Statistical Inference Course Project - PART 1: Exponential Distribution & Central Limit Theorem

Overview

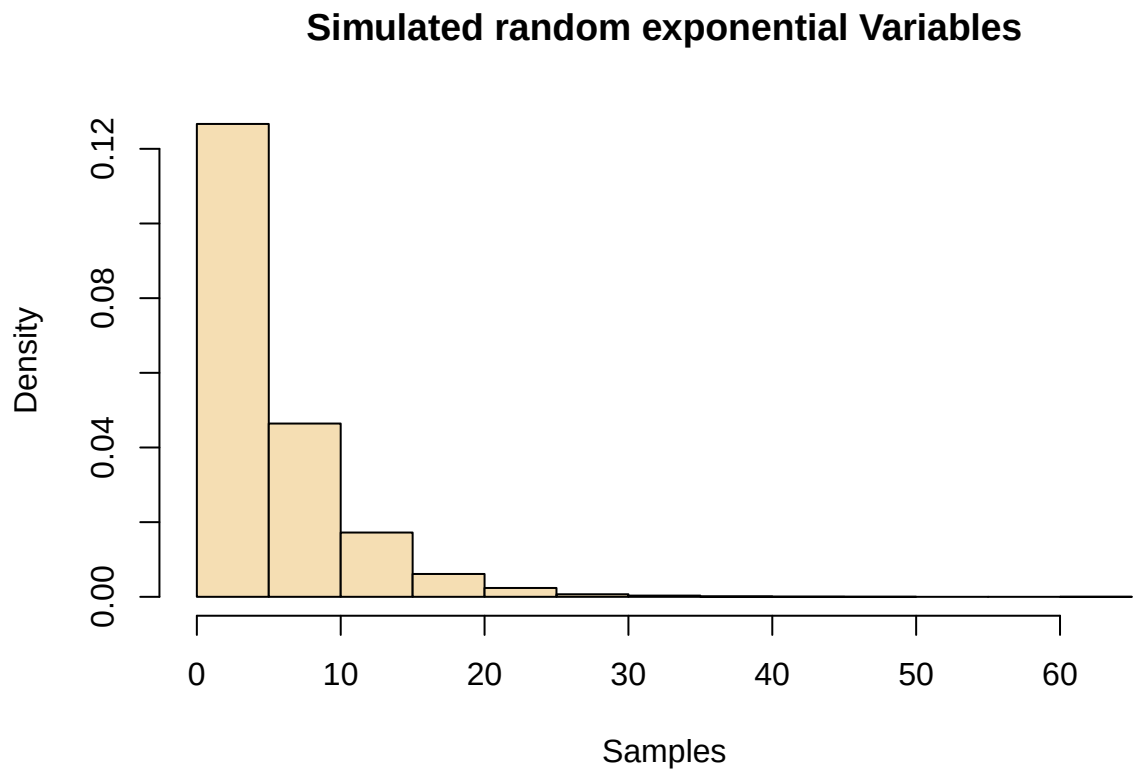
In this analysis a set of exponentially distributed variables and a set of averages of exponentially distributed variables is compared in order to demonstrate the Central Limit Theorem (CLL).

Simulation

Simulation Parameters:

```
set.seed(576)
n <- 40 # number of exponential variables for averaging
N <- 1000 # number of simulations / averages
lambda <- .2 ## lambda is kept constant for the whole simulation study
```

```
cexp <- rexp(n*N,lambda)
hist(cexp, prob = TRUE, col = "wheat",
     main ="Simulated random exponential Variables", xlab ="Samples", ylab="Density")
```




```
mean_exp = NULL
for (i in 1 : N) mean_exp <- c(mean_exp, mean(rexp(n, lambda)))
```

Comparison of simulated sample mean and sample variance with theoretical values

Calculate sample mean, sample variance and theoretical variance. Note theoretical mean = 1/ lambda = 1/0.2 = 5

```
smean    <- mean(mean_exp) # sample mean
svar     <-  var(mean_exp)
var_theo <- (1/lambda)^2/n
```

Sample mean (distribution is centered at):

```
smean
```

```
## [1] 4.996073
```

Sample variance:

```
svar
```

```
## [1] 0.6159614
```

Theoretical variance:

```
var_theo
```

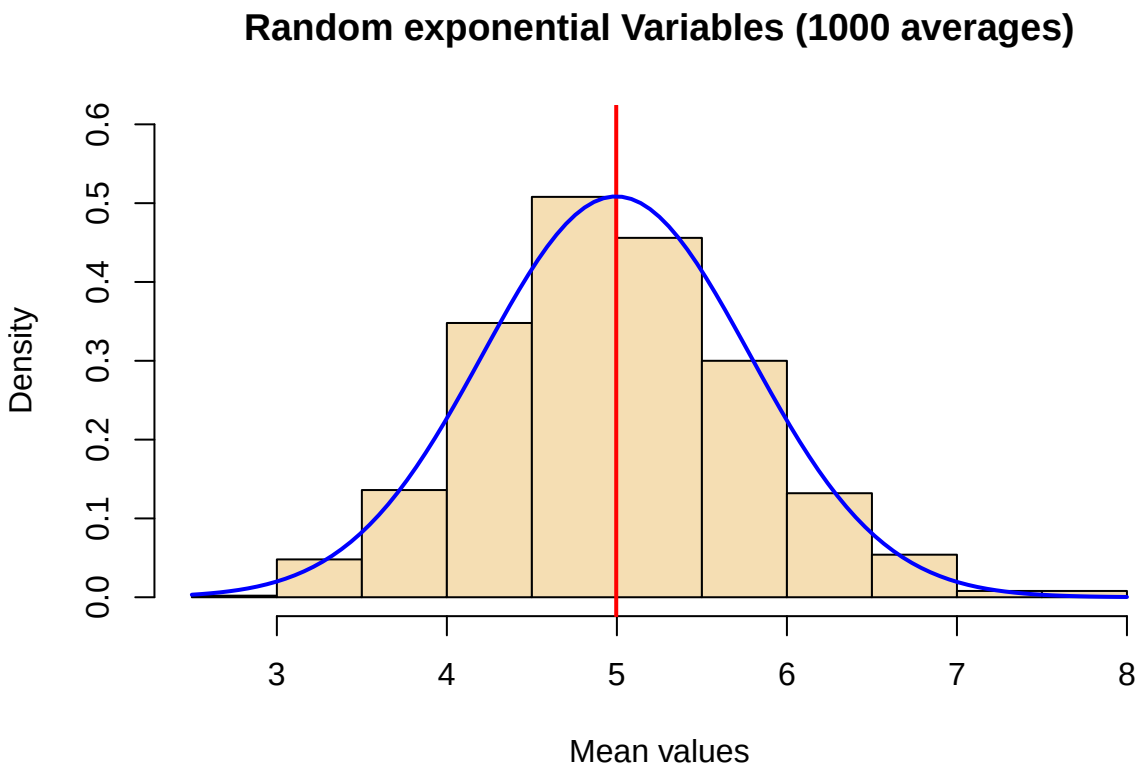
```
## [1] 0.625
```

Comparison sample distribution with theoretical distribution

Fit normal distribution to the simulated sample distribution

```
library(MASS)
fit <- fitdistr(mean_exp, "normal")
mean_fit <- fit$estimate[1]
sd_fit   <- fit$estimate[2]
```

```
hist(mean_exp, col = "wheat", prob = TRUE,
     main = "Random exponential Variables (1000 averages)", xlab = "Mean values",
     ylab="Density", ylim = c(0,0.6) )
abline(v = smean, col = "red", lwd = 2) # marked mean value by red line
x <- mean_exp
curve(dnorm(x, mean_fit, sd_fit), col = "blue", lwd = 2, add = TRUE) # estimated standard error
```



Conclusion

As expected for a sample size of 1000 the sample mean and the sample variance are very close to the theoretical mean of **5** and the theoretical variance of **0.625**. Furthermore obeying the Central Limt Theorem the sample distribution of the averaged set of exponential random variables can be described by a normal distribution.