Simulating the Exponential Distribution

Konstantinos Patelis

26/01/2021

Overview

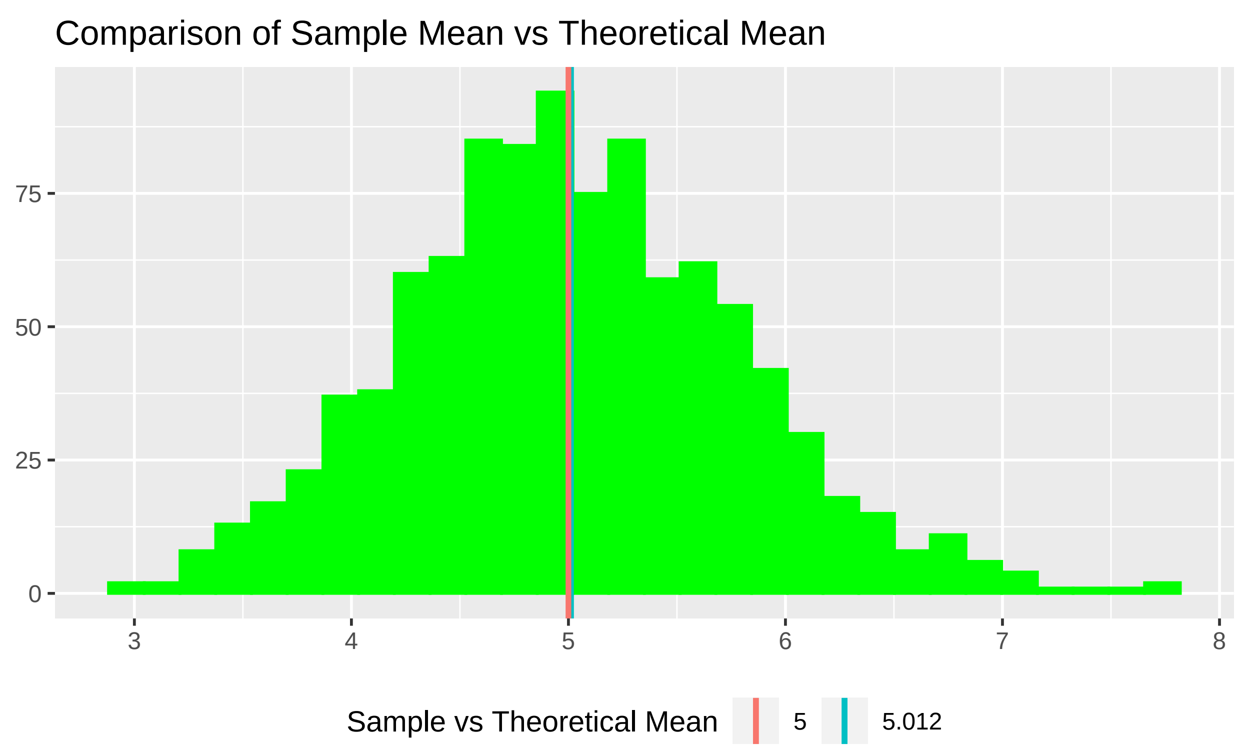
In this report, I am investigating the exponential distribution, simulating a distribution with lambda = 0.2. I will perform 1000 simulations of drawing 40 exponentials and calculating their average. I will compare the sample mean and variance to the theoretical values and show that the distribution is approximately normal. For all the code used in the assignment, please check the appendix.

Simulations

Using a lambda value of 0.2, we simulate the draw of 40 values from an exponential distribution. Using a loop, we will perform 1000 simulations. For each draw, the mean and variance are calculated. The below code accomplishes this (setting a seed for reproducibility):

Sample Mean versus Theoretical Mean

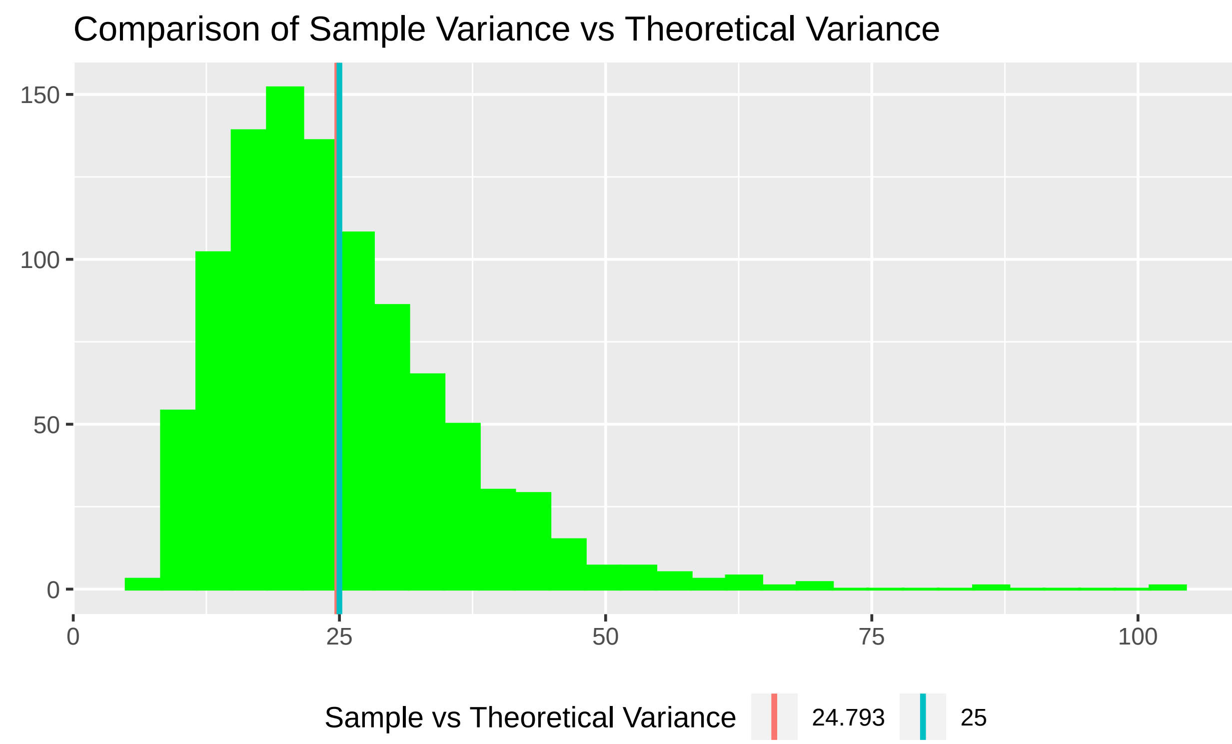
Based on theory, the mean of an exponential distribution is 1/lambda. In this example, mean = 1/0.2 = 5. Below we show the distribution of sample means, and compare the average to the theoretical value.



The above figure shows in green the distribution of sample means from each simulation. The blue vertical line shows the theoretical value of the distribution, while the red line is the average of the sample mean. The two vertical lines look very close to one another; according to CLT, if we continued to perform simulations, the average sample mean would become equal to the theoretical value. The difference between the two is only 0.238%.

Sample Variance versus Theoretical Variance

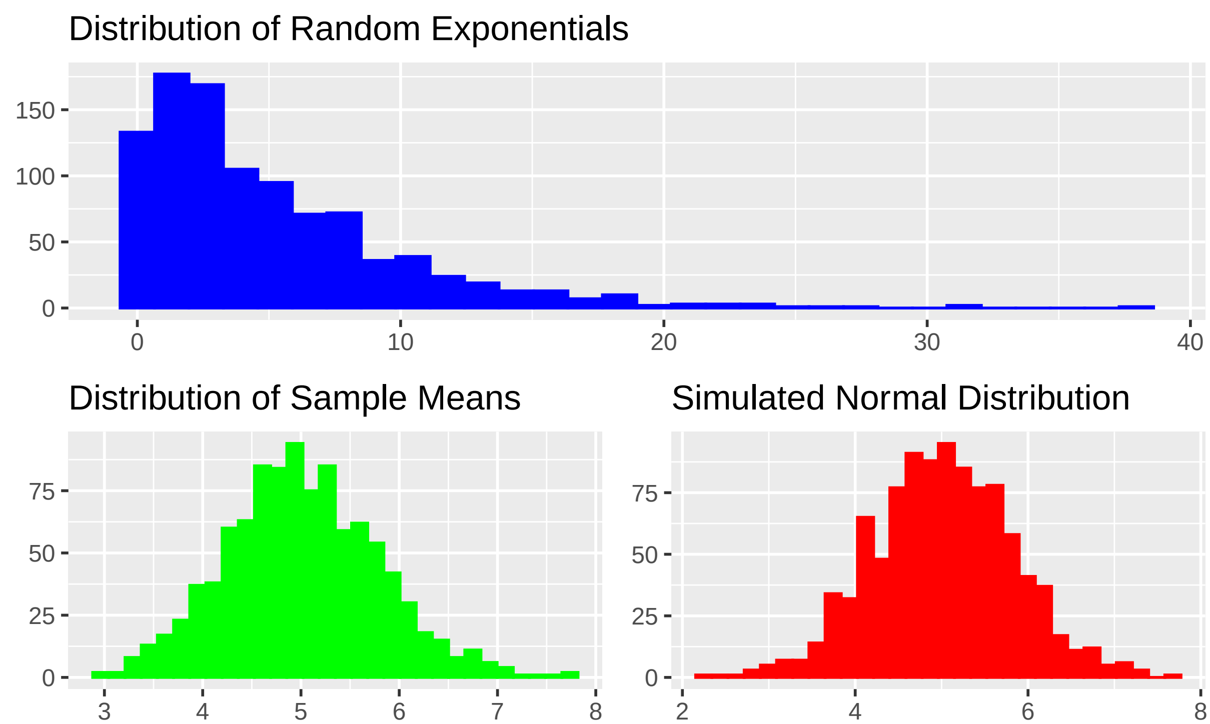
I will now perform the same analysis for the variance. Theoretical variance is the square of standard deviation, which is again 1/lambda = 5. Therefore, theoretical variance is 5ˆ2 = 25. Below code snippet calculates the sample variance distribution, its average and compares to the theoretical value:



This figure shows in green the distribution of sample variances from each simulation. The blue vertical line shows the theoretical value of the distribution, while the red line is the average of the sample variance. The two vertical lines look very close to one another; once again, according to CLT, eventually the average sample variance would become equal to the theoretical value. The difference between the two is -0.828%.

Distribution

Finally, we can see how the distribution of sample means approximates a normal distribution. The easiest way to do this is overlay a simulated normal distribution, with mean equal to the sample mean, which should be approximately 5, and standard deviation equal to the standard error of the sample mean, which should be equal to 5 divided by the square root of 40; the sample size. At the same time, we can look at the distribution of a large collection of random exponentials. We present this using below code to simulate and plot:



The blue histogram is representative of a drawing from an exponential distribution, we see many values surrounding the mean, that decay the further we move away on the positive x-axis. Note that a characteristic of this distribution is that it is bounded to zero. Looking at the green distribution, which corresponds to the distribution of the means of samples of size 40 from the exponential distribution. When comparing with the red distribution, which is actually a normal distribution with same mean and standard deviation as that of our sample means distribution, it is evident that the two are largely similar, meaning that the green is also a normal distribution, confirming the validity of the Central Limit Theorem.

Appendix

```
set.seed(123)
library(tidyverse)
library(patchwork)

lambda <- 0.2

sim_mean <- NULL
for (i in 1 : 1000) sim_mean <-  c(sim_mean, mean(rexp(40, lambda)))

sim_var <- NULL
for (i in 1 : 1000) sim_var <-  c(sim_var, var(rexp(40, lambda)))

mean_exp <- mean(sim_mean)

ggplot(data.frame(sims = sim_mean), aes(x = sims)) +
  geom_histogram(colour = "green", fill = "green", bins = 30) +
  geom_vline(aes(xintercept = mean_exp, colour = factor(round(mean_exp, 3))),
             show.legend = TRUE, size = 1) +
  geom_vline(aes(xintercept = 1/lambda, colour = factor(1/lambda)),
```

```r
             show.legend = TRUE, size = 1) +
  labs(title = "Comparison of Sample Mean vs Theoretical Mean",
       x = NULL,
       y = NULL,
       colour = "Sample vs Theoretical Mean") +
  theme(legend.position = "bottom")

var_exp <- mean(sim_var)

ggplot(data.frame(sims = sim_var), aes(x = sims)) +
  geom_histogram(colour = "green", fill = "green", bins = 30) +
  geom_vline(aes(xintercept = var_exp, colour = factor(round(var_exp, 3))),
             show.legend = TRUE, size = 1) +
  geom_vline(aes(xintercept = (1/lambda)^2, colour = factor((1/lambda)^2)),
             show.legend = TRUE, size = 1) +
  labs(title = "Comparison of Sample Variance vs Theoretical Variance",
       x = NULL,
       y = NULL,
       colour = "Sample vs Theoretical Variance") +
  theme(legend.position = "bottom")

sim_exp <- rexp(1000, lambda)

p3 <- ggplot(data.frame(sims = sim_exp), aes(x = sims)) +
  geom_histogram(colour = "blue", fill = "blue", bins = 30) +
  labs(title = "Distribution of Random Exponentials",
       x = NULL,
       y = NULL)

p4 <- ggplot(data.frame(sims = sim_mean), aes(x = sims)) +
  geom_histogram(colour = "green", fill = "green", bins = 30) +
  labs(title = "Distribution of Sample Means",
       x = NULL,
       y = NULL)

distr <- data.frame(x = rnorm(1000, 5, 5 / sqrt(40)))

p5 <- ggplot(data = distr, mapping = aes(x)) +
  geom_histogram(colour = "red", fill = "red", bins = 30) +
  labs(title = "Simulated Normal Distribution",
       x = NULL,
       y = NULL)

p3 /  (p4 + p5)
```