MAP433 - Projet 4

EL KHADIR Bachir, BIAD Anas

April 28, 2014

Le code utilisé pour trouver les valeurs numériques et sa sortie sont disponibles en annexe.

Question 1) - a

Pour $mois \in \{mai, juin\}$, on a:

$$\nabla_\vartheta M^{(mois)}(\hat{\vartheta}^{(mois)}) = \nabla_\vartheta M^{(mois)}(\vartheta^{(mois)}) + H_e.(\hat{\vartheta}^{(mois)} - \vartheta^{(mois)}) + \mathrm{O}\left(\left\|\hat{\vartheta}^{(mois)} - \vartheta^{(mois)}\right\|^2\right)$$

où H_e est la hessienne de $M^{(mois)}$ au point $\vartheta^{(mois)}$..

$M^{(mois)}$ atteint son minimum en $\hat{\vartheta}^{(mois)}$, donc: $\nabla_\vartheta M^{(mois)}(\hat{\vartheta}^{(mois)}) = 0$. Par conséquent:

$$\sqrt{n}(\hat{\vartheta}^{(mois)} - \vartheta^{(mois)}) = -\sqrt{n}\, H_e^{-1}\nabla_\vartheta M^{(mois)}(\vartheta^{(mois)}) + \mathrm{O}\left(\sqrt{n}\,\left\|H_e^{-1}(\hat{\vartheta}^{(mois)} - \vartheta^{(mois)})\right\|^2\right)$$

Or

$$\begin{aligned}(\forall \vartheta \in \mathbb{R})\, \nabla_\vartheta M^{(mois)}(\vartheta) &= \frac{-2}{n}\sum_{i=1}^{k}\sum_{j=1}^{r}(Y_{i,j}^{(mois)} - f(x_i,\vartheta))\nabla_\vartheta f(x_i,\vartheta)\\ &= \frac{-2}{n}\sum_{i=1}^{k}\sum_{j=1}^{r}\varepsilon_{i,j}\,\nabla_\vartheta f(x_i,\vartheta)\end{aligned}$$

$$\mathbb{E}[\sqrt{n}\nabla_\vartheta M^{(mois)}(\vartheta^{(mois)})] = 0$$

car les $\varepsilon_{i,j}$ sont centrées.

$$\begin{aligned}\mathbb{V}[\sqrt{n}\nabla_\vartheta M^{(mois)}(\vartheta^{(mois)}] &= 4\sum_{i=1}^{k}\sum_{j=1}^{r}\mathbb{E}[\varepsilon_{i,j}^2]\,\nabla_\vartheta f(x_i,\vartheta^{(mois)})\nabla_\vartheta f(x_i,\vartheta^{(mois)})^T \qquad \text{(car les } \varepsilon_{i,j} \text{ sont iid)}\\ &= \frac{4\sigma^2}{n}\sum_{i=1}^{k}\sum_{j=1}^{r}\nabla_\vartheta f(x_i,\vartheta^{(mois)})\nabla_\vartheta f(x_i,\vartheta^{(mois)})^T\\ &= \frac{4\sigma^2}{k}\sum_{i=1}^{k}\nabla_\vartheta f(x_i,\vartheta^{(mois)})\nabla_\vartheta f(x_i,\vartheta^{(mois)})^T\\ &\to 4\sigma^2\int_0^1\nabla_\vartheta f(x,\vartheta^{(mois)})\nabla_\vartheta f(x,\vartheta^{(mois)})^T\,\mathrm{d}x\\ &= 4\sigma^2 H(\vartheta^{(mois)})\end{aligned}$$

Par TCL :

$$\sqrt{n}\nabla_\vartheta M^{(mois)}(\vartheta^{(mois)} \xrightarrow{d} \mathcal{N}(0, 4\sigma^2 H(\vartheta))$$

On a aussi:

$$\begin{aligned}H_e &= \frac{-2}{n}\sum_{i,j} -\nabla_\vartheta f(x_i,\vartheta^{(mois)})\nabla_\vartheta f(x_i,\vartheta^{(mois)})^T + \varepsilon_{i,j}H(f)\\ &= \frac{2}{n}\sum_{i,j} -\nabla_\vartheta f\nabla_\vartheta f^T + \frac{-2}{n}\sum_{i,j}\varepsilon_{i,j}H(f)\end{aligned}$$

$\frac{-2}{n}\sum_{i,j}\varepsilon_{i,j}H(f) \longrightarrow 0$ par LGN, et $\frac{2}{n}\sum_{i,j} -\nabla_\vartheta f \nabla_\vartheta f^T \longrightarrow H\left(\vartheta^{(mois)}\right)$
d'où

$$H_e \xrightarrow{\mathbb{P}_\vartheta} 2H\left(\vartheta^{(mois)}\right)$$

On en déduit par Slutsky que:

$$\sqrt{n}(\hat{\vartheta}^{(mois)} - \vartheta^{(mois)}) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 H\left(\vartheta^{(mois)}\right)^{-1}\right)$$

Comme $\hat{\vartheta}^{(mai)}$ et $\hat{\vartheta}^{(juin)}$ sont indépendants, alors:

$$\sqrt{n}(\hat{\vartheta} - \vartheta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 H\left(\vartheta\right)^{-1}\right)$$

où

$$H\left(\vartheta\right) = \begin{pmatrix} H\left(\vartheta^{(mai)}\right) & 0 \\ 0 & H\left(\vartheta^{(juin)}\right)\end{pmatrix}$$

Question b)

Posons $A = (\mathbb{I}_3\, 0\ -\mathbb{I}_3\, 0) \in \mathcal{M}_{3,8}(\mathbb{R})$ de rang 3, alors $A\vartheta = 0$.
$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 H\left(\vartheta\right)^{-1}\right)$ Si $A\vartheta = 0$, alors par Slutsky:

$$\sqrt{n}A(\hat{\vartheta} - \vartheta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 AH\left(\vartheta\right)^{-1} A^T\right)$$

Question c)

On propose

$$\sigma_n^2 = \frac{\sum_{i,j}(\varepsilon_{i,j}^{(mai)})^2 + \sum_{i,j}(\varepsilon_{i,j}^{(juin)})^2}{2n} \xrightarrow{ps} \sigma^2$$

Par LGN parceque $\mathbb{V}[\epsilon_{i,j}^{(mois)}] = \sigma^2$.
Comme $\sqrt{n}(\hat{\vartheta} - \vartheta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 H\left(\vartheta\right)^{-1}\right)$, alors $\hat{\vartheta} - \vartheta \xrightarrow{\mathbb{P}_\vartheta} 0$.
Si $A\vartheta \neq 0$, alors

$$(A\hat{\vartheta}_n)^T \hat{V}(\hat{\vartheta}_n)^{-1} A\hat{\vartheta}_n \xrightarrow{\mathbb{P}_\vartheta} (A\vartheta)^T \hat{V}(\vartheta)^{-1} A\vartheta \neq 0$$

donc

$$T_n \xrightarrow{\mathbb{P}_\vartheta} +\infty$$

Si $A\vartheta = 0$ alors comme on a:

- $\sqrt{n}A(\hat{\vartheta} - \vartheta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 V(\vartheta)\right)$
- $\hat{V}(\hat{\vartheta}_n) \xrightarrow{\mathbb{P}_\vartheta} V(\vartheta)$
- $\sigma_n^2 \xrightarrow{\mathbb{P}_\vartheta} \sigma^2$

Alors par Slutsky:

$$T_n = \frac{1}{\hat{\sigma}_n^2}(\sqrt{n}A\hat{\vartheta}_n)^T \hat{V}(\hat{\vartheta}_n)(\sqrt{n}A\hat{\vartheta}_n) \xrightarrow{d} \frac{1}{\sigma^2} Z^T V(\vartheta) Z$$

où $Z \sim \mathcal{N}\left(0, \sigma^2 V(\vartheta)\right)$.
Or $V(\vartheta)$ est définie positive, en effet:

$$\begin{aligned}(\forall X \in \mathbb{R}^3 - \{0\})\; X^T V(\vartheta) X > 0 &\Leftrightarrow X^T AH(\vartheta)A^T X > 0 \\ &\Leftrightarrow X^T AH(\vartheta)A^T X > 0 \\ &\Leftrightarrow (A^T X)^T H(\vartheta) A^T X > 0 \qquad \text{(car } H(\vartheta) > 0 \text{ et } A^T X \neq 0)\end{aligned}$$

Donc:

$$\frac{1}{\sigma^2} Z^T V(\vartheta) Z = \left\| \frac{V(\vartheta)^{-\frac{1}{2}}}{\sigma} Z \right\|^2 \sim \chi^2(3)$$

Question d)

On propose le test suivant:

$$\varphi(\hat{\vartheta}_n) = \begin{cases} 1 & T_n > \beta \\ 0 & \text{sinon} \end{cases}$$

Sous H_0, si ϑ et tel que $A\vartheta = 0$:

$$\begin{aligned} \mathbb{P}_\vartheta(\varphi = 1) &= \mathbb{P}_\vartheta(T_n > \beta) \\ &\to 1 - \mathbb{P}_\vartheta(\chi^2(3) \leq \beta) \\ &= 5\% \end{aligned}$$

Où $\beta = 7.81$ est le quantile de $\chi 2(3)$ à 95%.

Question 2 a)

$\hat{\vartheta}_n = \left(\begin{smallmatrix} 0.042 & 0.059 \\ 1.94 & 1.91 \\ 2.56 & 2.84 \\ 3.47 & 3.25 \end{smallmatrix} \right)$, $T_n = 2.1 < \beta$ donc $\varphi = 0$. **On ne peut rien dire**.

Question 2 b)

On trouve $\hat{\rho} = 0.6$

Question 3 a)

$I_B(\alpha = 5\%) = [0.48, 0.75]$

Question 3 b)

Posons $g(x) = 10^x$, de dérivé $g'(x) = ln(10)g(x)$.

$$\begin{aligned} \sqrt{n}(\hat{\rho} - \rho) &= \sqrt{n}\left(10^{\tilde{\vartheta}_4^{(juin)} - \tilde{\vartheta}_4^{(mai)}} - 10^{\vartheta_4^{(juin)} - \vartheta_4^{(mai)}}\right) \\ &= \sqrt{n}\left(g\left(\tilde{\vartheta}_4^{(juin)} - \tilde{\vartheta}_4^{(mai)}\right) - g\left(\vartheta_4^{(juin)} - \vartheta_4^{(mai)}\right)\right) \end{aligned}$$

Par le théorème des accroissement finis, il existe $\tilde{\eta} \in \left(\tilde{\vartheta}_4^{(juin)} - \tilde{\vartheta}_4^{(mai)}, \vartheta_4^{(juin)} - \vartheta_4^{(mai)}\right)$ tel que

$$\sqrt{n}(\hat{\rho} - \rho) = g'(\tilde{\eta})\left(\tilde{\vartheta}_4^{(juin)} - \vartheta_4^{(juin)} + \tilde{\vartheta}_4^{(mai)} - \vartheta_4^{(mai)}\right)$$

Sous H_0 on a:

$$\sqrt{n}(\tilde{\vartheta} - \vartheta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 H(\vartheta)^{-1}\right)$$

donc par projection sur la quatrième coordonnée:

$$\tilde{\vartheta}_4^{(juin)} - \vartheta_4^{(juin)} \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \nu_1\right)$$

et

$$\tilde{\vartheta}_4^{(mai)} - \vartheta_4^{(mai)} \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \nu_2\right)$$

où ν_1 (resp. ν_2) est la composante $(4,4)$ de la matrice $H(\vartheta^{(juin)})^{-1}$ (resp. $H(\vartheta^{(mai)})^{-1}$)

Par indépendance:

$$\tilde{\vartheta}_4^{(juin)} - \vartheta_4^{(juin)} + \tilde{\vartheta}_4^{(mai)} - \vartheta_4^{(mai)} \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(\nu_1 + \nu_2)\right)$$

Et comme:

$$g'(\tilde{\eta}) \xrightarrow{\mathbb{P}_\vartheta} g'(\vartheta_4^{(juin)} - \vartheta_4^{(mai)})$$

Alors par Slutsky:

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} \mathcal{N}\left(0, s^2\right)$$

avec $s = g'(\vartheta_4^{(juin)} - \vartheta_4^{(mai)})\, \sigma \sqrt{\nu_1 + \nu_2}$.

On estime s par $\hat{s} = g'(\tilde{\vartheta}_4^{(juin)} - \tilde{\vartheta}_4^{(mai)})\, \sigma_n \sqrt{\hat{\nu}_1 + \hat{\nu}_2}$, avec $\hat{\nu}_1$ (resp. $\hat{\nu}_2$) est la composante $(4,4)$ de la matrice $H(\hat{\vartheta}^{(juin)})^{-1}$ (resp. $H(\hat{\vartheta}^{(mai)})^{-1}$).

On a $\hat{s} \xrightarrow{\mathbb{P}_\vartheta} s$.

Question b)

$$\begin{aligned}
1-\alpha &= \mathbb{P}\left(\hat{Z}^b \in [G^*(\alpha/2), G^*(1-\alpha/2)]\right)\\
&= \mathbb{P}\left(G^*(\alpha/2) < \sqrt{\frac{n}{\hat{s}_b}}(\hat{\rho}_b - \rho) < G^*(1-\alpha/2)\right)\\
&= \mathbb{P}\left(\hat{\rho}_b - \sqrt{\frac{\hat{s}_b}{n}}G^*(1-\alpha/2) < \rho < \hat{\rho}_b - \sqrt{\frac{\hat{s}_b}{n}}G^*(\alpha/2)\right)
\end{aligned}$$

Donc $I'_B(\alpha) = [\rho_b - \sqrt{\frac{\hat{s}_b}{n}}G^*(1-\alpha/2), \hat{\rho}_b - \sqrt{\frac{\hat{s}_b}{n}}G^*(\alpha/2)]$ est un intervalle de confiance (non asymptotique) de niveau α.

Code du programme:

```
# setwd("D:\\Ecole\\Projet MAP\\map432")
rm(list=ls())

library(numDeriv);
library(Matrix);

#helper function
integrate_matrix <- function(A, steps=1000) {
        res=0
        for (i in seq(0,1,(1/steps)))
                res=res+A(i)
        res=res/steps
        return(res)
}

z <- read.table("data.txt", header=T)
# calcul de x = log(1/d)
z$x <- sapply(strsplit(as.character(z$d), split = "/"), function(x) -log(as.numeric(x[1]) / as.numeric(x[2]),
print(z)

k <- length(z$d)
n <- 2 * k
message(sprintf("n = %d", n))

f <- function(x, theta)
theta[1] + (theta[2]-theta[1])/(1+exp(theta[3]*(x-theta[4])))

Mn_mai <- function(theta)
(1/n) *  sum( (z$mai1 - f(z$x, theta))**2 + (z$mai2 - f(z$x, theta))**2 )

Mn_juin <- function(theta)
(1/n) *  sum( (z$juin1 - f(z$x, theta))**2 + (z$juin2 - f(z$x, theta))**2 )

theta_mai <- nlm(Mn_mai, p = c(0,2,2.5,3.5), hessian=TRUE)$estimate
theta_juin <- nlm(Mn_juin, p = c(0,2,2.5,3.5), hessian=TRUE)$estimate
theta <- c(theta_mai, theta_juin)

message("theta = ")
print(matrix(c(theta_mai, theta_juin), nrow=4, ncol=2, byrow=F))

A = matrix(rep(0, 3*8), nrow=3, ncol=8, byrow=T)
for (i in 1:3) {
        A[i, i] <- 1
        A[i, i+4] <- -1
```

```
}
message("A = ")
print(A)

sigma2 =  (Mn_mai(theta_mai) + Mn_juin(theta_juin))/2
message( sprintf("Sigma^2 = %e", sigma2) )

grad_f <- function(x, theta) {
        gradient <- grad( (function(t)  f(x, t) ) , theta )
}
H_theta_mois = function(theta_mois)
                                        integrate_matrix(function(x){
                                                gradient <-grad_f(x, theta_mois);
                                                return ( gradient %*% t(gradient) );
                                        })

H_theta = bdiag( H_theta_mois(theta_mai), H_theta_mois(theta_juin))
message("H(theta) = ")
print(H_theta)

V_theta = A  %*% solve(H_theta)  %*% t(A)
A_theta = A %*% theta
T_n =  as.numeric(t(A_theta) %*% solve(V_theta) %*% A_theta) * (n/sigma2)


alpha <- 0.05
beta <- qchisq(1-alpha, 3)
message("Tn = ", T_n, " beta = ", beta)


# Calcul de rho

theta_tilde <- nlm(function(t) Mn_mai(t[1:4]) + Mn_juin(c(t[1:3], t[5])), p = c(0,2,2.5,3.5, 3.5), hessian=TR
theta_tilde <- c( theta_tilde[1:4], theta_tilde[1:3], theta_tilde[5] )
theta_tilde <- matrix(theta_tilde, nrow=4, ncol=2, byrow=F)
colnames(theta_tilde) <- c("mai", "juin")
message("theta tilde = ")
print(theta_tilde)

rho <- 10**(theta_tilde[4,"juin"] - theta_tilde[4,"mai"])
message("rho = ", rho)


# Partie 3, bootstrap -----------------------------------------------------

# mai
epsilon_tilde_mai <- c( z$mai1 - f(z$x, theta_tilde[,1]), z$mai2 - f(z$x, theta_tilde[,1]))
# juin
epsilon_tilde_juin <- c( z$juin1 - f(z$x, theta_tilde[,1]), z$juin2 - f(z$x, theta_tilde[,1]))

epsilon_tilde <- cbind(epsilon_tilde_mai, epsilon_tilde_juin)
epsilon_tilde <- epsilon_tilde  - mean(epsilon_tilde)
print(epsilon_tilde)

Mn <- function(theta, Y) (1./n) *  sum( (Y - f(c(z$x, z$x), theta))**2 )
B <- 1000
rho_b <- rho
for( i in 1:B ) {
        Y_mai <- f(z$x, theta_tilde[,"mai"]) + sample(epsilon_tilde, n, replace=T)
```

```
        Y_juin <- f(z$x, theta_tilde[,"juin"]) + sample(epsilon_tilde, n, replace=T)
        theta_b <- nlm(function(t)
                Mn(t[1:4], Y_mai) + Mn(c(t[1:3], t[5]), Y_juin),
                p = c(0,2,2.5,3.5, 3.5),
                hessian=TRUE)$estimate
        theta_b <- c( theta_b[1:4], theta_b[1:3], theta_b[5] )
        theta_b <- matrix(theta_b, nrow=4, ncol=2, byrow=F)
        colnames(theta_b) <- c("mai", "juin")
        rho_b <- c(rho_b, 10**(theta_b[4,"juin"] - theta_b[4,"mai"]))
}

FB <- function(alpha) quantile(rho_b, probs=alpha)

message(sprintf("Intervalle de confiance a 5 = [ %f, %f ]\n", FB(alpha/2), FB(1-alpha/2)))

# Calcul de la variance de rho
nu_1 <- solve(H_theta_mois(theta_tilde[,"juin"]))[4,4]
nu_2 <- solve(H_theta_mois(theta_tilde[,"mai"]))[4,4]

s_chapeau <- log(10) * (10**(theta_tilde[4,"juin"] - theta_tilde[4,"mai"])) * sqrt(sigma2 *(nu_1 + nu_2) )

message(sprintf("s^ = %f", s_chapeau))
```

Sortie du programme:

```
Données:
        d  mai1  mai2 juin1 juin2        x
1    1/30 1.909 1.956 1.886 1.880 1.477121
2    1/90 1.856 1.876 1.853 1.870 1.954243
3   1/270 1.838 1.841 1.747 1.772 2.431364
4   1/810 1.579 1.584 1.424 1.406 2.908485
5  1/2430 1.057 1.072 0.781 0.759 3.385606
6  1/7290 0.566 0.561 0.377 0.376 3.862728
7 1/21869 0.225 0.229 0.153 0.138 4.339829
8 1/65609 0.072 0.114 0.053 0.058 4.816963

n = 16

theta =
           [,1]       [,2]
[1,] 0.04239493 0.05808544
[2,] 1.93667722 1.90883944
[3,] 2.56495531 2.83651139
[4,] 3.46717460 3.25072103

A =
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    1    0    0    0   -1    0    0    0
[2,]    0    1    0    0    0   -1    0    0
[3,]    0    0    1    0    0    0   -1    0

Sigma^2 = 5.602085e-04

H(theta) =
8 x 8 sparse Matrix of class "dgCMatrix"
[1,] 6.177337e-07 0.0006422048 3.104112e-06 2.997823e-06 .
[2,] 6.422048e-04 0.9997149727 3.371012e-03 3.117314e-03 .
[3,] 3.104112e-06 0.0033710119 1.566996e-05 1.506447e-05 .
[4,] 2.997823e-06 0.0031173144 1.506447e-05 1.454825e-05 .
```

```
[5,] .            .            .            .            5.008486e-07
[6,] .            .            .            .            5.601068e-04
[7,] .            .            .            .            2.243690e-06
[8,] .            .            .            .            2.626329e-06

[1,] .            .            .
[2,] .            .            .
[3,] .            .            .
[4,] .            .            .
[5,] 0.0005601068 2.243690e-06 2.626329e-06
[6,] 0.9998792856 2.631622e-03 2.937758e-03
[7,] 0.0026316219 1.009912e-05 1.176567e-05
[8,] 0.0029377577 1.176567e-05 1.377184e-05

Tn = 2.10878561107025 beta = 7.81472790325118

theta tilde =
            mai       juin
[1,] 0.05020411 0.05020411
[2,] 1.92393686 1.92393686
[3,] 2.68886823 2.68886823
[4,] 3.46988033 3.24714608
rho = 0.598777887065384

Partie bootstrap

      epsilon_tilde_mai epsilon_tilde_juin
 [1,]        0.04780591        0.024805907
 [2,]        0.01731985        0.014319850
 [3,]        0.07620256       -0.014797444
 [4,]        0.04818855       -0.106811455
 [5,]        0.01819419       -0.257805810
 [6,]        0.08630643       -0.102693570
 [7,]        0.06399939       -0.008000614
 [8,]        0.02698314        0.007983137
 [9,]        0.09480591        0.018805907
[10,]        0.03731985        0.031319850
[11,]        0.07920256        0.010202556
[12,]        0.05318855       -0.124811455
[13,]        0.03319419       -0.279805810
[14,]        0.08130643       -0.103693570
[15,]        0.06799939       -0.023000614
[16,]        0.06898314        0.012983137

Intervalle de confiance a 5% = [ 0.489483, 0.740462 ]

s^ = 332325.262823
```