampling 40, 100, and 1000 Exponential Distribution And How Their Means Are Distributed




```
theoretical <- c(5,5,5)
df2 = data.frame(test_mean, test_mean_sd, test_mean_var,row.names = c("test","mn40","mn100", "mn1000","
df2
```

```
##             test_mean test_mean_sd test_mean_var
## test         4.772709    4.7551784   22.61172185
## mn40         4.987935    0.8029365    0.64470705
## mn100        5.005662    0.4937184    0.24375790
## mn1000       5.008146    0.1526023    0.02328746
## theoretical  5.000000    0.0000000    0.00000000
```

probablitity the normal mean distribution is similar to the exponential mean distributions using the t-test.

When we evaluate our set of examples for the normal mean distribution and the exponential mean distribution, we can see by the t-test, the difference of means = 0 has the probability of 2.2 e-16.

```
t.test(normal, exponential, mu = 0, paired=F, conf.level = 0.99)
```

```
##
##  Welch Two Sample t-test
##
## data:  normal and exponential
## t = -179.81, df = 1006.3, p-value < 2.2e-16
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  -4.561869 -4.432773
## sample estimates:
## mean of x mean of y
## 0.5009211 4.9982419
```

Conclusion

The central limit theory holds true for exponential distribution means, standard deviation of 1000 exponential distributions and variance of 1000 exponential distributions, where the theoretical is the same as the calculated, plus the more iteration the samller the standard deviation and variance of the samples become.