

Corrections des exercices

Voici les corrections, plus ou moins détaillées suivant leur intérêt, des différents exercices du livre.

1 Corrections des exercices du chapitre 1

Exercice 1.1

On a $SCR'(m) = 2 \sum_{i=1}^n (m - y_i)$ et $SCR''(m) = 2 \cdot n$. Donc $SCR'(\hat{m}) = 0 \iff \hat{m} = \frac{1}{n} \sum_{i=1}^n y_i$ et $SCR''(m) > 0$ pour tout $m \in \mathbb{R}$: \hat{m} est un minimum local d'une fonction convexe, c'est donc aussi un minimum global. De plus \hat{m} est la moyenne empirique des (y_i) .

Exercice 1.2

1. pour $n = 2p + 1$, on montre que $\tilde{m} = y_{(p+1)}$. En effet, dans un tel cas

$$SVA(y_{(p+1)}) = \sum_{i=1}^p (y_{(p+1)} - y_{(i)}) + \sum_{i=p+2}^{2p+1} (y_{(i)} - y_{(p+1)}) = \sum_{i=p+2}^{2p+1} y_{(i)} - \sum_{i=1}^p y_{(i)}.$$

Pour $m \in]y_{(j)}, y_{(j+1)}]$ et $j = p + 1, \dots, 2p + 1$ (avec, par convention, $y_{(2p+2)} = +\infty$ et les sommes vides nulles),

$$\begin{aligned} SVA(m) &= \sum_{i=1}^j (m - y_{(i)}) + \sum_{i=j+1}^{2p+1} (y_{(i)} - m) \\ &= \sum_{i=1}^p (m - y_{(i)}) + \sum_{i=p+2}^{2p+1} (y_{(i)} - m) + \sum_{i=p+1}^j (y_{(i)} - m) - \sum_{i=p+2}^j (y_{(i)} - m) \\ &= \sum_{i=p+2}^{2p+1} y_{(i)} - \sum_{i=1}^p y_{(i)} + 2 \sum_{i=p+2}^j (m - y_{(i)}) + (m - y_{(p+1)}) \\ &> SVA(y_{(p+1)}). \end{aligned}$$

Par symétrie, il en est de même si $m \in [y_{(j)}, y_{(j+1)})$ et $j = 0, \dots, p$. On en déduit que le minimum de $SVA(\cdot)$ est atteint en $\tilde{m} = y_{(p+1)}$, qui est la médiane des (y_j) .

2. Si $n = 2p$, on déduit de ce qui précède que le minimum de SVA doit se trouver entre $y_{(p)}$

et $y_{(p+1)}$. En effet, pour $m \in]y_{(p)}, y_{(p+1)}]$,

$$\begin{aligned} SVA(m) &= \sum_{i=1}^p (m - y_{(i)}) + \sum_{i=p+1}^{2p} (y_{(i)} - m) \\ &= \sum_{i=p+1}^{2p} y_{(i)} - \sum_{i=1}^p y_{(i)}, \end{aligned}$$

qui est indépendant de m . Maintenant, pour $m \in [y_{(j)}, y_{(j+1)}[$ et $j = p+1, \dots, 2p+1$, on a :

$$\begin{aligned} SVA(m) &= \sum_{i=1}^j (m - y_{(i)}) + \sum_{i=j+1}^{2p+1} (y_{(i)} - m) \\ &= \sum_{i=1}^p (m - y_{(i)}) + \sum_{i=p+1}^{2p} (y_{(i)} - m) + 2 \sum_{i=p+1}^j (m - y_{(i)}) \\ &> \sum_{i=p+1}^{2p} y_{(i)} - \sum_{i=1}^p y_{(i)}. \end{aligned}$$

On retrouve bien que le minimum est atteint entre $y_{(p)}$ et $y_{(p+1)}$, donc en la médiane des (y_i) .

Exercice 1.3

On a $SCR'(a) = 2 \sum_{i=1}^n x_i (a \cdot x_i - y_i)$ et $SCR''(a) = 2 \sum_{i=1}^n x_i^2$. Donc $SCR'(\hat{a}) = 0 \iff \hat{a} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$

et $SCR''(a) > 0$ pour tout $a \in \mathbb{R}$: \hat{a} est un minimum local d'une fonction convexe, c'est donc aussi un minimum global. On remarque que si la série a été recentrée (à savoir que l'on a retranché la moyenne des x_i à chacun des x_i , et la moyenne des y_i à chacun des y_i), alors $\hat{a} = \hat{\beta}$, pente de la régression simple.

Exercice 1.4

i. En notant $f_{\mu, \beta, \sigma^2}(Y)$ la densité de $Y = (Y_1, \dots, Y_n)'$, comme $\mathbb{E}(Y) = \mu + \beta \cdot Z$ et $\text{Var}(Y) = \sigma^2 \cdot I_n$ et que le bruit est supposé gaussien, alors

$$f_{\mu, \beta, \sigma^2}(Y) = \frac{1}{(2\pi \cdot \sigma^2)^{n/2}} \exp \left\{ \frac{1}{2\sigma^2} (Y - \mu - \beta \cdot Z)' \cdot (Y - \mu - \beta \cdot Z) \right\}.$$

Par suite, avec pour $Y \in \mathbb{R}^n$, $L(\mu, \beta, \sigma^2) = -2 \times \log(f_{\mu, \beta, \sigma^2}(Y))$, et on a bien :

$$L(\mu, \beta, \sigma^2) = n \cdot \log(2\pi) + n \cdot \log(\sigma^2) + \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \mu - \beta \cdot Z_i)^2.$$

ii. Pour déterminer un estimateur du maximum de vraisemblance des 3 paramètres, on peut chercher un minimum local de $L(\mu, \beta, \sigma^2)$, c'est-à-dire un point critique de cette fonction. On obtient ainsi :

$$\begin{pmatrix} \tilde{\mu} \\ \tilde{\beta} \\ \tilde{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \hat{\mu} \\ \hat{\beta} \\ \frac{1}{n} \sum_{i=1}^n (Y_i - \tilde{\mu} - \tilde{\beta} \cdot Z_i)^2 \end{pmatrix}$$

(on vérifie également que la matrice hessienne de l'application $L(\mu, \beta, \sigma^2)$ est définie positive et donc que l'on a bien un minimum). Finalement, seul l'estimateur de la variance est différent de

celui obtenu par la méthode des moindres carrés.

iii. On peut comparer la vitesse de convergence de ces deux estimateurs en calculant le risque quadratique de chacun d'eux. Ainsi, pour l'estimateur obtenu par moindres carrés, qui est non biaisé,

$$R(\hat{\sigma}^2) = \text{Var}(\hat{\sigma}^2) = \frac{2}{n-2} \sigma^4,$$

car c'est un estimateur dont la loi suit $\frac{\sigma^2}{n-2} \cdot \chi^2(n-2)$ et la variance d'une loi $\chi^2(n-2)$ est $2(n-2)$. Pour l'estimateur obtenu par maximum de vraisemblance, on a :

$$\begin{aligned} R(\tilde{\sigma}^2) &= \mathbb{E}\left((\tilde{\sigma}^2 - \sigma^2)^2\right) \\ &= \text{Var}(\tilde{\sigma}^2) + \left(\frac{n-2}{n} \sigma^2 - \sigma^2\right)^2 \quad \text{car l'espérance de } \tilde{\sigma}^2 \text{ est } \frac{n-2}{n} \sigma^2 \\ &= \left(\frac{n-2}{n}\right)^2 \frac{2}{n-2} \sigma^4 + \frac{4\sigma^4}{n^2} \quad \text{car } \tilde{\sigma}^2 = \frac{n-2}{n} \hat{\sigma}^2 \\ &= \frac{2}{n} \sigma^4. \end{aligned}$$

Ainsi, le risque quadratique de $\tilde{\sigma}^2$, pourtant biaisé, est plus petit que celui de $\hat{\sigma}^2$.

iv. On suppose donc que $Y_{ij} = \mu_j + \varepsilon_{ij}$ pour $i = 1, \dots, I$ et $j = 1, \dots, n_i$, avec les hypothèses faites en (1.6) sur le bruit. Alors, dans ce cas, si on note toujours $L(\mu_1, \dots, \mu_I, \sigma^2)$ la fonction $-2 \times \log$ -vraisemblance, on a :

$$L(\mu_1, \dots, \mu_I, \sigma^2) = n \cdot \log(2\pi) + n \cdot \log(\sigma^2) + \frac{1}{\sigma^2} \sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \mu_j)^2.$$

En minimisant cette expression, on montre (après calcul de l'unique point critique et de la matrice hessienne associée) que les estimateurs $\tilde{\mu}_i$ sont les mêmes que $\hat{\mu}_i = Y_{i \cdot}$, mais que l'estimateur de la variance est $\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \tilde{\mu}_j)^2$, soit $\frac{n-I}{n} \hat{\sigma}^2$. Enfin, comme précédemment, on montre que :

$$R(\hat{\sigma}^2) = \text{Var}(\hat{\sigma}^2) = \frac{2}{n-I} \sigma^4,$$

alors que :

$$R(\tilde{\sigma}^2) = \frac{2n + I^2 - 2I}{n^2} \sigma^4,$$

et ainsi, pour tout $n \in \mathbb{N}^*$, pour tout $I \in \mathbb{N}^*$, $R(\tilde{\sigma}^2) < R(\hat{\sigma}^2)$: l'estimateur par maximum de vraisemblance, même s'il est biaisé, converge plus vite (au sens quadratique) que celui obtenu par moindres carrés.

Exercice 1.5

i. On cherche donc $\sup_{(\mu, \beta, \sigma^2) \in \Theta_1} \log(f_{\mu, \beta, \sigma^2}(Y))$, où Θ_1 est l'ensemble inclus dans $\mathbb{R}^2 \times \mathbb{R}_+^*$ correspondant à l'hypothèse H_1 , soit $\Theta_1 = \mathbb{R} \times \mathbb{R}^* \times \mathbb{R}_+^*$. Du fait de la continuité de la fonction $(\mu, \beta, \sigma^2) \mapsto f_{\mu, \beta, \sigma^2}(Y)$, on en déduit que

$$\sup_{(\mu, \beta, \sigma^2) \in \Theta_1} \log(f_{\mu, \beta, \sigma^2}(Y)) = \log(f_{\tilde{\mu}, \tilde{\beta}, \tilde{\sigma}^2}(Y)),$$

avec $(\tilde{\mu}, \tilde{\beta}, \tilde{\sigma}^2)$ l'unique estimateur du maximum de vraisemblance (voir l'exercice précédent). En conséquence,

$$\sup_{(\mu, \beta, \sigma^2) \in \Theta_1} f_{\mu, \beta, \sigma^2}(Y) = \frac{1}{(2\pi\tilde{\sigma}^2)^{n/2}} \exp\left\{-\frac{1}{2}n\right\}.$$

et donc :

$$\sup_{(\mu, \beta, \sigma^2) \in \Theta_1} \log(f_{\mu, \beta, \sigma^2}(Y)) = -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\tilde{\sigma}^2) - \frac{1}{2}n.$$

ii. Sous l'hypothèse H_0 , on cherche $\sup_{(\mu, \beta, \sigma^2) \in \Theta_0} \log(f_{\mu, \beta, \sigma^2}(Y))$, avec $\Theta_0 = \mathbb{R} \times \{0\} \times \mathbb{R}_+^*$. Il revient

donc de chercher les paramètres μ et σ^2 maximisant la fonction $-\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \mu)^2$.

Après, recherche du point critique, on montre qu'un maximum local et global est atteint en $\bar{\mu} = \frac{1}{n} \sum_{i=1}^n Y_i$ et en $\bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{\mu})^2$, soit la moyenne empirique et la variance empirique usuelles.

En conséquence,

$$\sup_{(\mu, \beta, \sigma^2) \in \Theta_0} \log(f_{\mu, \beta, \sigma^2}(Y)) = -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\bar{\sigma}^2) - \frac{1}{2}n.$$

iii. On a $U = \frac{\sup_{(\mu, \beta, \sigma^2) \in \Theta_0} f_{\mu, \beta, \sigma^2}(Y)}{\sup_{(\mu, \beta, \sigma^2) \in \Theta_1} f_{\mu, \beta, \sigma^2}(Y)}$, donc grâce au fait que la fonction logarithme est croissante, on obtient que :

$$\log(U) = \sup_{(\mu, \beta, \sigma^2) \in \Theta_0} \log(f_{\mu, \beta, \sigma^2}(Y)) - \sup_{(\mu, \beta, \sigma^2) \in \Theta_1} \log(f_{\mu, \beta, \sigma^2}(Y)) = -\frac{n}{2} \log\left(\frac{\bar{\sigma}^2}{\tilde{\sigma}^2}\right).$$

La région d'acceptation du test est donc de la forme : $\frac{\bar{\sigma}^2}{\tilde{\sigma}^2} < K_\alpha$. Avec ces mêmes notations, le test de Fisher associé s'écrit $\hat{F} = (n-2) \cdot \frac{\bar{\sigma}^2 - \tilde{\sigma}^2}{\tilde{\sigma}^2}$ et la région d'acceptation est de la forme $\hat{F} < K'_\alpha$, ce qui revient à écrire que $\frac{\bar{\sigma}^2}{\tilde{\sigma}^2} < (1 + K'_\alpha)$. Ces deux tests (du rapport de vraisemblance et de Fisher) sont donc les mêmes (et donc les mêmes que le test de Student associé).

iv. Pour le modèle d'analyse de la variance, on a comme précédemment :

$$\sup_{(\mu_1, \mu_2, \mu_3, \sigma^2) \in \Theta_1} \log(f_{\mu_1, \mu_2, \mu_3, \sigma^2}(Y)) = \log(f_{\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3, \tilde{\sigma}^2}(Y)) = -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\tilde{\sigma}^2) - \frac{1}{2}n,$$

en notant avec un tilde les estimateurs du maximum de vraisemblance. Sous l'hypothèse H_0 , $\Theta_0 = \{(\mu, \mu, \mu), \mu \in \mathbb{R}\} \times \mathbb{R}_+^*$, et :

$$\begin{aligned} \sup_{(\mu_1, \mu_2, \mu_3, \sigma^2) \in \Theta_0} \log(f_{\mu_1, \mu_2, \mu_3, \sigma^2}(Y)) &= \sup_{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*} \left\{ -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \mu)^2 \right\} \\ &= -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\bar{\sigma}^2) - \frac{1}{2}n. \end{aligned}$$

Par suite, on trouve, comme précédemment, que le test du rapport de vraisemblance est le même que celui de Fisher.

3 Corrections des exercices du chapitre 3

Exercice 3.1

$$\begin{aligned}
 \text{On a :} \quad \mathbb{E}(\|Y - T\|^2) &= \mathbb{E}(\|(Y - X \cdot \theta) + (X \cdot \theta - T)\|^2) \\
 &= \mathbb{E}(\|(Y - X \cdot \theta)\|^2) + \|T - X \cdot \theta\|^2 \\
 &= \sum_{i=1}^n \mathbb{E}((Y - (X \cdot \theta)_i)^2) + \|T - X \cdot \theta\|^2 \\
 &= n \cdot \sigma^2 + \|T - X \cdot \theta\|^2
 \end{aligned}$$

Exercice 3.2

(i) $\tilde{\theta}$ est sans biais donc $\forall \theta, \mathbb{E}(\tilde{\theta}) = \theta$. C'est à dire : $M \cdot X \cdot \theta = \theta$. Donc $M \cdot X = I_n$

(ii) $\hat{\theta} = (X' \cdot X)^{-1} X' \cdot X (X' \cdot X)^{-1} X' Y$ donc $T = (X' \cdot X)^{-1} X'$.

$$M \cdot P_{[X]} = M \cdot X (X' \cdot X)^{-1} X' = (X' \cdot X)^{-1} X'$$

$$T \cdot P_{[X]} = (X' \cdot X)^{-1} X' \cdot X (X' \cdot X)^{-1} X' = (X' \cdot X)^{-1} X'.$$

(iii) $\tilde{\theta} = M(P_{[X]} \cdot Y + P_{[X]^\perp} \cdot Y) = \hat{\theta} + M \cdot P_{[X]^\perp} \cdot Y$, la somme étant non-corrélée car $P_{[X]} \cdot Y$ et $P_{[X]^\perp} \cdot Y$ le sont. On en déduit

$$\text{Var}(\tilde{\theta}) = \text{Var} \hat{\theta} + \text{Var}(M \cdot P_{[X]^\perp} \cdot Y).$$

ce qui donne le résultat.

Exercice 3.3

Il suffit de remarquer que

$$\begin{pmatrix} Y_1 \\ Y_2/2 \\ Y_3/3 \end{pmatrix}$$

suit un modèle linéaire non gaussien au sens de l'exercice précédent. En appliquant la théorème de Gauss-Markov

$$\hat{\theta} = \frac{4}{29} \begin{pmatrix} 9/4 & -5/2 \\ -5/2 & 6 \end{pmatrix} \begin{pmatrix} Y_1 + Y_2/2 + 2Y_3/3 \\ Y_1 - Y_2/4 + Y_3/3 \end{pmatrix} = \frac{1}{29} \begin{pmatrix} -Y_1 + 7Y_2 + 8/3Y_3 \\ 14Y_1 - 11Y_2 + 4/3Y_3 \end{pmatrix}.$$

Exercice 3.4

(i) Evident : $\varepsilon' \cdot P_{[X]^\perp} \cdot \varepsilon$ est scalaire.

(ii) Il suffit de commuter dans l'argument de trace, puis de commuter espérance et trace.

$$(iii) (n - k) \mathbb{E}(\hat{\sigma}^2) = \sigma^2 \text{Tr}(P_{[X]^\perp}) = \sigma^2 \text{rang}(P_{[X]^\perp}) = \sigma^2 (n - k).$$

Exercice 3.5

En reprenant l'exercice du chapitre 1, on se place cette fois-ci dans le cadre général du modèle linéaire gaussien, soit $Y = X \cdot \theta + \varepsilon$, en supposant que la matrice X , de taille (n, k) est également de rang k .

i. En notant $f_{\theta, \sigma^2}(Y)$ la densité de $Y = (Y_1, \dots, Y_n)'$, comme $\mathbb{E}(Y) = X \cdot \theta$ et $\text{Var}(Y) = \sigma^2 \cdot I_n$ et

comme le bruit est supposé gaussien, alors $f_{\theta, \sigma^2}(Y) = \frac{1}{(2\pi \cdot \sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} (Y - X \cdot \theta)' \cdot (Y - X \cdot \theta) \right\}$.

Par suite, avec pour $Y \in \mathbb{R}^n$, $L(\theta, \sigma^2) = -2 \times \log(f_{\theta, \sigma^2}(Y))$, on a :

$$L(\theta, \sigma^2) = n \cdot \log(2\pi) + n \cdot \log(\sigma^2) + \frac{1}{\sigma^2} (Y - X \cdot \theta)' \cdot (Y - X \cdot \theta).$$

ii. Pour déterminer un estimateur du maximum de vraisemblance des $k + 1$ paramètres, on peut chercher un minimum local de $L(\theta, \sigma^2)$, c'est-à-dire un point critique de cette fonction. On obtient ainsi, en différenciant par rapport à θ et à σ^2 :

$$\begin{cases} \frac{\partial L}{\partial \theta}(\theta, \sigma^2) = -\frac{1}{\sigma^2} (X' \cdot (Y - X \cdot \theta) + (Y - X \cdot \theta)' \cdot X) = -\frac{2}{\sigma^2} X' \cdot (Y - X \cdot \theta) \\ \frac{\partial L}{\partial \sigma^2}(\theta, \sigma^2) = \frac{n}{\sigma^2} - \frac{1}{\sigma^4} (Y - X \cdot \theta)' \cdot (Y - X \cdot \theta) \end{cases}$$

(noter que la différentiation par rapport à θ se fait matriciellement) et ainsi, en résolvant $\frac{\partial L}{\partial \theta}(\theta, \sigma^2) = 0$ et $\frac{\partial L}{\partial \sigma^2}(\theta, \sigma^2) = 0$, on obtient :

$$\begin{pmatrix} \tilde{\theta} \\ \tilde{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \hat{\theta} \\ \frac{1}{n} \|Y - X \cdot \hat{\theta}\|^2 \end{pmatrix}$$

(on vérifie également que la matrice hessienne de l'application $L(\theta, \sigma^2)$ est définie positive et donc que l'on a bien un minimum). Comme dans les cas particuliers, seul l'estimateur de la variance est différent de celui obtenu par la méthode des moindres carrés.

iii. On calcule le risque quadratique de chacun des estimateurs, et :

$$R(\hat{\sigma}^2) = \text{Var}(\hat{\sigma}^2) = \frac{2}{n - k} \sigma^4,$$

tandis que

$$R(\tilde{\sigma}^2) = \left(\frac{n - k}{n}\right)^2 \frac{2}{n - k} \sigma^4 + \frac{k^2 \sigma^4}{n^2} = \frac{2n - 2k + k^2}{n^2} \sigma^4.$$

On montre ainsi que le risque quadratique de $\tilde{\sigma}^2$, pourtant biaisé, est toujours plus petit que celui de $\hat{\sigma}^2$.

Exercice 3.6

On traite ici le cadre général du modèle linéaire gaussien, correspondant aussi bien à la régression qu'à l'analyse de la variance. Ainsi, les hypothèses H_0 et H_1 que l'on teste sont :

$$H_0 : [X] = [X^{(0)}], \quad \text{contre} \quad H_1 : [X] \neq [X^{(0)}],$$

où $[X^{(0)}]$ est un sous-espace vectoriel inclus dans $[X]$, de dimension $k_0 < k$.

i. On cherche $\sup_{(\theta, \sigma^2) \in \Theta_1} \log(f_{\theta, \sigma^2}(Y))$, où Θ_1 est l'ensemble inclus dans $\mathbb{R}^k \times \mathbb{R}_+^*$ correspondant à l'hypothèse H_1 . Du fait de la continuité de la fonction $(\mu, \beta, \sigma^2) \mapsto f_{\mu, \beta, \sigma^2}(Y)$, comme $\Theta_1 = \mathbb{R}^k \times \mathbb{R}_+^*$ à un ensemble de mesure de Lebesgue (sur \mathbb{R}^{k+1}) nulle, on en déduit que

$$\sup_{(\theta, \sigma^2) \in \Theta_1} \log(f_{\theta, \sigma^2}(Y)) = \log(f_{\tilde{\theta}, \tilde{\sigma}^2}(Y)),$$

avec $(\tilde{\theta}, \tilde{\sigma}^2)$ l'unique estimateur du maximum de vraisemblance (voir l'exercice précédent). En conséquence,

$$\sup_{(\theta, \sigma^2) \in \Theta_1} f_{\theta, \sigma^2}(Y) = \frac{1}{(2\pi\tilde{\sigma}^2)^{n/2}} \exp\left\{-\frac{1}{2}n\right\}.$$

et donc :

$$\sup_{(\theta, \sigma^2) \in \Theta_1} \log(f_{\theta, \sigma^2}(Y)) = -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\tilde{\sigma}^2) - \frac{1}{2}n.$$

ii. Sous l'hypothèse H_0 , on cherche $\sup_{(\theta, \sigma^2) \in \Theta_0} \log(f_{\theta, \sigma^2}(Y))$, avec $\Theta_0 = \{\theta \in \mathbb{R}^k, X \cdot \theta \in [X^{(0)}]\} \times \mathbb{R}_+^*$.

Il revient donc de chercher les paramètres θ et σ^2 appartenant à Θ_0 maximisant la fonction

$$-\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\sigma^2) - \frac{1}{2\sigma^2} \|Y - X \cdot \theta\|^2.$$

Dans la dernière partie de cette expression, on peut remplacer $X \cdot \theta$ par $X^{(0)} \cdot \theta^{(0)}$. Pour cette dernière partie qui est la seule dépendant de θ , maximiser reviendra à chercher le minimum de $\|Y - X^{(0)} \cdot \theta^{(0)}\|^2$, qui est obtenu par projection orthogonale, et ainsi : $\hat{\theta}^{(0)} = (X'^{(0)} \cdot X^{(0)})^{-1} \cdot X'^{(0)} \cdot Y$. En ce qui concerne σ^2 , la recherche classique d'un point critique montre que le maximum local et global est atteint en $\bar{\sigma}_{(0)}^2 = \frac{1}{n} \|Y - X^{(0)} \cdot \hat{\theta}^{(0)}\|^2$. En conséquence,

$$\sup_{(\theta, \sigma^2) \in \Theta_0} \log(f_{\theta, \sigma^2}(Y)) = -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\bar{\sigma}_{(0)}^2) - \frac{1}{2}n.$$

iii. Comme dans l'exercice 5 du chapitre 1, on a :

$$\log(U) = \sup_{(\theta, \sigma^2) \in \Theta_0} \log(f_{\theta, \sigma^2}(Y)) - \sup_{(\theta, \sigma^2) \in \Theta_1} \log(f_{\theta, \sigma^2}(Y)) = -\frac{n}{2} \log\left(\frac{\bar{\sigma}_{(0)}^2}{\tilde{\sigma}^2}\right),$$

et comme dans cet exercice, la région d'acceptation du test est donc de la forme : $\frac{\bar{\sigma}_{(0)}^2}{\tilde{\sigma}^2} < K_\alpha$, qui est de la même forme que celle obtenu par le test de Fisher. Le test de Fisher (ou de Student) est le même que le test de rapport de vraisemblance, à ceci près que la loi du test de Fisher est connue, quant celle du rapport de vraisemblance n'est pas exprimable simplement (même si les régions d'acceptation sont les mêmes).

Exercice 3.7

i. On a toujours $\hat{\theta} = (X' \cdot X)^{-1} \cdot X' \cdot Y$ et donc $\mathbb{E}(\hat{\theta}) = \theta$. En revanche, comme $\hat{\theta} = \theta + (X' \cdot X)^{-1} \cdot X' \cdot \varepsilon$, on a $\text{Cov}(\hat{\theta}) = (X' \cdot X)^{-1} \cdot X' \cdot \Sigma \cdot X \cdot (X' \cdot X)^{-1}$.

ii. Grâce au produit scalaire lié à la norme $\|\cdot\|_\Sigma$, on peut définir une Σ -orthogonalité par le fait que $\langle u, v \rangle_\Sigma = u' \cdot \Sigma^{-1} \cdot v = 0$. Par suite, les moindres carrés généralisés reviendront à minimiser $\|Y - X \cdot \theta\|_\Sigma$, donc à trouver la projection Σ -orthogonale de Y sur $[X]$. Le projeté Σ -orthogonal $P_{[X]}(Y) = M \cdot Y$ se définit par le fait que $P_{[X]}(Y) \in [X]$ et $(Y - P_{[X]}(Y)) \perp [X]$. En écrivant que $P_{[X]}(Y) = X \cdot \hat{\theta}_G$ on a déjà $P_{[X]}(Y) \in [X]$. Ensuite, pour chaque colonne $Z^{(i)}$ de X (qui est supposé de rang k), ce qui définit un vecteur de \mathbb{R}^n , on doit avoir $Z^{(i)'} \cdot \Sigma^{-1} \cdot (Y - X \cdot \hat{\theta}_G) = 0$, donc en "concaténant" ces k équations, on retrouve les équations "normales" :

$$X' \cdot \Sigma^{-1} \cdot (Y - X \cdot \hat{\theta}_G) = 0, \quad \text{soit} \quad X' \cdot \Sigma^{-1} \cdot Y = X' \cdot \Sigma^{-1} \cdot X \cdot \hat{\theta}_G$$

On en déduit donc que $\widehat{\theta}_G = (X' \cdot \Sigma^{-1} \cdot X)^{-1} \cdot X' \cdot \Sigma^{-1} \cdot Y$.

iii. On a $\mathbb{E}(\widehat{\theta}_G) = (X' \cdot \Sigma^{-1} \cdot X)^{-1} \cdot X' \cdot \Sigma^{-1} \cdot X \cdot \theta = \theta$: l'estimateur $\widehat{\theta}_G$ est non biaisé. En conséquence,

$$\begin{aligned} \text{Cov}(\widehat{\theta}_G) &= (X' \cdot \Sigma^{-1} \cdot X)^{-1} \cdot X' \cdot \Sigma^{-1} \cdot \Sigma \cdot \left((X' \cdot \Sigma^{-1} \cdot X)^{-1} \cdot X' \cdot \Sigma^{-1} \right)' \\ &= (X' \cdot \Sigma^{-1} \cdot X)^{-1} \cdot X' \cdot \Sigma^{-1} \cdot \Sigma \cdot \Sigma^{-1} \cdot X \cdot (X' \cdot \Sigma^{-1} \cdot X)^{-1} \\ &= (X' \cdot \Sigma^{-1} \cdot X)^{-1}. \end{aligned}$$

On peut maintenant reprendre l'exercice 2, en notant $\tilde{\theta}$ l'estimateur par moindres carrés "ordinaires" et par $\widehat{\theta}_G$ l'estimateur par moindres carrés généralisés. La notion d'orthogonalité est maintenant celle de Σ -orthogonalité (les projections sont donc Σ -orthogonales). Toute la démarche peut être reproduite, et on aboutit à $\widehat{\theta}_G = \tilde{\theta} + M \cdot P_{[X]^\perp} \cdot Y$. Il reste à montrer que les deux parties de la somme sont non corrélées. Ici, $M = (X' \cdot X)^{-1} \cdot X'$, d'où

$$\begin{aligned} \text{Cov}(\widehat{\theta}_G, M \cdot P_{[X]^\perp} \cdot Y) &= \text{Cov}\left((X' \cdot \Sigma^{-1} \cdot X)^{-1} \cdot X' \cdot \Sigma^{-1} \cdot \varepsilon, (X' \cdot X)^{-1} \cdot X' \cdot P_{[X]^\perp} \cdot \varepsilon \right) \\ &= (X' \cdot \Sigma^{-1} \cdot X)^{-1} \cdot X' \cdot \Sigma^{-1} \cdot \Sigma \cdot P'_{[X]^\perp} \cdot X \cdot (X' \cdot X)^{-1} \\ &= 0 \quad \text{car} \quad X' \cdot \Sigma^{-1} \cdot \Sigma \cdot P'_{[X]^\perp} = X' \cdot P'_{[X]^\perp} = (P_{[X]^\perp} \cdot X)' = 0. \end{aligned}$$

En conséquence, $\text{Var}(C' \cdot \widehat{\theta}_G) \leq \text{Var}(C' \cdot \tilde{\theta})$ pour tout vecteur C .

iv. Dans le cadre gaussien, en reprenant l'exercice 5, et en ne considérant que θ comme paramètre, on a $f_\theta(Y) = \frac{1}{(2\pi)^{n/2} \cdot |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(Y - X \cdot \theta)' \cdot \Sigma^{-1} \cdot (Y - X \cdot \theta)\right\}$. Ainsi, en minimisant $\theta \in \mathbb{R}^k \mapsto L(\theta) = -2 \times f_\theta(Y)$, on doit donc minimiser

$$n \log(2\pi) + \log |\Sigma| + (Y - X \cdot \theta)' \cdot \Sigma^{-1} \cdot (Y - X \cdot \theta).$$

Comme dans l'exercice 5, pour minimiser cette expression, on différencie par rapport à θ , et le point critique doit maintenant vérifier $X' \cdot \Sigma^{-1} \cdot (Y - X \cdot \tilde{\theta}) = 0$, ce qui permet d'aboutir à $\tilde{\theta} = \widehat{\theta}_G$ (on vérifie également que la matrice hessienne, qui vaut $2 \cdot X' \cdot \Sigma^{-1} \cdot X = 2 \cdot (\Sigma^{-1/2} \cdot X)' \cdot (\Sigma^{-1/2} \cdot X)$ est une matrice positive).

Exercice 3.8

(i)

$$Y_i = \theta \Delta + \sigma W_{t_i} - W_{t_{i-1}},$$

où les $W_{t_i} - W_{t_{i-1}}$ sont des v.a.i.i.d. gaussiennes de variance Δ . Il suffit ensuite d'appliquer les formules du modèle linéaire.

(ii) En appliquant la loi des grands nombres, les $W_{t_i} - W_{t_{i-1}}$ étant indépendantes, on obtient la convergence de $\widehat{\theta}$. Pour ce qui est de $\widehat{\sigma}^2$, on peut remarquer que l'on peut remplacer le $1/(n-1)$ par $1/n$ et donc :

$$\widehat{\sigma}^2 \simeq \frac{1}{n\Delta} \left[\sum_{i=1}^n Y_i^2 - n(\bar{Y})^2 \right]$$

Il suffit d'appliquer la loi des grands nombres aux Y_i^2 .

4 Corrections des exercices du chapitre 4

Exercice 4.1

En reprenant l'exercice 3.7, on a $\hat{\theta}_\Omega = (X' \cdot \Omega \cdot X)^{-1} \cdot X' \cdot \Omega \cdot Y$. En conséquence, on a $\mathbb{E}(\hat{\theta}_\Omega) = \theta$ et $\text{Cov}(\hat{\theta}_\Omega) = \sigma^2 (X' \cdot \Omega \cdot X)^{-1} \cdot X' \cdot \Omega^2 \cdot X \cdot (X' \cdot \Omega \cdot X)^{-1}$. De plus, $\hat{\sigma}_\Omega^2 = \frac{1}{n-k} (Y - X \cdot \hat{\theta}_\Omega)' \cdot \Omega \cdot (Y - X \cdot \hat{\theta}_\Omega)$.

Exercice 4.2

i. La matrice X du modèle linéaire associé à cette régression polynômiale, est une matrice de taille $(n, 3)$, avec pour première colonne des 1, pour deuxième colonne les Z_i et pour troisième colonne les Z_i^2 . Son rang peut se déterminer comme la taille maximale d'une sous-matrice carrée de X de déterminant non nul. Or, pour que X soit régulière, il faut qu'il existe une sous-matrice carrée de taille 3 de déterminant non nul. Or un tel déterminant est de la forme :

$$\begin{vmatrix} 1 & Z_{i_1} & Z_{i_1}^2 \\ 1 & Z_{i_2} & Z_{i_2}^2 \\ 1 & Z_{i_3} & Z_{i_3}^2 \end{vmatrix}, \quad \text{avec } i_1, i_2 \text{ et } i_3 \text{ distincts dans } \{1, \dots, n\}.$$

Donc si les Z_i ne prennent que deux valeurs distinctes, tous ses déterminants seront nuls, et la matrice X sera de rang 2, donc non régulière.

ii. Sous ces hypothèses, il est clair qu'avec les notations de la question précédente, et en choisissant les i_1, i_2 et i_3 comme les indices des Z_i distincts, alors le déterminant est non nul, et la matrice X est donc de rang 3.

iii. De la même manière, si la régression est polynômiale de degré k , il faut qu'il existe $k+1$ valeurs distinctes des Z_i pour que la matrice X soit de rang plein. En particulier, si $k = n-1$, toutes les valeurs des Z_i doivent être distinctes et le polynôme obtenu par moindres carrés n'est autre que le polynôme d'interpolation de Lagrange passant par tous les points (Z_i, Y_i) .

Exercice 4.3

i. Le modèle de régression nous permet d'écrire $Y_{n+1} = \theta_0 + \theta_1 \cdot Z_{n+1}^{(1)} + \dots + \theta_k \cdot Z_{n+1}^{(k)} + \varepsilon_{n+1}$. Si l'on peut encore supposer les postulats **P1-3** sur ε_{k+1} , on utilisera comme valeur de prédiction pour Y_{n+1} ,

$$\hat{Y}_{n+1} = \hat{\theta}_0 + \hat{\theta}_1 \cdot Z_{n+1}^{(1)} + \dots + \hat{\theta}_k \cdot Z_{n+1}^{(k)}.$$

On a donc $\hat{Y}_{n+1} = X_{n+1} \cdot \hat{\theta}$ avec $X_{n+1} = (1, Z_{n+1}^{(1)}, \dots, Z_{n+1}^{(k)})$. Par conséquent, on obtient que $\text{Var}(\hat{Y}_{n+1}) = X_{n+1} \cdot \text{Cov}(\hat{\theta}) \cdot X_{n+1}' = \sigma^2 \cdot X_{n+1} \cdot (X' \cdot X)^{-1} \cdot X_{n+1}'$.

Comme σ^2 est inconnue, on utilisera la variance estimée, soit $\hat{\sigma}^2 = \frac{1}{n-k-1} \|Y - X \cdot \hat{\theta}\|^2$, et ainsi la variance "empirique" de \hat{Y}_{n+1} sera :

$$\hat{\sigma}^2 \cdot X_{n+1} \cdot (X' \cdot X)^{-1} \cdot X_{n+1}'.$$

ii. Le choix de p n'a aucune importance pour le calcul de $\hat{\theta}_\Omega$ où Ω est la matrice des poids. En effet, on a $\hat{\theta}_\Omega = (X' \cdot \Omega^{-1} \cdot X)^{-1} \cdot X' \cdot \Omega^{-1} \cdot Y$, et p^{-1} apparaît en haut et en bas dans cette expression, donc s'annule. En revanche, pour le calcul de $\hat{\sigma}_\Omega^2$, il est naturel d'imposer que la somme des poids fasse n (comme pour les poids de la matrice identité). Alors pour $\beta \neq 1$, on a $p \cdot \sum_{i=1}^n \beta^{n-i} = p(1 - \beta^n)\beta(1 - \beta)^{-1} = n$ donc $p = n(1 - \beta)((1 - \beta^n)\beta)^{-1}$. Si $\beta = 1$, alors $p = 1$

et on retrouve les moindres carrés "ordinaires".

On obtient maintenant comme prédiction de Y_{n+1} :

$$\widehat{Y}_{n+1}^{(\beta)} = X_{n+1} \cdot \widehat{\theta}_\Omega,$$

et $\text{Var}(\widehat{Y}_{n+1}) = \sigma^2 X_{n+1} \cdot (X' \cdot \Omega \cdot X)^{-1} \cdot X' \cdot \Omega^2 \cdot X \cdot (X' \cdot \Omega \cdot X)^{-1} \cdot X'_{n+1}$ et on approchera cette variance par :

$$\widehat{\sigma}_\Omega^2 X_{n+1} \cdot (X' \cdot \Omega \cdot X)^{-1} \cdot X' \cdot \Omega^2 \cdot X \cdot (X' \cdot \Omega \cdot X)^{-1} \cdot X'_{n+1},$$

où $\widehat{\sigma}_\Omega^2 = \frac{1}{n-k-1} (Y - X \cdot \widehat{\theta}_\Omega)' \cdot \Omega \cdot (Y - X \cdot \widehat{\theta}_\Omega)$.

Exercice 4.4

(i) On a : $\arcsin(x) = \arcsin(x_0) + \frac{x - x_0}{\sqrt{1 - x_0^2}} + o(x - x_0)$.

(ii) On développe autour de μ_i pour obtenir :

$$Z_{ij} = \arcsin(\mu_i) + \varepsilon_{ij}.$$

Dans notre modèle Y_{ij} est de loi $\text{Bin}(n, \mu_i)$. Par le théorème central limite sa loi peut être approchée par (4.1).

Exercice 4.5

(i) Le nombre de runs est égal à 1 plus le nombre de changements de signe. Chaque Z_i marque un changement de signe entre l'étape i et $i + 1$.

(ii) $\mathbb{E}(Z_i) = \mathbb{P}\{Y_i > 0, Y_{i+1} < 0\} + \mathbb{P}\{Y_i < 0, Y_{i+1} > 0\} + 1/4 + 1/4 = 1/2$.

(iii) Si $|i - j| > 1$, Z_i et Z_j dépendent des couples de Y différents donc indépendants, ils héritent donc de la propriété d'indépendance.

Les Z_i et Z_{i+1} ne peuvent prendre que 4 valeurs différentes. Il est facile vérifier qu'elles ont toutes probabilité 1/4, ce qui donne l'indépendance. Donc $\text{Var}(R) = (n - 1)/4$.

(iv) Tous calculs faits, il suffit de considérer

$$\frac{R - (n + 1)/2}{\sqrt{(n - 1)/4}}.$$

Exercice 4.6

i. Le modèle s'écrit donc $Y_i = a_1 \cdot g_1(i) + \dots + a_k \cdot g_k(i) + \varepsilon_i$ pour $i = 1, \dots, n \cdot T$. On est bien dans le cadre du modèle linéaire, et en appelant X la matrice $(g_j(i))_{1 \leq i \leq n \cdot T, 1 \leq j \leq k}$, on a

$$\widehat{t}(x) = (g_1(x), \dots, g_k(x)) \cdot (X' \cdot X)^{-1} \cdot X' \cdot Y.$$

On peut alors définir les résidus $\widehat{\varepsilon}_i = Y_i - \widehat{t}(i)$, ce qui permet d'obtenir l'estimateur de σ^2 ,

$$\widehat{\sigma}_1^2 = \frac{1}{n - k} \sum_{i=1}^{n \cdot T} (Y_i - \widehat{t}(i))^2.$$

ii. On a $Y_{i.T+k} = s_k + \varepsilon_{i.T+k}$ pour $i = 1, \dots, (n-1)$ et $k = 1, \dots, T$. Considérons que s_1, \dots, s_{T-1} sont les paramètres à estimer, et que $s_n = -s_1 - \dots - s_{T-1}$. Le modèle peut alors s'écrire :

$$Y = X \cdot \theta_s + \varepsilon, \text{ avec } Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \\ Y_{n+1} \\ \vdots \\ Y_{n.T} \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ -1 & -1 & \dots & -1 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ -1 & -1 & \dots & -1 \end{pmatrix}, \theta_s = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_{T-1} \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \\ \varepsilon_{n+1} \\ \vdots \\ \varepsilon_{n.T} \end{pmatrix}$$

Alors une estimation $\hat{\theta}_s = (\hat{s}_1, \dots, \hat{s}_{T-1})'$ sera donc $\hat{\theta}_s = (X' \cdot X)^{-1} \cdot X' \cdot Y$. Or

$$X' \cdot X = n \begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & \dots & 2 \end{pmatrix}, \quad (X' \cdot X)^{-1} = \frac{1}{n \cdot T} \begin{pmatrix} T-1 & -1 & \dots & -1 \\ -1 & T-1 & \dots & -1 \\ \vdots & \vdots & \dots & \vdots \\ -1 & -1 & \dots & T-1 \end{pmatrix}$$

$$\text{et } X' \cdot Y = \begin{pmatrix} Y_1 + Y_{T+1} + \dots + Y_{(n-1)T+1} - (Y_T + Y_{2T} + \dots + Y_{nT}) \\ Y_2 + Y_{T+2} + \dots + Y_{(n-1)T+2} - (Y_T + Y_{2T} + \dots + Y_{nT}) \\ \vdots \\ Y_{T-1} + Y_{2T-1} + \dots + Y_{nT-1} - (Y_T + Y_{2T} + \dots + Y_{nT}) \end{pmatrix}.$$

Maintenant, le calcul de $\hat{\theta}_s$ conduit à la formule demandée pour les \hat{s}_i . On a bien-sûr la formule habituelle pour l'estimation de la variance, soit :

$$\hat{\sigma}_2^2 = \frac{1}{n - T + 1} \|Y - X \cdot \hat{\theta}\|^2.$$

iii. On peut encore poser un modèle linéaire, soit $Y = X \cdot \theta + \varepsilon$ avec les mêmes expressions pour Y et ε que précédemment, mais avec :

$$X = \begin{pmatrix} g_1(1) & \dots & g_k(1) & 1 & 0 & \dots & 0 \\ g_1(2) & \dots & g_k(2) & 0 & 1 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots \\ g_1(T) & \dots & g_k(T) & -1 & -1 & \dots & -1 \\ g_1(T+1) & \dots & g_k(T+1) & 1 & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots \\ g_1(nT) & \dots & g_k(nT) & -1 & -1 & \dots & -1 \end{pmatrix}, \quad \text{et } \theta = \begin{pmatrix} a_1 \\ \vdots \\ a_k \\ s_1 \\ \vdots \\ s_{T-1} \end{pmatrix}.$$

On a toujours $\hat{\theta}_{st} = (X' \cdot X)^{-1} \cdot X' \cdot Y$, ce qui conduit à $\hat{t}(x) = (g_1(x), \dots, g_k(x), 0, \dots, 0) \cdot \hat{\theta}_{st}$ et à $\hat{s}(i) = (0, \dots, 0, 0, \dots, 1, 0, \dots, 0) \cdot \hat{\theta}_{st}$ pour $i = 1, \dots, T-1$ (et une nouvelle fois, $\hat{s}(T) = -\hat{s}(1) - \dots - \hat{s}(T-1)$). Ainsi \hat{s} dépend des $(g_j(k))$, même si c'est également une application linéaire des (Y_i) .

Maintenant, si on utilise $\hat{\theta}_s$ pour estimer d'abord s et qu'ensuite, on travaille avec $Y_i - \hat{s}(i)$ pour estimer $t(x)$, alors on ne pourra obtenir la même chose que si la tendance est une constante. En effet, tout ceci revient à résoudre l'équation : $P_{[S+T]}(X) = (P_{[T]} \circ P_{[S]^\perp} + P_{[S]})(X)$. avec $[S]$ et $[T]$ les sous-espace vectoriels de \mathbb{R}^{nT} correspondant respectivement à la partie composante saisonnière de dimension $T-1$ et à la partie tendance de dimension k , et $P_{[S]^\perp}$ la projection sur le sous-espace orthogonal de $[S]$. Si $[T]$ est orthogonal à $[S]$, ce qui est le cas lorsque $k=0$ et $g_1 = C$ (constante), il est facile de voir que $P_{[S+T]} = P_{[T]} \circ P_{[S]^\perp} + P_{[S]}$. En revanche, si $[T] \cap [S] = R$,

avec R de dimension ≥ 1 , si $[T]$ et $[R]$ ne sont pas orthogonaux, et si on choisit $X \in [T] \setminus [S]$, alors $P_{[S+T]}(X) = X$ alors que $P_{[T]} \circ P_{[S]^\perp} + P_{[S]}(X)$ n'aura pas une intersection vide avec $[S]$, donc sera différent de X . Ainsi cette méthode en deux temps n'équivaut pas à celle en une seule étape d'une manière générale.

Exercice 4.7

i. Il est clair que le polynôme de Lagrange, $P(x) = y_1 \cdot P_1(x) + \dots + y_n \cdot P_n(x)$ vérifie la propriété. Supposons qu'il existe un deuxième polynôme $Q(x)$ vérifiant cette même propriété. Alors, le polynôme $R(x) = P(x) - Q(x)$ est de degré au plus $n - 1$ et admet n racines distinctes, les x_i : il est donc nul et P est bien unique.

Le problème d'une telle estimation est qu'elle aboutit à un polynôme de degré élevé qui interdit toute bonne propriété de prédiction (ou même de lissage).

Comme on a pu le voir à l'exercice 2, une régression polynomiale de degré $n - 1$ de la forme $y_i = \theta_0 + \theta_1 \cdot x_1 + \dots + \theta_{n-1} \cdot x_1^{n-1} + \varepsilon_i$ amènerait à un polynôme passant par tous les x_i : ce serait donc également le polynôme d'interpolation de Lagrange.

ii. Montrons par récurrence que cette propriété est vraie. Choisissons par exemple que les conditions aux bords s'expriment en terme de dérivées d'ordre 1.

• Au rang 2, c'est-à-dire avec un seul intervalle $[x_1, x_2]$, existe-t-il un polynôme de degré 3 passant par (x_1, y_1) et (x_2, y_2) et tel que $f'(x_1) = y'_1$ et $f'(x_2) = y'_2$, avec y_1 et y_2 arbitraires ? La réponse est oui, car si l'on pose $P(x) = a_3x^3 + a_2x^2 + a_1x + a_0$ alors on vérifiera le système :

$$M(x_1, x_2) \cdot A = Y, \quad \text{avec} \quad M(x_1, x_2) = \begin{pmatrix} x_1^3 & x_1^2 & x_1 & 1 \\ 3x_1^2 & 2x_1 & 1 & 0 \\ x_2^3 & x_2^2 & x_2 & 1 \\ 3x_2^2 & 2x_2 & 1 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} a_3 \\ a_2 \\ a_1 \\ a_0 \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y'_1 \\ y_2 \\ y'_2 \end{pmatrix}.$$

On montre, après quelques soustractions entre lignes, que $\det(M(x_1, x_2)) = (x_1 - x_2)^4$. Donc, dès que $x_1 \neq x_2$, il existe bien une unique solution au système, donc un unique polynôme.

• Supposons que la proposition est vraie au rang n , donc sur les intervalles $[x_1, x_2], \dots, [x_{n-1}, x_n]$. Montrons qu'elle est alors vraie au rang $n + 1$. Pour les $n - 1$ premiers intervalles, si on fixe une valeur y'_n pour la dérivée en x_n , on a, d'après la propriété de récurrence, une unique suite de polynômes ainsi construite. La valeur de sa dérivée seconde en x_n est alors y''_n , qui vaut $6a_3x_n + 2$, avec les notations précédentes, soit encore

$$6 \cdot (1, 0, 0, 0) \cdot M^{-1}(x_{n-1}, x_n) \cdot (y_{n-1}, y'_{n-1}, y_n, y'_n)' \cdot x_n + 2,$$

avec y'_{n-1} ne dépendant que des x_i , des y_i et de y'_1 . De plus sur l'intervalle $[x_n, x_{n+1}]$, pour chaque y'_n choisi, on a deux conditions fixées pour les dérivées en x_n et x_{n+1} , donc un unique polynôme de degré 3. Ce polynôme admet une dérivée seconde en x_n , qui vaut cette fois-ci :

$$6 \cdot (1, 0, 0, 0) \cdot M^{-1}(x_{n+1}, x_n) \cdot (y_{n+1}, y'_{n+1}, y_n, y'_n)' \cdot x_n + 2, \quad (1)$$

où y_n, y_{n+1} et y'_{n+1} sont fixés. Donc si $x_n = 0$, la condition de continuité de la dérivée seconde est vérifiée. Si $x_n \neq 0$, comme $(1, 0, 0, 0) \cdot M^{-1}(x_{n+1}, x_n) \neq (1, 0, 0, 0) \cdot M^{-1}(x_{n-1}, x_n)$, alors l'équation (1) peut s'écrire $ay'_n + b = a'y'_n + b'$ avec $a \neq a'$, équation qui admet bien une unique solution. En conséquence la propriété est vraie au rang $n + 1$, donc vraie pour tout $n \geq 2$.

iii. Quand λ est proche de 0, alors la pénalisation, qui pénalise les brusques variations de la fonction

f , n'intervient plus, et la fonction \hat{f} ainsi trouvée se rapproche de la suite de polynômes de splines interpolateurs (passant donc par tous les points).

Quand λ devient grand, c'est le critère de pénalisation qui intervient donc principalement. Or minimiser ce critère uniquement revient à considérer une fonction affine. La fonction \hat{f} alors obtenue est une suite de polynômes de degré 3 se rapprochant de la droite des moindres carrés "ordinaires".

Les conditions aux bords (en x_1 et en x_n) sont fixées, c'est-à-dire que les dérivées première et seconde en ces points sont nulles. Les conditions aux bords et de raccordements fournissent au total $4 + 3 \cdot (n - 2)$ équations linéaires, sachant que l'on doit déterminer en tout $4 \cdot (n - 1)$ paramètres.

Enfin, on peut considérer $S_\lambda(f)$ comme $\|y - f\|^2$, où $\|f\|^2 = \frac{1}{n} \sum_{i=1}^n f^2(x_i) + \int_{-\infty}^{\infty} \left| \frac{\partial^2 f}{\partial x^2}(x) \right|^2 dx$. On peut donc associer le produit scalaire :

$$\langle f, g \rangle = \frac{1}{n} \sum_{i=1}^n f(x_i)g(x_i) + \int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial x^2}(x) \cdot \frac{\partial^2 g}{\partial x^2}(x) dx$$

(la démonstration du fait que ce soit un produit scalaire est aisée).

5 Corrections des exercices du chapitre 5

Exercice 5.1

On obtient : $\hat{\mu} = 10$, $\hat{\alpha}_1 = -\hat{\alpha}_2 = 2$, $\hat{\beta}_1 = -\hat{\beta}_2 = 3$, $\hat{\gamma}_{11} = -\hat{\gamma}_{12} = -\hat{\gamma}_{21} = \hat{\gamma}_{22} = 1$.

Exercice 5.2

On a

$$-2 \log(\text{vraisemblance}) = Cte + \sum_{i=1}^I n_i \log(\sigma_i^2) + \sum_{j=1}^{n_i} \frac{(Y_{ij} - \mu_i)^2}{\sigma_i^2}.$$

La maximisation se fait donc i par i . On peut donc se ramener au cas classique où $I = 1$ et on trouve

$$\hat{\mu}_i = Y_{i.} \quad ; \quad \hat{\sigma}_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - Y_{i.})^2.$$

Si on pose $\mu_i = \mu$ on a

$$-2 \log(\text{vraisemblance}) = Cte + \sum_{i=1}^I n_i \log(\sigma_i^2) + \sum_{j=1}^{n_i} \frac{(Y_{ij} - \mu)^2}{\sigma_i^2}.$$

La maximisation en σ_i^2 donne encore

$$\hat{\sigma}_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - Y_{i.})^2.$$

On réintroduit cette valeur dans la vraisemblance pour obtenir

$$-2 \log(\text{vraisemblance}) = Cte + \sum_{i=1}^I \left(n_i \log\left(\frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - Y_{i.})^2\right) + n_i \right) .$$

En dérivant par rapport à μ on obtient au terme du calcul

$$\sum_{i=1}^I \frac{Y_{i.} - \mu}{\hat{\sigma}_i^2} = 0.$$

Ce qui donne :

$$\hat{\mu} = \frac{\sum_{i=1}^I Y_{i.} / \hat{\sigma}_i^2}{\sum_{i=1}^I 1 / \hat{\sigma}_i^2}.$$

Exercice 5.3

$\alpha_i = 0 \Rightarrow \theta_{ij} = \mu + \beta_j + \gamma_{ij} \Rightarrow \theta_{i.} = \mu$ donc constant
 $\theta_{i.} = (cte) \Rightarrow \alpha_i = \theta_{i.} - \theta_{..} = 0.$

Exercice 5.4

Par cohérence avec la décomposition marginale utilisée dans le cas où il y a deux facteurs, on serait tenté de poser la première contrainte. Ce n'est pas la bonne réponse. En effet :

- L'estimation et donc la définition précise du paramètre μ importe peu, puisque ce paramètre de moyenne générale est souvent sans intérêt. N'oublions pas que le but d'une expérience est de comparer ; toute l'attention est donc focalisée sur les effets différentiels α_i . Le premier type de contraintes a peu d'intérêt.
- Le second type n'a d'autre intérêt que calculatoire. Si on l'utilise, l'estimateur $\hat{\mu}$ de μ est la moyenne générale : $Y_{..}$ (estimateur sous l'hypothèse nulle), nécessaire pour la construction du test de Fisher. En particulier, la formule $SC = \sum_{ik} (Y_{ik} - Y_{..})^2 = \sum_{ik} (\hat{\alpha}_i)^2$ ne serait pas vraie sinon.

Cette réponse est en contradiction avec celle que l'on fait pour deux facteurs. Nous voyons donc une fois de plus sur cet exemple, l'intérêt de travailler avec des modèles réguliers, comme l'est le modèle de notre première présentation.

Exercice 5.5

Comme dans l'exercice 5.3 on montre l'équivalence entre

$$\forall i, \alpha_i = 0 \text{ et } \sum_{j=1}^J \theta_{ij} = (cte).$$

Pour définir le test de type I du premier facteur on introduit l'espace

$$F := \{Y_{ijk} \in E : Y_{ijk} = a_i; \sum_{i=1}^I n_{i+} a_i = 0\}$$

Le test correspond à l'hypothèse nulle : $P_F(X.\theta) = 0$ ce qui est équivalent à "pour tout a_i vérifiant $\sum_{i=1}^I n_{i+} a_i = 0$ on a $\sum_{i,j} \theta_{ij} n_{ij} a_i = 0$ ". Cette condition implique

$$\sum_i a_i \sum_j \theta_{ij} n_{ij} = 0 = \sum_i a_i n_{i+} (\mu + \alpha_i).$$

Ce qui implique - puisque c'est la seule relation vérifiée par les a_i - que $\mu + \alpha_i$ est constant, donc que tous les α_i sont nuls.

La réciproque est directe.

Exercice 5.6

Avec les notations du chapitre :

$$-12 = \hat{\theta}_{11} - \hat{\theta}_{12} - \hat{\theta}_{21} - \hat{\theta}_{22} = 4 - 28 - 12 + 24 = -12.$$

7 Corrections des exercices du chapitre 7

Exercice 7.1

Soit X matrice d'ordre (n, k) avec $k \leq n$. On peut donc associer à X une application linéaire $u : \mathbb{R}^k \mapsto \mathbb{R}^n$ telle que X soit la matrice de u dans deux bases de \mathbb{R}^k et de \mathbb{R}^n . De même, à X' on peut également associer une application linéaire $v : \mathbb{R}^n \mapsto \mathbb{R}^k$. On rappelle qu'alors $X' \cdot X$ est la matrice de l'application $v \circ u$ (dans les bases adéquates). Ainsi :

$$\begin{aligned} X \text{ est de rang } k &\iff \dim(\text{Im}(u)) = k \text{ et } \dim(\text{Ker}(u)) = 0 \\ &\iff X' \text{ est de rang } k, \text{ donc } \dim(\text{Im}(v)) = k \text{ et } \dim(\text{Ker}(v)) = n - k \\ &\iff \text{Ker}(v) = [X]^\perp, \text{ et ainsi } \text{Ker}(v) \oplus \text{Im}(v) = \mathbb{R}^n \\ &\iff \text{Im}(v \circ u) = v(\text{Im}(u)) = \text{Im}(v) \\ &\iff \dim(\text{Im}(v \circ u)) = \dim(\text{Im}(v)) = k \\ &\iff \text{le rang de } X' \cdot X \text{ est } k, \text{ pour une matrice carrée de taille } k \\ &\iff X' \cdot X \text{ est une matrice inversible} \end{aligned}$$

Correction alternative : X est une matrice d'ordre (n, k) avec $k \leq n$, elle est donc de plein rang si et seulement si elle est injective, c'est-à-dire que son noyau est réduit à l'élément nul. Or $X' \cdot X$ est une matrice carrée de taille k , elle est donc inversible si et seulement si son noyau est réduit à l'élément nul. Maintenant pour tout $t \in \mathbb{R}^k$,

$$t' \cdot X' \cdot X \cdot t = 0 \iff \|X \cdot t\| = 0 \iff X \cdot t = 0.$$

Donc $X \cdot t = 0$ est équivalent à $X' \cdot X \cdot t = 0$ et donc les deux matrices ont même noyau.

Exercice 7.2

Pour le modèle additif la condition est identique puisque ce qui compte c'est l'orthogonalité entre les espaces notés $[\alpha]$ et $[\beta]$.

Pour ce qui concerne le modèle hiérarchisé :

$$Y_{ijk} = \mu + \alpha_i + \gamma_{ij} + \varepsilon_{ijk}$$

les contraintes :

$$\sum_{i=1}^I n_i \alpha_i = 0 \quad ; \quad \forall I; \sum_{j=1}^j n_{ij} = 0$$

donnent toujours l'orthogonalité.

8 Corrections des exercices du chapitre 8

Exercice 8.1

On a donc le modèle $Y_{ij} = \mu_i + \varepsilon_{ij}$, avec $i = 1, \dots, I$ et $j = 1, \dots, K$ et $n = K \cdot I$. Sous l'hypothèse **H**, les ε_{ij} sont des v.a.i.i.d. centrées de variance σ^2 . Or $\hat{\theta}^n = (Y_1, \dots, Y_I)'$, avec $Y_i = \frac{1}{K} \sum_{k=1}^K Y_{ik} = \frac{1}{K} \sum_{k=1}^K (\mu_i + \varepsilon_{ik})$. Le vecteur $(Y_i)_{1 \leq i \leq K}$ est constitué de moyennes empiriques de v.a.i.i.d. d'espérance μ_i et de variance σ^2 . On est bien sous les hypothèses du Théorème de la Limite Centrale multidimensionnel, et ainsi :

$$\sqrt{K} \left(\hat{\theta}^n - (\mu_1, \dots, \mu_I)' \right) \xrightarrow{K \rightarrow +\infty} \mathcal{N}_I((0, \dots, 0)', \sigma^2 \cdot I_I).$$

On peut également écrire que $\hat{\sigma}^{2,n} = \frac{1}{n-I} \sum_{i=1}^I \sum_{k=1}^K (Y_{ik} - Y_i)^2 = \frac{1}{I} \sum_{i=1}^I \frac{1}{K-1} \sum_{k=1}^K (Y_{ik} - Y_i)^2$. Or chaque élément de cette somme est indépendants des autres. De plus

$$\frac{1}{K-1} \sum_{k=1}^K (Y_{ik} - Y_i)^2 = \frac{1}{K-1} \left(-K \cdot (Y_i - \mu_i)^2 + \sum_{k=1}^K (Y_{ik} - \mu_i)^2 \right).$$

Or d'après la Loi Forte des Grands Nombres, $\frac{K}{K-1} (Y_i - \mu_i)^2 \xrightarrow{K \rightarrow +\infty} 0$. De plus, toujours d'après la Loi Forte des Grands Nombres, $\frac{1}{K-1} \sum_{k=1}^K (Y_{ik} - \mu_i)^2 = \frac{K}{K-1} \frac{1}{K} \sum_{k=1}^K \varepsilon_{ik}^2$ converge p.s. vers σ^2 comme moyenne empirique des $(\varepsilon_{ik}^2)_k$ qui sont des variables i.i.d. d'espérance σ^2 (le terme en $\frac{K}{K-1}$ convergeant vers 1). En conséquence, $\hat{\sigma}^{2,n} \xrightarrow{K \rightarrow +\infty} \sigma^2$.

Exercice 8.2

On a donc le modèle $Y_i = \mu + \beta \cdot Z_i + \varepsilon_i$ pour $i = 1, \dots, n$ que l'on analyse avec un modèle quadratique. Aussi obtient on l'estimateur $\hat{\theta} = (X^{(2)' \cdot X^{(2)}})^{-1} \cdot X^{(2)' \cdot Y}$, avec $X^{(2)}$ la matrice de colonnes $(1, (Z_i)$ et (Z_i^2) (de la même manière, on notera $X^{(1)}$ la matrice de colonnes $(1, (Z_i))$, et l'estimateur $\hat{\gamma}^{2,n} = \frac{1}{n-3} \|Y - X^{(2)} \cdot \hat{\theta}\|^2$. Donc $X^{(2)} \cdot \hat{\theta} = P_{[X^{(2)}]} Y$, projeté orthogonal sur un espace plus grand.

- On a $\mathbb{E}(\hat{\theta}_3^n) = (0, 0, 1) \cdot (X^{(2)' \cdot X^{(2)}})^{-1} \cdot X^{(2)' \cdot (X^{(2)} \cdot (\mu, \beta, 0)' + \varepsilon)$, donc $\mathbb{E}(\hat{\theta}_3^n) = (0, 0, 1) \cdot (\mu, \beta, 0)' = 0$.
- On a donc $\hat{\theta}_3^n = (0, 0, 1) \cdot (X^{(2)' \cdot X^{(2)}})^{-1} \cdot X^{(2)' \cdot \varepsilon$. Ainsi,

$$\begin{aligned} \text{Var}(\hat{\theta}_3^n) &= (0, 0, 1) \cdot (X^{(2)' \cdot X^{(2)}})^{-1} \cdot X^{(2)' \cdot \sigma^2 \cdot I_n \cdot X^{(2)} \cdot (X^{(2)' \cdot X^{(2)}})^{-1} \cdot (0, 0, 1)' \\ &= \sigma^2 \cdot (0, 0, 1) \cdot (X^{(2)' \cdot X^{(2)}})^{-1} \cdot (0, 0, 1)'. \end{aligned}$$

Ainsi $\hat{\theta}_3^n$ ne converge pas plus vers 0, que le troisième paramètre d'une régression quadratique classique : tout dépend de $X^{(2)}$.

• en ce qui concerne $\hat{\gamma}^{2,n}$, on a

$$\begin{aligned}\hat{\gamma}^{2,n} &= \frac{1}{n-3} \|Y - P_{[X^{(2)}]}Y\|^2 \\ &= \frac{1}{n-3} \|P_{[X^{(2)}]^\perp} \varepsilon\|^2 \\ &\stackrel{\mathcal{L}}{\sim} \frac{\sigma^2}{n-3} \chi^2(n-3).\end{aligned}$$

Par suite, $\hat{\gamma}^{2,n} \xrightarrow[n \rightarrow +\infty]{\mathcal{P}} \sigma^2$ (la seule différence en ce qui concerne l'estimation de la variance dans le fait d'avoir sur-ajuster, réside dans la perte d'un degré de liberté pour la loi de cet estimateur, et donc une convergence un tout petit peu plus lente).

• On a $\hat{Y} = (1, z, z^2) \cdot \hat{\theta}$, donc $\mathbb{E}(\hat{Y}) = \mu + \beta \cdot z$ et ainsi, $\text{Var}(\hat{Y}) = \sigma^2 \cdot (1, z, z^2) \cdot (X^{(2)' \cdot X^{(2)}})^{-1} \cdot (1, z, z^2)'$. Si le vrai modèle avait été connu, alors $\text{Var}(\hat{Y}) = \sigma^2 \cdot (1, z) \cdot (X^{(1)' \cdot X^{(1)}})^{-1} \cdot (1, z)'$: si z est grand cela peut conduire à des différences importantes.

Exercice 8.3

On utilise ici les notations de l'exercice précédent. On a donc le modèle $Y = X^{(2)} \cdot \theta + \varepsilon$, et on suppose que $\theta_3 \neq 0$. L'estimateur de la variance des (ε_i) est donc $\hat{\sigma}^2 = \frac{1}{n-2} \|Y - P_{[X^{(1)}]}Y\|^2$. Or

$$\mathbb{E}(P_{[X^{(1)}]}Y) = \mathbb{E}\left(X^{(1)} \cdot (X^{(1)' \cdot X^{(1)}})^{-1} \cdot X^{(1)' \cdot X^{(2)}} \cdot \theta\right).$$

Or $X^{(1)' \cdot X^{(1)}} = \begin{pmatrix} n & n(n+1)/2 \\ n(n+1)/2 & n(n+1)(2n+1)/6 \end{pmatrix}$, donc

$$(X^{(1)' \cdot X^{(1)}})^{-1} = \frac{12}{n(n^2-1)} \begin{pmatrix} (n+1)(2n+1)/6 & -(n+1)/2 \\ -(n+1)/2 & 1 \end{pmatrix}.$$

De plus,

$$X^{(1)' \cdot X^{(2)}} = \begin{pmatrix} n & n(n+1)/2 & n(n+1)(2n+1)/6 \\ n(n+1)/2 & n(n+1)(2n+1)/6 & n^2(n+1)^2/4 \end{pmatrix}$$

On en arrive alors à :

$$(X^{(1)' \cdot X^{(1)}})^{-1} \cdot X^{(1)' \cdot X^{(2)}} = \begin{pmatrix} 1 & 0 & -(n+1)(n+2)/6 \\ 0 & 1 & (n+1) \end{pmatrix}$$

(on pouvait se douter de voir apparaître la matrice identité de taille 2 car les deux premières colonnes de $X^{(2)}$ constituent $X^{(1)}$. Par conséquent,

$$\mathbb{E}(P_{[X^{(1)}]}Y) = \left(\theta_1 + \theta_2 \cdot i + \theta_3 \cdot \frac{(n+1)}{6} \cdot (6i - (n+2))\right)_{1 \leq i \leq n},$$

et donc :

$$\hat{\sigma}^2 = \frac{1}{n-2} \left\| \theta_3 \cdot \left(i^2 - \frac{(n+1)}{6} \cdot (6i - (n+2))\right)_{1 \leq i \leq n} + P_{[X^{(1)}]^\perp} \varepsilon \right\|^2.$$

Les résidus sont donc fortement biaisés, avec un terme de biais d'ordre n^2 pour i proche de 0 ou i proche de n . Ceci conduit à ce que l'on obtienne :

$$\hat{\sigma}^2 \xrightarrow[n \rightarrow +\infty]{\mathcal{P}} +\infty.$$

En utilisant le modèle d'analyse, on a : $\widehat{Y}_{n+1} = (1 \quad (n+1)) \cdot (X^{(1)'} \cdot X^{(1)})^{-1} \cdot X^{(1)'} \cdot Y$, alors que $Y_{n+1} = \theta_1 + (n+1) \cdot \theta_2 + (n+1)^2 \cdot \theta_3 + \varepsilon_{n+1}$. De tout ce qui précède, on peut encore écrire que :

$$\begin{aligned}\widehat{Y}_{n+1} &= \begin{pmatrix} 1 & (n+1) \end{pmatrix} \cdot \left[\begin{pmatrix} \theta_1 - \theta_3 \cdot (n+1)(n+2)/6 \\ \theta_2 + \theta_3 \cdot (n+1) \end{pmatrix} + (X^{(1)'} \cdot X^{(1)})^{-1} \cdot X^{(1)'} \cdot \varepsilon \right] \\ &= \theta_1 + (n+1) \cdot \theta_2 + \frac{1}{6} (n+1)(5n+4) \cdot \theta_3 + \frac{1}{6} (n+1) \sum_{i=1}^n (3i - n - 2) \varepsilon_i.\end{aligned}$$

Par suite,

$$\widehat{Y}_{n+1} - Y_{n+1} = -\frac{1}{6} (n+1)(n+2) \cdot \theta_3 + \left(\frac{1}{6} (n+1) \sum_{i=1}^n (3i - n - 2) \varepsilon_i - \varepsilon_{n+1} \right).$$

Finalement, on montre ainsi que $\mathbb{E}(\widehat{Y}_{n+1} - Y_{n+1}) \xrightarrow[n \rightarrow +\infty]{} +\infty$ ou $-\infty$ (suivant le signe de θ_3) et

$\text{Var}(\widehat{Y}_{n+1} - Y_{n+1}) \xrightarrow[n \rightarrow +\infty]{} +\infty$. Donc $|\widehat{Y}_{n+1} - Y_{n+1}| \xrightarrow[n \rightarrow +\infty]{\mathcal{P}} +\infty$.

11 Corrections des exercices du chapitre 11

Exercice 11.1

On a obtenu

$$\text{Var}(\varepsilon') = \sigma^2 I_n + \tau^2 (I_n - J_n/n),$$

où J_n est la matrice carrée de taille n composée de 1.

Comme $(Y_i - Y_{i'})$ est un "contraste", c'est-à-dire une combinaison linéaire de Y de somme de poids nulle, la contribution de J_n s'annule et on obtient :

$$\text{Var}(Y_i - Y_{i'}) = \frac{2}{r} (\sigma^2 + \tau^2).$$

Exercice 11.2

En appliquant la définition, on obtient (avant randomisation...)

Blocs	Trait. dans unité 1	Trait. dans unité 2	Trait. dans unité 3
Bloc 1 = répétition ∞ , bloc 0	$(0, 0) = \mathbf{1}$	$(1, 0) = \mathbf{4}$	$(2, 0) = \mathbf{7}$
Bloc 2 = répétition ∞ , bloc 1	$(0, 1) = \mathbf{2}$	$(1, 1) = \mathbf{5}$	$(2, 1) = \mathbf{8}$
Bloc 3 = répétition ∞ , bloc 2	$(0, 2) = \mathbf{3}$	$(1, 2) = \mathbf{6}$	$(2, 2) = \mathbf{9}$
Bloc 4 = répétition 0, bloc 0	$(0, 0) = \mathbf{1}$	$(0, 1) = \mathbf{2}$	$(0, 2) = \mathbf{3}$
Bloc 5 = répétition 0, bloc 1	$(1, 0) = \mathbf{4}$	$(1, 1) = \mathbf{5}$	$(1, 2) = \mathbf{6}$
Bloc 6 = répétition 0, bloc 2	$(2, 0) = \mathbf{7}$	$(2, 1) = \mathbf{8}$	$(2, 2) = \mathbf{9}$
Bloc 7 = répétition 1, bloc 0	$(0, 0) = \mathbf{1}$	$(1, 2) = \mathbf{6}$	$(2, 1) = \mathbf{8}$
Bloc 8 = répétition 1, bloc 1	$(1, 0) = \mathbf{4}$	$(0, 1) = \mathbf{2}$	$(2, 2) = \mathbf{9}$
Bloc 9 = répétition 1, bloc 2	$(2, 0) = \mathbf{7}$	$(1, 1) = \mathbf{5}$	$(0, 2) = \mathbf{3}$
Bloc 10 = répétition 2, bloc 0	$(0, 0) = \mathbf{1}$	$(1, 1) = \mathbf{5}$	$(2, 2) = \mathbf{9}$
Bloc 11 = répétition 2, bloc 1	$(1, 0) = \mathbf{4}$	$(0, 2) = \mathbf{3}$	$(2, 1) = \mathbf{8}$
Bloc 12 = répétition 2, bloc 2	$(0, 1) = \mathbf{2}$	$(1, 2) = \mathbf{6}$	$(2, 0) = \mathbf{7}$

Exercice 11.3

Il suffit de compléter le plan présenté dans le cours...

Exercice 11.4

Il suffit de le faire...

Exercice 11.5

(i) Il est direct de vérifier que le $2p$ ième élément vaut p . Comme n est supposé pair, le dernier élément vaut bien $n/2$.

(ii) Comme il y a n éléments il suffit de montrer qu'ils sont tous différents. Mais les éléments pairs vont de 1 à $n/2$ et les éléments impairs sont dans l'ordre : 0 (ou n), $(n-1), \dots, n/2+1$. Ils sont bien différents.

(iii) Les incréments $w_{i+1} - w_i$ valent dans l'ordre : $1, -2, 3, -4, \dots, (n-1)$. Les nombres négatifs sont pairs, les nombres positifs sont impairs. Ils sont donc tous différents et il y en a $(n-1)$ d'où le résultat.

(iv) Il suffit de le faire.

(v) Le rôle des lignes et colonnes est symétrique. Vérifions qu'à l'intérieur des lignes on trouve une fois et une seule chaque paire ordonnée $(a, b) \in (\mathbb{Z}/n\mathbb{Z})^2, a \neq b$. Une telle paire peut être également notée : $(a, a + \Delta), \Delta \neq 0$.

Δ est l'un des incréments $w_{i+1} - w_i$. Dans la colonne i on trouve l'image, par addition d'un terme, de la suite de Williams, on trouve donc une nouvelle permutation de $\mathbb{Z}/n\mathbb{Z}$. Il y a donc exactement une occurrence de a et l'élément voisin vaut $a + \Delta$.

12 Corrections des exercices du chapitre 12

Les exercices proposés sont essentiellement des développements relativement longs du chapitre. Il faut utiliser pas à pas la méthodologie de la section 12.3.3. On peut également consulter les articles en référence.

Dans l'exercice 12.1 les orbites doubles sont

- même unité,
- unités différentes, même ligne,
- unités différentes, même colonne,
- lignes et colonnes différentes.

La proposition 12.3 donne les strates.

Dans l'exercice 12.2, les orbites doubles sont

- même unité,
- unités différentes, même grande parcelle,
- grandes parcelles différentes, même bloc,
- blocs différents.

13 Corrections des exercices du chapitre 13

Exercice 13.1

Dans le cas de 3 facteurs, un élément de E peut être représenté par un tableau à 3 entrées, $2 \times 2 \times 2$. Nous allons l'écrire sous forme de deux tableaux 2×2 . Le premier facteur correspond aux lignes, le deuxième aux colonnes, le troisième sépare les deux tableaux. On a par exemple :

$$\begin{aligned}
 V_{\{1\}} &= \left\{ \begin{array}{|c|c|} \hline x & x \\ \hline y & y \\ \hline \end{array} ; \begin{array}{|c|c|} \hline x & x \\ \hline y & y \\ \hline \end{array} ; x, y \in \mathbb{R} \right\} \\
 V_{\{1,2\}} &= \left\{ \begin{array}{|c|c|} \hline x & y \\ \hline z & t \\ \hline \end{array} ; \begin{array}{|c|c|} \hline x & y \\ \hline z & t \\ \hline \end{array} ; x, y, z, t \in \mathbb{R} \right\} \\
 W_{\{1,2\}} &= \left\{ \begin{array}{|c|c|} \hline x & -x \\ \hline -x & x \\ \hline \end{array} ; \begin{array}{|c|c|} \hline x & -x \\ \hline -x & x \\ \hline \end{array} ; x \in \mathbb{R} \right\} \\
 W_{\{3\}} &= \left\{ \begin{array}{|c|c|} \hline x & x \\ \hline x & x \\ \hline \end{array} ; \begin{array}{|c|c|} \hline -x & -x \\ \hline -x & -x \\ \hline \end{array} ; x \in \mathbb{R} \right\} .
 \end{aligned}$$

Exercice 13.2

Il faut une résolution IV et donc le plan 2^{7-3} du tableau 13.1. Les relations complètes sont

$$1 = A \cdot B \cdot C \cdot E = B \cdot C \cdot D \cdot F = A \cdot C \cdot D \cdot G = A \cdot D \cdot E \cdot F = A \cdot B \cdot F \cdot G = B \cdot D \cdot E \cdot G = C \cdot E \cdot F \cdot G.$$

On a

$$A = B \cdot C \cdot E = A \cdot B \cdot C \cdot D \cdot F = C \cdot D \cdot G = D \cdot E \cdot F = B \cdot F \cdot G = A \cdot B \cdot D \cdot E \cdot G = A \cdot C \cdot E \cdot F \cdot G.$$

Exercice 13.3

Dans le premier cas, en comptant la moyenne générale il y a $1+5+4 = 10$ paramètres à estimer, on ne peut pas le faire avec 8 unités. Par contre avec 16, le plan 2^{5-1} du tableau 13.1 est de résolution V donc il convient.

Dans le second cas, le plan 2^{5-2} suivant, trouvé en tâtonnant, convient : B, C, D sont facteurs de base et les clefs sont $A = B \cdot C$, $E = C \cdot D$. Les relations complètes de définition sont

$$1 = A \cdot B \cdot C = C \cdot D \cdot E = A \cdot B \cdot D \cdot E$$

Exercice 13.4

Après recodage de A on obtient 6 facteurs. Le plan 2^{6-2} du tableau 13.1 a pour clefs, en changeant les noms : $D = A_1 \cdot A_2 \cdot B$; $E = A_2 \cdot B \cdot C$. Ce qui donne les relations complètes de définition suivantes :

$$1 = A_1 \cdot A_2 \cdot B \cdot D = A_2 \cdot B \cdot C \cdot E = A_1 \cdot C \cdot D \cdot E$$

La résolution n'est que de III! car $A_1 \cdot A_2 \cdot B \cdot D$ est une partie de l'interaction triple ABD .

On ne peut faire mieux : Un quart de plan est donné par deux clefs qui dans le cas le plus général s'écrivent

$$1 = A_1^{\alpha_1} \cdot A_2^{\alpha_2} \cdot B^{\alpha_3} \cdot C^{\alpha_4} \cdot D^{\alpha_5} \cdot E^{\alpha_6} ; \quad 1 = A_1^{\beta_1} \cdot A_2^{\beta_2} \cdot B^{\beta_3} \cdot C^{\beta_4} \cdot D^{\beta_5} \cdot E^{\beta_6}$$

où les α_i et β_i valent 0 ou 1. Dans les relation complètes de définition on trouve de plus la relation :

$$1 = A_1^{\alpha_1+\beta_1} \cdot A_2^{\alpha_2+\beta_2} \cdot B^{\alpha_3+\beta_3} \cdot C^{\alpha_4+\beta_4} \cdot D^{\alpha_5+\beta_5} \cdot E^{\alpha_6+\beta_6}$$

où la somme des exposants est exprimée modulo deux.

Pour avoir une résolution IV, il faudrait que parmi $\alpha_3, \alpha_4, \alpha_5$ et α_6 il y ait au moins trois 1. Il faut également la même propriété parmi $\beta_3, \beta_4, \beta_5$ et β_6 et parmi $\alpha_3 + \beta_3, \alpha_4 + \beta_4, \alpha_5 + \beta_5$ et $\alpha_6 + \beta_6$. Il est facile de voir que cela est impossible.

Exercice 13.5

32, 256

Exercice 13.6

- (i) En construisant le tableau on voit que la résolution est IV.
 (ii) Le plan 2^{5-1} à 16 unités donné par le tableau 13.1 est meilleur car de résolution V. On l'aurait donc choisi a priori plutôt que le plan replié.

Exercice 13.7

Ne sont pas estimables les effets confondus avec l'effet bloc B_l . Dans le plan 2^{7-1} de la table Bl est confondu avec $A_1 \cdot A_2 \cdot B \cdot C \cdot D \cdot E$ qui est en fait une interaction quintuple entre A, B, C, D, E et il n'y a pas de problèmes pour l'estimation des effets principaux et des interaction doubles.

Exercice 13.8

Cet exercice est particulièrement difficile quant à la définition de la matrice de variance-covariance des estimateurs. Deux stratégies se présentent

- utiliser un modèle avec les générateurs comme écrit à la fin de la section 13.3. Mais comme un des facteur a 3 niveaux cela impliquerait l'utilisation de générateurs à valeurs complexes. Ce n'est pas un problème en soi mais c'est à notre connaissance un problème informatique important. Nous ne connaissons pas de langage qui manipule aisément des matrices à valeurs complexes.
- Pour le facteur à trois niveaux, utiliser un modèle d'analyse de la variance avec une contrainte. C'est cette dernière voie que nous avons choisie.

Le premier plan construit avec la clef $B \cdot C \cdot D = 1$ est

A	B	C	D
1	1	1	1
1	1	1	1
1	1	-1	-1
1	-1	1	-1
1	-1	-1	1
2	1	1	1
2	1	-1	-1
2	-1	1	-1
2	-1	-1	1
3	1	1	1
3	1	-1	-1
3	-1	1	-1
3	-1	-1	1

Le second est construit par exemple avec la clef $A1 \cdot A2 \cdot B \cdot C \cdot D = 1$ en oubliant le dernier niveau de $(A1, A2)$

A	B	C	D
1	1	1	1
1	1	1	1
1	1	-1	-1
1	-1	1	-1
1	-1	-1	1
2	1	1	-1
2	1	-1	1
2	-1	1	1
2	-1	-1	-1
3	1	1	-1
3	1	-1	1
3	-1	1	1
3	-1	-1	-1

Ils diffèrent très peu ! Pour le premier plan la matrice X vaut

$$X = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & -1 & -1 \\ 1 & 0 & 0 & -1 & 1 & -1 \\ 1 & 0 & 0 & -1 & -1 & 1 \\ 0 & 1 & 0 & 1 & 1 & -1 \\ 0 & 1 & 0 & 1 & -1 & 1 \\ 0 & 1 & 0 & -1 & 1 & 1 \\ 0 & 1 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 1 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 & 1 & 1 \\ 0 & 0 & 1 & -1 & -1 & -1 \end{bmatrix}.$$

On doit rajouter la contrainte d'identifiabilité $K\theta = 0$ avec $K = (111000)$. La matrice de variance de $\hat{\theta}$ vaut donc

$$\text{Var} \hat{\theta} = (X' \cdot X + K'K)^{-1} X' \cdot X (X' \cdot X + K' \cdot K)^{-1}$$

Dans ce cas simple les deux méthodes donnent le même résultat : dans les deux cas, par R, on trouve

$$\begin{aligned} \det((X' \cdot X + K' \cdot K)^{-1} X' \cdot X (X' \cdot X + K' \cdot K)^{-1}) &\simeq 2.952610^{-6} \\ \text{Tr}((X' \cdot X + K' \cdot K)^{-1} X' \cdot X (X' \cdot X + K' \cdot K)^{-1}) &\simeq 0.8316 \end{aligned}$$

14 Corrections des exercices du chapitre 14

Exercice 14.1

Dans un tel cadre, si on considère la D-optimalité, on a $\det(X' \cdot X) = n \cdot \sum_{i=1}^{2n} Z_i^2 - \left(\sum_{i=1}^{2n} Z_i \right)^2$. Ce critère sera donc maximal si les (Z_i) sont tels que $\sum_{i=1}^{2n} Z_i^2$ que soit maximal et $\left(\sum_{i=1}^{2n} Z_i \right)^2$ minimal.

Comme les Z_i sont dans $[-a, a]$, $\sum_{i=1}^{2n} Z_i^2$ est maximal lorsque tous les Z_i valent $+a$ ou $-a$. De plus, au minimum, $\left(\sum_{i=1}^{2n} Z_i\right)^2$ peut-être nul. Cela est le cas si la somme des Z_i s'annule. De ces deux optimisations, on en déduit que l'unique solution qui optimise ce plan d'expérience, est de fixer n valeurs de Z_i en a et n valeurs en $-a$.

Exercice 14.2

Si la partie factorielle est de résolution V, tout quadruplet de facteurs : $A \cdot B \cdot C \cdot D$ a ses 16 combinaisons qui sont observées le même nombre de fois (en effet $A \cdot B \cdot C \cdot D$ est non confondu avec II donc il est orthogonal). Cela implique que la matrice d'information $X' \cdot X$ garde la même forme (14.5). En particulier dans la partie "produit par produit", les termes extra-diagonaux sont nuls à cause de la propriété ci-dessus. Le reste de la démonstration est inchangée.

Exercice 14.3

Il convient donc de dénombrer les monômes composant un polynôme à plusieurs (m) variables et de degré q . On peut procéder par récurrence sur le nombre de variables : soit $q \in \mathbb{N}$, fixé, et soit la propriété $\mathcal{P}(m)$: "le nombre de monômes est C_{m+q}^m ". Montrons que $\mathcal{P}(m)$ est vraie pour tout $m \in \mathbb{N}$:

- si $m = 1$, alors le nombre de monômes est $q + 1$ soit C_{1+q}^1 : $\mathcal{P}(1)$ est donc vraie.
- Supposons que $\mathcal{P}(m)$ est vraie pour un m fixé. Montrons alors que $\mathcal{P}(m + 1)$ est vraie. On travaille donc avec $m + 1$ variables, on a donc rajouté une variable que l'on peut appeler X . Le nombre total de monôme est donc la somme de ce qui se passe sans X (soit C_{m+q}^m), puis quand X est de degré 1, ce qui se passe sans X mais de degré $q - 1$ (soit C_{m+q-1}^m, \dots , jusqu'à ce que X soit de degré q (soit C_m^m). Ainsi, pour $m + 1$ variables, le nombre de monômes est :

$$\begin{aligned} C_{m+q}^m + C_{m+q-1}^m + \dots + C_m^m &= \sum_{k=0}^q C_{m+k}^m \\ &= C_{m+q+1}^{m+1} \end{aligned}$$

(ce qui peut se montrer par récurrence par exemple). Ainsi la propriété $\mathcal{P}(m + 1)$ est vraie.

La propriété $\mathcal{P}(m)$ est donc vraie pour tout $m \in \mathbb{N}$ et tout $q \in \mathbb{N}$.