Rapport_TP1

BON, BARON, FOUSSE

9/26/2020

#Question 1

On constate que la loi $p(y, \theta)$ se factorise en deux lois normales, dont une ne dépendant que de θ :

$p(y, \theta) \propto exp(-\frac{1}{4}[(y - \frac{\theta}{4})^2) + \frac{\theta^2}{16}) \propto exp(-\frac{1}{4}(y - \frac{\theta}{4})^2) \times exp(-\frac{\theta^2}{64})$

On reconnaît les lois $p(y|\theta) = \mathcal{N}(\frac{\theta}{4}; 2)$ et $p(\theta) = \mathcal{N}(0; 32)$

On identifie alors les constantes de normalisations grâce à la formule de la loi normale, qui valent respectivement $c_1 = \frac{1}{2\sqrt{\pi}}$ et $c_2 = \frac{1}{8\sqrt{\pi}}$

#Question 2

On peut également trouver une autre factorisation de la loi $p(y, \theta)$ en deux lois normales, dont une ne dépendant que de y :

$p(y, \theta) \propto exp(-\frac{1}{32}(\theta - 2y)^2 + 4y^2) \propto exp(-\frac{(\theta-2y)^2}{32}) \times exp(-\frac{y^2}{8})$

On reconnaît les lois $p(\theta|y) = \mathcal{N}(2y, 16)$ et $p(y) = \mathcal{N}(0, 4)$.

On identifie les constantes de normalisations, sui valent respectivement $c_3 = \frac{1}{4\sqrt{2\pi}}$ et $c_4 = \frac{1}{2\sqrt{2\pi}}$.

#Question 3

Le programme suivant permet de simuler des tirages de la dimension voulue de la loi de densité $p(y, \theta)$:

```r
taille <- 1000

#On tire les thetas
theta <- rnorm(taille, 0, 32)

#on tire les y sachant theta

y <- rnorm(taille, theta/4, 2)
```

##Question 4

On a vu que la loi conditionelle $p(\theta|y)$ est une loi normale d'espérance $2y$.

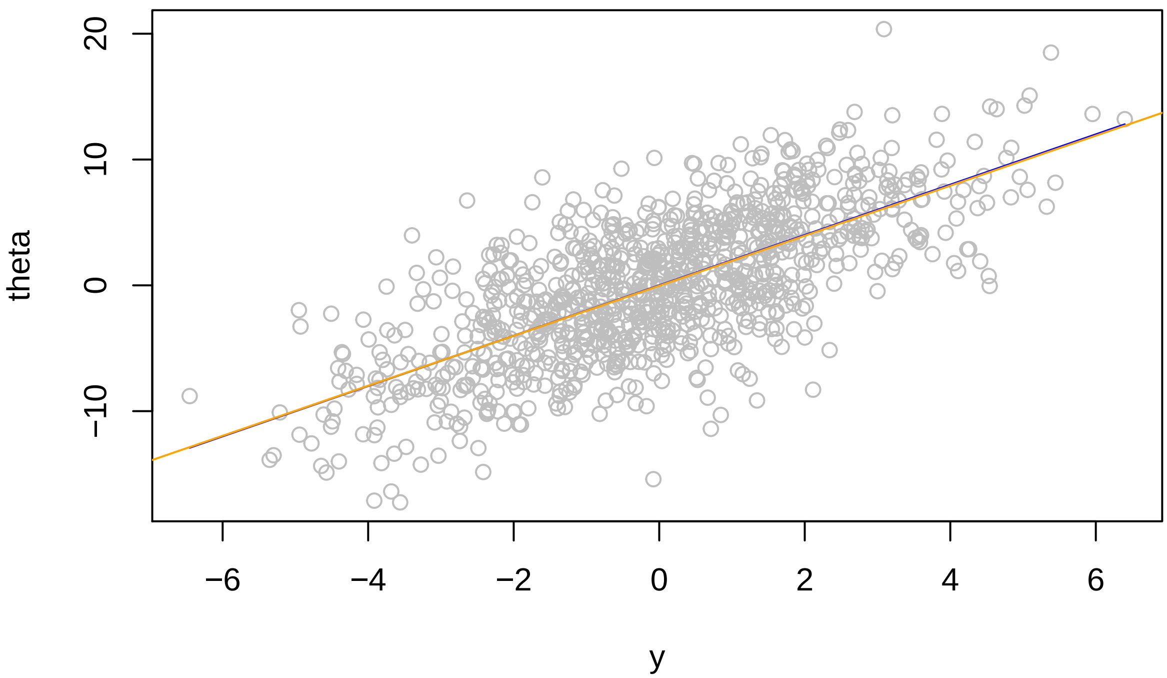
Donc $\mathbb{E}(\theta|y) = 2y$.

##Question 5

On effectue 1000 tirages et on fait apparaître en bleu l'espérance :

```r
taille <- 1000

#On tire les thetas
theta <- rnorm(taille, 0, sqrt(32))
```

```r
#on tire les y sachant theta

y <- rnorm(taille, theta/4, sqrt(2))

#on trace les points
plot(y, theta, col="grey")

#on trace la droite de l'espérance

curve(2*x, add=TRUE, col="blue")

#on trace la droite de la régression linéaire

abline(lm(theta ~ y), col="orange")
```



##Question 6

On trace en orange la droite de la régression linéaire et on constate qu'elle correspond à la droite de l'espérance conditionelle.(voir graphique précédent)

##Question 7

On réutilise le code précédent avec 100000 tirages, puis on élimine les valeurs qui ne correspondent pas à l'intervalle.

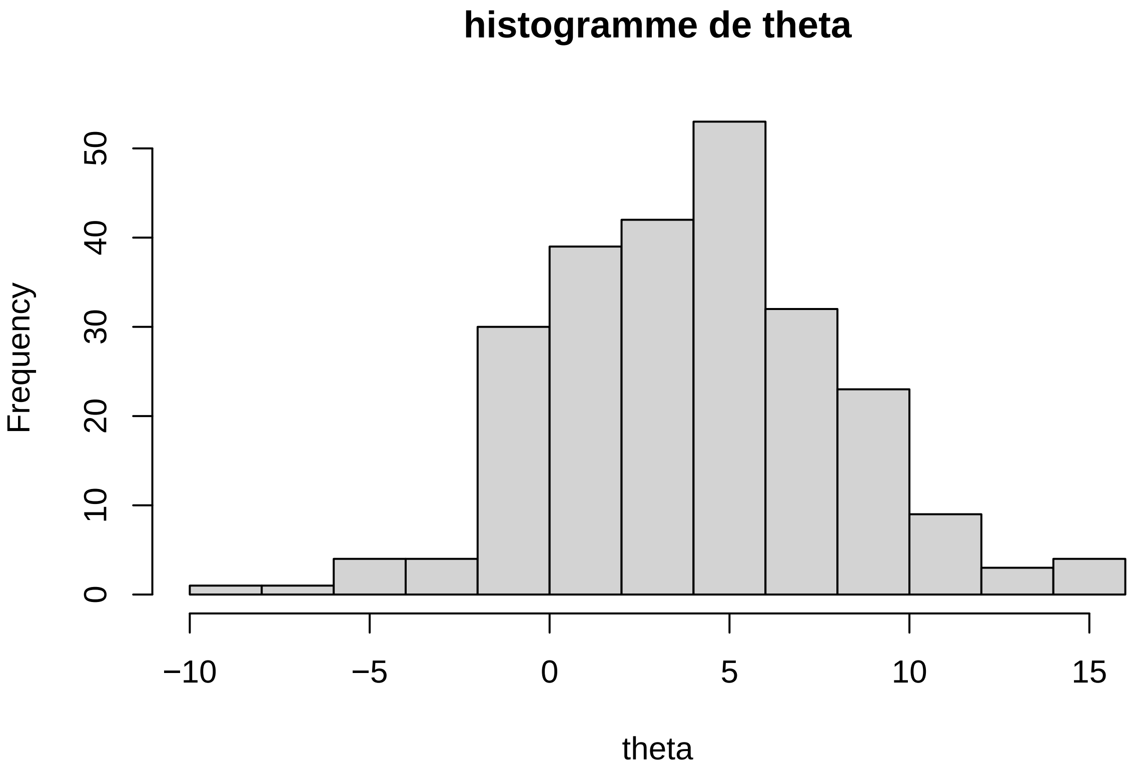
```r
taille <- 100000

#On tire les thetas
theta <- rnorm(taille, 0, sqrt(32))

#on tire les y sachant theta

y <- rnorm(taille, theta/4, sqrt(2))
ytheta <- data.frame(theta, y)
```

```r
ytheta = ytheta[ytheta[,2] < 2.01 & ytheta[,2] > 1.99,]
length(ytheta)
```

```
## [1] 2
```

#Question 8 Voici l'histogramme pour theta avec y dans l'intervalle [1.99, 2.01].

```r
hist(ytheta[,1], main = 'histogramme de theta', xlab = 'theta')
```

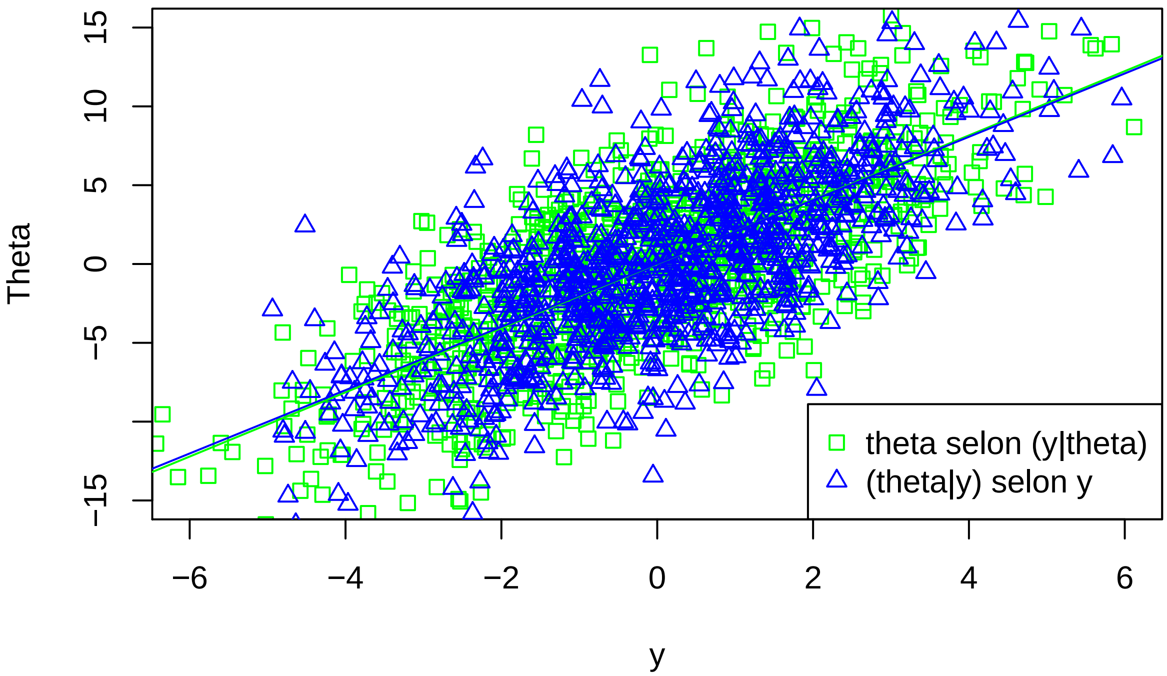


On constate que theta suit une loi normale de valeur moyenne E[2*y] avec y=2, c'est le résultat auquel on s'attendait pour y suffisamment proche du centre du nuage de point de la question Q5.

#Question 9 On représente la loi $p(y, \theta)$ en utilisant la loi conditionnelle $p(y|\theta)$ (carrés vert sur le graphe), puis en utilisant la loi conditionnelle $p(\theta|y)$ (triangles bleus sur le graphe). On les superpose et on obtient le graphe ci-dessous:

```r
taille <- 1000

#On tire les thetas
theta <- rnorm(taille, 0, sqrt(32))

#on tire les y sachant theta

y_theta <- rnorm(taille, theta/4, sqrt(2))

#On tire les y²
y <- rnorm(taille, 0, sqrt(4))

#on tire les thetas sachant y

theta_y <- rnorm(taille, 2*y, sqrt(16))
```

```r
#on trace les deux nuages de points pour les comparer
plot(y_theta,theta,xlab="y",ylab="Theta",col="green",pch=0,xlim=c(-6,6),ylim=c(-15,15),cex=1)

points(y,theta_y,col="blue",pch=2,cex=1)
legend(x="bottomright", legend=c("theta selon (y|theta)","(theta|y) selon y"), col=c("green","blue"), p

#on trace les droites de régressions linéaires

abline(lm(theta ~ y_theta), col="green")
abline(lm(theta_y ~ y), col="blue")
```



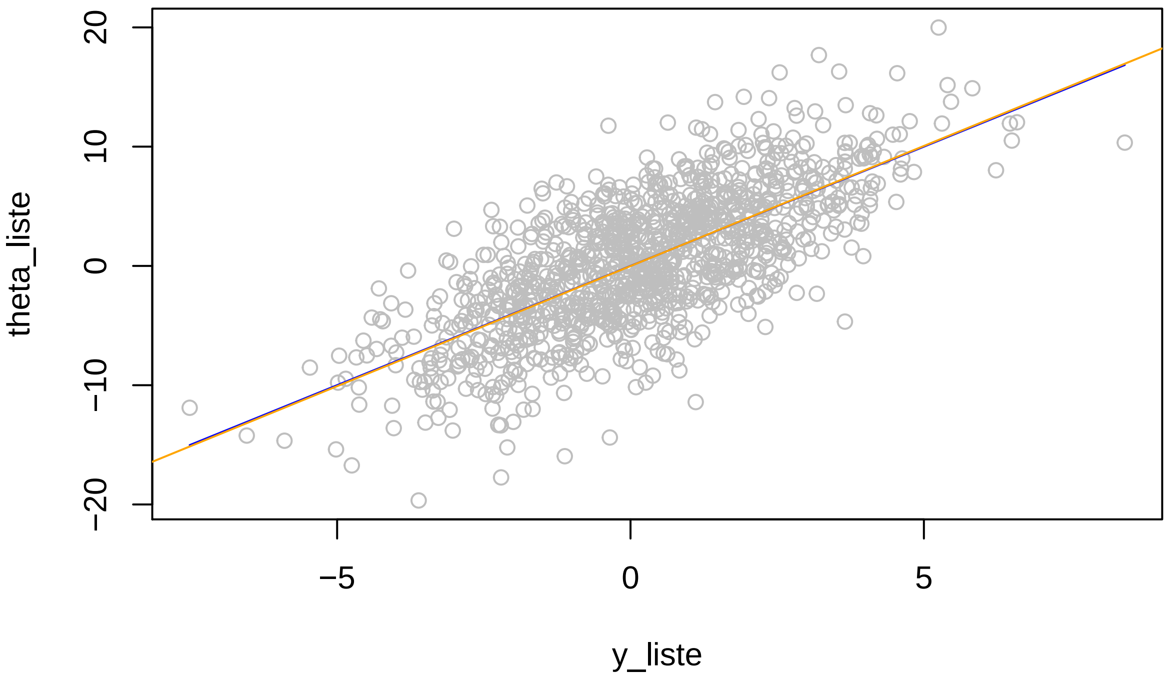
On constate que les nuages de points se superposent, la loi $p(y, \theta)$ est donc bien caractérisée par ses deux lois conditionelles.

#Question 10

On fait un échantillonage de Gibbs avec 1000 valeurs et on obtient un résultat similaire :

```r
theta_t = 0
t = 1100
y_liste = c()
theta_liste = c()

while (t > 0) {
  y_t = rnorm(1, theta_t/4, sqrt(2))
  theta_t = rnorm(1, 2*y_t, sqrt(16))
  y_liste = c(y_liste, y_t)
  theta_liste = c(theta_liste, theta_t)
  t = t-1
  }

ytheta= data.frame(y_liste, theta_liste)
ytheta = tail(ytheta, 1000)
```

```r
#on trace les points
plot(y_liste, theta_liste, col="grey")

#on trace la droite de l'espérance

curve(2*x, add=TRUE, col="blue")

#on trace la droite de la régression linéaire

abline(lm(theta_liste ~ y_liste), col="orange")
```



En faisant varier le nombre de valeurs initiales qu'on élimine, on s'aperçoit que ce paramètre n'a pas d'influence sur la qualité de l'échantillon, à moins que celui-ci soit vraiment petit.