Do exponential distributions become Gaussian with increasing samples?

@tribetect

August 18, 2015

Overview

We investigate the exponential distribution and compare it with the Gaussian distribution to test the Central Limit Theorem (CLT) by simulating random exponential values, and contrasting the characteristics of such values against a normal distribution.

Simulations

A base set of random exponential values is generated by repeatedly sampling, 40 values at a time, from the exponential distribution using rexp(samplesize, lambda) where lambda is the rate parameter, given as 0.2 for this assignment. For comparision, we bear in mind that the mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda, or $1/.2 = 5$ for this assignment.

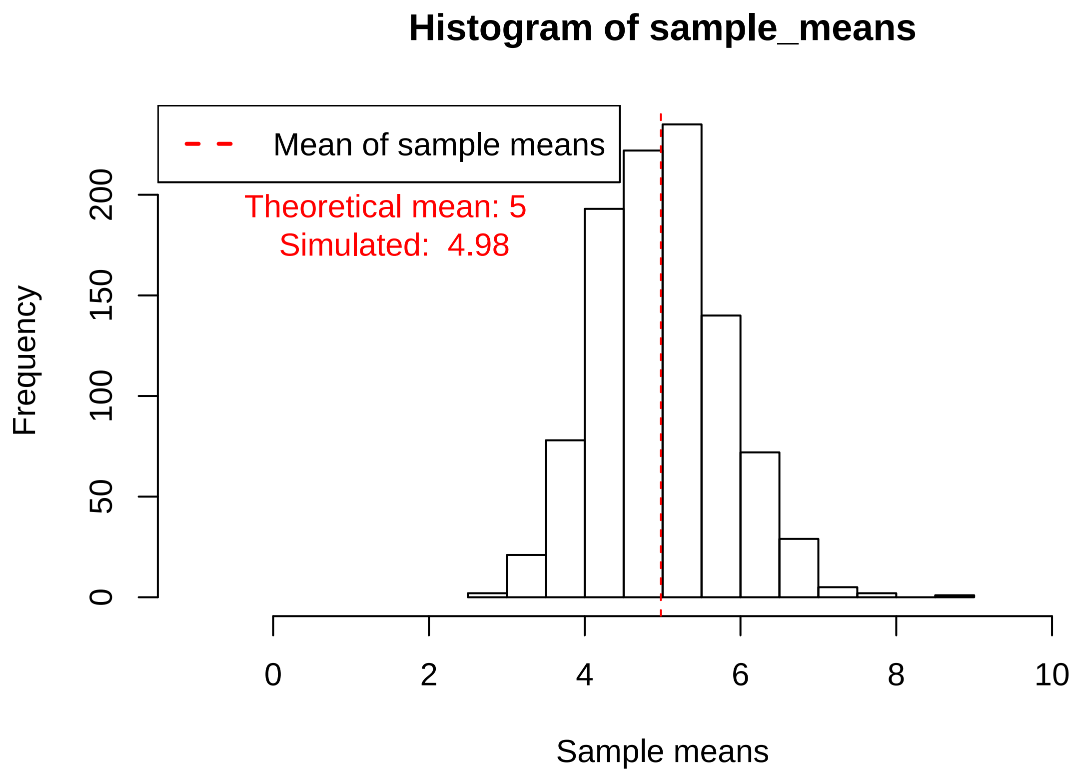
```
my_seed = 512 #for random number generators based simulations
set.seed(my_seed)

# Setup the supplied quantities as variables
lambda <- 0.2
samplesize <- 40 #size of random exponential samples
nosim = 1000 #number of simulations
sample_means = NULL #variable to store means of samples
sample_var = NULL #variances of samples
```

Comparing Sample Mean and Theoretical Mean

We plot the distribution of a 1000 simulated means of sampletaken from an exponential distribution. Size of each sample was 40.

The theoratical mean of the distribution, shown using a vertical red line, is close to the center of the distribution of means of the samples.

```
for (i in 1 : nosim) sample_means = c(sample_means, mean(rexp(samplesize, lambda)))
hist(sample_means, xlab = "Sample means",xlim = c(-1,11))
main = "Distribution of sample means"
theo_mean <- mean(sample_means)
abline(v = theo_mean, col = "red", lty = 2)
text(x = 1.5, y = 185, labels = paste("Theoretical mean: 5", "\n",
                                      "Simulated: ", round(theo_mean,2)), col = "red", lwd = 2)
legend("topleft", legend = "Mean of sample means", lty = 2, col = "red", lwd = 2)
```



Comparing Distributions

The distribution of means is approximately symmetrical around the mean, as shown from the figure above and the close values of mean and median:

```
mean(sample_means)
```

```
## [1] 4.97779
```

```
median(sample_means)
```

```
## [1] 4.94445
```

Sample Variance versus Theoretical Variance

Theoretical variance

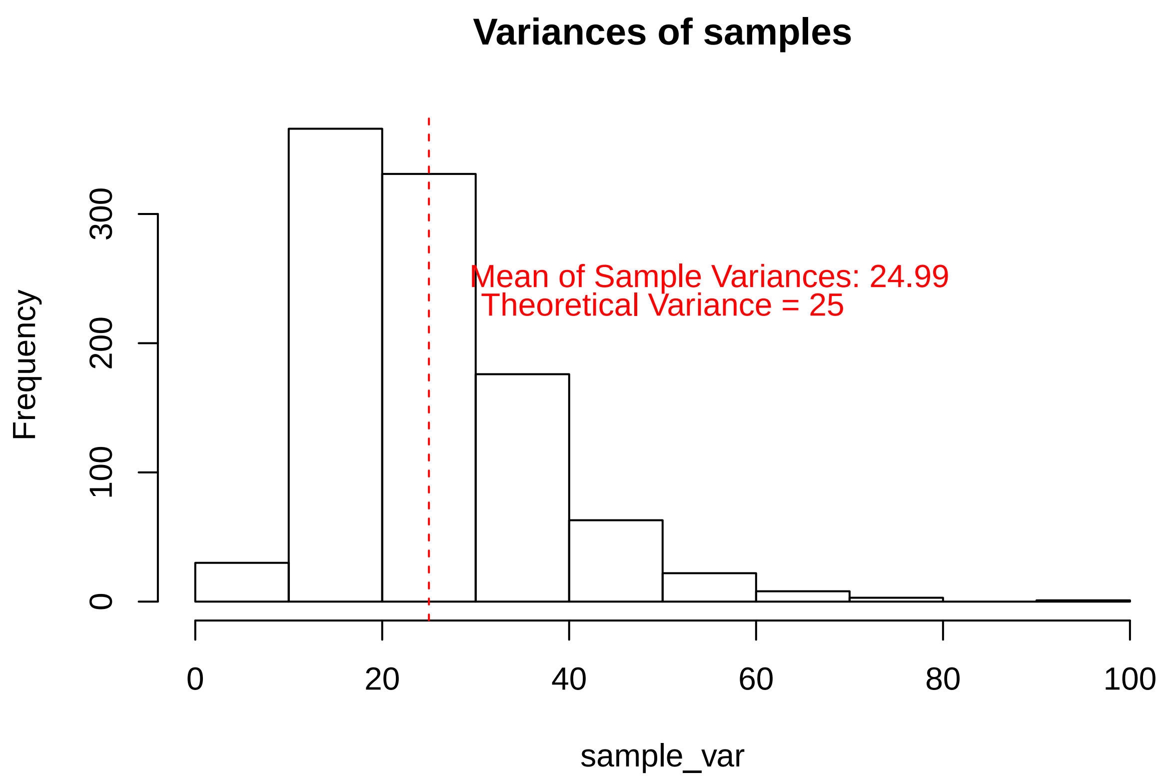
Variance is (standard deviation)^2 For the given distribution, standard deviation = 1/lambda, i.e. 1/0.2 = 5

```
theo_variance = (1/lambda)^2
theo_variance
```

```
## [1] 25
```

Variance from simulation

```
set.seed(my_seed)
for (i in 1 : nosim) sample_var = c(sample_var, var(rexp(samplesize, lambda)))
hist(sample_var, main = "Variances of samples")
abline(v = theo_variance, col = "red", lty = 2)
text(theo_variance + 25, y = 230, labels = paste("Theoretical Variance = 25"), col = "red")
text(theo_variance + 30, y = 250, labels = paste("Mean of Sample Variances:", round(mean(sample_var),2)
```



Conclusion: Answering the big question

The mean of sample variances is very close to the theoretical variance of 25

The shape, mean, distribution of mass around median, and variances for the sampled data and the theoretical values support the Central Limit Theorem.