Statistical inference w4 report

Part 1 Simulation exercice

Stephane Marceau

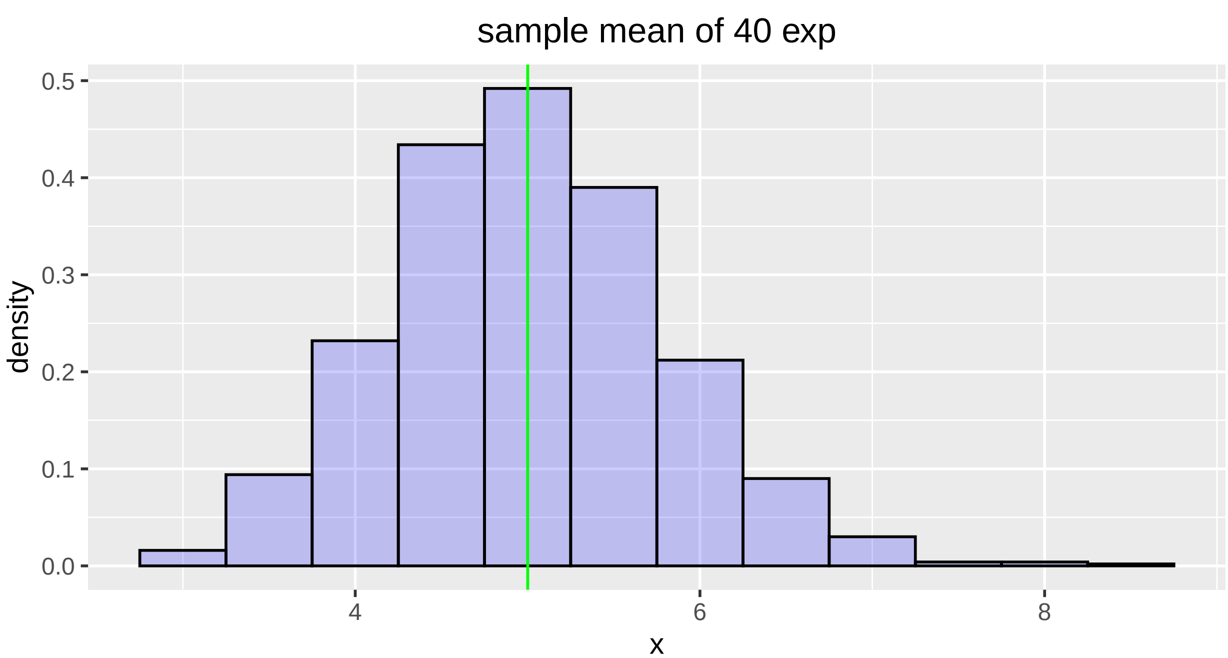
Summary

During this study, we will investigate the exponential distribution in R and compare it with the Central Limit Theorem.
We will simulate its distribution by a getting 1000 means of 40 random exponential values. We will then, compare the sample mean and sample mean sd with the Normal distribution $\mathcal{N}(ţ, \sigma/\sqrt(n))$ to verify the CLT. Finally, we will carry out a quantile to quantile check in order to visualize their matching.

Simulation

We review an exponential mean random variable by simulating a thousand times the average of 40 exponentials distribution with lambda = 0.2.

```
library(ggplot2)
lambda<-0.2
mns <- sapply(1:1000,function(i) mean(rexp(40,lambda)))
sample_mean <- mean(mns)
sample_sd <- sd(mns)
mns_dat <- data.frame(x=mns, conf=mns>quantile(mns,probs = .025) & mns<quantile(mns,probs = .975))
g<-ggplot(data=mns_dat,aes(x=x))+geom_histogram(alpha = .2, binwidth=.5, color = "black",fill="blue", a
g+ggtitle("sample mean of 40 exp")+theme(plot.title = element_text(hjust = 0.5))+ geom_vline(xintercept
```



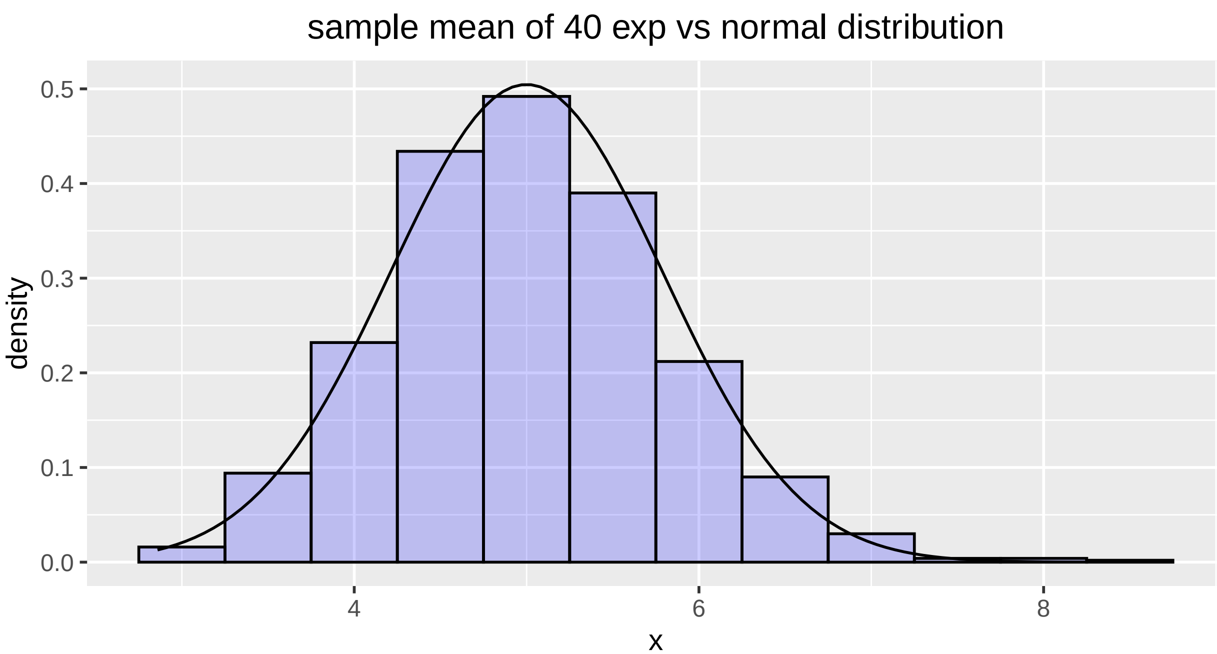
From this simulation we get

- a mean of 5.0006416 which is very close to the theoretical distribution mean of ţ $= 1/\lambda = 5$
- a standard deviation of 0.8054984 similar to $sd = \sigma/\sqrt(n) = 0.7905694$

plotting the Normal distribution

As per the central limit theorem, $\bar{x} = \mathcal{N}($ţ$, \sigma/\sqrt(n)$. Then, we plot the normal distribution $\mathcal{N}(5, 5/\sqrt(40)$ on top of our simulation histogram.

```
g<-g+stat_function(fun = dnorm,args=list(mean=5,sd=5/sqrt(40)))
g+ggtitle("sample mean of 40 exp vs normal distribution")+theme(plot.title = element_text(hjust = 0.5))
```

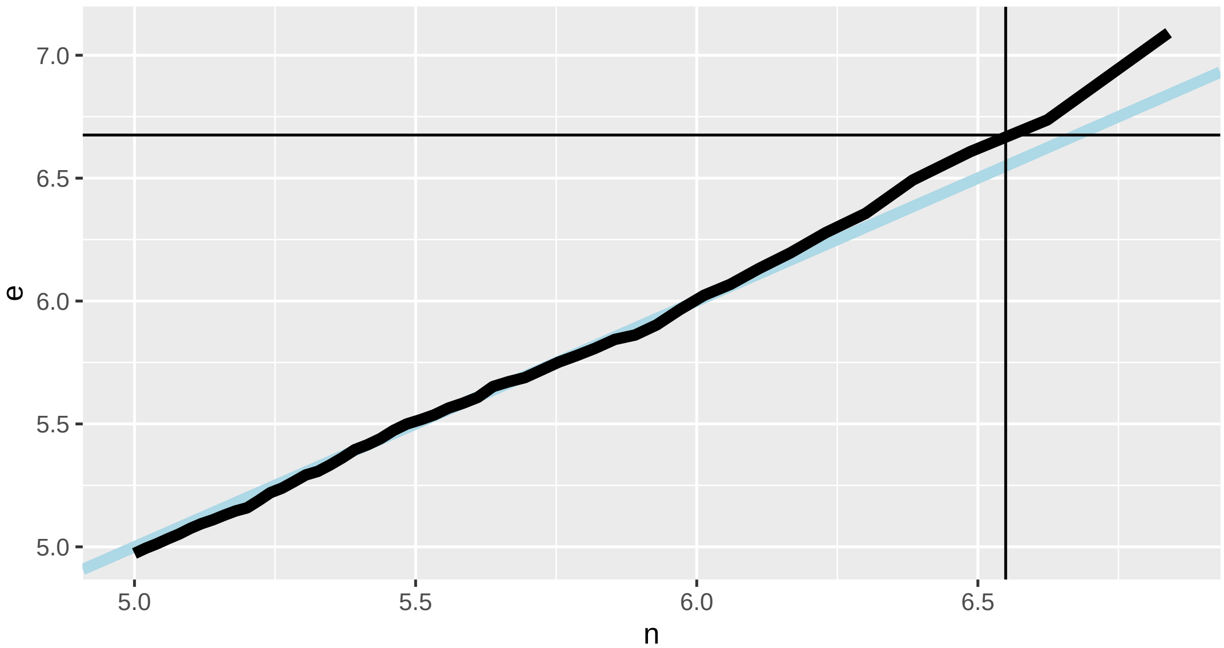


We see that both simulations and the normal distribution look very similar.

plotting quantile to quantile chart

We use a qqplot chart to assess our simulation output with the normal distribution

```
pvals <- seq(.5, .99, by = .01)
d <- data.frame(n= qnorm(pvals,mean = 5,sd = 5/sqrt(40)),e=quantile(x=mns,probs = pvals), p = pvals)
g <- ggplot(d, aes(x= n, y = e))
g <- g + geom_abline(size = 2, col = "lightblue")
g <- g + geom_line(size = 2, col = "black")
g <- g + geom_vline(xintercept = qnorm(0.975,mean = 5,sd = 5/sqrt(40)))
g <- g + geom_hline(yintercept = quantile(x=mns,probs = 0.975))
g
```



Quantile to quantile plot shows that the sample means and the normal distribution $\mathcal{N}(ţ, \sigma/\sqrt(n)$ are very close. We can conclude than the sample mean of 40 exp is approximately normal with $\mathcal{N}(5, 5/\sqrt(40)$.