

Comparison of Exponential Distribution vs. Central Limit Theorem

Fraser Stark

07/04/2020

Synopsis

The purpose of this example is to demonstrate the exponential distribution in R while comparing it with the Central Limit Theorem (CLT). The CLT specifies that the distribution of averages of iid variables become that of a standard normal, as the sample size increases. The exponential distribution can be simulated in R by using `rexp(n, lambda)` where `lambda` is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. The rate parameter (`lambda`) will be set to 0.2 for all of the simulations. The example will compare the distribution of averages of 40 exponentials, over 1000 simulations.

Simulation exercise

Firstly we need to set the seed parameter for reproducibility, and the rate parameter (`lambda`), the sample size (40) and the number of simulations (1000).

```
set.seed(765)
lambda <- 0.2      # Rate parameter
n <- 40            # Sample size
sims <- 1000       # Number of simulations
```

To generate the random exponential variables we can use the R function `rexp()` which takes the sample size and rate as arguments. The R function `apply()` lets us take the mean for each of the samples, and we store the results in a data.frame object.

```
sample_means <- data.frame(x = apply(1:sims, function(x) mean(rexp(n, lambda))))
```

Sample Mean vs. Theoretical Mean

The sample mean is an unbiased estimator for the population mean. The CLT specifies that the mean of any sampling distribution will converge towards the true mean of its population. Here we can calculate the theoretical mean of the exponential distribution which is specified as $1/\lambda$:

```
1 / lambda
```

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

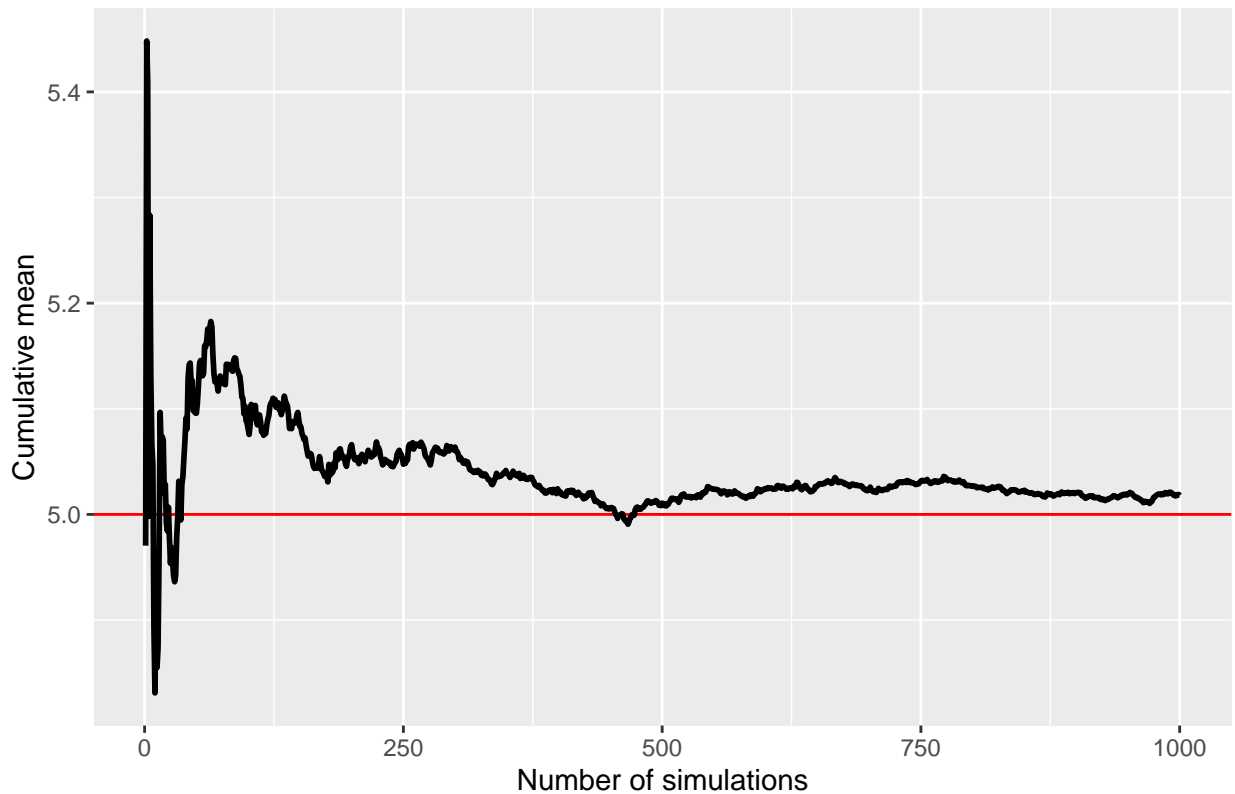
And if we compare this with the actual mean of the simulation which provided our sample, we can see it converges to the theoretical mean.

```
mean(sample_means$x)
```

```
## [1] 5.017824
```

We can see that the sample mean (5.0178245) is a good approximation of the theoretical mean (5) and will continue to converge as the sample size increases, which can be demonstrated by the plot below.

Sample Means of 1000 Samples of 40 Random Exponentials



Sample Variance vs. Theoretical Variance

The sample variance is the estimator of the population variance. Note that the population variance is the expected squared deviation around the population mean.

So therefore we can calculate the Theoretical Variance and the sample variance as follows:

```
(1/lambda)^2 / n
```

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

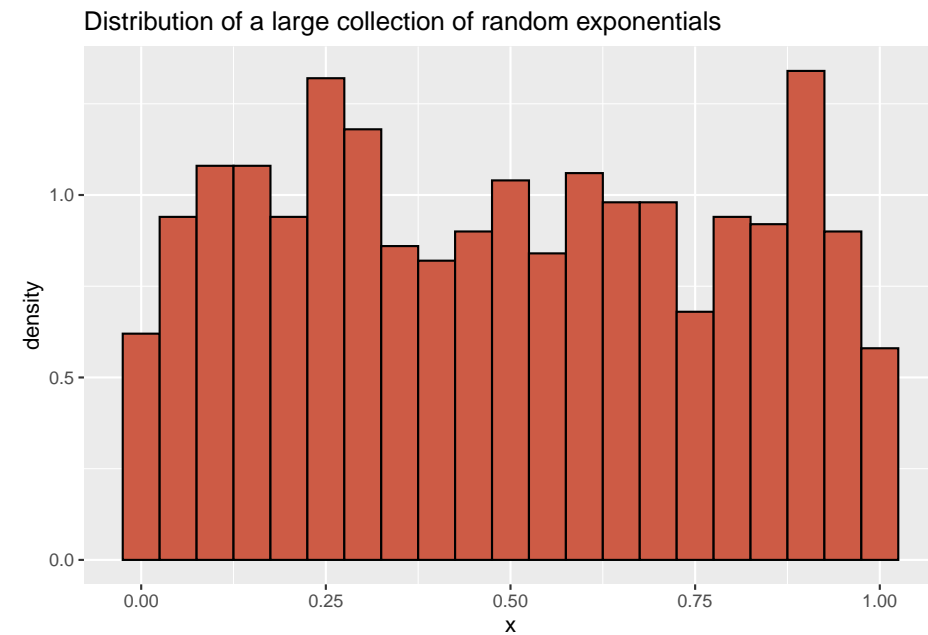
```
var(sample_means$x)
```

```
## [1] 0.5919431
```

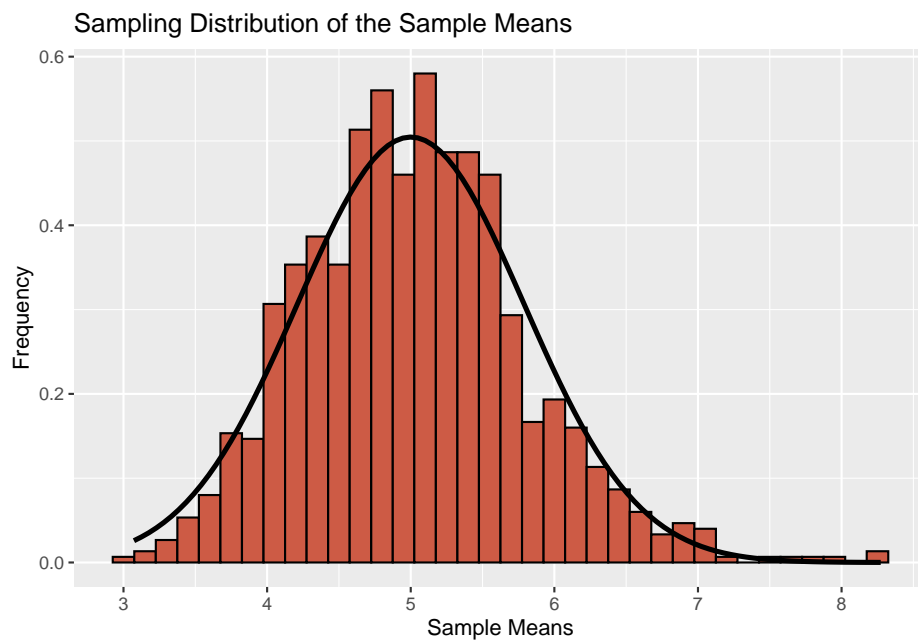
As we can see, the sample variance (0.5919431) provides a good approximation to the theoretical variance (0.625), as both have values close to 0.6. Note that as we collect more data, the distribution of the sample variance gets more concentrated around the population variance that it's estimating.

Distribution Shape

Let's first compare the distribution of a large collection of random exponentials (using `runif(1000)`) and the distribution of a large collection of the mean of 40 exponentials.



As we can see from the above, the distribution of 1000 random uniforms are more or less equally distributed



And as we can see from this plot, this distribution of means from our sample looks much more like the standard normal distribution.

Appendix: Code used to produce the plots found above

Plot 1 - Sample Mean vs. Theoretical Mean

```
library(ggplot2)

means <- cumsum(sample_means)/(1:sims)

ggplot(data.frame(x = 1:sims, y = means), aes(x = x, y = x.1)) +
  geom_hline(yintercept = t_Mean, colour = 'red') + geom_line(size = 1) +
  labs(x = "Number of simulations", y = "Cumulative mean") +
  ggtitle("Sample Means of 1000 Samples of 40 Random Exponentials")
```

Plot 2 - Showing the distribution of a large collection of random exponentials

```
library(ggplot2)

e <- data.frame(x = runif(1000))

ggplot(data = e, aes(x = x)) +
  geom_histogram(aes(y = ..density..),
    binwidth = 0.05,
    color = I("black"),
    fill = I("coral3")) +
  labs(title = 'Distribution of a large collection of random exponentials')
```

Plot 3 - Showing the standard normal distribution of means.

```
library(ggplot2)

ggplot(data = sample_means, aes(x = x)) +
  geom_histogram(
    aes(y = ..density..),
    binwidth = 0.15,
    color = I("black"),
    fill = I("coral3")) +
  stat_function(
    fun = dnorm,
    args = list(mean = t_Mean, sd = sqrt(t_Variance)),
    size = 1.2) +
  labs(
    title = "Sampling Distribution of the Sample Means",
    x = "Sample Means",
    y = "Frequency")
```