An Emperical Look at the Central Limit Theorem

Buel Chandler

2016-06-27

Overview

In this paper we experiment with sampling from the Exponential distribution $\lambda e^{-\lambda x}$ to show how close we come to the actual distribution mean ($\mu = {}^{1}/_{\lambda}$) and the actual distribution variance ($\sigma^2 = {}^{1}/_{\lambda^2}$) of the Exponential distribution. Additionally we show how well the experiment supports the Central Limit Theorem – i.e., convergence to a Normal distribution.

Initialize the Experiment

```
library(pastecs, quietly = TRUE) ## we want to use the "stat.desc" function

## prime the pump of the Random Number Generator for reproducability
set.seed(1234567)

## set our constants
n.obs <- 40 ## each sample will have 40 obs pulled from exponential distribution
n.samples <- 1000 ## we'll take 1,000 such samples
lambda <- 0.2 ## Exponential distribution "rate" used in this study
exp.mean <- 1/lambda ## the actual mean of the population distribution
exp.var <- 1/(lambda^2)/n.obs ## the actual variance of the population distribution
```

Simulation – Get the Samples

Here we use the *pastecs::stat.desc()* function to compute the statistics on each of the `n.samples` columns, where each column represents one unique sample. Each column contains 14 statistics of the single sample pulls of `n.obs` from the exponential distribution with parameter lambda.

```
samples <- replicate(n.samples, stat.desc(rexp(n.obs, rate = lambda)))
dim(samples)
```

```
## [1]   14 1000
```

```
row.names(samples)
```

```
##  [1] "nbr.val"      "nbr.null"     "nbr.na"       "min"
##  [5] "max"          "range"        "sum"          "median"
##  [9] "mean"         "SE.mean"      "CI.mean.0.95" "var"
## [13] "std.dev"      "coef.var"
```

How Close to the Population Mean and Varience?

```
samp.mean <- mean(samples["mean",]) ## take the mean of all the sample means
samp.var <- mean(samples["var",])/n.obs ## take the mean of all the sample variances
```

In Table 1 we show how close our sampled means and variance come to the idealized $N(5, 0.625)$.

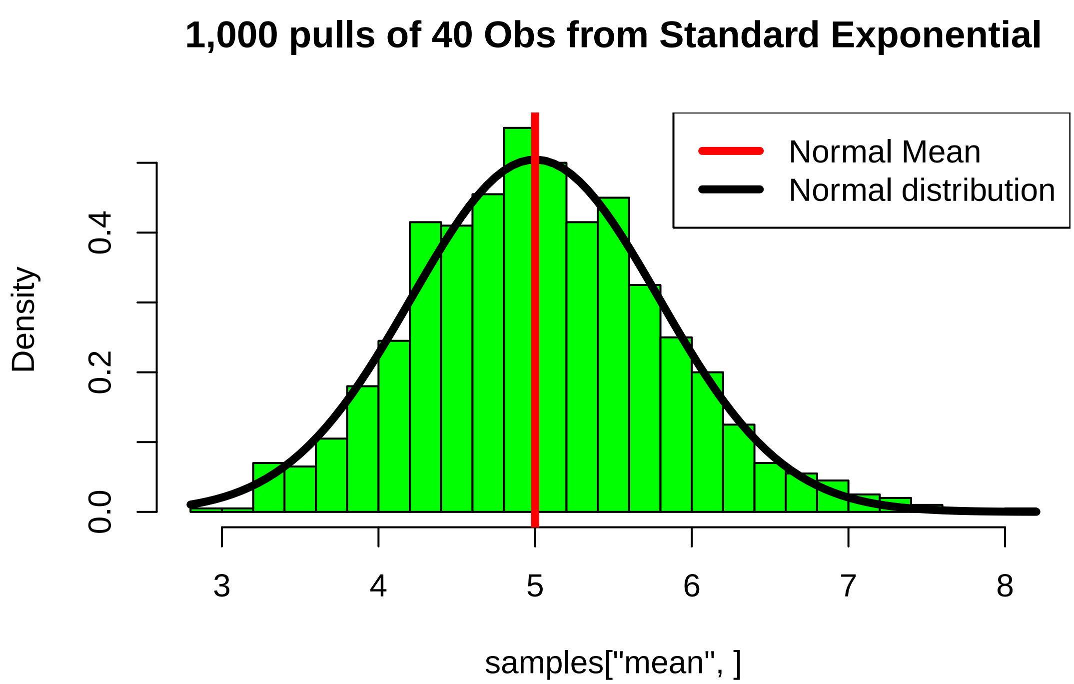
Table 1: Exponential Theoretical vs. Empirical

	Mean	Variance
Theoretical	5	0.625
Samples	5.039	0.633
Absolute Deltas	0.039	0.008

Empirical Look at Central Limit Theorem

The Central Limit Theorem for this case states that as we increase number of samples from any distribution, the mean of each sample considered as a distribution should, as number of samples gets bigger, trend to a Normal distribution with mean $\mu = {}^1/_\lambda = 5$ and standard deviation $\sigma = {}^1/_{\lambda\sqrt{n}} = 0.79$ where n is number observations per sample (40 in this case). As we see from the graphic this seems to be the case.

```
x <- seq(1,9,0.01)
hist(samples["mean",], breaks = 20, col = "green", freq = FALSE,
     main = "1,000 pulls of 40 Obs from Standard Exponential")
curve(dnorm(x, mean = exp.mean, sd = sqrt(exp.var)), add = TRUE, lwd = 4, col = "black")
abline(v = exp.mean, col = "red", lwd = 4)
legend("topright", legend = c("Normal Mean", "Normal distribution"), col = c("red", "black"),
       lty = c(1,1), lwd = c("4", "4"))
```



One Large Sample Versus Many Small Samples

When we look at one large sample of the Exponential distribution, we end up with data that obviously tends to the population (or theoretical) Exponential distribution.

```
## we will compare the distribution of a sample of 1000 pulls from the Exponential distribution
big.sample <- rexp(n.samples, rate = lambda) ## used in CLT portion for comparison

x <- seq(0,40,1)
hist(big.sample, breaks = 20, col = "green", freq = FALSE,
     main = "1,000 pulls from Standard Exponential")
curve(dexp(x, rate = 0.2), add = TRUE, lwd = 4, col = "black")
abline(v = exp.mean, col = "red", lwd = 4)
legend("topright", lty = c(1,1), lwd = c("4", "4"),
       legend = c("Exponential Dist mean", "Exponential distribution"), col = c("red", "black"))
```