Application of Central Limit Theorem to the exponential distribution

Alex Gaggin

Overview

Central Limit Theorem (CLT) states that if we take sufficiently large number of samples of some distribution, then means of these samples will be distributed in a different way - they will follow normal distribution, and their average (mean of means) will be centered on theoretical mean of the original distribution. The following simulation demonstrates how this applies to exponential distribution. Exponential distribution is described by a rate parameter lambda, and then its theoretical mean and standard deviation will both be 1/lambda.

Simulations

We use rexp() function in R to simulate random observations.

Let's initialize parameters: lambda (property of the exponential distribution), sample size and number of samples.

```
lambda      <- 0.2
sample.size <- 40
numsim      <- 1000
```

Let's initialiaze a list to store observations, and initialize seed for random numbers generator (for reproducibility).

```
dat <- list()
set.seed(1)
```

Let's generate observations and take their means.

```
for(i in 1:numsim) dat[[i]] <- rexp(sample.size, lambda)
means <- sapply(dat, mean)
```

Sample Mean versus Theoretical Mean

The CLT suggests that mean of means of many samples should converge to theoretical mean of the original distribution, i.e. 1/lambda.

```
(sample.mean <- mean(means))
```

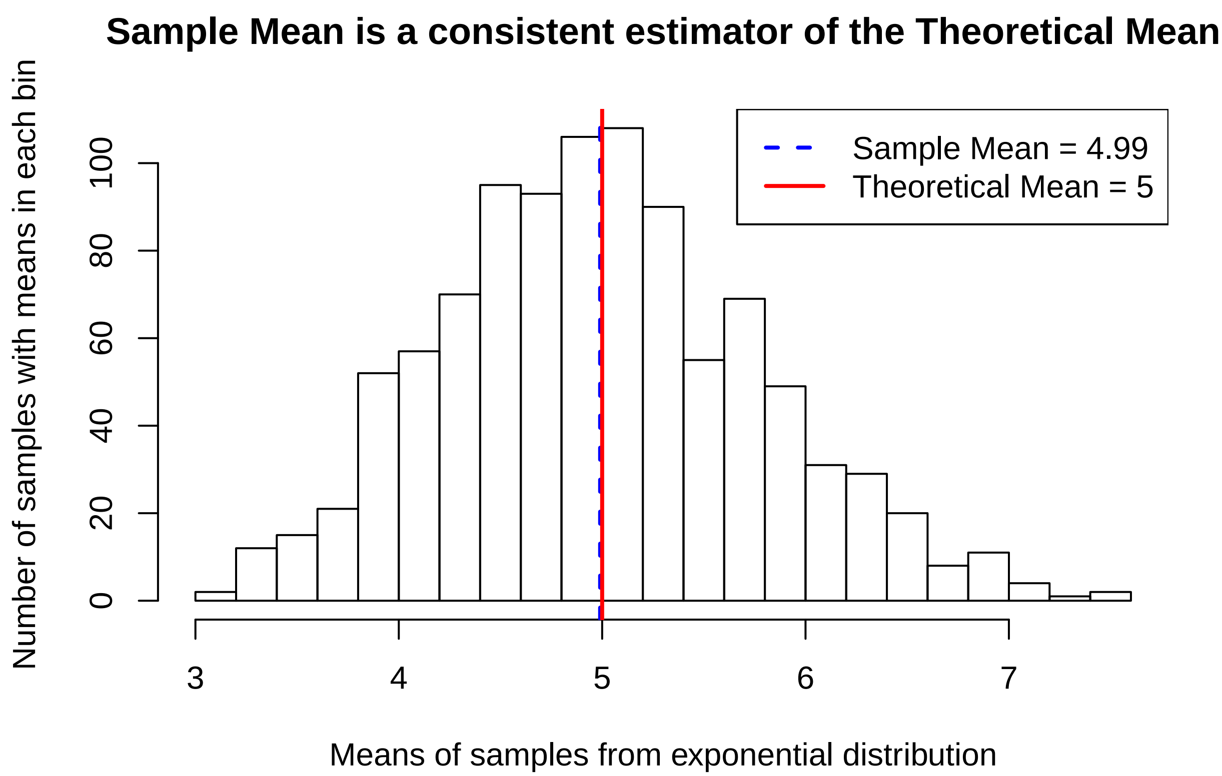
```
## [1] 4.990025
```

```
(theoretical.mean <- 1/lambda)
```

```
## [1] 5
```

Sample and Theoretical means are almost indistinguishable on a histogram of the sample means.

```r
breaks <- 20
hist(means, breaks=breaks,
     main="Sample Mean is a consistent estimator of the Theoretical Mean",
     xlab="Means of samples from exponential distribution",
     ylab="Number of samples with means in each bin")
abline(v=sample.mean, lwd=2, lty=2, col="blue")
abline(v=theoretical.mean, lwd=2, col="red")
legend("topright", c(paste0("Sample Mean = ", round(sample.mean, 2)),
                     paste0("Theoretical Mean = ", theoretical.mean)),
       lty=c(2,1), lwd=c(2,2),col=c("blue","red"))
```



Sample Variance versus Theoretical Variance

Variance of sample means should be equal to variance of original population divided by sample size (sigma squared over n). By collecting more data from the large number of samples, we decrease variance in distribution of the means and can estimate variance of the original distribution more precisely.

```r
(sample.var <- var(means))
```

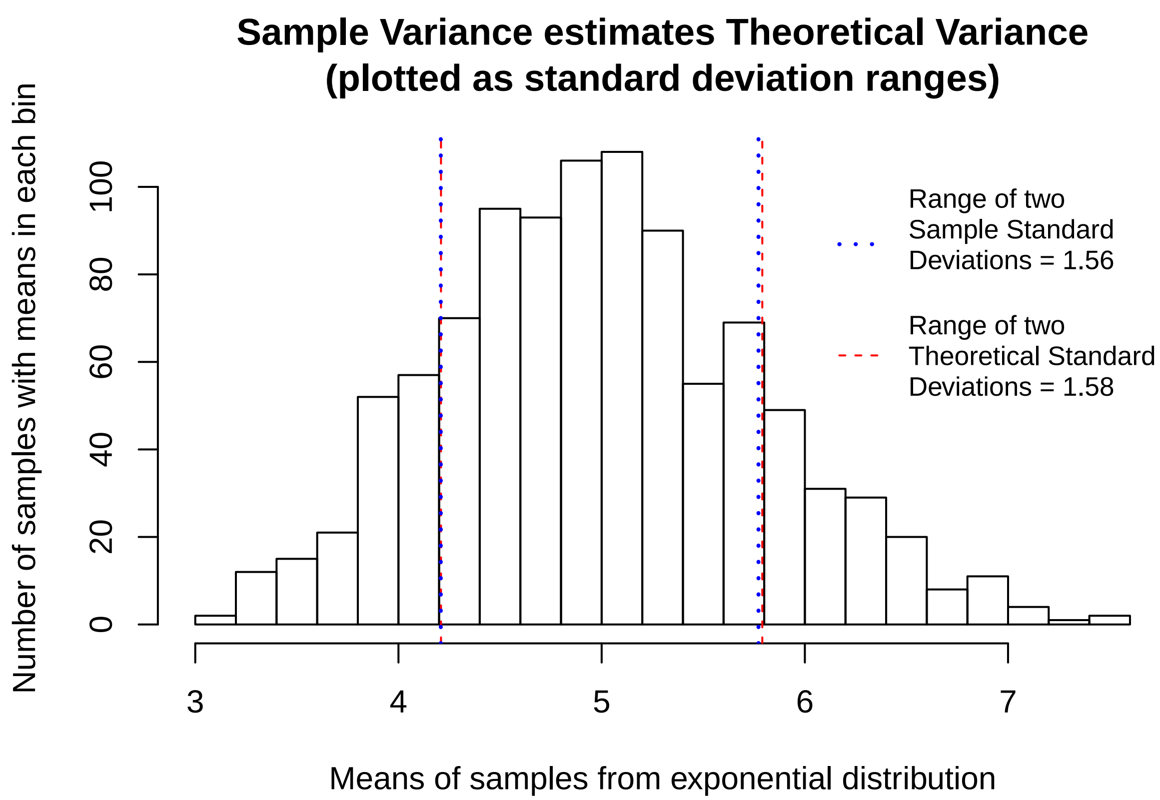
```
## [1] 0.6111165
```

```r
(theoretical.var <- 1/lambda^2/sample.size)
```

```
## [1] 0.625
```

Because variance is in squared units, let's plot standard deviation instead.

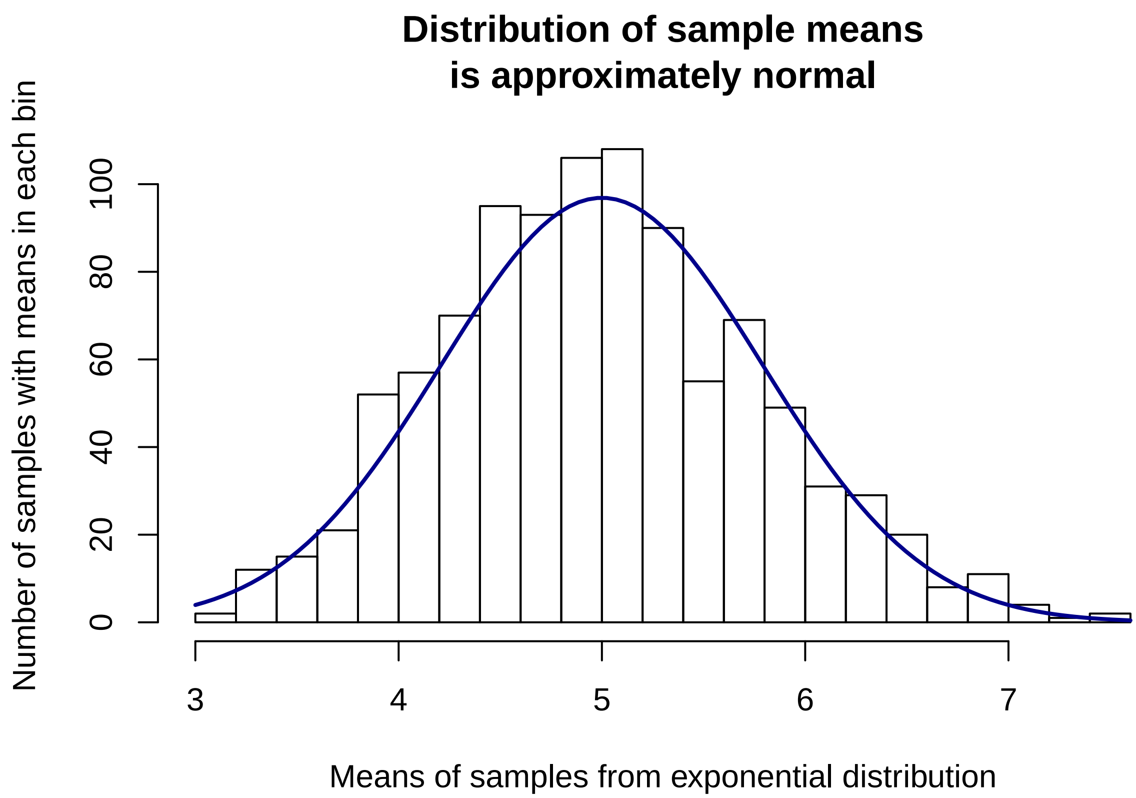
```
sample.sd <- sqrt(sample.var)
theor.sd <- sqrt(theoretical.var)
hist(means, breaks=breaks,
     main="Sample Variance estimates Theoretical Variance
(plotted as standard deviation ranges)",
     xlab="Means of samples from exponential distribution",
     ylab="Number of samples with means in each bin")
abline(v=theoretical.mean - theor.sd, lwd=1, lty=2, col="red")
abline(v=theoretical.mean + theor.sd, lwd=1, lty=2, col="red")
abline(v=sample.mean - sample.sd, lwd=2, lty=3, col="blue")
abline(v=sample.mean + sample.sd, lwd=2, lty=3, col="blue")
legend("topright", c(paste0("Range of two\nSample Standard\nDeviations = ",
                            round(2 * sample.sd, 2), "\n"),
                     paste0("Range of two\nTheoretical Standard\nDeviations = ",
                            round(2 * theor.sd, 2))),
       lty=c(3,2), lwd=c(2,1),col=c("blue","red"), bty="n", cex=0.8)
```



Distribution

Histogram's shape is bell-like, let's draw a normal distribution's bell with the theoretical mean and standard deviation on top of it to illustrate this.

```r
hist(means, breaks=breaks, main="Distribution of sample means
is approximately normal",
     xlab="Means of samples from exponential distribution",
     ylab="Number of samples with means in each bin")
curve(192*dnorm(x, mean=1/lambda, sd=1/lambda/sqrt(40)),
      col="darkblue", lwd=2, add=TRUE, yaxt="n")
```

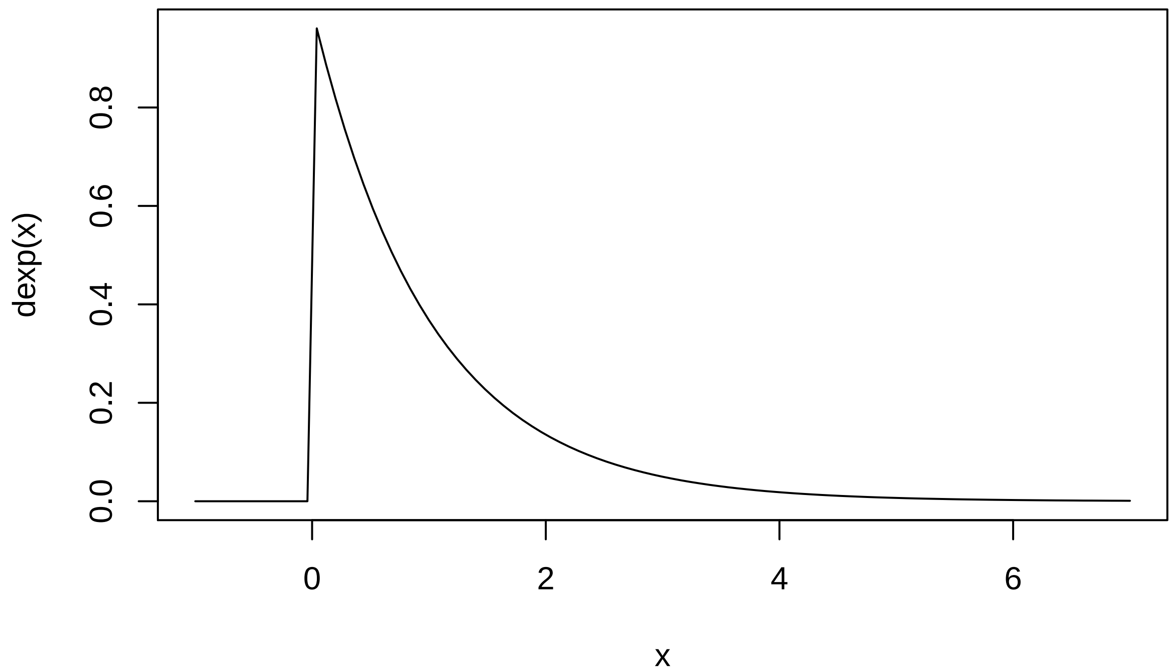


Appendices

Appendix 1. Shape of exponential distribution

This is the shape of exponential distribution. We can look at it by plotting its density function.

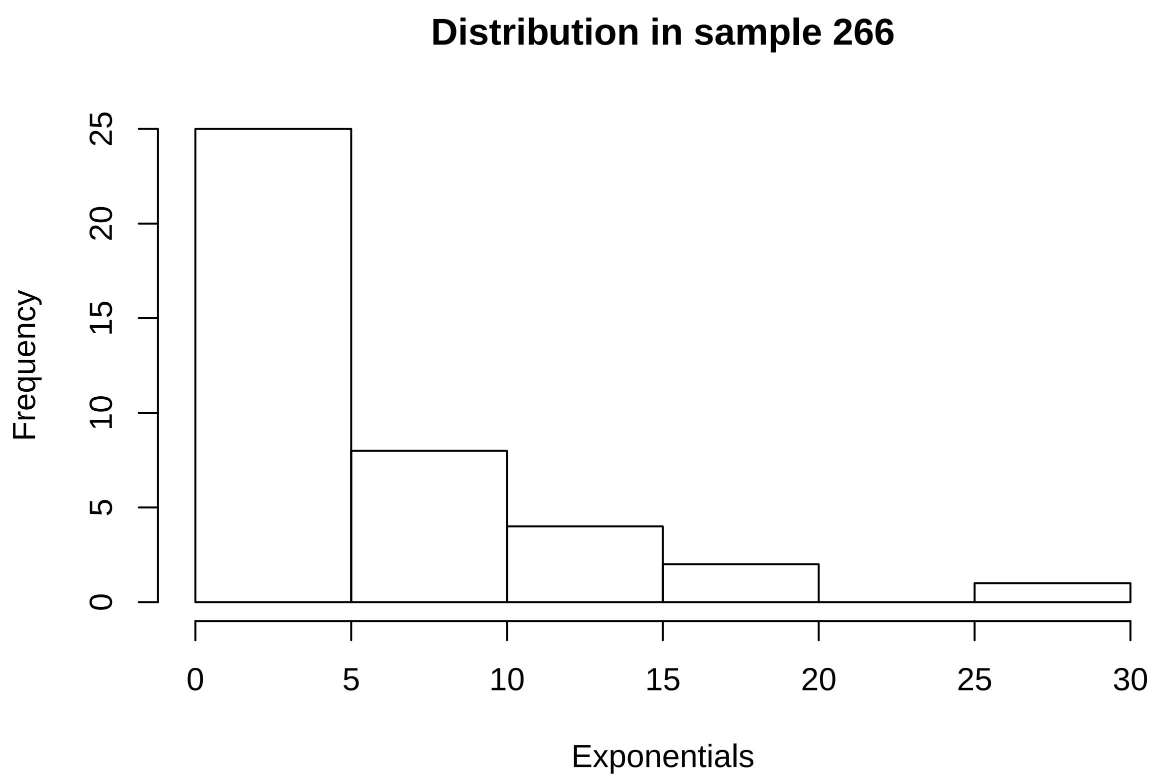
```r
curve(dexp(x),-1,7)
```



Appendix 2. Look at a random sample

If we take a random sample and see it, we should see something similar to the previous plot of the distribution function for the exponential distribution. And its mean and standard deviation should be close to 5 both.

```
set.seed(1)
random.sample <- sample(1:numsim,1)
hist(dat[[random.sample]],
     main=paste("Distribution in sample", random.sample),
     xlab="Exponentials")
```




```
(random.mean <- mean(dat[[random.sample]]))
```

```
## [1] 5.232086
```

```
(random.sd <- sd(dat[[random.sample]]))
```

```
## [1] 5.491864
```

Appendix 3. Determining scale factor for norm plot over histogram

Because normal distribution line at the figure above is drawn on the same vertical scale as the histogram, it had to be vertically scaled. Let's use another simulation to determine the scale, assuming that hist()'s plot function algorithm isn't transparent to us, and so we don't know a way to calculate this theoretically. In order to make it more symmetrical, number of observations should be largely increased.

```
larger <- 1000
round(max(hist(rnorm(numsim * larger), breaks=breaks, plot=FALSE)$counts)/larger)
```

```
## [1] 192
```