A Simulation of Exponential Distribution

Zhongyu

Overview

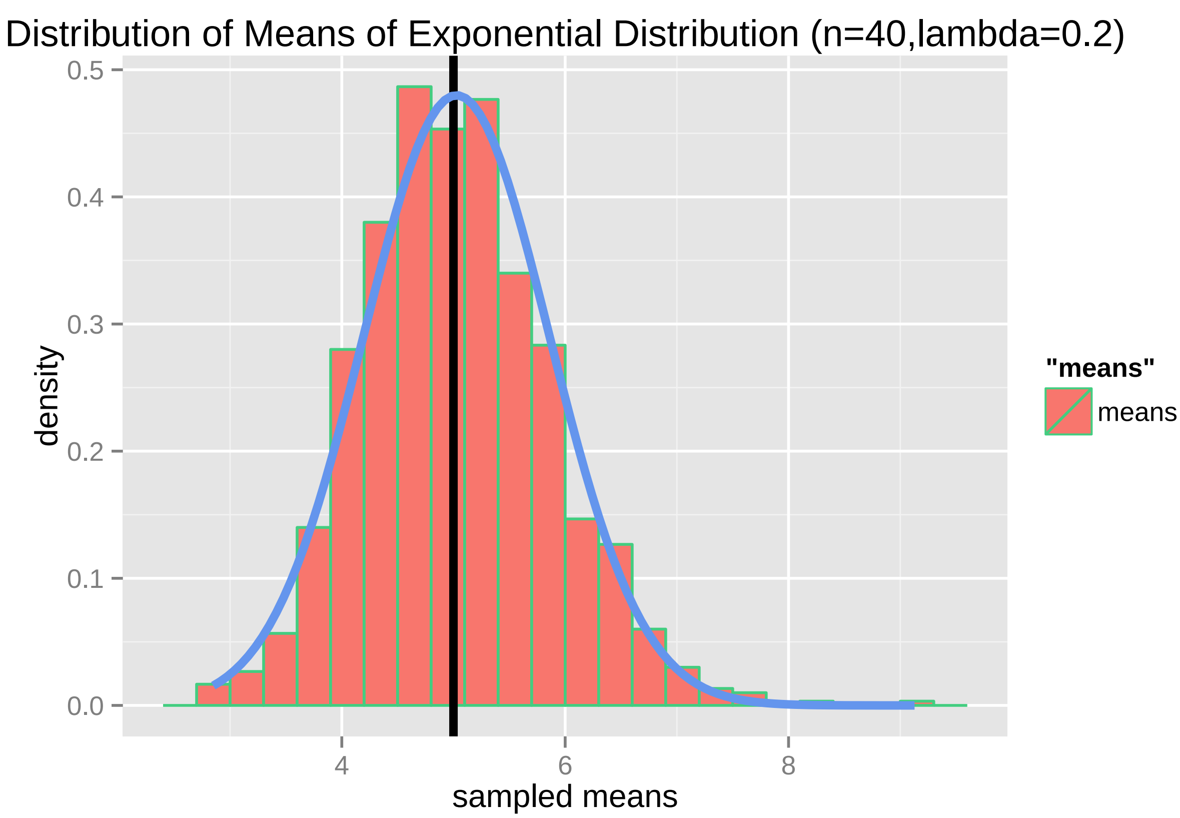
This project verified that Central Limit Theorem is applicable to exponential distribution.

Simulation

Both the mean and standard deviation of exponential distribution are determined by parameter lambda, and can be calculated by 1/lambda. For each simulation, lambda is fixed to 0.2, and 40 observations are collected. The mean and variance for each simulation are collected.

```
expdistr <- rexp(1000, rate=.2);expmeans <- NULL; expvars <- NULL;
for (i in 1:1000){expmeans <- c(expmeans, mean(rexp(40, rate = .2)));
                  expvars <- c(expvars, var(rexp(440, rate = .2)))}
simdata <- data.frame(expdistr, expmeans, expvars)
```

Sample Mean versus Theoretical Mean

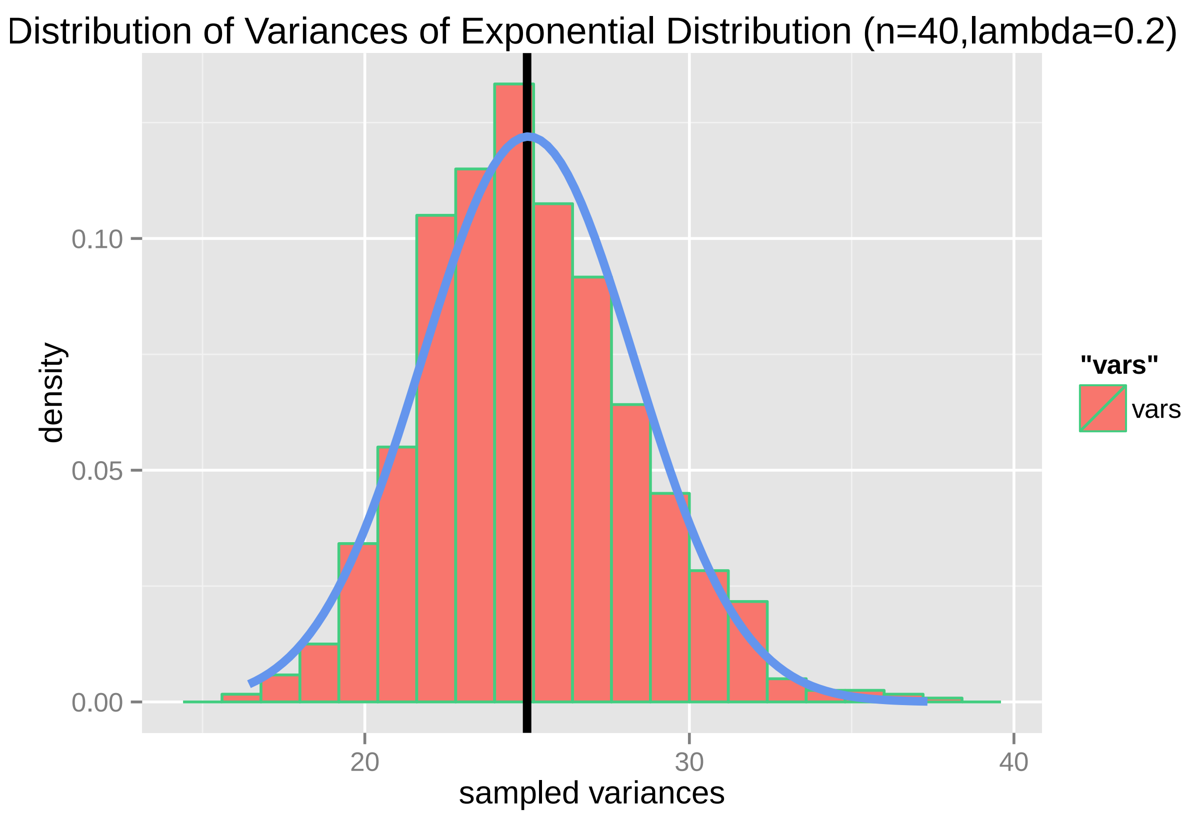


```
## [1] "the mean of simulated means is: 5.02707527951095"
```

The sampling distribution of the 1000 simulated means of exponential distribution (n = 40 observations, lambda = 0.2) is plotted through following code, with the vertical black line denoting the theoretical mean 1/lambda = 5, and the blue bell curve depicting the normal distribution.The 1000 sampled/simulated means forms a normal distribution, and center around the black vertical line, x=1/lambda=5, which is the theoretical mean of the simulated exponential distribution. The mean of the sampled distribution of means (who is a set of realizations of a random variable) is exactly what the means are trying to estimate, therefore it is an unbiased estimate

Sample Varance versus Theoretical Varance

The sampling distribution of the 1000 simulated variances of exponential distribution (n = 40 observations, lambda = 0.2) is plotted through following code, with the vertical black line denoting the theoretical variance (1/lambda)ˆ2 = 25, and the blue bell curve depicting the normal distribution.




```
## [1] "the mean of simulated variances is: 25.0338211812899"
```

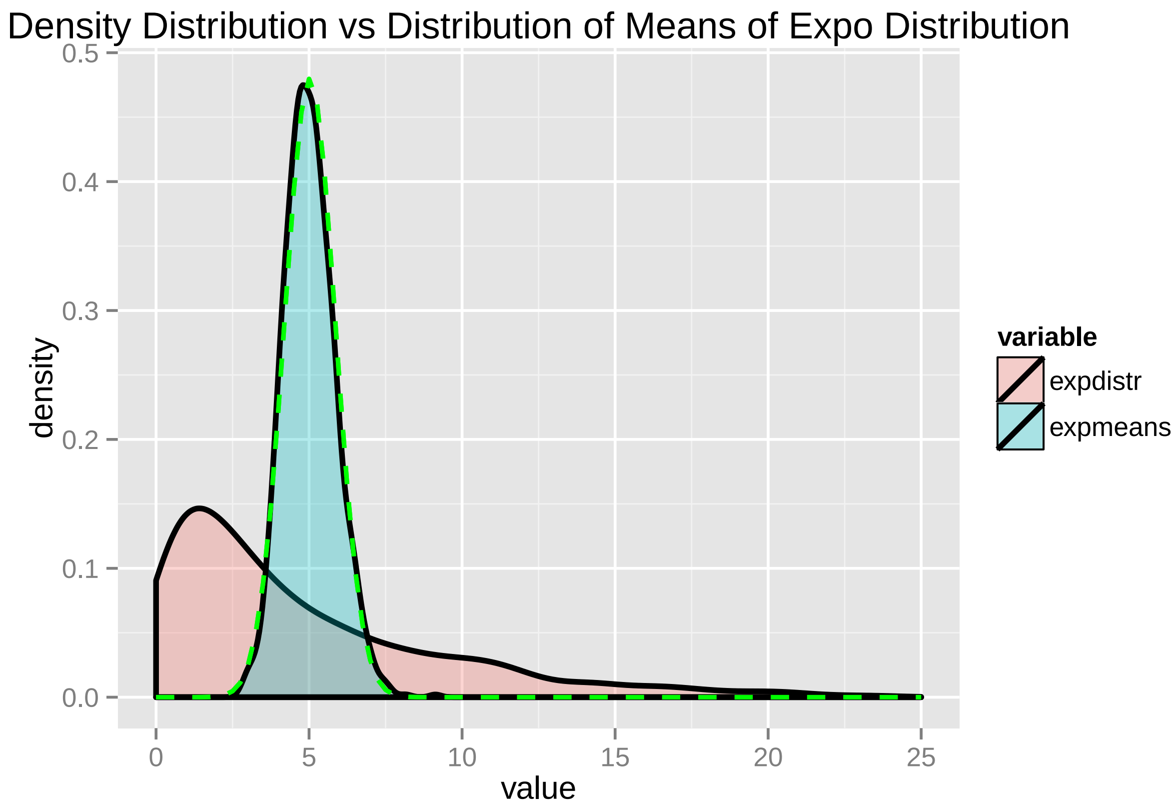
The 1000 sampled/simulated variances forms a normal distribution, and center around the black vertical line, x=(1/lambda)ˆ2=25, which is the theoretical variance of the simulated exponential distribution. The mean of the sampled distribution of variances (who is a set of realizations of one random variable) is exactly what the variances are trying to estimate, therefore it is an unbiased estimate.

Distribution

A large collection of random exponentials is samples collected from the exponential distribution itself, therefore it is likely to see the values in the high probability density region appear more frequent in the collection. However, a large collection of averages of 40 exponentials is likely to present a normal distribution, which is stated in Central Limit Theorem, which is also verified by this study.

```
## Warning: package 'reshape2' was built under R version 3.1.3
```

```
## Warning in loop_apply(n, do.ply): Removed 6 rows containing non-finite
## values (stat_density).
```



The figure showing above plots both the distribution of the 1000 simulated exponentials (the pink shaded area) and the distribution of the 1000 simulated distribution of averages/means of 40 exponentials (the blue shaded area). As it is indicating, the blue shaded area is pretty much overlapping the normal distribution curve which is plotted in the green dashed line, just as stated by Central Limit Theorem.