Simulation Exercise

Sam Edwardes; April 5, 2019

Overview

In this document we will investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLT). The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter:

- mean of exponential distribution is 1/lambda,
- standard deviation is also 1/lambda,
- lambda = 0.2 for all of the simulations.

We will investigate the distribution of averages of 40 exponentials by running **1,000 simulations**.

This document will illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Simulations

To compare the exponential distribution with the CLT, the exponential distribution will be simulated 1,000 times. For each simulation, the sample mean and variance will be calculated.

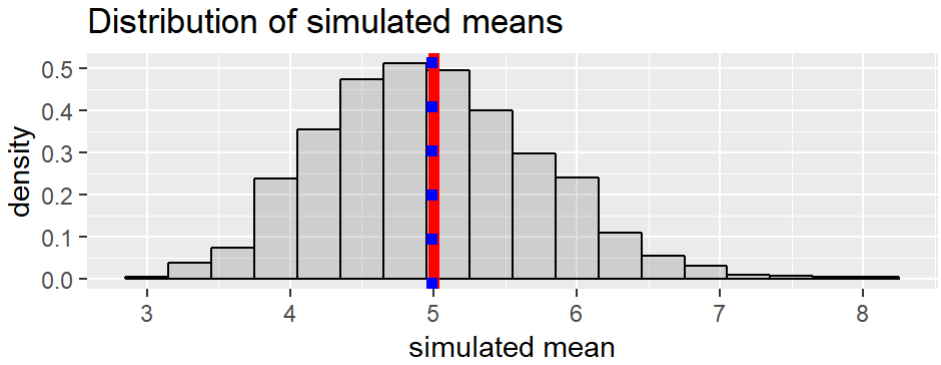
```
# load required libraries
library(ggplot2);library(gridExtra)
```

```
# define simulation peramters
n_sims <- 1000; n_obs <- 40; lambda <- 0.2; u <- 1/lambda; sd <- 1/lambda; var <- sd^2
# create an empty vector to store the simiulation results
sim_means <- c(); sim_var <- c()
# run the simulation n_sims times, and store the mean of each simulation in the sim_means variab
le
for (i in 1:n_sims){
    temp <- rexp(n_obs, lambda)
    temp_mean <- mean(temp)
    temp_var <- var(temp)
    sim_means <- append(sim_means, temp_mean)
    sim_var <- append(sim_var, temp_var)}
# store the results in a dataframe
df <- data.frame(simulated.mean = sim_means, simulated.var = sim_var)
```

Sample mean vs. theoretical mean

Based on the results of the simulation, we can see the means are distributed as follows.

```
# Create a histogram of the means
g <- ggplot(df, aes(x = simulated.mean)) + geom_histogram(alpha = .20, binwidth=.3, colour = "bl
ack", aes(y = ..density..))
g <- g + geom_vline(xintercept = u, color = 'red', size = 2.0)
g <- g + geom_vline(xintercept = mean(sim_means), color = 'blue', size = 2.0, linetype = "dotte
d")
g <- g + labs(title = "Distribution of simulated means", x = "simulated mean")
g
```

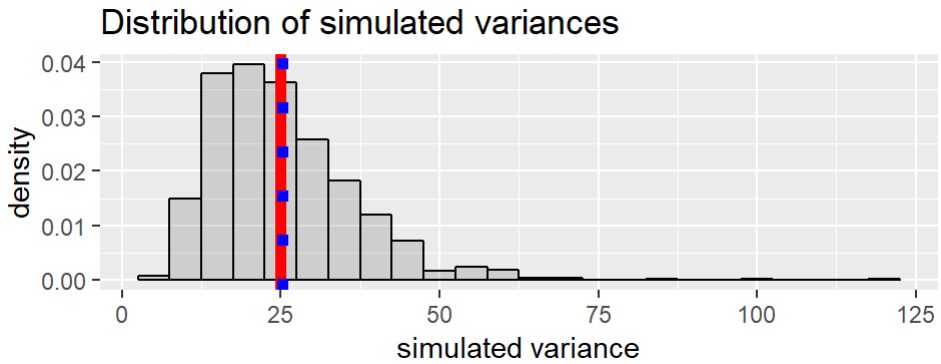


The histogram above shows the distribution of simulated means. The red vertical line shows the theoretical mean of the exponential distribution 5 (1/lambda or 1/0.2). The dotted blue line shows the mean of the simulated means: 4.9867079. As can be seen in the chart, the simulated mean approximates the theoretical mean. It is also interesting to note, that the distribution is Gaussian (normal) in nature.

Sample variance vs. theortical variance

Based on the results of the simulation, we can see the variances are distributed as follows.

```
# Create a histogram of the variances
g <- ggplot(df, aes(x = simulated.var)) + geom_histogram(alpha = .20, binwidth=5, colour = "blac
k", aes(y = ..density..))
g <- g + geom_vline(xintercept = var, color = 'red', size = 2.0)
g <- g + geom_vline(xintercept = mean(sim_var), color = 'blue', size = 2.0, linetype = "dotted")
g <- g + labs(title = "Distribution of simulated variances", x = "simulated variance")
g
```



The histogram above shows the distribution of simulated variances The red vertical line shows the theoretical variance of the exponential distribution 25 ((1/lambda)^2 or (1/0.2)^2). The dotted blue line shows the mean of the simulated means: 25.3286158. As can be seen in the chart, the simulated variances approximates the theoretical

variance It is also interesting to note, that the distribution is Gaussian (normal) in nature once again.

Distributions

As the charts demonstrated above, the distribution of sample means and variances approximates a normal distribution. This is demonstrated below, where we re-plot the charts above, but this time overlay a normal distribution so we can visually compare.

```
# Create a histogram of the means
g1 <- ggplot(df, aes(x = simulated.mean)) + geom_histogram(alpha = .20, binwidth=.3, colour = "b
lack", aes(y = ..density..))
g1 <- g1 + geom_vline(xintercept = u, color = 'red', size = 2.0)
g1 <- g1 + geom_vline(xintercept = mean(sim_means), color = 'blue', size = 2.0, linetype = "dott
ed")
g1 <- g1 + labs(title = "Distribution of means", x = "simulated mean")
g1 <- g1 + stat_function(fun = dnorm, args = list(mean = mean(sim_means), sd = sd(sim_means)))
# Create a histogram of the variances
g2 <- ggplot(df, aes(x = simulated.var)) + geom_histogram(alpha = .20, binwidth=5, colour = "bla
ck", aes(y = ..density..))
g2 <- g2 + geom_vline(xintercept = var, color = 'red', size = 2.0)
g2 <- g2 + geom_vline(xintercept = mean(sim_var), color = 'blue', size = 2.0, linetype = "dotte
d")
g2 <- g2 + labs(title = "Distribution of variance", x = "simulated variance")
g2 <- g2 + stat_function(fun = dnorm, args = list(mean = mean(sim_var), sd = sd(sim_var)))
grid.arrange(g1, g2, ncol=2)
```