Simulation Statistical inference

Overview In the first part I will investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLT). I Should use the knowledge learnt in class to relate the normal and gaussian distributions, using averages of random numbers of thosse distributions. In the second part I will perform a basic inferential data analysis.

Part 1: Simulation Exercise

For this simulation exercise, we first start loading libraries and initializing variables, setting lambda to 0.2 and n to 40. Then we compute 1000 random exponentials and save it as a vector into dexp variable. Also, we run simulations of the means of 40 exponentials, 1000 times.

```
library(knitr)
lambda <- 0.2
n <- 40
total <- data.frame(estimator=character(), theoretical=numeric(), sample=numeric(), sample_means=numeri
# 1000 exponentials
dexp <- rexp(1000, lambda)
# 1000 simulations of the mean of 40 exponentials
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(n, lambda)))
```

With our values computed and stored, we can proceed to compute the mean.

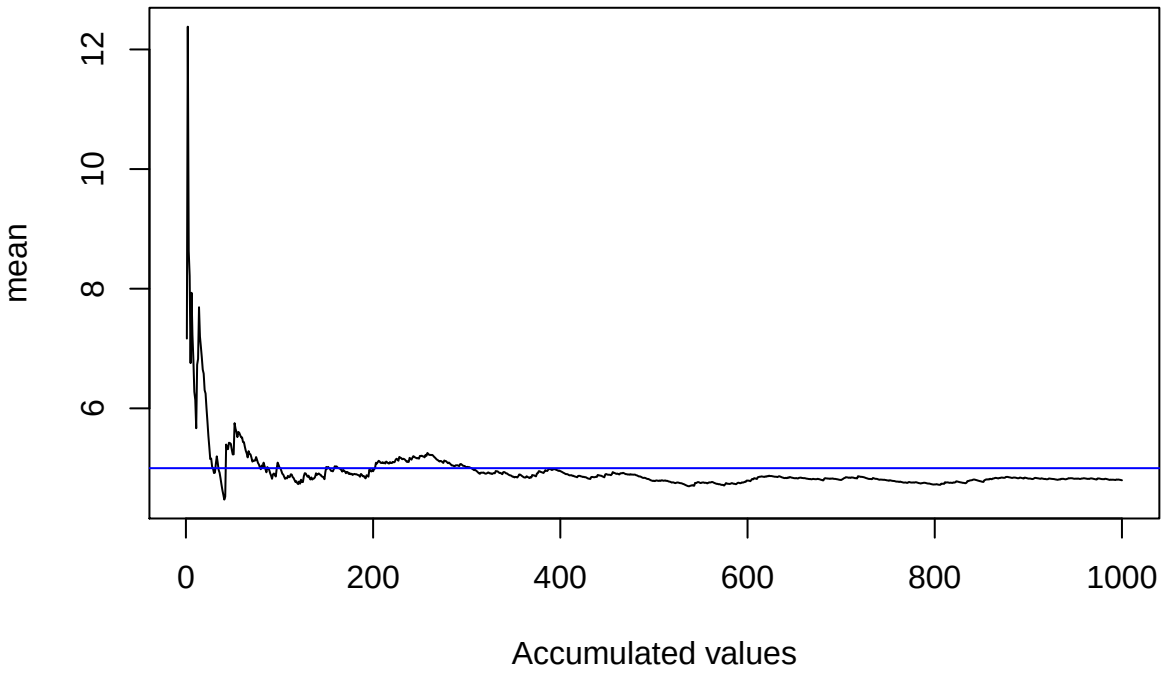
```
# Compute mean
theorethical_mean <- 1/lambda
sample_mean <- mean(dexp)
sample_means_1000mean <- mean(mns)
new_row <- data.frame("Mean", theorethical_mean, sample_mean, sample_means_1000mean)
names(new_row) <- c("estimator", "theoretical", "sample", "sample_means" )
total <- rbind(total, new_row)
kable(total, caption="Theoretical vs sample mean")
```

Table 1: Theoretical vs sample mean

estimator	theoretical	sample	sample_means
Mean	5	4.84481	4.982926

We can see that the sample mean is almost near the theoretical mean. But as the number of values increases, the mean gets closer to the asympthotic representing the theorethical mean. The sample means of the means of 1000 simulations, is even closer.

```
cum <- cumsum(rexp(1000, lambda))/(1:1000)
plot(cum, type="l", xlab = "Accumulated values", ylab = "mean")
abline(h=theorethical_mean, col="blue")
```



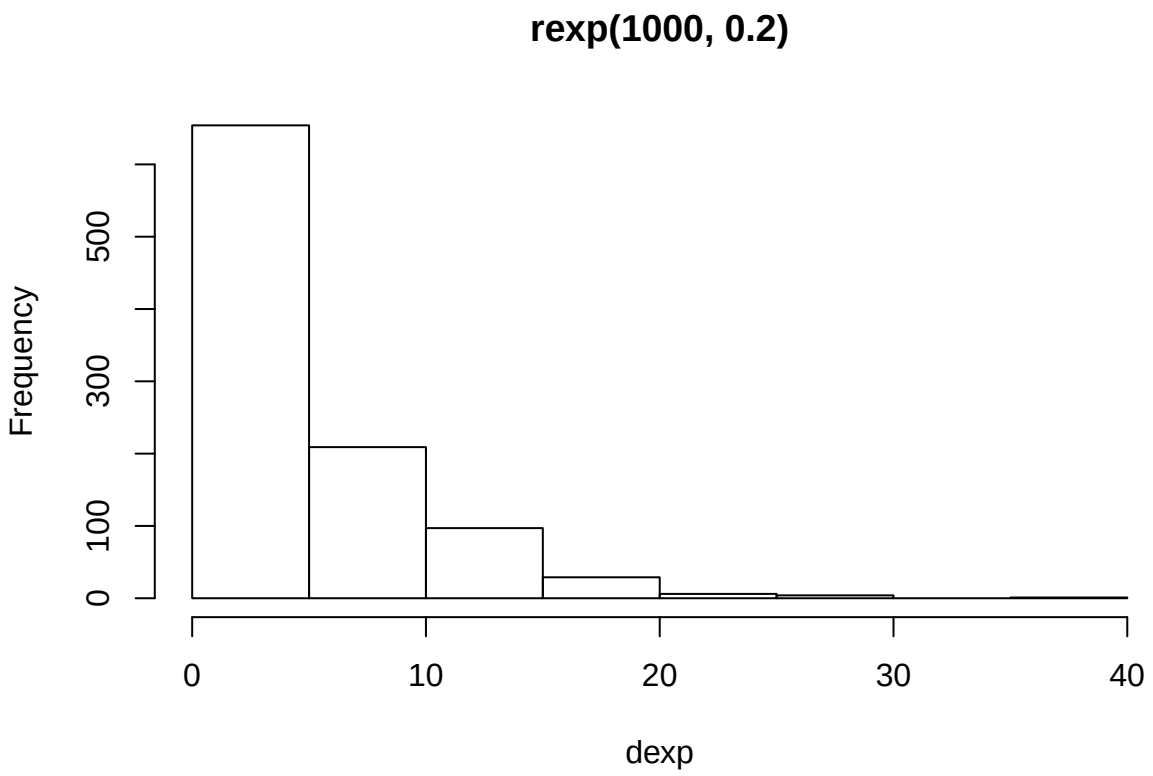

```
theorethical_sd <- 1/lambda
sample_sd <- sd(dexp)
sample_means_1000sd <- sd(mns)
new_row <- data.frame("Standard deviation", theorethical_sd, sample_sd, sample_means_1000sd)
names(new_row) <- c("estimator", "theoretical", "sample", "sample_means")
total <- rbind(total, new_row)
theorethical_var <- theorethical_sd^2
sample_var <- var(dexp)
sample_means_1000var <- var(mns)
new_row <- data.frame("Variance", theorethical_var, sample_var, sample_means_1000var)
names(new_row) <- c("estimator", "theoretical", "sample", "sample_means")
total <- rbind(total, new_row)
kable(total, caption="Theoretical vs sample")
```

Table 2: Theoretical vs sample

estimator	theoretical	sample	sample_means
Mean	5	4.844810	4.9829261
Standard deviation	5	4.780426	0.8086827
Variance	25	22.852478	0.6539677

Then calculating the standard deviation and variance, shows the theoretical and sample mean are very close, given the high number of values accumulated. While the variance of the means of 1000 simulations is different, due to it being converted into a standard normal, as the two following figures shows. This shows what the CLT states that the distribution of iid variables becomes that of a standard normal as the sample size increase.

```
hist(dexp,  main = 'rexp(1000, 0.2)')
```




```
h <- hist(mns, main = '1000 simulations of the means of 40 random exponential')
xfit <- seq(min(mns), max(mns), length = 40)
yfit <- dnorm(xfit, mean = mean(mns), sd = sd(mns))
yfit <- yfit * diff(h$mids[1:2]) * length(mns)
lines(xfit, yfit, col = "blue", lwd = 2)
```