CLT Demo using Exponential Distribution

April 24, 2015

Overview

The goal of this report is to investigate the Exponential distribution and explain Central Limit Theorem (CLT) using the same.

The exponential distribution is simulated in R using the function rexp(n, rate) where n is the vector containing the values of the distribution whose mean and standard deviation are 1/rate.

The CLT states that the distribution of averages of independently and identically distributed random variables becomes that of a standard normal as the sample size increases.

Simulations

To illustrate, let us take two distributions. The first is a distribution of 1000 random exponentials with rate = 0.2.

```
set.seed(1061)
rate <- 0.2
num <- 1000
t <- rexp(num, rate)
meanT <- mean(t)
sdT <- sd(t)
varT <- var(t)
```

The second is a distribution of 1000 averages of 40 random exponentials (using the same rate = 0.2).

```
set.seed(1061)
nosim <- 1000
n <- 40
rate <- 0.2
mns <- NULL
for (i in 1:nosim)
  mns = c(mns, mean(rexp(n, rate)))
meanMNS <- mean(mns)
sdMNS <- sd(mns)
varMNS <- var(mns)
```

Sample mean vs. Theoretical mean

Given the value of rate is 0.2, we know that the mean of the first distribution is 1/0.2 = 5. Indeed, we notice that the mean is close to 5 (though not exactly equal to 5 because of the randomness factor).

Per CLT, the second distribution should be centered around the same mean (5). Again, it is close to 5 but not exactly equal to 5 because of a Monte Carlo simulation error, that is, we ran the simulation a finite number of times.

Sample variance vs. Theoretical variance

Given the value of rate is 0.2, we know that the standard deviation of the first distribution is 1/0.2 = 5 and hence the variance is 5^2 = 25. For our sample, the observed variance is close to 25 (~25.16).

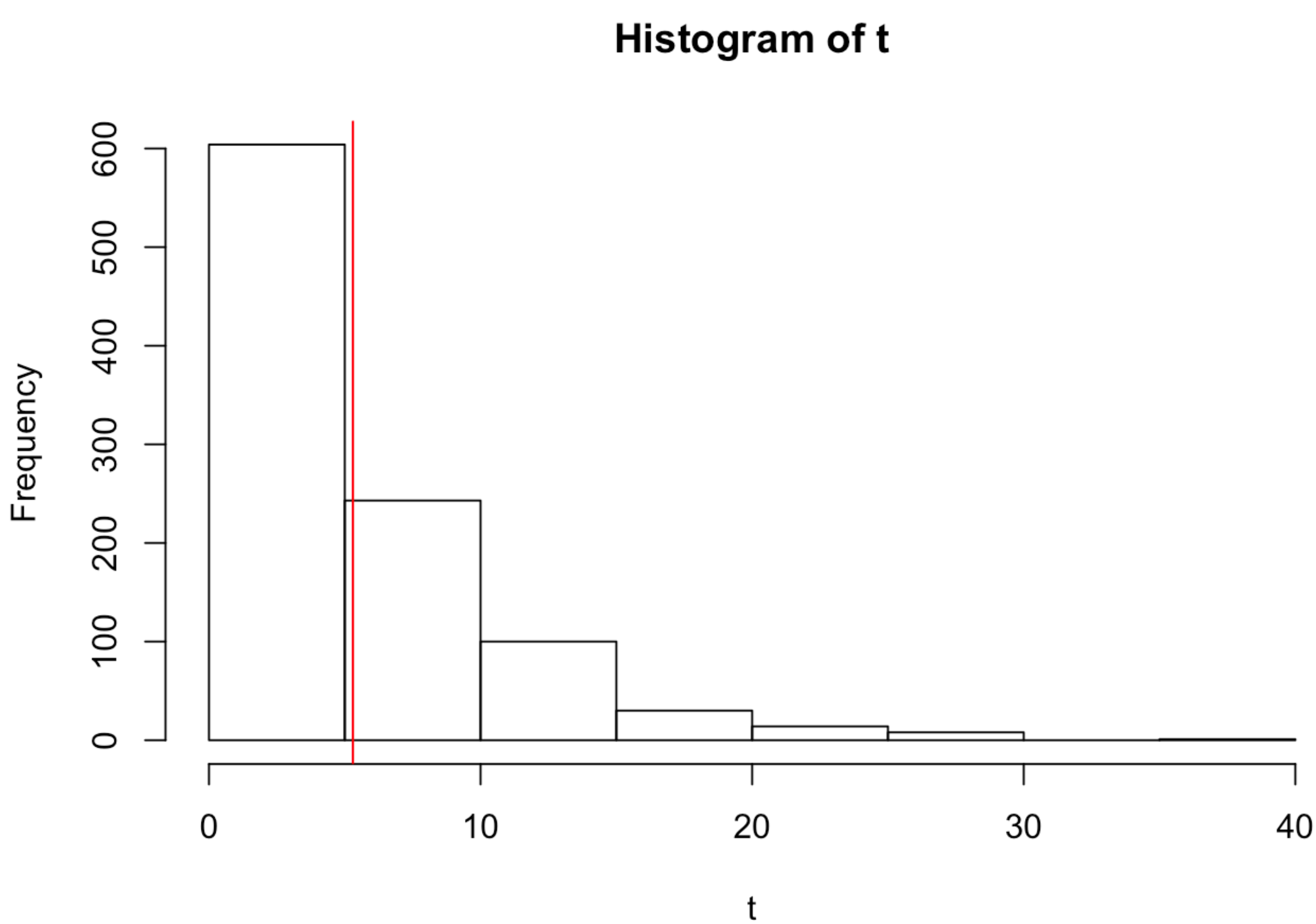
Per CLT, the second distribution should have a variance equal to variance of the underlying distribution / n where n is the sample size. In this case, that would be equal to 25/40 = 0.625.

The observed value of the variance is 0.61 which is close to 0.625.

Distribution

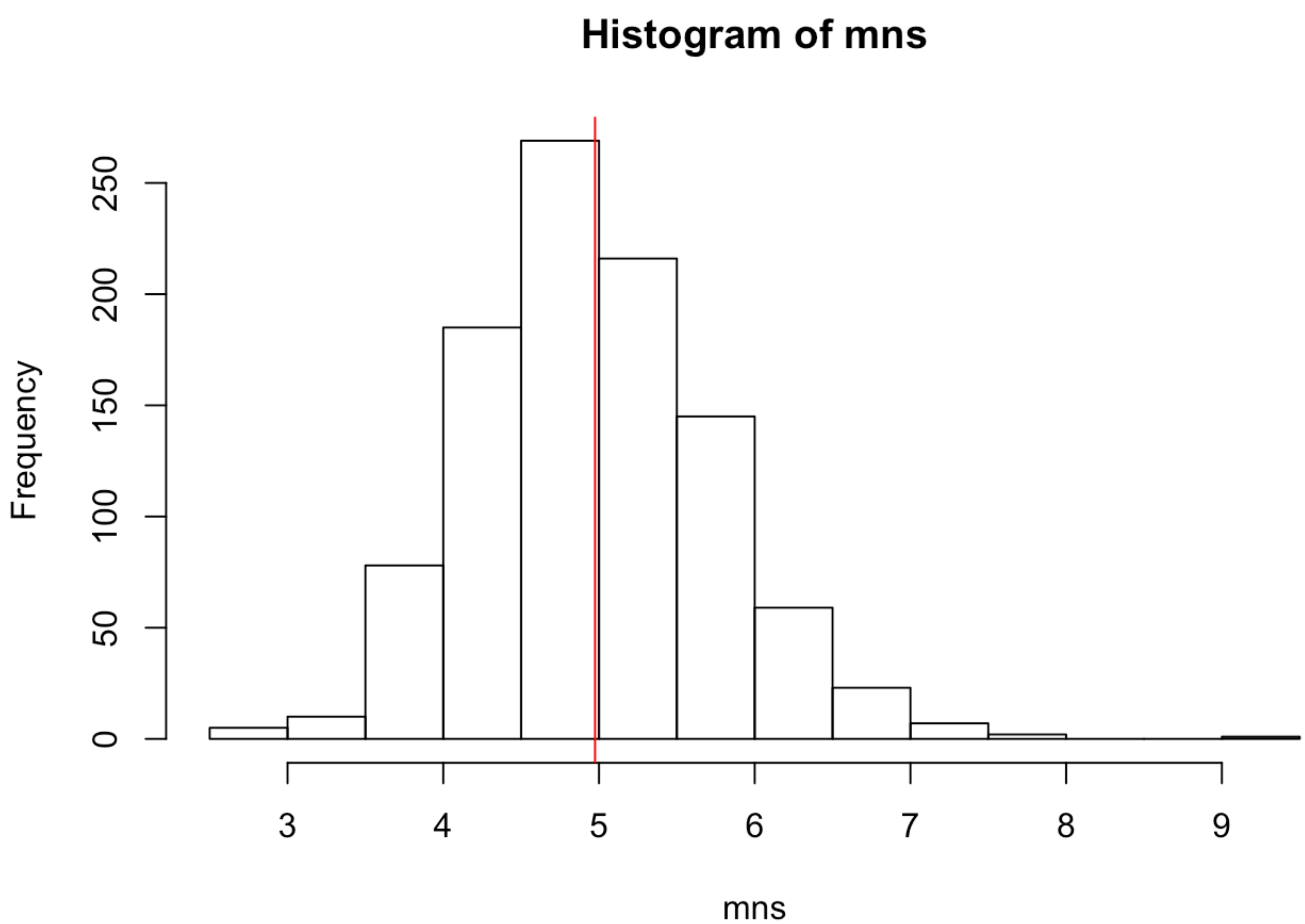
The first distribution is a typical exponential distribution with mean around 5, denoted by the red vertical line.

```
hist(t)
abline(v=meanT, col="red")
```



The second distribution is normal per CLT with mean around 5, denoted by the red vertical line.

```
hist(mns)
abline(v=meanMNS, col="red")
```



Conclusion

By simulating an exponential distribution of 40 variables 1000 times, we are able to quickly demonstrate the central limit theorem. We see that the resulting distribution is normal centered around the mean of the underlying distribution with a variance equal to the variance of the underlying distribution divided by the sample size.

Appendix

Here are the list of the variables used in the illustration.

Distribution #1:

```
str(t)
```

```
##  num [1:1000] 0.0769 1.5864 3.5284 0.2226 5.2375 ...
```

```
meanT
```

```
## [1] 5.302947
```

```
sdT
```

```
## [1] 5.015502
```

```
varT
```

```
## [1] 25.15527
```

Distribution #2:

```
str(mns)
```

```
##  num [1:1000] 4.84 5.42 6.32 3.96 6.12 ...
```

```
meanMNS
```

```
## [1] 4.97467
```

```
sdMNS
```

```
## [1] 0.7776964
```

```
varMNS
```

```
## [1] 0.6048117
```