A Simulation of Exponential Distribution

Zhongyu

Overview

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

Simulation

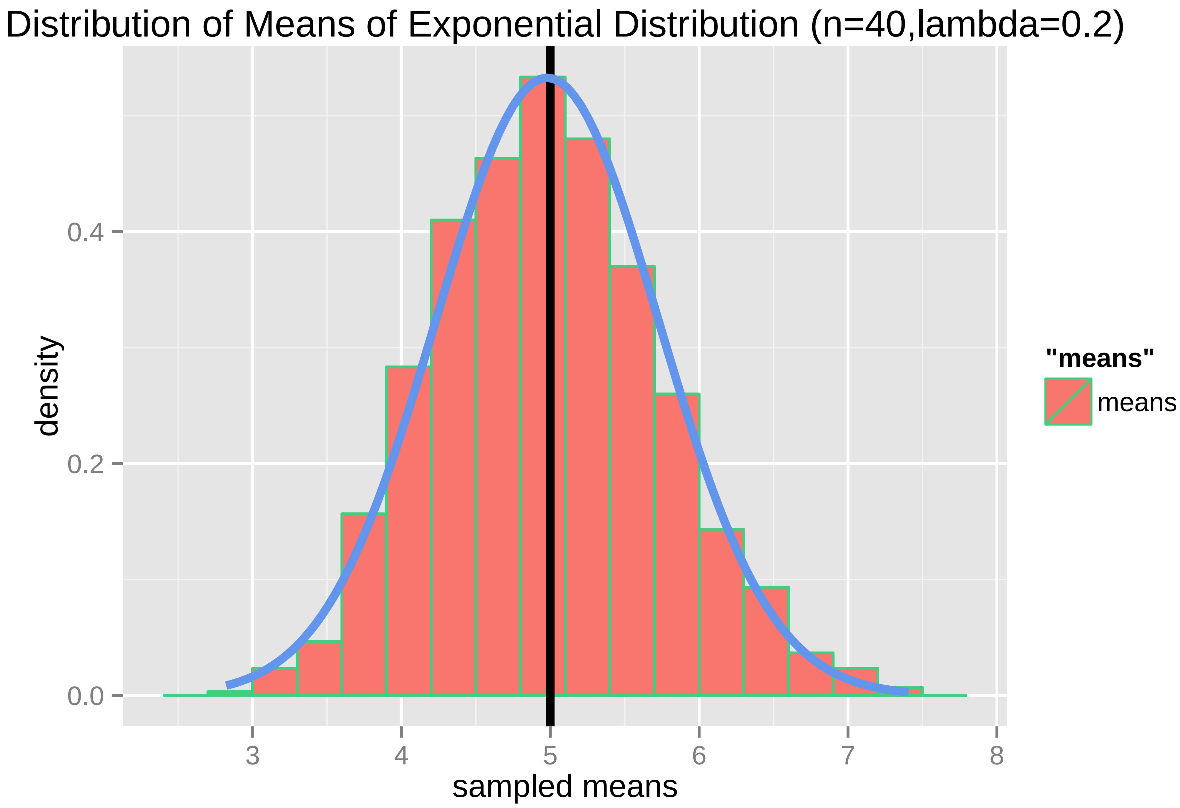
Both the mean and standard deviation of [exponential distribution] are determined by parameter lambda, and can be calculated by 1/lambda. For each simulation, lambda is fixed to 0.2, and 40 observations are collected. The mean and variance for each simulation are collected.

```
lambda=0.2; n=40;
expdistr <- rexp(1000, rate=lambda);expmeans <- NULL; expvars <- NULL;
for (i in 1:1000){expmeans <- c(expmeans, mean(rexp(n, rate = lambda)))}
simdata <- data.frame(expdistr, expmeans)
```

Sample Mean versus Theoretical Mean

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.

```
library(ggplot2); p <- ggplot(simdata, aes(x=expmeans, fill="means"))
p <- p + geom_histogram(binwidth=.3,colour="seagreen3",aes(y=..density..))
p <- p + geom_vline(xintercept=5, size=1.5, colour="black")
p <- p + stat_function(fun=dnorm,
        args=list(mean=mean(simdata$expmeans),sd=sd(simdata$expmeans)),size=1.5, colour="cornflowerblue
p + labs(title="Distribution of Means of Exponential Distribution (n=40,lambda=0.2)") + xlab("sampled me
```



The mean of the sampled distribution of means (who is a set of realizations of a random variable) is computed and printed out, with the theoretical mean is computed and printed out following it. As it is shown, the sample mean and the theoretical mean is fairly close to each other.

```
simulatedmean <- mean(simdata$expmeans)
print(paste0("the mean of simulated means is: ", as.character(simulatedmean)))
```

```
## [1] "the mean of simulated means is: 4.97897841163578"
```

```
print(paste0('the theoretical mean is: 1/lambda=',as.character(1/lambda)))
```

```
## [1] "the theoretical mean is: 1/lambda=5"
```

Sample Variance versus Theoretical Varance

In this section, the varibility of the sample is computed and compared with the theoretical variance of the distribution through following code.

```
simulatedvar <- var(simdata$expmeans)
print(paste0("the variance of the simulated sample is: ", simulatedvar))
```

```
## [1] "the variance of the simulated sample is: 0.560784156258281"
```

```
print(paste0("the theoretical variance of the distribution is the population variance/n:",as.character(
```

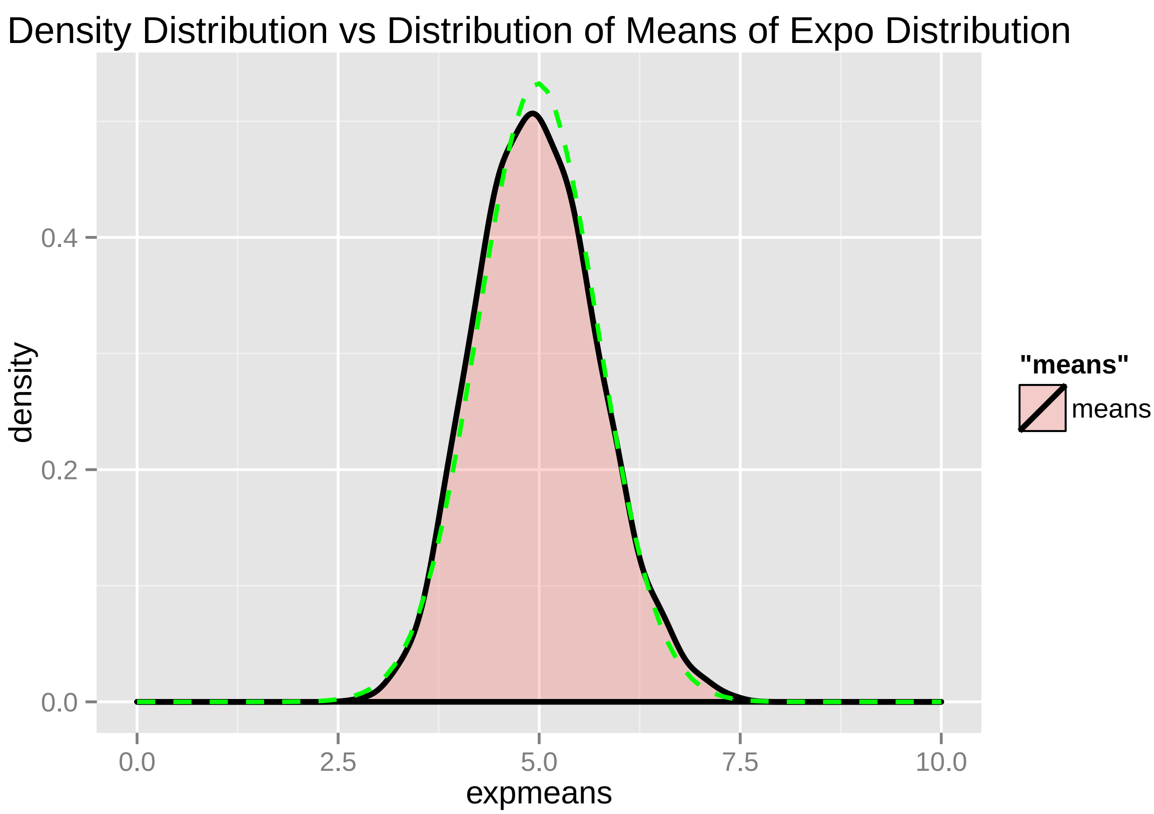
```
## [1] "the theoretical variance of the distribution is the population variance/n:0.625"
```

The Central Limit Theorem states that the variance of the sampling distribution of the mean is theoretically approximates to the true population variance/n. As it is computed, the variance of the sample is slightly different from the theoretical variance of the distribution of 0.625. However the difference is not very big. And this difference is expected to decrease with the increasing simulation runs (i.e. increase to 5000 simulations)

Distribution

A large collection of means of 40 sample from exponential distributions is likely to present a normal distribution shape bell curve, which is stated in Central Limit Theorem. This is studied by plotting the sample density distribution, and then overlay the normal bell curve (green dashed curve) on the density distribution.

```
p <- ggplot(simdata,aes(x=expmeans,fill="means"))
p <- p + geom_density(alpha=.3,size=1)
p <- p + stat_function(fun=dnorm,
                  args=list(mean=mean(simdata$expmeans),sd=sd(simdata$expmeans)),
                  size=0.8, colour="green", linetype="dashed")
p + labs(title="Density Distribution vs Distribution of Means of Expo Distribution") + xlim(0,10)
```



As it is shown in the figure, there is some disagreement between the pink shaded area and the green dashed bell curve, but overall they display similar curve shape. Just as the Central Limit Theorem stated, the sample distribution is approximately normal.