

BasicDiff

NithaDuff

7/23/2020

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

Central Limit Theorem

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. Set `lambda = 0.2` for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

```
lambda <- 0.2  
mean <- sd <- 1/lambda  
n <- 40  
nosim <- 1000  
x <- rexp(n,lambda)
```

```
samples <- matrix(sample(x,n*nosim,replace = T),nosim,n)  
means <- apply(samples,1,mean)  
vars <- apply(samples, 1, var)
```

The variation between the theoretical mean and the mean obtained from the simulation show a small variation which can be narrowed further on increasing either the sample size or the number of simulations.

```
mu0 <- mean(means)  
print(c(mean,mu0))
```

```
## [1] 5.000000 5.768084
```

The variance change from the theoretical value and the simulated value are also converging to the expected value.

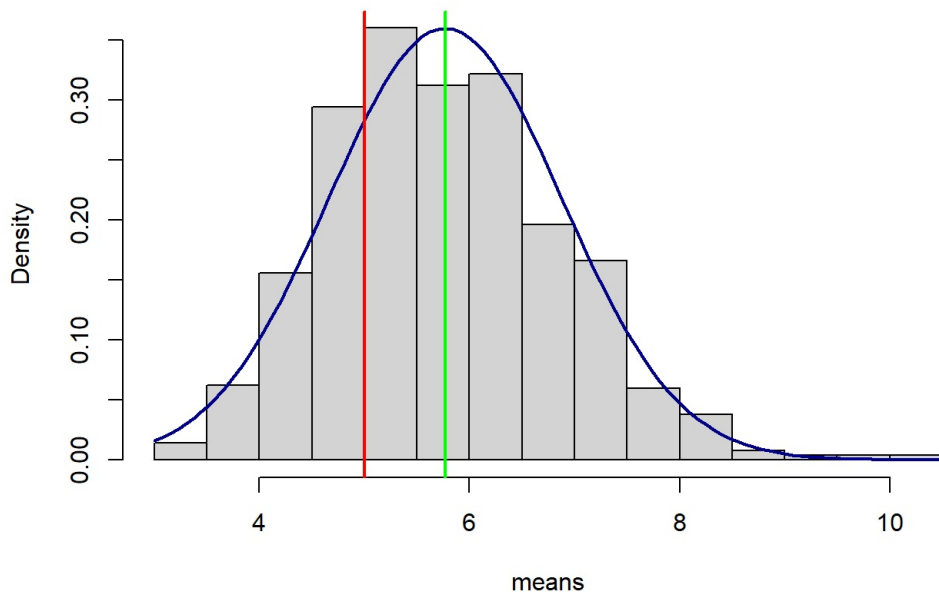
```
print(c(sd^2,mean(vars)))
```

```
## [1] 25.00000 46.42699
```

Distribution of means

```
hist(means, probability = TRUE)  
curve(dnorm(x, mean=mean(means), sd=sd(means)), col="darkblue", lwd=2, add=TRUE, yaxt="n")  
abline(v = mean, col = "red", lwd = 2)  
abline(v = mean(means),col = "green", lwd = 2)
```

Histogram of means

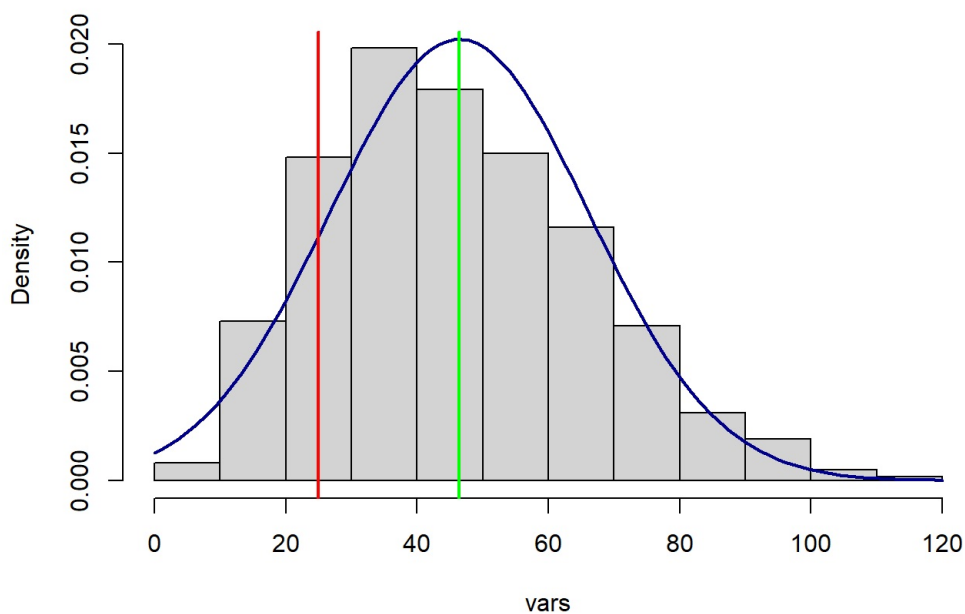


The above plot is overlaid with a curve that is approximately normal or is almost normal. The means of the two distributions vary by a certain quantity. The green line here shows the theoretical mean whereas the red line is the observed mean.

Distribution of variance

```
hist(vars, probability = TRUE)
curve(dnorm(x, mean=mean(vars), sd=sd(vars)), col="darkblue", lwd=2, add=TRUE, yaxt="n")
abline(v = sd^2, col = "red", lwd = 2)
abline(v = mean(vars), col = "green", lwd = 2)
```

Histogram of vars



The green line here shows the theoretical variance and the red line shows the observed variance. Moreover the distributions obtained resembles a normal distribution as shown by the overlaid curve.