Exponential Distribution

Enelen

Overview

This project will investigate the exponential distribution through simulation, and compare it with the Central limit theorem. More specifically, the distribution of means of 40 exponentials will be simulated.

Simulations

R has been used to perform the simulations, and here is the code first.

```
lambda = 0.2  # setting lambda to the specified value
mns <- vr <- vector()  # creating empty vectors for mean and variance
for (i in 1:1000) {
    temp <- rexp(40, lambda)  # create a sample of random exponentials
    mns <- c(mns, mean(temp))  # compute mean of sample
    vr <- c(vr, var(temp))  # compute variance of sample
}
dat = data.frame(means = mns, variance = vr)
```

The lambda has been set to the specified value of 0.2. Thus, the mean and standard deviation of our exponential distribution would be $^1/_\lambda = 5$

A thousand samples of 40 random exponentials were created, and for each, the mean of the sample and its variance were stored in arrays, later combined into a data frame.

Sample Mean and Variance vs their Theoretical values

Let us now look at the result of our simulations, namely, the average value estimated by the simulations for the mean and variance of the exponential distribution.

```
simMean <- mean(mns)  # mean of distribution via simulation
simVar <- mean(vr)  # variance of the distribution
require(pander)  # easy and pretty markdown printing
pander(data.frame(Mean = simMean, Std.Dev. = sqrt(simVar), Variance = simVar))
```

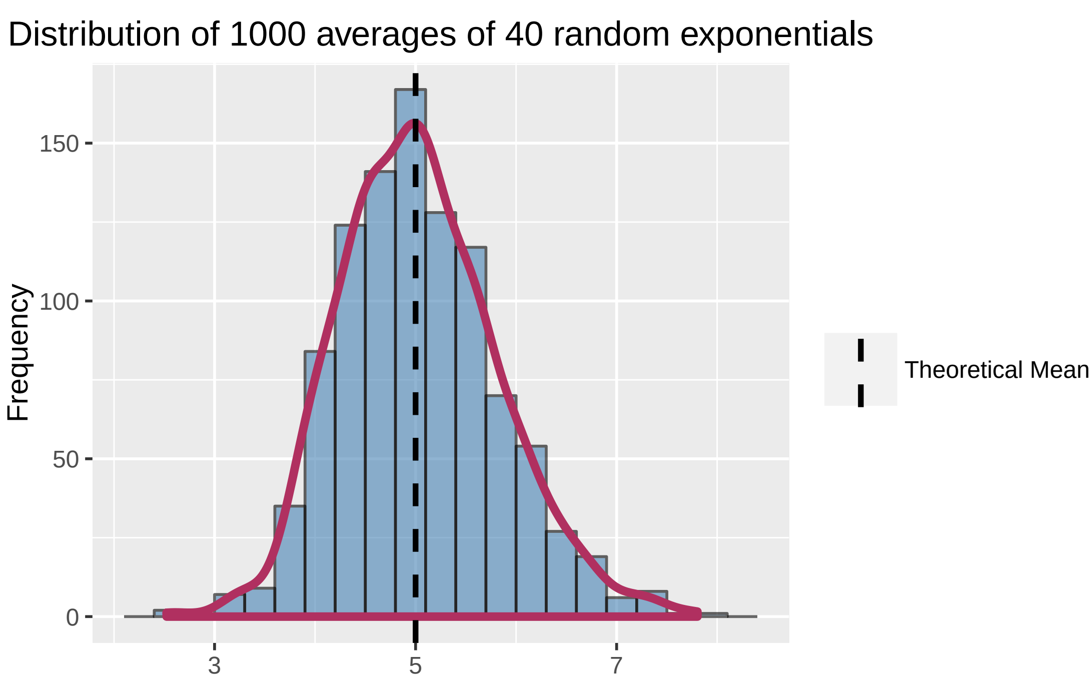
Mean	Std.Dev.	Variance
5.032	5.015	25.15

Both are very close to our theoretical mean and standard deviation of 5 (variance 25). They would have been even closer if we had taken larger samples.

Distribution

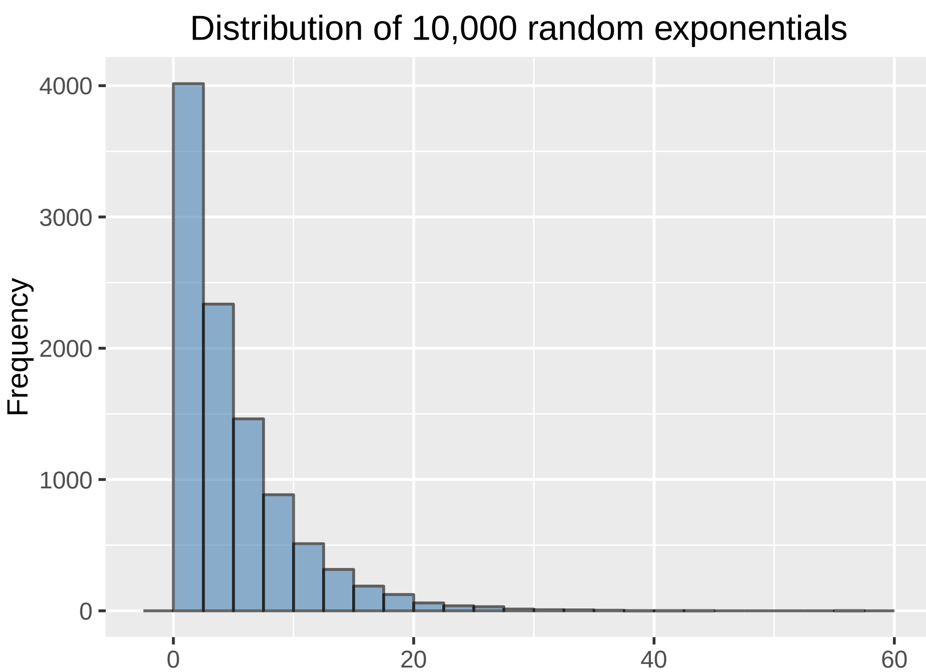
Let us now look at the distribution of our sample means.

```
# Density curve scaling formula (histogram with density curve with correct height)
# bin count = density(N*BinWidth)
# So setting y = ..density..*N*Binwidth
# (where Binwidth has been set in histogram option)
require(ggplot2)
m <- ggplot(dat, aes(means)) +
     geom_histogram(fill = "steelblue", alpha = .6, colour = "black", binwidth = 0.3) +
     geom_density(size = 1.5, colour = "maroon",
                  aes(y = ..density.. * 1000 * 0.3)) +
     geom_vline(aes(xintercept = 1/lambda, linetype = "dashed"), size = 1) +
     scale_x_continuous(breaks = seq(1, 9, 2)) +
     scale_linetype_manual(labels = "Theoretical Mean",values = "dashed",name = "") +
     labs(x = "", y = "Frequency",
          title = "Distribution of 1000 averages of 40 random exponentials") +
     theme(legend.key.size = unit(1,"cm"))
m
```



Sample means, as can be seen in the figure above, are normally distributed. For a similar figure for variance, see appendix. In comparison, here is the distribution of 10,000 random exponentials.

```
dat2 <- data.frame(dist = rexp(10000,lambda))
ggplot(dat2, aes(dist)) +
     geom_histogram(fill = "steelblue", alpha = .6, colour = "black", binwidth = 2.5) +
     labs(x = "", y = "Frequency",
          title = "Distribution of 10,000 random exponentials")
```



The distribution of random exponentials is not normal, but highly positively skewed. The distribution of sample means however, is Gaussian by the virtue of the Law of Large Numbers, and the mean of the large number of samples is close to the expected value as it should be.

Appendix

Distribution of sample standard deviation

```
v <- ggplot(dat,aes(sqrt(variance))) +
     geom_histogram(fill = "steelblue",alpha = .6,colour = "black",binwidth = 1/2)  +
     geom_density(size = 1.5,colour = "maroon",aes(y = ..density..*1000/2)) +
     geom_vline(aes(xintercept = 1/lambda, linetype = "dashed"), size = 1) +
     scale_x_continuous(breaks = seq(1, 9, 2)) +
     scale_linetype_manual(labels = "Theoretical SD",values = "dashed",name = "") +
     labs(x = "", y = "Frequency",
          title = "Distribution of sample standard deviation") +
     theme(legend.key.size = unit(1,"cm"))
v
```