The distribution of sample means obtained from an exponential distribution

Jae Wook Lee

October 23, 2015

Overview

In this document, the properties of sample means obtained from an exponential distribution are analyzed in terms of their relationship to the population mean and variance.

Simulations

The following R code simulates sampling of 1000 sets of 40 exponentials. The rate parameter for the exponential distribution (lambda) is set to 0.2.

```
lambda = 0.2   # rate for the exponential distribution
nosim=1000     # number of simulations
n = 40         # size of a sample
samples = replicate(nosim, mean(rexp(n, rate=lambda)))  # a vector of 1000 sample means
```

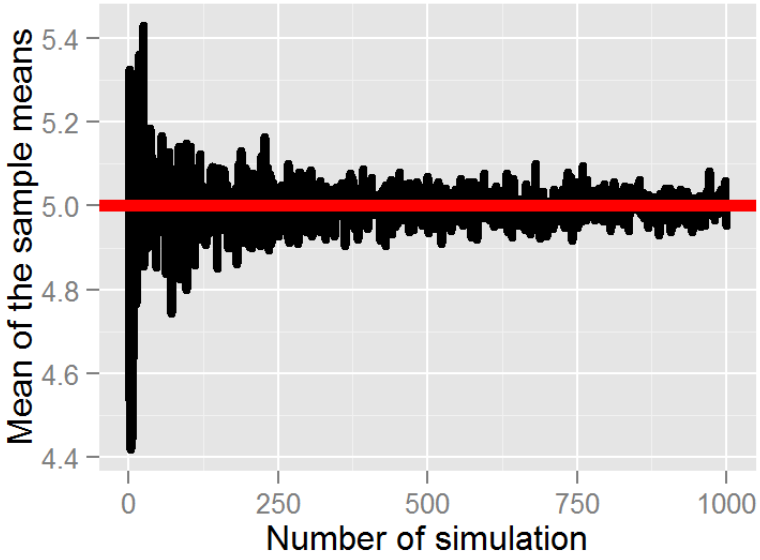
Sample mean versus Theoretical Mean

The mean of the sample means are 4.9886349. It is almost equal to the population mean 5. In the following code, we can see that the mean of the sample means approaches the theoretical mean as the number of simulation grows.

```
means <- sapply(1:nosim, function(x) mean(replicate(x, mean(rexp(n, lambda)))))
require(ggplot2)
```

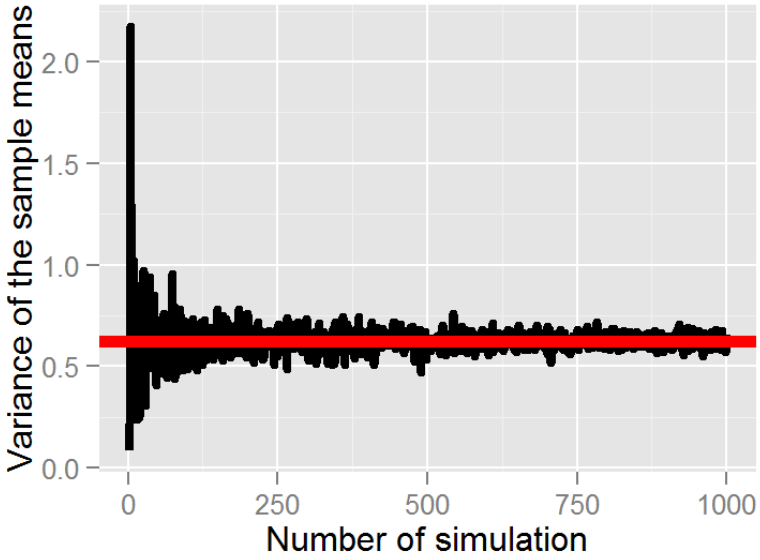
```
## Loading required package: ggplot2
```

```
g <- ggplot(data.frame(x = 1 : nosim, y = means), aes(x = x, y = y)) + geom_line(siz
e=1.5)
g <- g + geom_hline(yintercept = 1/lambda, col="red", size=2)
g <- g + labs(x = "Number of simulation", y = "Mean of the sample means")
g
```



Sample variance vs Theoretical variance

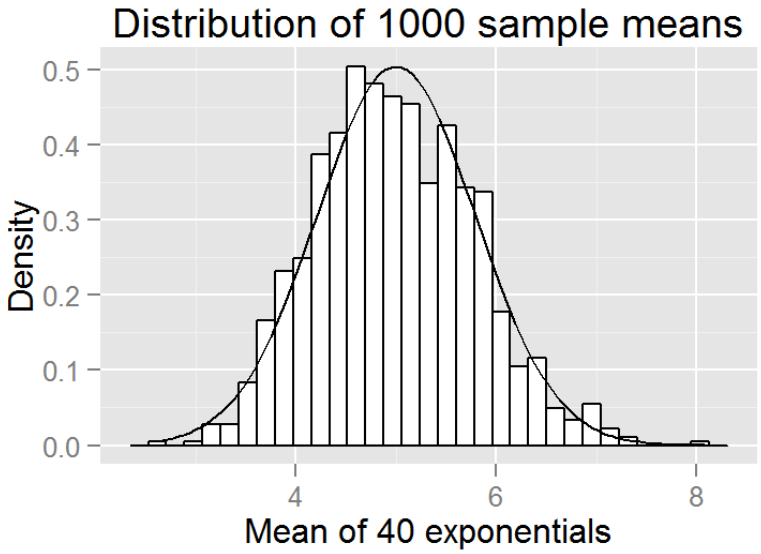
The theoretical variance for the exponential distribution with rate = 0.2 is (1/0.2)^2 = 25. The variance of 1000 sample means is 0.6290536, which is very close to the population variance divided by the size of the sample (40) = 25/40 = 0.625. Similar to the mean of the sample means, the variance of the sample means approaches population variance/sqrt(40) as the number of simulation grows.



The distribution of 1000 sample means is approximately normal.

The histogram for the sample means shows that they approximate a normal distribution with a mean of 1/lambda and standard deviation of (1/lambda)/(sqrt(n)).

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



The quantile-quantile plot below also shows that 1000 samples means are normally distributed.