Statistical Inference Course Project - Part 1

Jón Guttesen
15 mai 2017

This project consists of two parts:

1. a simulation exercise.
2. a basic inferential data analysis. ## Part 1 - The simulation ### Overview This part investigates the exponential distribution and compare it with the Central Limit Theorem. A exponential distribution will be simulated with rexp(n, lambda) where lambda, the rate parameter, will be set to lambda = 0.2. One thousand simualations will take the mean of 40 exponentials. ### The simulations

```
library(ggplot2)
lambda <- 0.2
n <- 40
dexp <- NULL
nosim <- 1000
set.seed(1)
for( i in 1:nosim) dexp <- c(dexp, mean(rexp(n,lambda)))
rm(i)
```

1. The sample mean and the theoretical mean of the distribution

```
#sample mean
mean_s <- mean(dexp)
# theoretical mean
mean_t <- 1/lambda
sprintf( "The theoretical mean is %f", mean_t)

## [1] "The theoretical mean is 5.000000"

sprintf( "The sample mean is %f", mean_s)

## [1] "The sample mean is 4.990025"
```

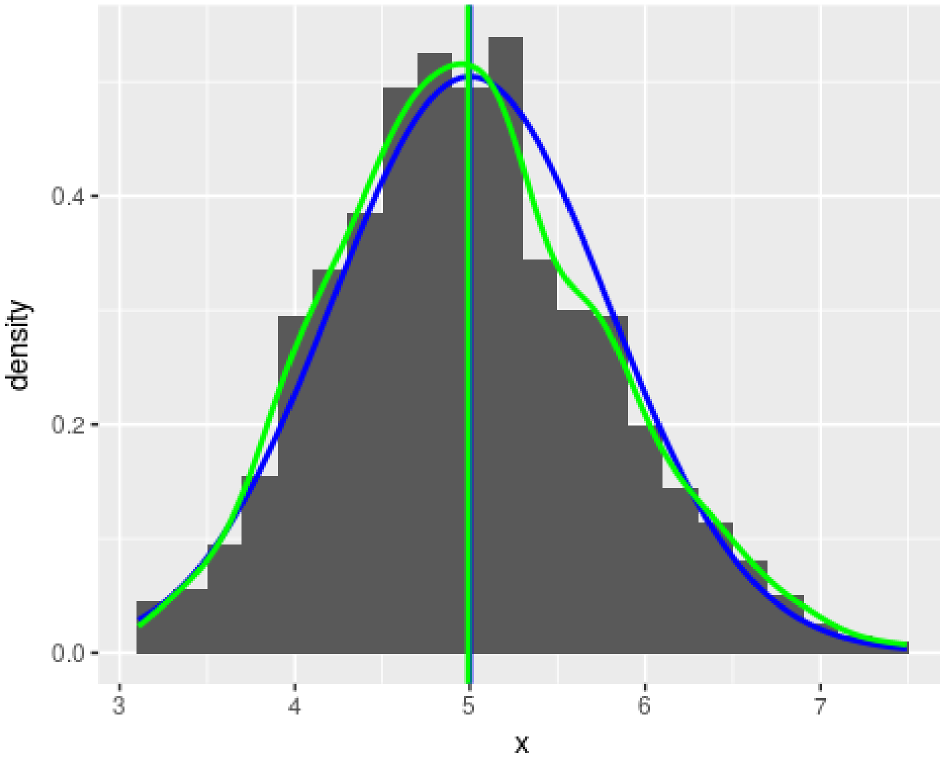
The sample mean aproximates the theoretical mean. #### 2. The sample variance and the theoretical variance of the distribution

```
# sample sd
sd_s <- sd(dexp)
# theoretical sd
sd_t <- (1/lambda)/sqrt(n)
#sample variance
```

```
var_s <- var(dexp)
# theoretical variance
var_t <- ((1/lambda)/sqrt(n))^2
sprintf("The sample variance is %f", var_s)

## [1] "The sample variance is 0.611116"

sprintf("The theoretical variance is %f", var_t)

## [1] "The theoretical variance is 0.625000"
```

The sample variance aproximates the theoretical variance. To summarize the following plot shows the sample mean and variance versus the theoretical mean and variance. The theoretical values ar plotted in blue while the sample values are plotted in green.

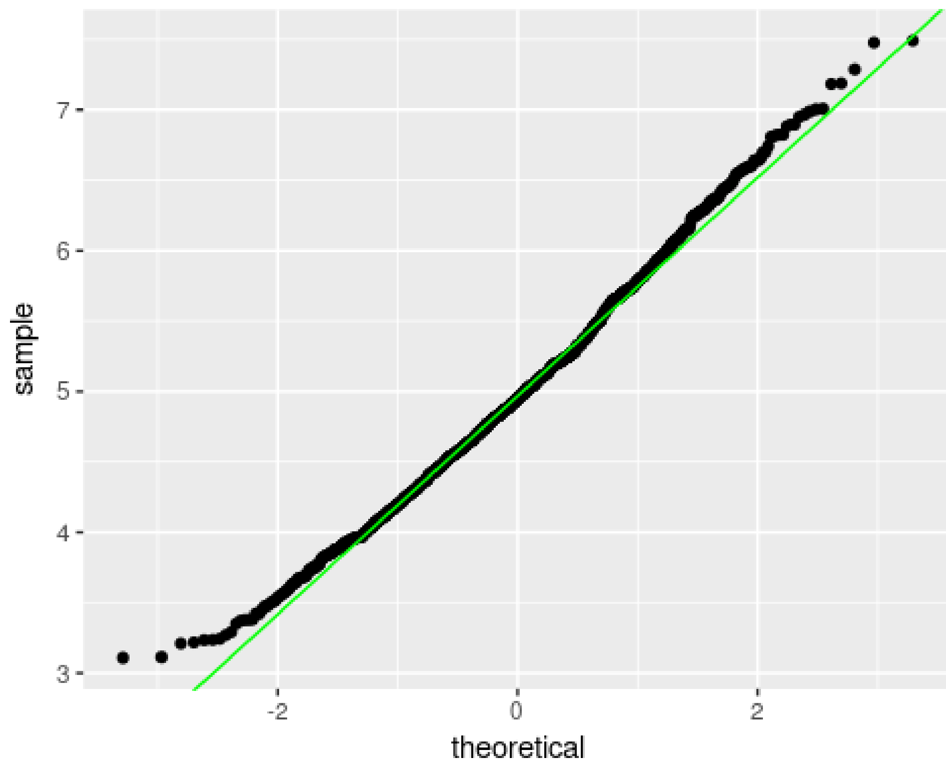
```
g <- ggplot(data = data.frame(x = dexp), aes(x = x)) +
geom_histogram(aes(y = ..density..), binwidth = lambda )
g <- g + stat_function(fun=dnorm, args=list(mean=mean_t,
sd=sd_t), color="blue", size=1)
g <- g + geom_vline(xintercept=mean_t, color="blue", size=1)
g <- g + stat_density(geom = "line", color = "green", size=1)
g <- g + geom_vline(xintercept=mean_s, color="green", size=1)
g <- g + scale_colour_manual(name="Line Color",
values=c(myline1="blue", myline2="green"))
g
```



###

3.The distribution is approximately normal The following plot show that the distribution is approximately normal.

```
tmpy <- quantile(dexp[!is.na(dexp)], c(0.25, 0.75))
tmpx <- qnorm(c(0.25, 0.75))
tmpslope <- diff(tmpy)/diff(tmpx)
tmpint <- tmpy[1L] - tmpslope * tmpx[1L]
tmpdst <- data.frame(resids = dexp)
g <- ggplot(tmpdst, aes(sample = resids)) + stat_qq() +
geom_abline(slope = tmpslope, intercept = tmpint, col="green")
g
```



sample
3
4
5
6
7
-2
0
2
theoretical