Institut national de l'information géographique et forestière
École Nationale des Sciences Géographiques
mars 2016

Initiations théorique et pratique à la méthode d'estimation par moindres carrés.

Notes de cours

David COULOT

IGN/LAREG

E-mail :David.Coulot@ign.fr

Table des matières

Introduction

Dans nos domaines métrologiques (astronomie, géodésie, topométrie,...), nous sommes tributaires de mesures diverses. C'est en effet à partir de ces mesures que l'on tente de modéliser les phénomènes physiques observés afin d'en améliorer la compréhension.

Notons b le vecteur de telles observations. Dans le cadre de toute méthode de compensation, on modélise b à l'aide d'un vecteur de paramètres x par l'intermédiaire d'une fonction $f : y \simeq f(x)$. Le problème est alors de trouver les paramètres x permettant de vérifier "au mieux" cette relation suivant un critère donné. Plusieurs problèmes se présentent alors. Le modèle n'est en général qu'approximatif et les mesures sont très souvent entachées d'erreurs aléatoires. De plus, il est bon de pouvoir juger de la qualité des estimations des paramètres. Le cadre mathématique idéal est donc celui de l'estimation statistique. La méthode des moindres carrés s'inscrit parfaitement dans cet axe mais elle n'est pas exempte de défauts. Le principal est son manque de robustesse [1] qui peut lui faire préférer des méthodes d'estimation en norme $\mathcal{L}1$ telles que le simplexe modifié. Néanmoins, la méthode des moindres carrés présente une formulation analytique simple et facile d'utilisation. De plus, sous certaines hypothèses, on peut construire des tests statistiques utiles pour juger de la qualité des paramètres estimés et donc améliorer l'estimation des dits paramètres.

Dans le cas où la fonction f n'est pas différentiable, ni-même analytique, on peut préférer à cette méthode l'utilisation d'algorithmes génétiques. Ces algorithmes présentent en effet l'avantage de ne pas nécessiter la linéarisation de la fonction f autour d'une valeur approchée x_0 des paramètres recherchés.

1. La robustesse d'une méthode d'estimation mesure le degré de sensibilité de cette méthode aux fautes de mesure.

Première partie

Méthode des moindres carrés

La méthode d'estimation par moindres carrés nous vient principalement des travaux de Gauss [2], Legendre [3] et de Laplace [4]. Elle est donc née il y a près de deux siècles. Elle repose sur la minimisation du carré de la norme [5] ($<,>$ désigne le produit scalaire dont dérive $\|.\|$)

$$\|b - f(x)\|^2 =< b - f(x), b - f(x) >$$

Dans cette première partie, les bases de la méthode ainsi que les définitions utiles sont présentées à partir de l'exemple de la droite de régression linéaire. Ensuite, la formulation statistique pour le modèle linéaire est décrite ainsi que les tests pouvant être effectués sur la base d'hypothèses relatives à la densité de probabilité des observations. Quelques techniques de calculs complémentaires sont également exposées (opérations sur les équations normales, équations de contraintes, réduction par rapport à certains paramètres, méthodes d'analyse des composantes de la variance). Enfin, dans un dernier temps, on présente les problèmes liés à l'utilisation de cette méthode.

Une deuxième partie donne deux exemples de méthodes alternatives à la méthode des moindres carrés : une méthode d'estimation robuste (le simplexe modifié) et une méthode d'estimation stochastique (les algorithmes génétiques).

Toutes les bases mathématiques nécessaires à la compréhension de la théorie sont rappelées en annexe. J'ai souhaité que cette annexe soit la plus complète possible afin que toutes les notions théoriques (mêmes les plus simples) nécessaires à la bonne compréhension du corps de ce document soient immédiatement accessibles au lecteur.

Certaines démonstrations de ce document sont inspirées de [15] dont je recommande la lecture à tout lecteur souhaitant approfondir ses connaissances sur la méthode des moindres carrés.

2. Carl Friedrich Gauss (1777-1855).
3. Adrien Marie Legendre (1752-1833).
4. Pierre Simon, marquis de Laplace (1749-1827).
5. C'est d'ailleurs de ce critère qu'elle tire son nom...

Chapitre 1

Exemple introductif : la droite de régression linéaire

On considère n ($n \geq 2$) points du plan $\mathbb{R}^2$ et on cherche à faire passer "au mieux" une droite par ces points.

1.1 Résolution analytique

Dans un premier temps, $n = 2$.

Dans ce cas, le problème se résout très facilement. Les deux inconnues sont la pente a_1 de la droite et son ordonnée à l'origine a_2. Le système à résoudre est donc

$$\begin{pmatrix} t_1 & 1 \\ t_2 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

et la solution cherchée bien connue est[1]

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \end{pmatrix} = \begin{pmatrix} \frac{b_1 - b_2}{t_1 - t_2} \\ \frac{t_1 b_2 - t_2 b_1}{t_1 - t_2} \end{pmatrix}$$

Supposons maintenant que $n > 2$.

On peut toujours poser le problème sous la forme matricielle $B \simeq AX$ avec

$$A = \begin{pmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix},$$

1. Les deux points étant distincts, la matrice du système est inversible.

$X^T = (a_1, a_2)$ et $B^T = (b_1, \cdots, b_n)$. Par contre, la matrice A n'est plus inversible et on ne peut donc pas résoudre facilement le système comme précédemment. Il faut trouver un moyen de calculer $\hat{a}_1$ et $\hat{a}_2$ en tenant compte de la redondance d'information dont on dispose.

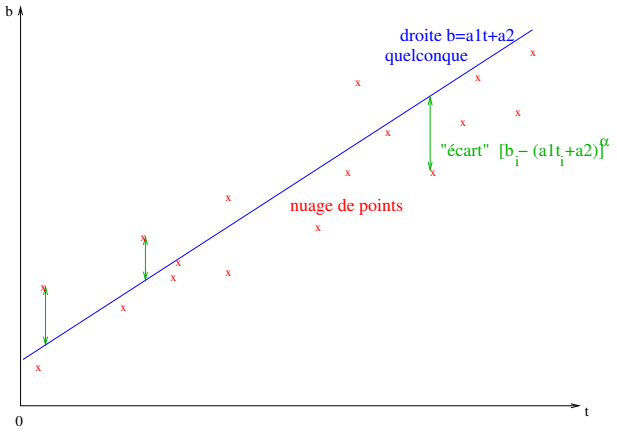

FIGURE 1.1 – "Écarts" de la droite au nuage de points.

Pour que la droite cherchée passe "au mieux" par tous les points du nuage, on peut minimiser la somme des écarts Δb_i d'une droite quelconque $b = a_1 t + a_2$ à chacun des points (t_i, b_i) (cf. figure (1.1)).

Ces écarts peuvent être quantifiés de diverses façons. On peut en effet poser $\Delta b_i = (a_1 t_i + a_2 - b_i)^\alpha$, $\alpha \in \mathbb{N}$. Dans le cadre de la méthode des moindres carrés, on prend $\alpha = 2$. Il nous faut donc minimiser la fonction numérique

$$\left\{ \begin{array}{rcl} \varphi & : & \mathbb{R}^2 \to \mathbb{R}_+ \\ & & (a_1, a_2) \mapsto \varphi(a_1, a_2) = \sum_{i=1}^n (a_1 t_i + a_2 - b_i)^2 \end{array} \right.$$

On sait que si φ (qui est clairement de classe C^∞ sur $\mathbb{R}^2$) admet un minimum local en (a_1, a_2) alors (a_1, a_2) est un point critique de φ et la forme quadratique fondamentale associée à φ en (a_1, a_2) $Q_{\varphi,(a_1,a_2)}$ est positive. On sait également que si (a_1, a_2) est un point critique non dégénéré, il suffit que $Q_{\varphi,(a_1,a_2)}$ soit définie positive pour que φ admette un minimum local en (a_1, a_2). Ce minimum est alors strict (cf. annexe).

Les dérivées partielles d'ordres 1 et 2 de la fonction φ sont

$$\frac{\partial \varphi}{\partial a_1} = 2 \sum_{i=1}^n t_i (a_1 t_i + a_2 - b_i)$$

$$\begin{aligned}
\frac{\partial \varphi}{\partial a_2} &= 2\sum_{i=1}^{n}(a_1 t_i + a_2 - b_i) \\
\frac{\partial^2 \varphi}{\partial a_1^2} &= 2\sum_{i=1}^{n} t_i^2 \\
\frac{\partial^2 \varphi}{\partial a_2^2} &= 2n \\
\frac{\partial^2 \varphi}{\partial a_1 \partial a_2} &= 2\sum_{i=1}^{n} t_i
\end{aligned}$$

La matrice de la forme quadratique fondamentale $Q_{\varphi,(a_1,a_2)}$

$$\begin{pmatrix} 2\sum_{i=1}^{n} t_i^2 & 2\sum_{i=1}^{n} t_i \\ . & 2n \end{pmatrix}$$

a donc pour déterminant $4[n\sum t_i^2 - (\sum t_i)^2]$. *Les vecteurs* $(1,\cdots,1)^T$ *et* $(t_1,\cdots,t_n)^T$ *étant indépendants, ce déterminant est strictement positif en vertu de l'inégalité de Cauchy-Schwartz (cf. annexe). En tout point* (a_1,a_2) *de* $\mathbb{R}^2$, $Q_{\varphi,(a_1,a_2)}$ *est donc non dégénérée.*

$\forall h = (h_1,h_2) \in \mathbb{R}^2$,

$$Q_{\varphi,(a_1,a_2)}(h) = 2\sum_{i=1}^{n}(t_i h_1 + h_2)^2$$

donc $Q_{\varphi,(a_1,a_2)}(h) \geq 0$. *De plus,*

$$Q_{\varphi,(a_1,a_2)}(h) = 0 \Rightarrow \forall i \in \{1\cdots n\},\ t_i h_1 + h_2 = 0$$

Comme au moins deux des t_i *sont distincts, ceci implique* $h_1 = h_2 = 0$. $Q_{\varphi,(a_1,a_2)}$ *est donc définie positive pour tout point* (a_1,a_2) *de* $\mathbb{R}^2$.

Il faut également que le point (a_1,a_2) *soit un point critique de* φ. *Les deux conditions suivantes doivent donc être remplies*

$$\begin{aligned}
\sum_{i=1}^{n} t_i(a_1 t_i + a_2 - b_i) &= 0 \\
\sum_{i=1}^{n}(a_1 t_i + a_2 - b_i) &= 0
\end{aligned}$$

La première relation donne

$$an\bar{t} + na_2 - n\bar{b} = 0 \Longleftrightarrow a_2 = \bar{b} - a_1\bar{t}$$

Utilisant cette expression et la seconde relation, on obtient l'équation

$$a_1 n\sigma_t^2 + a_1\bar{t}^2 + \bar{t}\bar{b} - a_1\bar{t}^2 - n\sigma_{tb} - \bar{t}\bar{b} = 0$$

On aboutit finalement à

$$\begin{aligned} \hat{a}_1 &= \frac{\sigma_{tb}}{\sigma_t^2} \\ \hat{a}_2 &= \bar{b} - \frac{\sigma_{tb}}{\sigma_t^2}\bar{t} \end{aligned}$$

où [2]

$$\begin{aligned} \bar{t} &= \frac{1}{n}\sum_{i=1}^{n} t_i \\ \bar{b} &= \frac{1}{n}\sum_{i=1}^{n} b_i \\ \sigma_t^2 &= \frac{1}{n}\sum_{i=1}^{n}(t_i - \bar{t})^2 \\ \sigma_{tb} &= \frac{1}{n}\sum_{i=1}^{n}(t_i - \bar{t}).(b_i - \bar{b}) \end{aligned}$$

φ admet donc un minimum local strict en $(\hat{a}_1, \hat{a}_2)$ qui constitue la solution des moindres carrés au problème posé. On peut même aller plus loin. En effet, le minimum global de φ est avant tout un minimum local. Le calcul précédent nous fournit tous les minima locaux de φ sur $\mathbb{R}^2$. Nous n'en avons trouvé qu'un seul : c'est donc le minimum global de φ sur $\mathbb{R}^2$.

1.2 Formulation matricielle

On peut également raisonner de façon matricielle en remarquant que $a_1 t_i + a_2 - b_i$ n'est autre que le $i^{ème}$ terme du vecteur $AX - B$. Ainsi, on peut écrire

$$\varphi(a_1, a_2) = (AX - B)^T.(AX - B) = \phi(X).$$

En développant,

$$\phi(X) = X^T A^T AX - X^T A^T B - B^T AX + B^T B$$

2. Ce sont les opérateurs statistiques empiriques bien connus, [6] par exemple.

ϕ étant une fonction numérique, l'application linéaire qui lui est tangente en X, $\phi'(X)$, est une forme linéaire et on peut écrire, en différenciant la précédente relation, $\forall dX \in \mathbb{R}^2$,

$$\begin{aligned}\phi'(X).dX &= dX^T A^T AX + X^T A^T AdX - dX^T A^T B - B^T AdX \\ &= 2(X^T A^T A - B^T A).dX\end{aligned}$$

soit $\phi'(X) = 2(X^T A^T A - B^T A)$

$\hat{X}$ est un point critique de ϕ ssi $A^T A\hat{X} = A^T B$. Dans le cas présent, $A^T A$ est une matrice inversible (puisque de déterminant $4n^2\sigma_t^2 \neq 0$). On a donc

$$\hat{X} = (A^T A)^{-1} A^T B$$

En différenciant de nouveau $\phi'(X)$, on obtient la forme quadratique fondamentale associée à ϕ en X :

$$\phi''(X) = 2A^T A$$

.

Cette forme quadratique est non dégénérée en tout point X et

$$\forall U \in \mathbb{R}^2,\ U^T.\phi''(X).U = 2U^T.A^T A.U = 2(AU)^T.(AU) \geq 0$$

De plus[3]

$$U^T.\phi''(X).U = 0 \Rightarrow (AU)^T.(AU) = 0 \Rightarrow AU = 0 \Rightarrow U = 0.$$

La forme quadratique fondamentale $\phi''(X)$ est donc définie positive. ϕ admet donc un minimum local strict en $\hat{X}$ qui, comme précédemment, est le minimum global de ϕ. En appliquant la formule matricielle définissant $\hat{X}$, on retrouve les précédentes expressions de $\hat{a}_1$ et $\hat{a}_2$.

3. Puisqu'au moins deux des t_i sont distincts, la matrice A est de plein rang et son noyau est donc réduit à 0.

Chapitre 2

Méthode des moindres carrés : présentation générale

L'exemple précédent est linéaire par rapport aux paramètres recherchés. En général, ce n'est pas le cas et il faut se placer sous certaines conditions pour se ramener à un problème linéaire.

2.1 Linéarisation et solution générale

On considère un $\mathbb{R}$ espace vectoriel normé E de dimension p et un $\mathbb{R}$ espace euclidien F de dimension n $(n > p)$ de produit scalaire $<,>$. On note P sa matrice relativement à une base donnée de F et $\|.\|$ la norme qu'il induit. On cherche à modéliser le vecteur de mesures $b \in F$ à l'aide de paramètres $x \in E$ par l'intermédiaire de la fonction [1]

$$f : \Omega \subset E \to F.$$

Comme précédemment, on introduit la fonction

$$\left\{ \begin{array}{lcl} \varphi & : & \Omega \subset E \to \mathbb{R} \\ & & x \mapsto \varphi(x) = \|f(x) - b\|^2 \end{array} \right. \tag{2.1}$$

Le problème est à nouveau de minimiser φ. On suppose que l'on connaît une solution approchée x_0 à l'équation $\varphi'(x) = 0$, que f est linéarisable au voisinage de ce point et que la matrice A de l'application linéaire tangente $f'(x_0)$ relativement à des bases données de E et F est de plein rang. On peut ainsi écrire le développement de Taylor de f au voisinage de x_0 [2]

$$f(x) = f(x_0) + f'(x_0).(x - x_0) + \|x - x_0\|\varepsilon(x - x_0)$$

1. Ω est un ouvert de E.
2. $\lim_{x \to x_0} \varepsilon(x - x_0) = 0$

Ainsi, au voisinage de x_0,

$$\varphi(x) \simeq \|f(x_0) + f'(x_0).(x - x_0) - b\|^2 \tag{2.2}$$

Adoptant une formulation matricielle, on introduit de nouveau la fonction ϕ :

$$\phi(X) = (AX - B)^T.P.(AX - B) \tag{2.3}$$

où $X = x - x_0$ et $B = b - f(x_0)$.

Reprenant le raisonnement de la précédente partie, en supposant que la matrice A^TPA est inversible, on montre que φ admet un minimum local strict en $\hat{x}$ donné par

$$\hat{x} = x_0 + (A^TPA)^{-1}A^TP(b - f(x_0)) \tag{2.4}$$

$\hat{x}$ est la solution générale au problème par la méthode des moindres carrés. Contrairement à l'exemple précédent, on ne peut affirmer que $\hat{x}$ correspond à un minimum global de φ. En effet, la linéarisation de la fonction f nous oblige d'emblée à nous restreindre à un voisinage de la solution approchée x_0.

2.2 Définitions et remarques

Le processus de résolution par moindres carrés peut être itératif. En effet, disposant d'une estimation $\hat{x}$ des paramètres, on peut itérer le processus en considérant cette estimation comme la valeur approchée x_0 de la solution. Si la valeur approchée initiale est proche de la solution, on peut se permettre de ne pas recalculer la matrice des dérivées partielles A entre deux itérations.

Reprenant la forme générale de la solution (2.4), on s'aperçoit que cette solution $\hat{x}$ est invariante par homothétie sur la matrice de produit scalaire P. Cette dernière est donc définie à un facteur multiplicatif près. Elle est appelée ***matrice de poids***.

Le vecteur $X = x - x_0$ est le ***vecteur des paramètres***. Le vecteur $B = b - f(x_0)$ est appelé ***vecteur des observations***, dénomination abusive (à cause du terme $-f(x_0)$) à laquelle on préfère généralement le terme de ***pseudo-observations***. La matrice A des dérivées partielles est la ***matrice modèle***. La matrice A^TPA est la ***matrice normale*** ; on l'a supposée inversible, ce n'est pas toujours vrai. Dans ce cas, il manque de l'information pour résoudre le problème, information pouvant être fournie sous la forme d'équations de contraintes, par exemple.

Dans le calcul précédent, tout se passe comme si l'on partait de l'équation approchée non inversible

$$b = f(x) \simeq f(x_0) + f'(x_0)(x - x_0) \Leftrightarrow B \simeq AX$$

On multiplie alors chaque membre de l'équation par la matrice $A^T P$ (c'est le processus de ***normalisation***) pour obtenir l'***équation normale***

$$A^T P A\hat{X} = A^T P B$$

qui est inversible.

Enfin, on définit le ***vecteur des résidus*** $\hat{V} = A\hat{X} - B$, vecteur fondamental de la théorie des moindres carrés. C'est en effet ce vecteur qui contient le signal résiduel dans B après retrait du modèle estimé $A\hat{X}$. On va notamment y trouver les éventuelles fautes de mesures ainsi que les lacunes du modèle. Le vecteur des pseudo-observations $b - f(x_0)$ est parfois appelé vecteur des résidus *a priori* puisqu'il représente en effet le vecteur des résidus avant l'estimation.

2.3 Interprétation géométrique

Reprenant les précédentes notations, on introduit le vecteur $\hat{B} = A\hat{X}$ appartenant à F. Par définition de $\hat{X}$,

$$\hat{B} = \hat{M}B = [A(A^T PA)^{-1} A^T P]B \tag{2.5}$$

$\hat{M}$ est appelée ***matrice chapeau*** car son "rôle" est de "mettre un chapeau" sur les vecteurs. On montre facilement que $\hat{M}$ est la matrice d'un projecteur ($\hat{M}^2 = \hat{M}$). De plus, $\forall \vec{v} \in Im(\hat{M})$, $\exists \vec{u}/\ \vec{v} = \hat{M}\vec{u}$ soit $v = A[(A^T PA)^{-1} A^T P\vec{u}]$ donc $\vec{v} \in Im(A)$. Réciproquement, $\forall \vec{v} \in Im(A)$ $\exists \vec{u}/\ \vec{v} = A\vec{u}$ et $\hat{M}A\vec{u} = A\vec{u} = \vec{v}$ donc $\vec{v} \in Im(\hat{M})$. $\hat{M}$ est donc la matrice d'une projection sur $Im(A)$.

Soit $\hat{N} = I_n - \hat{M}$ la matrice du projecteur associé à celui de matrice $\hat{M}$. On sait que $E = Im(A) \oplus Im(\hat{N})$. Cette somme directe est également orthogonale.

En effet, $\forall B \in E$, $B = \hat{M}B + \hat{N}B$ et

$$(\hat{M}B)^T.P.\hat{N}B = B^T \hat{M}^T PB - B^T \hat{M}^T P\hat{M}B = 0$$

car $\hat{M}^T P\hat{M} = \hat{M}^T P$.

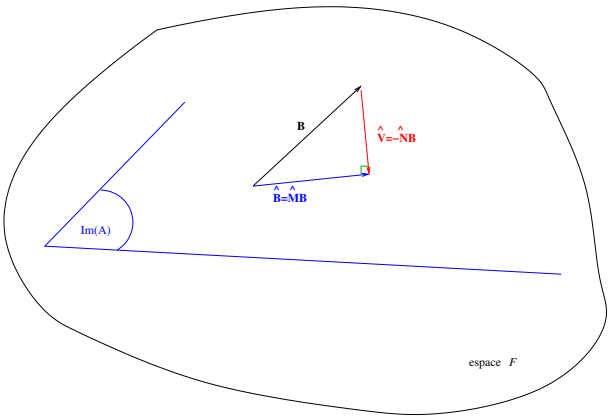

FIGURE 2.1 – Interprétation géométrique de la méthode des moindres carrés.

En remarquant que $\hat{V} = -\hat{N}B$, on constate que $\hat{B}$ est la projection orthogonale de B sur $Im(A)$ qui minimise donc la norme du vecteur des résidus (cf. figure (2.1)).

2.4 Exemple numérique

On mesure la largeur l, la longueur L et la diagonale $d = \sqrt{l^2 + L^2}$ d'un rectangle. De précédentes mesures avaient donné les valeurs $l = 1,60$m et $L = 2,08$m. Le rectangle, en bois, a été soumis à des contraintes thermiques depuis ces derniers relevés. Les nouvelles mesures effectuées sont destinées à vérifier si les dimensions du rectangle ont évolué de façon significative. Elles sont fournies dans le tableau (2.1).

l	L	d
1,634	2,078	2,609
1,643	2,065	2,619
1,659	2,044	2,631
1,656	2,036	2,642
1,621	2,018	2,606

TABLE 2.1 – Mesures effectuées.

Toutes les fonctions mises en jeu dans le modèle sont clairement différentiables sur $\mathbb{R}^2\backslash\{(0,0)\}$. Supposant que les dimensions n'ont pas évolué de façon excessive entre les deux époques de mesure, on peut linéariser le modèle autour des valeurs $l_0 = 1,60$ et $L_0 = 2,08$, ce qui donne les équations d'observations

$$\left\{\begin{array}{rcl} l & = & l_0 + \delta l \\ L & = & L_0 + \delta L \\ d & \simeq & d_0 + \frac{l_0}{d_0}.\delta l + \frac{L_0}{d_0}.\delta L \end{array}\right.$$

Ainsi, les équations d'observations sont données sous forme matricielle par

$$\begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ \frac{l_0}{d_0} & \frac{L_0}{d_0} \\ \vdots & \vdots \\ \frac{l_0}{d_0} & \frac{L_0}{d_0} \end{pmatrix} \cdot \begin{pmatrix} \delta l \\ \delta L \end{pmatrix} \simeq \begin{pmatrix} l_1 - l_0 \\ \vdots \\ l_5 - l_0 \\ L_1 - L_0 \\ \vdots \\ L_5 - L_0 \\ d_1 - d_0 \\ \vdots \\ d_5 - d_0 \end{pmatrix}$$

Numériquement, on obtient

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0,610 & 0,793 \\ 0,610 & 0,793 \\ 0,610 & 0,793 \\ 0,610 & 0,793 \\ 0,610 & 0,793 \end{pmatrix} \cdot \begin{pmatrix} \delta l \\ \delta L \end{pmatrix} \simeq \begin{pmatrix} 0,034 \\ 0,043 \\ 0,059 \\ 0,056 \\ 0,021 \\ -0,002 \\ -0,015 \\ -0,036 \\ -0,044 \\ -0,062 \\ -0,015 \\ -0,005 \\ 0,007 \\ 0,018 \\ -0,018 \end{pmatrix}$$

La matrice modèle de rang 2. Ne disposant d'aucune information particulière sur la nature de la matrice de poids, on la prend égale à la matrice identité I_{15}. Le système normal est donc

$$\begin{pmatrix} 6,860 & 2,419 \\ . & 8,144 \end{pmatrix} \cdot \begin{pmatrix} \delta\hat{l} \\ \delta\hat{L} \end{pmatrix} = \begin{pmatrix} 0,205 \\ -0,169 \end{pmatrix}$$

La matrice normale est bien inversible et l'inversion du système précédent donne la solution

$$\begin{pmatrix} \delta\hat{l} \\ \delta\hat{L} \end{pmatrix} = \begin{pmatrix} 0,042 \\ -0,033 \end{pmatrix} \text{ soit } \begin{pmatrix} \hat{l} \\ \hat{L} \end{pmatrix} = \begin{pmatrix} 1,642 \\ 2,047 \end{pmatrix}$$

La largeur du rectangle a donc augmenté de 4,2 cm et sa longueur a diminué de 3,3 cm. On voit sur cet exemple que l'on aimerait pouvoir pousser le calcul plus loin et, notamment, disposer d'indicateurs statistiques de la qualité des estimations $\delta\hat{l}$ et $\delta\hat{L}$. Ces indicateurs font l'objet du chapitre suivant consacré au modèle linéaire.

Chapitre 3

Modèle linéaire et formulation statistique

L'utilisation du modèle linéaire n'est pas une obligation. En effet, on peut parfaitement imaginer chercher directement le minimum de la fonction φ définie par (2.1) à l'aide d'une méthode de minimisation quelconque.

Mais on a précédemment vu que l'on peut facilement se ramener au modèle linéaire. De plus, ce modèle donne une formulation analytique simple de la solution. Enfin, nous allons voir dans cette partie que le modèle linéaire permet d'obtenir des propriétés statistiques de l'estimation du vecteur des paramètres et que, moyennant un bon choix de la matrice de poids P, l'estimateur des moindres carrés ainsi obtenu est optimal (théorème de Gauss-Markov).

3.1 Propriétés de l'estimation pour le modèle linéaire

On se place dans le cadre de l'estimation statistique et on considère le système linéaire

$$B = A\dot{X} + V \tag{3.1}$$

Dans ce système [1], les variables B et V sont des variables aléatoires vectorielles d'espérances respectives $\dot{B}$ et [2] 0. Le vecteur déterministe $\dot{X}$ est le vecteur des paramètres cherchés. De (3.1), on déduit $\dot{B} = A\dot{X}$. On suppose de plus que la matrice de variance-covariance Σ_B du vecteur erreur

1. La matrice modèle A est déterministe.
2. Cette hypothèse revient à considérer les observations non biaisées ; nous reviendrons sur ce point un peu plus loin dans le cours.

aléatoire V est connue à un facteur σ_0^2 près : $\mathbb{E}(VV^T) = \sigma_0^2 \Sigma_B$.

Dans la pratique, on n'a accès qu'à une réalisation de la variable aléatoire B à partir de laquelle on détermine une réalisation de l'estimateur $\hat{X}$ des moindres carrés. Dans la suite, je confonds (abusivement !) les variables aléatoires et leurs réalisations.

3.1.1 Estimateur des moindres carrés

L'estimateur des moindres carrés solution de (3.1) pour la matrice de poids P est (cf. (2.4))

$$\hat{X} = (A^T P A)^{-1} A^T P B \tag{3.2}$$

À ce niveau, on peut préciser quelque peu l'usage de la formulation aléatoire. On peut en effet distinguer deux signaux dans les mesures B : un signal déterministe modélisé sous la forme $A\dot{X}$ et un signal purement aléatoire modélisé sous la forme de l'erreur V. Ainsi, les paramètres $\dot{X}$ sont des constantes ; par contre, en tant qu'estimateur, $\hat{X}$ est une variable aléatoire, fonction linéaire des mesures B. Le caractère aléatoire du vecteur des observations B ne provient d'ailleurs en fait que de la présence de V. De plus, (3.2) est valable à la fois pour les variables aléatoires elles-mêmes et pour leurs réalisations.

L'estimateur des moindres carrés $\hat{X}$ est un estimateur linéaire en B non biaisé [3]. En effet,

$$\mathbb{E}(\hat{X}) = (A^T P A)^{-1} A^T P \mathbb{E}(B) = (A^T P A)^{-1} A^T P A \dot{X} = \dot{X}$$

On connaît l'espérance de l'estimateur $\hat{X}$, déterminons maintenant sa matrice de variance-covariance. Par définition,

$$var(\hat{X}) = \mathbb{E}((\hat{X} - \dot{X}).(\hat{X} - \dot{X})^T)$$

soit

$$var(\hat{X}) = \mathbb{E}(\hat{X}\hat{X}^T - \dot{X}\hat{X}^T - \hat{X}\dot{X}^T + \dot{X}\dot{X}^T)$$

À l'aide de (3.2), on obtient

$$\begin{aligned} var(\hat{X}) &= (A^T P A)^{-1} A^T P \mathbb{E}(BB^T) P A (A^T P A)^{-1} \\ &- \dot{X} \underbrace{\dot{B}^T P A (A^T P A)^{-1}}_{\dot{X}^T} - \underbrace{(A^T P A)^{-1} A^T P \dot{B}}_{\dot{X}} \dot{X}^T \\ &+ \dot{X}\dot{X}^T \end{aligned}$$

3. C'est la première propriété importante du modèle linéaire. En effet, dans le cas général, les estimateurs des moindres carrés sont souvent biaisés [11].

De plus, la relation $var(B) = \mathbb{E}[(B - A\dot{X}).(B - A\dot{X})^T] = \sigma_0^2\Sigma_B$ donne $\mathbb{E}(BB^T) = \sigma_0^2\Sigma_B + A\dot{X}\dot{X}^TA^T$ et, finalement,

$$var(\hat{X}) = \sigma_0^2(A^TPA)^{-1}A^TP\Sigma_BPA(A^TPA)^{-1} \tag{3.3}$$

Le caractère non-biaisé de l'estimateur des moindres carrés ainsi que la relation (3.3) sont valables quelle que soit la matrice de poids P. Le théorème de Gauss-Markov précise la valeur à donner à cette matrice de poids.

Théorème 3.1 (Théorème de Gauss-Markov) *Lorsque la matrice de poids P est prise égale à Σ_B^{-1}, l'estimateur des moindres carrés est optimal, i.e. de variance minimale, dans la classe des estimateurs sans biais linéaires en B.*

Démonstration

Considérons donc deux estimateurs linéaires en B : l'estimateur des moindres carrés $\hat{X}$ et un estimateur $\hat{X}' = CB$.

Reprenant la relation (3.3) précédente, par hypothèse,

$$var(\hat{X}) = \sigma_0^2(A^T\Sigma_B^{-1}A)^{-1} \tag{3.4}$$

$\hat{X}'$ doit être sans biais donc

$$\mathbb{E}(\hat{X}') = \dot{X} = C\dot{B} = CA\dot{X} \Rightarrow (CA - I_p)\dot{X} = 0$$

Cette relation doit être valable quel que soit le vecteur des paramètres $\dot{X}$ choisi donc $\ker(CA - I_p) = E$ d'où $CA = I_p$.

On sait que $var(\hat{X}') = \sigma_0^2C\Sigma_BC^T$ et on peut écrire $\hat{X}' = RB + \hat{X}$ avec $R = C - (A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1}$ soit

$$var(\hat{X}') = \sigma_0^2\left[(R + (A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1})\Sigma_B(R + (A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1})^T\right]$$

En développant et en utilisant le fait que $CA = I_p$, on aboutit finalement à la relation

$$var(\hat{X}') = \sigma_0^2R\Sigma_BR^T + \sigma_0^2(A^T\Sigma_B^{-1}A)^{-1} = \sigma_0^2R\Sigma_BR^T + var(\hat{X})$$

Enfin, $\forall U \in \mathbb{R}^n$,

$$U^TR\Sigma_BR^TU = (R^TU)^T\Sigma_B(R^TU) \geq 0$$

car Σ_B, en tant que matrice de variance-covariance, est positive. Ainsi, $var(\hat{X}') \succeq var(\hat{X})$ et $var(\hat{X})$ est donc minimale dans la classe des estimateurs sans biais linéaires en B pour la relation d'ordre $\succeq$ sur les matrices

symétriques (cf. annexe).

D'après [15], l'estimateur des moindres carrés est optimal mais aussi efficace. D'après [11], si la matrice normale reste non-singulière quand on fait tendre la dimension n du vecteur des observations vers l'infini, l'estimateur des moindres carrés est également convergent. Ainsi, l'estimateur des moindres carrés $\hat{X}$ est un estimateur convergent efficace (donc optimal) dans la classe des estimateurs sans biais linéaires en B.

Dans toute la suite, on utilise donc $P = \Sigma_B^{-1}$ comme matrice de poids.

La matrice de variance-covariance (3.4) de l'estimateur $\hat{X}$ des moindres carrés pour le modèle linéaire est donc l'inverse de la matrice normale multipliée par σ_0^2.

On avait précédemment remarqué que $\hat{X}$ est invariant par homothétie sur la matrice de poids. La présence de σ_0^2 ne change donc rien à l'estimation et permet en plus de remettre en cause la connaissance que l'on a des erreurs de mesure V. Supposons que les variables aléatoires V_i, composantes de la variable aléatoire vectorielle V, soient deux à deux indépendantes. On peut alors écrire

$$\Sigma_B^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n^2} \end{pmatrix}$$

et

$$\Sigma_B^{-1} B = \begin{pmatrix} \frac{b_1}{\sigma_1^2} \\ \vdots \\ \frac{b_n}{\sigma_n^2} \end{pmatrix}$$

Ainsi, plus une observation b_i est précise, plus son écart-type σ_i est petit et plus elle a de "poids" dans l'estimation. En prenant l'inverse d'une matrice de variance-covariance complète des observations, on tient non seulement compte des variances de chaque observation mais aussi des corrélations existant entre ces observations.

3.1.2 Vecteur des résidus

Le vecteur des résidus $\hat{V} = A\hat{X} - B$ est une variable aléatoire vectorielle, elle-aussi fonction linéaire du vecteur des observations. Son espérance est

$$\mathbb{E}(\hat{V}) = A\dot{X} - \dot{B} = 0$$

Comme $\hat{V} = -\hat{N}B$,

$$\mathbb{E}(\hat{V}\hat{V}^T) = [A(A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1} - I_n]\mathbb{E}(BB^T)[A(A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1} - I_n]^T$$

Or $\mathbb{E}(BB^T) = \sigma_0^2\Sigma_B + A\dot{X}\dot{X}^TA^T$ et

$$[A(A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1} - I_n]A\dot{X}\dot{X}^TA^T[A(A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1} - I_n]^T = 0$$

$\hat{V}$ a donc pour matrice de variance-covariance

$$var(\hat{V}) = \sigma_0^2[\Sigma_B - A(A^T\Sigma_B^{-1}A)^{-1}A^T] \tag{3.5}$$

$\hat{V}$ et $\hat{X}$ étant tous deux des estimateurs linéaires en B, on peut soupçonner l'existence de corrélations entre ces deux vecteurs. Il n'en est rien. En effet,

$$\begin{aligned}
cov(\hat{V},\hat{X}) &= \mathbb{E}[\hat{V}.(\hat{X}-\dot{X})^T] = \mathbb{E}(\hat{V}\hat{X}^T) - \mathbb{E}(\hat{V}\dot{X}^T) \\
&= \mathbb{E}(\hat{V}\hat{X}^T) - \mathbb{E}(\hat{V})\dot{X}^T = \mathbb{E}(\hat{V}\hat{X}^T) \\
&= [A(A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1} - I_n]\mathbb{E}(BB^T)\Sigma_B^{-1}A(A^T\Sigma_B^{-1}A)^{-1} \\
&= \sigma_0^2[A(A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1} - I_n]\Sigma_B\Sigma_B^{-1}A(A^T\Sigma_B^{-1}A)^{-1} \\
&+ [A(A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1} - I_n]A\dot{X}\dot{X}^TA^T = 0
\end{aligned}$$

Enfin, on peut définir un vecteur de résidus normalisés. Si L est une racine carrée de la matrice de variance-covariance (à σ_0^2 près) des observations B [4] : $\Sigma_B = LL^T$, le vecteur des résidus normalisés $\hat{V}_{nor}$ est défini par

$$\hat{V}_{nor} = L^{-1}\hat{V}$$

3.1.3 Facteur unitaire de variance

On a supposé connaître la matrice de variance-covariance des observations à σ_0^2 près. Ce facteur multiplicatif est donc une inconnue du problème d'estimation. On l'appelle facteur unitaire de variance ; en effet, dans le cas idéal, $\sigma_0^2 = 1$.

Reprenons l'équation (3.5) et multiplions chacun de ses membres par Σ_B^{-1}. On obtient

$$\mathbb{E}(\hat{V}\hat{V}^T\Sigma_B^{-1}) = \sigma_0^2[I_n - A(A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1}]$$

4. Une telle matrice peut être déterminée par l'algorithme de Cholesky [15].

Prenant alors la trace de chaque membre, on a

$$\begin{aligned} tr(\mathbb{E}(\hat{V}\hat{V}^T\Sigma_B^{-1})) &= \mathbb{E}(tr(\hat{V}^T\Sigma_B^{-1}\hat{V})) = \mathbb{E}(\hat{V}^T\Sigma_B^{-1}\hat{V}) \\ &= \sigma_0^2[n - tr(A(A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1})] \\ &= \sigma_0^2[n - tr(A^T\Sigma_B^{-1}A(A^T\Sigma_B^{-1}A)^{-1})] = \sigma_0^2(n-p) \end{aligned}$$

Un estimateur sans biais du facteur unitaire de variance σ_0^2 est donc

$$\hat{\sigma}_0^2 = \frac{\hat{V}^T\Sigma_B^{-1}\hat{V}}{n-p} \tag{3.6}$$

$n - p$ est appelé facteur de redondance ou nombre de degrés de liberté du problème.

$\hat{\sigma}_0^2$ dépend directement de $\hat{V}$. Ainsi, un facteur unitaire de variance fort peut être à la fois signe d'une très mauvaise connaissance de la matrice de variance-covariance des erreurs de mesures et signe de résidus forts donc de la présence de problèmes (fautes de mesures, erreurs du modèle, etc).

3.1.4 Résumé des résultats

Le tableau (3.1) donne les estimateurs des moindres carrés du modèle linéaire ainsi que leurs caractéristiques statistiques.

variable	***expression***	***espérance***	***variance***
erreurs de mesure V	/	0	$\sigma_0^2\Sigma_B$
observations B (n)	/	$A\dot{X}$	$\sigma_0^2\Sigma_B$
paramètres $\hat{X}$ (p)	$(A^T\Sigma_B^{-1}A)^{-1}A^T\Sigma_B^{-1}B$	$\dot{X}$	$\sigma_0^2(A^T\Sigma_B^{-1}A)^{-1}$
résidus $\hat{V}$	$A\hat{X} - B$	0	$\sigma_0^2(\Sigma_B - A(A^T\Sigma_B^{-1}A)^{-1}A^T)$
facteur unitaire $\hat{\sigma}_0^2$	$\frac{\hat{V}^T\Sigma_B^{-1}\hat{V}}{n-p}$	σ_0^2	/

TABLE 3.1 – Résultats pour le modèle linéaire.

3.2 Propriétés supplémentaires pour une distribution normale des observations

Dans cette partie, on suppose que le vecteur des observations B suit une loi normale multidimensionnelle de moyenne $A\dot{X}$ et de matrice de variance-covariance $\sigma_0^2\Sigma_B$ et on note

$$B \hookrightarrow \mathcal{N}(A\dot{X}, \sigma_0^2\Sigma_B)$$

Cette hypothèse trouve sa justification dans les théorèmes de la limite centrale (cf. annexe). [15] montre que l'on peut trouver une matrice unitaire U telle que les composantes du vecteur $U\hat{V}_{nor}$ soient indépendantes et de variances égales à 0 ou 1. En appliquant le théorème de la limite centrale, [15] déduit alors que les p premières composantes de $U\hat{V}_{nor}$ suivent asymptotiquement la loi normale standard. Ceci justifie la comparaison systématique des résidus à la loi normale (voir plus loin) et il devient alors naturel de supposer que les observations sont elles-mêmes gaussiennes, ne serait-ce que pour pouvoir établir les tests présentés ici. Ce résultat étant asymptotique, il est valable en toute rigueur pour une redondance $n-p$ élevée.

Est également évoqué dans cette partie le lien entre estimation par moindres carrés et estimation par maximum de vraisemblance.

3.2.1 Propriétés des estimateurs et tests statistiques

Une combinaison linéaire de variables aléatoires suivant des lois normales est elle-même une variable aléatoire suivant une loi normale donc

$$\hat{V} \hookrightarrow \mathcal{N}(0, \sigma_0^2[\Sigma_B - A(A^T\Sigma_B^{-1}A)^{-1}A^T]) \text{ et } \hat{X} \hookrightarrow \mathcal{N}(\dot{X}, \sigma_0^2(A^T\Sigma_B^{-1}A)^{-1})$$

On montre de plus que [5] [15] :

(i) $\frac{\hat{\sigma}_0^2}{\sigma_0^2}(n-p)$ suit une loi du χ^2 à $n-p$ degrés de liberté ;

(ii) $\frac{\Gamma_i(\hat{V})}{\sqrt{\hat{\sigma}_0^2}\sqrt{\Gamma_{ii}(\Sigma_B - A(A^T\Sigma_B^{-1}A)^{-1}A^T)}}$ suit une loi de Student à $n-p$ degrés de liberté.

(iii) $\frac{\Gamma_i(\hat{X})-\Gamma_i(\dot{X})}{\sqrt{\hat{\sigma}_0^2\Gamma_{ii}((A^T\Sigma_B^{-1}A)^{-1})}}$ suit une loi de Student à $n-p$ degrés de liberté, sous l'hypothèse H_0 du test correspondant (cf. ci-après).

Sur la base de ces propriétés, on peut bâtir des tests pour juger de la qualité de l'estimation effectuée.

Tests sur les résidus

Le premier test pouvant être effectué sur les résidus est de vérifier qu'ils suivent bien une loi normale. Pour ce faire, on peut utiliser des tests d'ajustement à la loi normale tels que le test du χ^2 ou le test de Kolmogorov

5. $\Gamma_i(X)$ désigne la $i^{\text{ème}}$ composante du vecteur X et $\Gamma_{ij}(A)$ désigne la composante a_{ij} de la matrice A.

[6].

La deuxième caractéristique importante du vecteur des résidus est qu'il est de moyenne nulle à condition que les observations soient non biaisées. Ainsi, on peut bâtir un test de détection des erreurs de mesure sur la propriété (ii) précédente. Pour ce test, les hypothèses nulle et alternative H_0 et H_1 sont définies par

$$\begin{cases} H_0 & : & \mathbb{E}(\Gamma_i(\hat{V})) = 0 \\ H_1 & : & \mathbb{E}(\Gamma_i(\hat{V})) \neq 0 \end{cases}$$

Ainsi, se fixant un risque de première espèce α, si $t_{n-p}(1-\frac{\alpha}{2})$ désigne le fractile $1-\frac{\alpha}{2}$ de la loi de Student à $n-p$ degrés de liberté [6], on accepte H_0 au seuil α si

$$\frac{|\Gamma_i(\hat{V})|}{\sqrt{\hat{\sigma}_0^2}\sqrt{\Gamma_{ii}(\Sigma_B - A(A^T\Sigma_B^{-1}A)^{-1}A^T)}} \leq t_{n-p}(1-\frac{\alpha}{2})$$

Tests sur les paramètres compensés

On peut tester la significativité des paramètres à partir de la propriété (iii) précédente. En effet, on peut considérer qu'un paramètre estimé $\Gamma_i(\hat{X})$ n'est pas significatif (donc, n'apporte rien au modèle) si

- $\Gamma_i(\hat{X}) = 0$ dans le cas d'un modèle directement linéaire ;

- $\Gamma_i(\hat{X}) = \Gamma_i(X_0)$ dans le cas d'un modèle obtenu par linéarisation.

Ainsi, pour ce test, les hypothèses nulle et alternative H_0 et H_1 sont définies par [7]

$$\begin{cases} H_0 & : & \Gamma_i(\hat{X}) = \Gamma_i(X_0) \\ H_1 & : & \Gamma_i(\hat{X}) \neq \Gamma_i(X_0) \end{cases}$$

Ainsi, se fixant un risque de première espèce α, on accepte H_0 au seuil α si

$$\frac{|\Gamma_i(\hat{X}) - \Gamma_i(X_0)|}{\sqrt{\hat{\sigma}_0^2\Gamma_{ii}((A^T\Sigma_B^{-1}A)^{-1})}} \leq t_{n-p}(1-\frac{\alpha}{2})$$

6. La loi de Student étant symétrique ([6]), on coupe le risque de première espèce en deux.

7. $\Gamma_i(X_0) = 0$ dans le cas d'un modèle directement linéaire.

Test sur le facteur unitaire de variance

On sait que, sous l'hypothèse gaussienne pour les observations, $\frac{\hat{\sigma}_0^2}{\sigma_0^2}(n-p)$ suit une loi du χ^2 à $n-p$ degrés de liberté. On sait également que, dans le cas idéal, $\sigma_0^2 = 1$ et qu'alors $\hat{\sigma}_0^2(n-p)$ suit une loi du χ^2 à $n-p$ degrés de liberté. Au seuil α fixé, on peut donc déterminer les deux quantités χ^2_{min} et χ^2_{max} telles que $P(\chi^2 \leq \chi^2_{min}) = P(\chi^2 \geq \chi^2_{max}) = \alpha/2$ et vérifier que la valeur estimée $\hat{\sigma}_0^2$ se trouve ou non dans l'intervalle $[\chi^2_{min}/(n-p), \chi^2_{max}/(n-p)]$. Si tel est le cas, on considérera que $\hat{\sigma}_0^2$ n'est pas significativement différent de 1 et qu'il n'est pas nécessaire de l'appliquer aux matrices de variance-covariance concernées ; sinon, on considérera qu'il est significativement différent de 1 et qu'il faut au contraire l'appliquer auxdites matrices.

3.2.2 Moindres carrés et maximum de vraisemblance

Dans le cas d'une distribution normale du vecteur des observations, l'estimateur des moindres carrés est l'estimateur du maximum de vraisemblance.

En effet, $B \hookrightarrow \mathcal{N}(A\dot{X}, \sigma_0^2\Sigma_B)$ donc la densité de probabilité de B est une loi normale multidimensionnelle à n dimensions dépendant des paramètres $\dot{X}$ soit

$$p_B(B, \dot{X}) = \frac{1}{(\sqrt{2\pi})^n \sigma_0^{2n} \det \Sigma_B} e^{-\frac{1}{2\sigma_0^2}(B-A\dot{X})^T.\Sigma_B^{-1}.(B-A\dot{X})}$$

La fonction log-vraisemblance est alors

$$\mathcal{L}(B, \dot{X}) = cte - \frac{1}{2\sigma_0^2}(B - A\dot{X})^T.\Sigma_B^{-1}.(B - A\dot{X})$$

Rendre $\mathcal{L}$ maximale revient donc à rendre minimale la quantité

$$(B - A\dot{X})^T.\Sigma_B^{-1}.(B - A\dot{X})$$

C'est le critère de l'estimation par moindres carrés.

3.3 Exemple numérique

On reprend le précédent exemple du rectangle en bois.

On suppose que les mesures que l'on a effectuées sont indépendantes les unes des autres et toutes de même écart-type inconnu σ_0, indicateur de la précision de l'instrument de mesure de longueur utilisé. On peut donc prendre pour matrice de poids I_{15} et reprendre les résultats de nos précédents calculs.

l	L	d
0,008	-0,031	0,014
-0,001	-0,018	0,004
-0,017	0,003	-0,008
-0,014	0,011	-0,019
0,021	0,029	0,017

TABLE 3.2 – Résidus de mesures.

Le vecteur des résidus est donné dans le tableau (3.2)
À partir de ces résidus, on peut calculer le facteur unitaire de variance :

$$\hat{\sigma}_0^2 = \frac{\hat{V}^T.\hat{V}}{n-p} \simeq \frac{0,0042}{13} \simeq 3,22.10^{-4}$$

L'inverse de la matrice normale est

$$\begin{pmatrix} 0,163 & -0,048 \\ . & 0,137 \end{pmatrix}$$

Ainsi, la précision sur l'estimation $\delta\hat{l}$ est de $\sqrt{0,163 \times 3,22.10^{-4}}$ soit environ 7,2 mm et celle de $\delta\hat{L}$ est de $\sqrt{0,137 \times 3,22.10^{-4}}$ soit environ 6,6 mm, soit

$$\begin{pmatrix} \hat{l} \\ \hat{L} \end{pmatrix} \simeq \begin{pmatrix} 1,642 \pm 0,0072 \\ 2,047 \pm 0,0066 \end{pmatrix}$$

Reste à voir si ces valeurs sont significatives. On se fixe un risque de première espèce de 4%. Le fractile $t_{13}(0.98)$ de la loi de Student à 13 degrés de liberté vaut 2,650 [6]. Les valeurs tests pour l et L sont respectivement $\frac{0,042}{0,0072} \simeq 5,83$ et $\frac{0,033}{0,0066} \simeq 5,00$ donc supérieures au fractile $t_{13}(0.98)$. Ainsi les variations des dimensions du rectangle sont significatives au regard du test (sévère !) effectué.

On peut enfin remarquer que la corrélation entre les deux estimations $\hat{l}$ et $\hat{L}$,

$$\frac{-0,048}{\sqrt{0,163 \times 1,137}} \simeq -0,32$$

est faible. Les variables estimées $\hat{l}$ et $\hat{L}$ sont donc bien décorrélées lors de l'estimation. Cette remarque est importante. En effet, si deux variables estimées sont corrélées, si l'une des variables est mal estimée, il y a de fortes chances que l'autre le soit également. Dans la pratique, il est important, si

c'est possible, de décorréler les paramètres estimés. Dans notre cas, la corrélation entre l et L n'a aucune nature physique : elle provient uniquement des mesures de longueur de la diagonale d.

Chapitre 4

Techniques et remarques complémentaires

Sont présentées dans ce chapitre des techniques de calculs supplémentaires, propres à la méthode de compensation par moindres carrés, qui peuvent s'avérer très utiles en géodésie, voire dans d'autres disciplines.

Quelques remarques sur les limites d'utilisation de la méthode des moindres carrés sont également formulées.

4.1 Techniques de calculs complémentaires

4.1.1 Cumul d'équations normales. Équations de contraintes

Est présentée ici une méthode permettant de cumuler les observations relatives à un même ensemble de paramètres et donc d'augmenter la redondance, soit la quantité d'information utilisée pour déterminer les dits paramètres.

Cumul d'équations normales

On considère m équations d'observations indépendantes relatives au même vecteur de paramètres X :

$$AX = \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix} X \simeq B = \begin{pmatrix} B_1 \\ \vdots \\ B_m \end{pmatrix}$$

Les systèmes d'observations étant supposés indépendants, la matrice de variance-covariance du vecteur d'observations global est

$$\Sigma_B = \begin{pmatrix} \Sigma_{B_1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma_{B_m} \end{pmatrix}$$

chaque matrice Σ_{B_i} étant la matrice de variance-covariance du vecteur des observations B_i.

La matrice normale du système est

$$A^T\Sigma_B^{-1}A = \left(A_1^T \cdots A_m^T\right) . \begin{pmatrix} \Sigma_{B_1}^{-1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma_{B_m}^{-1} \end{pmatrix} . \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix}$$

$$A^T\Sigma_B^{-1}A = \left(A_1^T \cdots A_m^T\right) . \begin{pmatrix} \Sigma_{B_1}^{-1}A_1 \\ \vdots \\ \Sigma_{B_m}^{-1}A_m \end{pmatrix}$$

soit finalement

$$A^T\Sigma_B^{-1}A = \sum_{i=1}^{n} A_i^T\Sigma_{B_i}^{-1}A_i \tag{4.1}$$

On montre de la même façon que le second membre du système normal est

$$A^T\Sigma_B^{-1}B = \sum_{i=1}^{n} A_i^T\Sigma_{B_i}^{-1}B_i \tag{4.2}$$

de sorte que le système normal global s'écrive

$$\left(\sum_{i=1}^{n} A_i^T\Sigma_{B_i}^{-1}A_i\right)\hat{X} = \sum_{i=1}^{n} A_i^T\Sigma_{B_i}^{-1}B_i \tag{4.3}$$

Ce système résulte donc du "cumul" des sous-systèmes normaux. Cette méthode est très employée dans le cadre des calculs de géodésie spatiale, pour le positionnement ou pour déterminer les coefficients du champ de gravité, par exemple. En effet, ces calculs nécessitent d'avoir une bonne répartition spatio-temporelle des observations (liées directement au mouvement de satellites) pour bien décorréler les différents paramètres. Pour avoir cette répartition, il est nécessaire de cumuler les mesures effectuées sur plusieurs satellites différents et sur une certaine période de temps.

Équations de contraintes

Soit à résoudre le système linéaire $AX \simeq B$ avec $var(B) = \Sigma_B$. On suppose que l'on dispose d'une information *a priori* sur le vecteur des paramètres X (information indépendante des observations B) de la forme [1]

$$A_0 X = X_0 \text{ où } var(X_0) = \Sigma_{X_0}$$

On peut alors considérer que l'on dispose du système d'observations

$$\begin{pmatrix} A \\ A_0 \end{pmatrix} X \simeq \begin{pmatrix} B \\ X_0 \end{pmatrix} \text{ de poids } \begin{pmatrix} \Sigma_B^{-1} & 0 \\ 0 & \Sigma_{X_0}^{-1} \end{pmatrix}$$

Par cumul des équations normales (4.3), on obtient le système normalisé

$$(A^T \Sigma_B^{-1} A + A_0^T \Sigma_{X_0}^{-1} A_0) X = (A^T \Sigma_B^{-1} B + A_0^T \Sigma_{X_0}^{-1} X_0) \qquad (4.4)$$

L'équation normale $A_0^T \Sigma_{X_0}^{-1} A_0 X = A_0^T \Sigma_{X_0}^{-1} X_0$ est appelée équation de contrainte.

Exemple

Dans le cadre de la détermination d'un repère de référence terrestre, on peut utiliser des contraintes simples sur les positions des stations. On peut en effet contraindre une station à une position donnée par les équations

$$\begin{cases} x &= x_0 \pm \sigma_x \\ y &= y_0 \pm \sigma_y \\ z &= z_0 \pm \sigma_z \end{cases}$$

On distingue alors dans ce cas les contraintes lâches (σ grand) des contraintes fortes (σ faible).

Les équations de contraintes [2] peuvent être utilisées pour "compléter" des systèmes normaux non inversibles pour lesquels l'information dont on dispose est insuffisante pour déterminer les paramètres. Il faut tout de même être vigilant quant à leur utilisation : elles doivent rester compatibles avec les observations qui permettent de déterminer les paramètres et la physique des phénomènes observés. Par exemple, on ne peut contraindre une station terrestre à rester fixe au mm près sur un an si l'on sait que la tectonique des plaques fait bouger cette station de quelques cm par an !

1. On "contraint" X selon X_0.
2. On parle même couramment de contraintes.

4.1.2 Réduction de paramètres

On considère le système d'observations dont le vecteur des paramètres est scindé en deux :

$$(A_1|A_2)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \simeq B \text{ avec } var(B) = \Sigma_B$$

On cherche à résoudre ce système par moindres carrés selon la seule inconnue X_2 en obtenant la même solution que celle résultant de l'inversion du système pour le vecteur des paramètres global. Déterminons le système normal associé.

$$\left(\frac{A_1^T}{A_2^T}\right)\Sigma_B^{-1}(A_1|A_2)\begin{pmatrix} \hat{X}_1 \\ \hat{X}_2 \end{pmatrix} = \left(\frac{A_1^T}{A_2^T}\right)\Sigma_B^{-1}B$$

$$\left(\frac{A_1^T}{A_2^T}\right)(\Sigma_B^{-1}A_1|\Sigma_B^{-1}A_2)\begin{pmatrix} \hat{X}_1 \\ \hat{X}_2 \end{pmatrix} = \left(\frac{A_1^T\Sigma_B^{-1}B}{A_2^T\Sigma_B^{-1}B}\right)$$

$$\left(\begin{array}{c|c} A_1^T\Sigma_B^{-1}A_1 & A_1^T\Sigma_B^{-1}A_2 \\ \hline A_2^T\Sigma_B^{-1}A_1 & A_2^T\Sigma_B^{-1}A_2 \end{array}\right)\begin{pmatrix} \hat{X}_1 \\ \hat{X}_2 \end{pmatrix} = \left(\frac{A_1^T\Sigma_B^{-1}B}{A_2^T\Sigma_B^{-1}B}\right)$$

La première équation nous donne

$$\hat{X}_1 = (A_1^T\Sigma_B^{-1}A_1)^{-1}A_1^T\Sigma_B^{-1}B - (A_1^T\Sigma_B^{-1}A_1)^{-1}A_1^T\Sigma_B^{-1}A_2\hat{X}_2$$

En injectant ce résultat dans la deuxième équation, on obtient alors le système normal permettant d'estimer directement $\hat{X}_2$:

$$A_2^T\Sigma_B^{-1}[I-A_1(A_1^T\Sigma_B^{-1}A_1)^{-1}A_1^T\Sigma_B^{-1}]A_2\hat{X}_2 = A_2^T\Sigma_B^{-1}[I-A_1(A_1^T\Sigma_B^{-1}A_1)^{-1}A_1^T\Sigma_B^{-1}]B$$

Cette méthode est abondamment utilisée dans le cadre des calculs dynamiques de géodésie spatiale. En effet, lors de ces calculs, on est amené à estimer en même temps des paramètres caractéristiques de l'orbite d'un satellite ainsi que des paramètres purement géodésiques (positions de stations terrestres, coefficients du champ de gravité, etc). En général, seuls les paramètres de nature géodésique nous intéressent. On effectue alors une réduction par rapport aux paramètres orbitaux, ce qui présente en plus l'avantage de diminuer l'ordre des systèmes à inverser.

4.1.3 Analyse des composantes de la variance

Les méthodes d'analyse des composantes de la variance (méthodes du degré de liberté [12] et de Helmert [13]) sont utilisées lorsque l'on veut compenser des observations statistiquement hétérogènes.

Dans le contexte d'utilisation de ces méthodes [3], on suppose que le vecteur des observations est scindé en m vecteurs d'observations $B^T = (B_1^T, ..., B_m^T)$, ces m vecteurs étant deux à deux indépendants et leurs matrices de variance-covariance connues à un facteur multiplicatif près.

$$var[(B_1^T, ..., B_m^T)^T] = \Sigma_B = \begin{pmatrix} \sigma_1^2\Sigma_{B_1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_m^2\Sigma_{B_m} \end{pmatrix}$$

Le système à résoudre par la méthode des moindres carrés a donc la forme

$$AX + V = \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix} X + V = \begin{pmatrix} B_1 \\ \vdots \\ B_m \end{pmatrix}$$

La méthode de cumul des équations normales précédente permet de déterminer la matrice normale du système

$$N = \sum_{i=1}^{m} A_i^T \Sigma_{B_i}^{-1} A_i$$

ainsi que le vecteur second membre

$$K = \sum_{i=1}^{m} A_i^T \Sigma_{B_i}^{-1} B_i$$

On pose $\Omega = N^{-1}$.

Méthode du degré de liberté

Avec les notations précédentes, le vecteur des résidus se met sous la forme

$$\hat{V} = \begin{pmatrix} A_1\Omega K \\ \vdots \\ A_m\Omega K \end{pmatrix} - B = (D' - I)B = DB$$

où

$$D' = \begin{pmatrix} A_1\Omega A_1^T\Sigma_{B_1}^{-1} & \cdots & A_1\Omega A_m^T\Sigma_{B_m}^{-1} \\ \vdots & \ddots & \vdots \\ A_m\Omega A_1^T\Sigma_{B_1}^{-1} & \cdots & A_m\Omega A_m^T\Sigma_{B_m}^{-1} \end{pmatrix}$$

3. La formulation peut être encore plus générale dans le cadre de la méthode de Helmert [15].

et [4]

$$I = \begin{pmatrix} I_{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & I_{n_m} \end{pmatrix}$$

De plus, si

$$C_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_{n_i} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$\forall i$, $C_i^T \hat{V} = \hat{V}_i$, vecteur des résidus correspondant au vecteur des observations B_i.

On cherche alors à calculer

$$\mathbb{E}(\Sigma_{B_i}^{-1} C_i^T \hat{V} \hat{V}^T C_i) = \Sigma_{B_i}^{-1} C_i^T D \mathbb{E}(BB^T) D^T C_i$$

$C_i^T D \mathbb{E}(BB^T) D^T C_i$ n'est autre que le bloc (i,i) de la matrice $D\mathbb{E}(BB^T)D^T$. Or

$$\mathbb{E}(BB^T) = \Sigma_B + A\dot{X}\dot{X}^T A^T$$

et [5]

$$D = (A(A^T \Sigma_{B_u}^{-1} A)^{-1} A^T Sigma_{B_u}^{-1} - I_n)$$

soit $DA\dot{X}\dot{X}^T A^T = 0$. Ainsi, il faut déterminer le bloc (i,i) de la matrice $D\Sigma_B D^T$. On montre que ce bloc est égal à

$$\Gamma_{(i,i)}(D\Sigma_B D^T) = \sum_{k=1}^{m} \sigma_k^2 A_i \Omega A_k^T \Sigma_{B_k}^{-1} A_k \Omega A_i^T + \sigma_i^2 \Sigma_{B_i} - 2\sigma_i^2 A_i \Omega A_i^T$$

A ce stade du calcul, on fait alors l'hypothèse restrictive suivante. On suppose que les composantes de la variance σ_i^2 diffèrent très peu d'un même facteur de variance σ_0^2. Ainsi,

4. $\forall i$, $n_i = \dim(B_i)$.

5. $n = \sum_i n_i$. De plus, Σ_{B_u} désigne la matrice déduite de la matrice Σ_B en égalant tous les σ_i à 1.

$$\Gamma_{(i,i)}(D\Sigma_B D^T) = \sigma_0^2[\Sigma_{B_i} - A_i \Omega A_i^T]$$

soit

$$\Sigma_{B_i}^{-1} C_i^T D \mathbb{E}(BB^T) D^T C_i = \sigma_0^2[I_{n_i} - \Sigma_{B_i}^{-1} A_i \Omega A_i^T]$$

et, en prenant la trace des deux membres de l'égalité [6],

$$\mathbb{E}(\hat{V}_i^T \Sigma_{B_i}^{-1} \hat{V}_i) = \sigma_0^2[n_i - tr(\Omega N_i)]$$

Un estimateur sans biais de σ_0^2 fondé sur le $i^{\text{ème}}$ groupe d'observations B_i est donc

$$\hat{\sigma_0^2}^i = \frac{\hat{V}_i^T \Sigma_{B_i}^{-1} \hat{V}_i}{n_i - tr(\Omega N_i)}$$

Cette expression est analogue à l'expression (3.6) définissant le facteur unitaire de variance. Le nombre de degrés de liberté "$n - p$" vaut ici $n_i - tr(\Omega N_i)$.

Du fait de l'hypothèse restrictive qui est faite sur les composantes de la variance, cette méthode est surtout utilisée pour des tests d'égalité de ces composantes [15]. Néanmoins, dans la pratique, on constate que les valeurs fournies par cette méthode sont cohérentes avec celles données par la méthode de Helmert ; on peut donc l'utiliser comme méthode d'estimation des composantes de la variance.

Méthode de Helmert

On résout le système $B = A\dot{X} + V$ par la méthode des moindres carrés qui donne l'estimation $\hat{X} = \Omega A^T \Sigma_{B_u}^{-1} B$ soit

$$\hat{X} - \dot{X} = \Omega A^T \Sigma_{B_u}^{-1} V$$

et le vecteur des résidus est donné par

$$\hat{V} = (A\Omega A^T \Sigma_{B_u}^{-1} - I_n)V$$

Tenant compte de la décomposition du vecteur des observations, on a alors

$$\begin{pmatrix} \hat{V}_1 \\ \vdots \\ \hat{V}_m \end{pmatrix} = \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix} \Omega A^T \Sigma_{B_u}^{-1} V - \begin{pmatrix} V_1 \\ \vdots \\ V_m \end{pmatrix}$$

6. N_i désigne la matrice normale associée au vecteur d'observations B_i.

soit $\forall i$, $\hat{V}_i = A_i \Omega A^T \Sigma_{B_u}^{-1} V - V_i$.

$$\begin{aligned} \hat{V}_i^T \Sigma_{B_i}^{-1} \hat{V}_i &= (A_i \Omega A^T \Sigma_{B_u}^{-1} V - V_i)^T \Sigma_{B_i}^{-1} (A_i \Omega A^T \Sigma_{B_u}^{-1} V - V_i) \\ \hat{V}_i^T \Sigma_{B_i}^{-1} \hat{V}_i &= V^T \Sigma_{B_u}^{-1} A \Omega N_i \Omega A^T \Sigma_{B_u}^{-1} V + V_i^T \Sigma_{B_i}^{-1} V_i - 2 V_i^T \Sigma_{B_i}^{-1} A_i \Omega A^T \Sigma_{B_u}^{-1} V \end{aligned}$$

En prenant la trace des deux membres de l'égalité, on obtient

$$\hat{V}_i^T \Sigma_{B_i}^{-1} \hat{V}_i = tr(V_i^T \Sigma_{B_i}^{-1} V_i) + tr(\Omega A^T \Sigma_{B_u}^{-1} V V^T \Sigma_{B_u}^{-1} A \Omega N_i) - 2tr(\Omega A^T \Sigma_{B_u}^{-1} V V_i^T \Sigma_{B_i}^{-1} A_i)$$

Tenant compte de la structure de la matrice de variance-covariance des observations et prenant l'espérance de chaque membre de l'équation précédente, on obtient finalement

$$\mathbb{E}(\hat{V}_i^T \Sigma_{B_i}^{-1} \hat{V}_i) = [n_i - 2tr(\Omega N_i) + tr(\Omega N_i \Omega N_i)] \sigma_i^2 + \sum_{j=1, j \neq i}^{m} tr(\Omega N_j \Omega N_i) \sigma_j^2$$

Considérant alors $\hat{c}_i = \hat{V}_i^T \Sigma_{B_i}^{-1} \hat{V}_i$ et H la matrice symétrique de terme général $h_{i,j} = \delta_i^j [n_i - 2tr(\Omega N_i)] + tr(\Omega N_j \Omega N_i)$, la résolution de l'équation de Helmert

$$H. \begin{pmatrix} \hat{\sigma_1}^2 \\ \vdots \\ \hat{\sigma_m}^2 \end{pmatrix} = \begin{pmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_m \end{pmatrix}$$

fournit des estimateurs sans biais des composantes de la variance σ_i^2.

L'équation de Helmert n'impose pas aux composantes de la variance d'être positives. Elle peut donc fournir des composantes négatives. On peut dans ce cas lui préférer la précédente méthode du degré de liberté ou résoudre l'équation de Helmert sous contraintes de positivité, en utilisant l'algorithme du simplexe modifié, par exemple (voir la deuxième partie).

Ces méthodes d'analyse des composantes de la variance sont utilisées, par exemple, dans le processus de calcul du repère de référence terrestre international. En effet, ce calcul repose sur la combinaison de jeux de coordonnées (positions/vitesses) de stations terrestres issus de techniques d'observations et de calculs différentes donc statistiquement hétérogènes. L'analyse des composantes de la variance permet d'obtenir une meilleure connaissance *a posteriori* des matrices de variance-covariance de chaque lot d'observations et donc d'opérer une repondération fine du vecteur des observations global.

4.2 Limites de la méthode d'estimation par moindres carrés

4.2.1 Linéarisation et minima locaux

La fonction φ (2.1) admet un minimum global sur Ω. Or, la linéarisation et les hypothèses utilisées permettent uniquement de raisonner sur des minima locaux (stricts en l'occurrence). Le minimum global de φ est certes un minimum local avant tout mais si la linéarisation s'effectue au voisinage d'un autre minimum local (s'il en existe), la méthode risque de fournir comme solution ce minimum local ou de ne pas converger en cas d'itérations successives.

Néanmoins il existe des hypothèses rigoureuses permettant de s'assurer de l'obtention d'un minimum global et, donc, de la linéarisation autour d'une solution approchée correspondant à ce minimum global. Ces hypothèses supposent que l'on se restreigne à un hyperrectangle convexe et compact de $\mathbb{R}^p$. Pour plus de détails, consulter [2].

4.2.2 Erreurs de mesures

Les exemples qui suivent illustrent le comportement erratique de l'estimateur des moindres carrés en présence d'observations erronées. Ces erreurs peuvent venir d'un biais sur les observations (on parle alors d'erreur systématique) ou d'une erreur purement aléatoire, c'est-à-dire, une observation peu probable au regard de la loi de probabilité supposée des observations.

Observations biaisées

On veut faire passer une fonction de la forme

$$x \mapsto \frac{1}{a_1 t + a_2}$$

au travers d'un nuage de points donné. Les observations que l'on utilise ne sont pas de moyenne nulle : elles sont biaisées. On effectue alors deux estimations différentes. Dans la première estimation, on estime les deux seuls paramètres a_1 et a_2. Dans la seconde estimation, on ajoute un biais inconnu au modèle. Les résultats des deux estimations sont donnés à la figure (4.1). Il est clair que l'estimation d'un biais améliore nettement les résultats de la compensation par moindres carrés.

La figure (4.2) montre les résidus pour les deux estimations précédentes. Là-encore, on note une nette amélioration du calcul dans le cas où l'on ajoute un biais au modèle. De plus, le fait que les résidus soient de moyenne nulle est plus flagrant dans le deuxième cas.

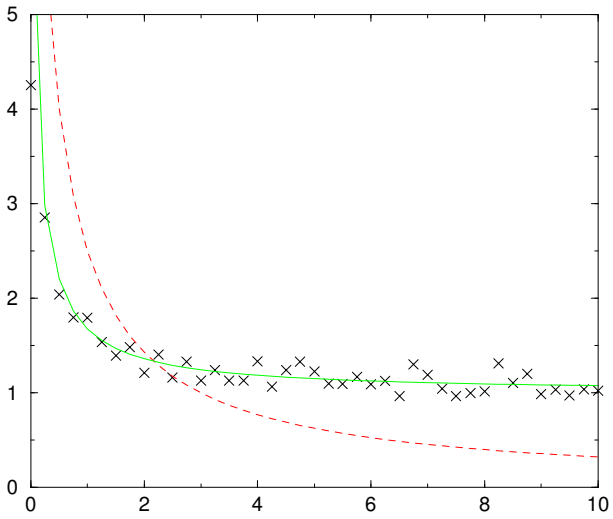

FIGURE 4.1 – Résultats des estimations. *Courbe discontinue* : sans estimation d'un biais. *Courbe continue* : avec estimation d'un biais.

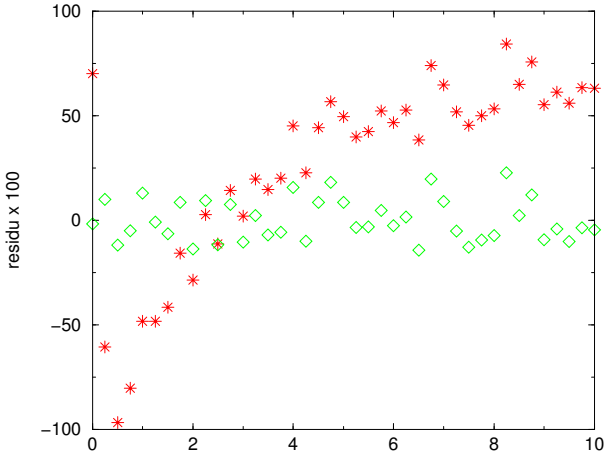

FIGURE 4.2 – Résidus des estimations. *Etoiles* : sans estimation d'un biais. *Losanges* : avec estimation d'un biais.

Faute de mesure

On reprend l'exemple introductif de la droite de régression linéaire. On cherche donc à faire passer au mieux une droite d'équation $b = a_1 t + a_2$ au travers d'un nuage de points. On effectue deux estimations des paramètres a_1 et a_2. Dans la deuxième estimation, une faute de mesure s'est "glissée" dans les observations.

Les résultats (figure (4.3)) montrent bien le comportement erratique de l'estimation par moindres carrés en présence d'une observation aberrante. Ceci est confirmé par la comparaison des valeurs absolues des résidus de chaque compensation donnée en figure (4.4).

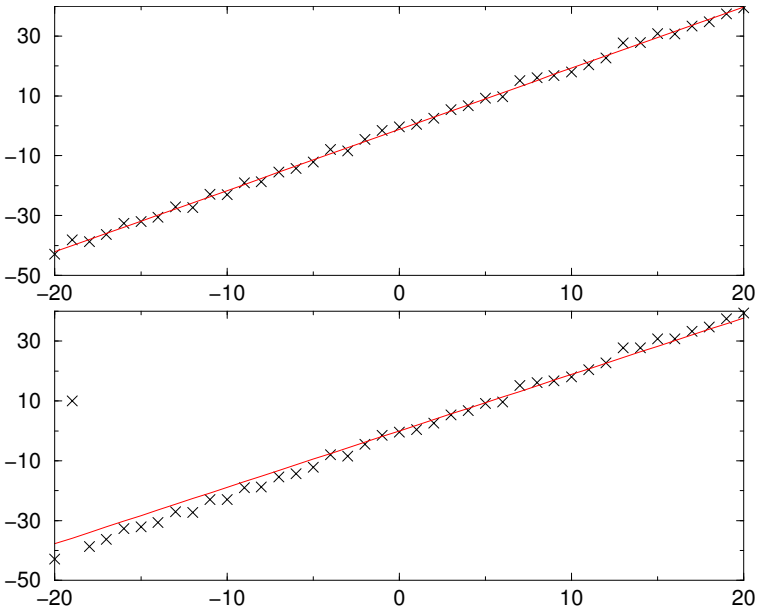

FIGURE 4.3 – Deux estimations. *En haut* : sans erreur. *En bas* : avec erreur.

Remarques

Généralement, on dit d'un estimateur qu'il est robuste s'il est peu sensible à la présence de valeurs aberrantes dans l'échantillon dont il est fonction. Les deux exemples précédents montrent assurément que l'estimateur des moindres carrés n'est pas un estimateur robuste.

Dans le cas de la présence d'un biais dans les observations, tous les résidus de mesures sont dégradés. Dans le cas de la faute de mesure, le résidu de l'observation comportant la faute "explose" littéralement et entraîne les

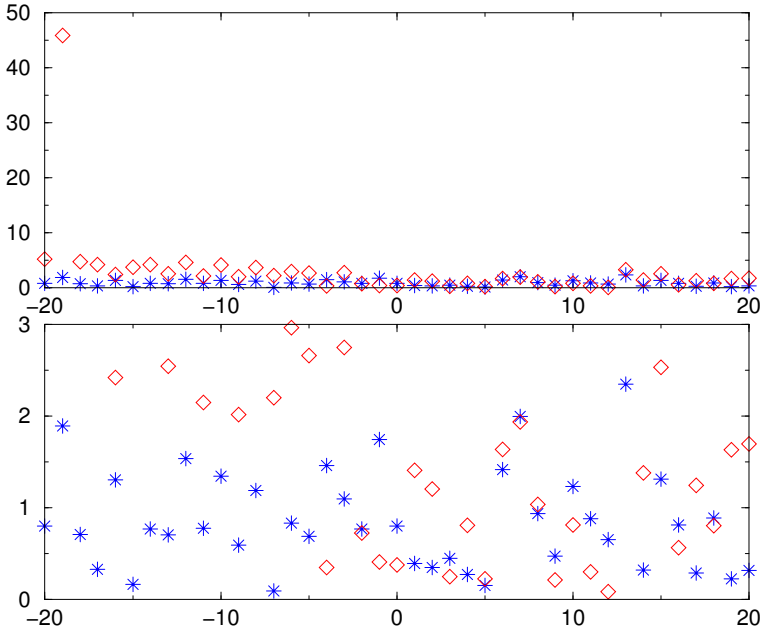

FIGURE 4.4 – Résidus des estimations. *Etoiles* : sans erreur. *Losanges* : avec erreur.

autres résidus proches de lui.

Les erreurs de mesure peuvent ainsi être détectées par l'intermédiaire du test statistique sur les résidus précédemment évoqué. Mais il faut bien garder à l'esprit qu'un résidu dégradé peut avoir été contaminé par une erreur n'affectant pas l'observation à laquelle il est lié.

Pour des observations biaisées, l'estimation peut être améliorée en considérant le(s) biais comme inconnue(s) du problème de compensation. Dans le cas d'une valeur aberrante, il faut ôter cette dernière du vecteur des observations.

Dans tous les cas, il faut bien garder à l'esprit que des lacunes du modèle peuvent également être sources d'erreurs détectables au travers des résidus. Il peut alors se poser le problème de distinguer les erreurs provenant des observations et celles provenant des manques du modèle, ce qui n'est pas toujours aisé.

Deuxième partie

Autres méthodes d'estimation : deux exemples

Chapitre 5

Estimation robuste : exemple du simplexe

Comme nous l'avons précédemment vu, l'un des problèmes majeurs du critère des moindres carrés est sa grande sensibilité aux erreurs de mesure. Il existe différentes méthodes d'optimisation robuste (cf. [15]). J'ai choisi de présenter ici l'estimation en norme $\mathcal{L}_1$ reposant sur l'algorithme du simplexe modifié, méthode fiable et assez rapide d'après [3].

5.1 Présentation générale de l'algorithme du simplexe

L'algorithme du simplexe nous vient de G.B. Dantzig, qui l'a créé en 1947 pour résoudre un problème de programmation de l'armée américaine.

Dans le contexte d'utilisation de cet algorithme, on se pose le problème suivant. On veut trouver le vecteur X tel que la fonction $\psi(X) = C^T X$ soit maximale [1] sous les conditions $AX = B$ et $X \geq 0$. C et X sont deux vecteurs de $\mathbb{R}^n$, B un vecteur de $\mathbb{R}^m$ $(m \leq n)$ et A une matrice de $\mathcal{M}_{(m,n)}(\mathbb{R})$.

5.1.1 Définitions et remarques

La fonction ψ est appelée fonction objectif ou fonction coût (c'est l'équivalent de la fonction (2.1) pour la méthode des moindres carrés). Le vecteur X est appelé vecteur des paramètres et tout vecteur X^* vérifiant les conditions $AX = B$, $X \geq 0$ est une solution réalisable.

L'ensemble des contraintes $K = \{X \ / AX = B, X \geq 0\}$ est l'ensemble des solutions réalisables. Cet ensemble est un polytope. S'il est borné, on

1. ou minimale, ce qui revient alors à maximiser la fonction opposée $-\psi$...

parle de polyèdre. Si K n'est pas un polyèdre, on dit que le problème n'est pas borné et si $K = \emptyset$, on dit que le problème est infaisable (dans ce cas, en effet, il n'y a pas de solution réalisable).

Les ensembles $H_\psi = \{X \in \mathbb{R}^n \ / C^T X = \psi\}$ constituent une famille d'hyperplans. Le problème est donc de trouver la valeur extrêmale de ψ telle que l'hyperplan d'équation $C^T X = \psi$ ait une intersection non vide avec le polyèdre des contraintes. Si ce polyèdre est convexe [2], on montre [3] que cette intersection se trouve obligatoirement sur la surface du polyèdre ; c'est donc un sommet, une arête ou même une face. Ainsi, l'algorithme du simplexe est né des deux constatations suivantes :

- si le problème posé admet une solution optimale, au moins un des sommets de K est solution ;

- le polyèdre convexe K possède un nombre fini de sommets.

Un algorithme fini pour résoudre le problème existe donc : il suffit de comparer les valeurs de la fonction ψ en un nombre fini de points, les sommets de K.

5.1.2 Algorithme

On cherche donc à résoudre le problème tel qu'il a été précédemment posé.

On suppose que la matrice A est de plein rang ($rg(A) = m$). On peut alors en extraire une matrice carrée $M \in \mathcal{M}_{(m,m)}(\mathbb{R})$ inversible de sorte que l'on puisse écrire [3]

$$A = (\underbrace{M}_{m} \mid \underbrace{N}_{n-m})$$

et

$$X = (\underbrace{X_M^T}_{m} \mid \underbrace{X_N^T}_{n-m})^T$$

La condition $AX = B$ donne alors

$$X_M = M^{-1}B - M^{-1}NX_N = X_M^* - M^{-1}NX_N$$

En adoptant la même décomposition pour le vecteur C, on aboutit à

2. **Rappel** : un ensemble K est convexe si $\forall (A, B) \in K,\ [A, B] \subset K$.

3. On peut éventuellement permuter les colonnes de la matrice A et les lignes du vecteur X.

$$\psi(X) = C_M^T X_M + C_N^T X_N = C_M^T X_M^* + (C_N^T - C_M^T M^{-1} N) X_N$$

Les composantes du vecteur X_M sont les variables en base et celles du vecteur X_N les variables hors-base.

Une condition nécessaire pour que ψ soit maximale en X_0 est que X_0 soit un point critique de ψ, soit

$$\forall i \in \{1, \cdots, n\},\ \frac{\partial \psi}{\partial x_i}(X_0) = 0$$

C'est bien le cas pour les variables en base ; par contre, ce n'est pas le cas pour les variables hors-base. On peut alors imaginer faire entrer toutes les variables en base. C'est impossible dans la mesure où c'est le rang de la matrice A qui définit le nombre de variables en base. Une solution plus radicale est alors d'annuler toutes les variables hors-base. On montre alors que ψ est bien maximale ([3]) en X_0 et alors $\psi(X) = C_M^T X_M^*$.

Si ψ n'est pas encore maximisée, il existe une composante du vecteur $(C_N^T - C_M^T M^{-1} N)^T$ strictement positive. En effet, si ce n'est pas le cas, les composantes de ce vecteur sont toutes négatives ou nulles ; si on augmente la valeur d'une variable hors-base, soit cela ne change pas la valeur de ψ, soit cela diminue la valeur de ψ. Toutes les variables hors-base sont donc nulles et ψ est maximale, ce qui est contradictoire.

Parmi les composantes strictement positives de $(C_N^T - C_M^T M^{-1} N)^T$, on choisit la plus grande (on suppose que c'est la $j^{\text{ème}}$). Ainsi la variable hors-base correspondante (x_j) que l'on va augmenter en l'entrant en base est celle qui va donner l'accroissement maximal à ψ. Il faut donc donner à x_j la valeur maximale qui lui est permise.

On note $y = M^{-T} C_M$. Si a_j désigne le vecteur colonne de la matrice N correspondant à la variable x_j, la composante strictement positive de $(C_N^T - C_M^T M^{-1} N)^T$ correspondante est [4] $\Gamma_j(C_N) - y^T a_j$.

L'accroissement de la variable x_j est limité par les m conditions $X_M \geq 0$ qui donnent

$$X_M = X_M^* - (M^{-1} a_j) x_j - \sum_{1 \leq i \leq n-m,\, i \neq j} (M^{-1} a_i) x_i \geq 0$$

4. $\Gamma_i(Y)$ désigne la $i^{\text{ème}}$ composante du vecteur Y.

Pour tout indice $i \neq j$, on pose $x_i = 0$: on se place ainsi dans des conditions proches de celles correspondant à l'extremum de ψ et les m conditions deviennent alors

$$X_M^* - (M^{-1}a_j)x_j \geq 0$$

On prend alors la valeur

$$x_j^* = \inf_{k\in\{1,\cdots,m\}} \left\{ \frac{\Gamma_k(X_M^*)}{\Gamma_k(M^{-1}a_j)},\ \frac{\Gamma_k(X_M^*)}{\Gamma_k(M^{-1}a_j)} > 0 \right\}$$

En effet, le vecteur X_M^* constitue, à chaque itération, une solution admissible donc toutes ses composantes sont positives. Ainsi, si $\Gamma_k(M^{-1}a_j) > 0$, la valeur obtenue est une limite supérieure positive pour x_j, sinon une limite inférieure négative dont on n'a pas à tenir compte.

La variable en base sortante (qui va donc remplacer x_j hors de la base) est la composante de X_M pour laquelle [5]

$$\Gamma_k(X_M^*) - x_j^*\Gamma_k(M^{-1}a_j) = 0$$

La contribution numérique de la variable sortante x_k à la fonction coût ψ demeure inchangée. Par contre, la valeur de cette fonction est augmentée de $(\Gamma_j(C_N) - y^T a_j)x_j^*$.

Pour préparer l'itération suivante, il faut actualiser les données du problème, soit :

(1) Permuter les colonnes k et j respectivement des matrices M et N.

(2) Modifier le vecteur des variables en base selon

$$\left\{ \begin{array}{rcl} \Gamma_k(X_M^*) & \rightarrow & x_j^* \\ \Gamma_{i\neq k}(X_M^*) & \rightarrow & \Gamma_i(X_M^*) - x_j^*\Gamma_i(M^{-1}a_j) \end{array} \right.$$

L'algorithme prend fin lorsque toutes les composantes de $(C_N^T - C_M^T M^{-1}N)^T$ sont négatives.

5.1.3 Exemple numérique

Prenons un exemple numérique simple pour expliciter un peu plus l'algorithme du simplexe.

5. En effet, l'extremum est atteint en un vecteur pour lequel toutes les variables hors-base sont nulles.

On veut minimiser la fonction des deux variables x et y, $\varphi(x,y) = x - 3y$ sous les conditions

$$\begin{cases} x & \geq & 0 \\ y & \geq & 0 \\ x+y & \leq & 40 \\ x+3y & \geq & 30 \\ 3x+y & \geq & 50 \end{cases}$$

Résolution graphique

Le problème précédent se résout facilement de manière graphique.

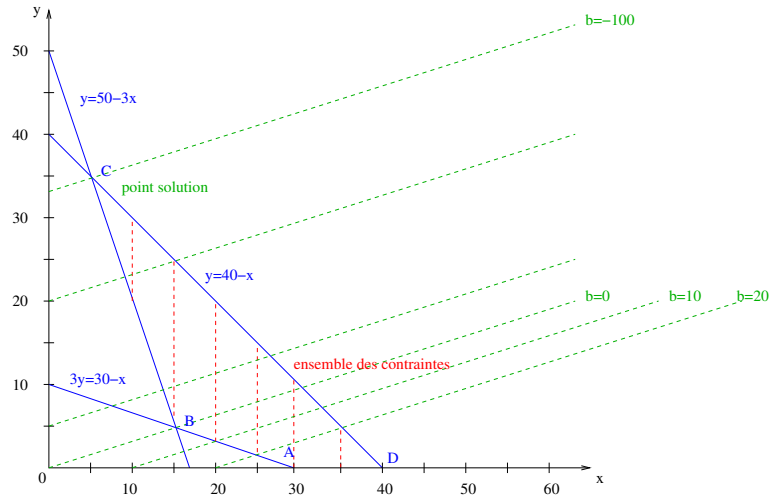

FIGURE 5.1 – Solution graphique du problème.

On trace les droites d'équations $x = 0$, $y = 0$, $y = 40 - x$, $3y = 30 - x$ et $y = 50 - 3x$. Ces droites déterminent l'ensemble des contraintes qui est bien un polyèdre convexe. Les droites d'équation $x = 3y + b$ sont des droites parallèles dont l'abscisse à l'origine b augmente à droite de la droite passant par le point B (pour laquelle $b = 0$) et diminue à gauche de cette droite. Le point solution au problème est donc le point C pour lequel $b = -100$.

Résolution par l'algorithme du simplexe

Minimiser la fonction $\varphi(x,y) = x - 3y$ revient à maximiser la fonction $\psi(x,y) = -x + 3y$. On introduit de plus trois nouvelles variables u, v et w vérifiant

$$\left\{\begin{array}{rcl} x+y+u & = & 40 \\ x+3y-v & = & 30 \\ 3x+y-w & = & 50 \end{array}\right.$$

Ainsi, le problème revient à minimiser $\psi(X) = C^T X$, avec $X^T = (x, y, u, v, w)$ et $C^T = (-1, 3, 0, 0, 0)$ sous les conditions $AX = B$, $X \geq 0$ où

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 3 & 0 & -1 & 0 \\ 3 & 1 & 0 & 0 & -1 \end{pmatrix} \text{ et } B = \begin{pmatrix} 40 \\ 30 \\ 50 \end{pmatrix}$$

Etape 1

On prend x, u et w comme variables en base et, donc, y et v comme variables hors-base. Les matrices M et N correspondantes sont donc

$$M = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 3 & 0 & -1 \end{pmatrix} \text{ et } N = \begin{pmatrix} 1 & 0 \\ 3 & -1 \\ 1 & 0 \end{pmatrix}$$

Le vecteur solution admissible $X_M^* = M^{-1}B$ est $X_M^* = (30, 10, 40)^T$. Les variables hors-base y et v sont nulles et $\psi(X) = -30$: on se trouve au point A (cf. figure (5.1)).

Le calcul donne $(C_N^T - C_M^T M^{-1} N)^T = (6, -1)$; c'est donc la variable y qui doit entrer en base. L'accroissement de cette variable est limité par les conditions $X_M^* - (M^{-1}a_1)y \geq 0$ qui donnent

$$\left\{\begin{array}{rcl} 30-3y & \geq & 0 \\ 10+2y & \geq & 0 \\ 40-8y & \geq & 0 \end{array}\right.$$

On prend donc $y = 5$ et c'est la variable w qui doit sortir de la base. La nouvelle solution admissible est donc

$$X_M^* = \begin{pmatrix} x & \rightarrow & 30 - 3 \times 5 = 15 \\ u & \rightarrow & 10 + 2 \times 5 = 20 \\ y & \rightarrow & 5 \end{pmatrix}$$

Pour les variables hors-base, v reste nulle et w passe à 0. Quant à la fonction objectif ψ, elle passe de la valeur -30 à $-30 + 6 \times 5 = 0$. On se trouve donc au point B du polyèdre des contraintes (cf. figure (5.1)).

Etape 2

Les variables en base sont x, u et y et w et v constituent les variables hors-base. Les matrices M et N correspondantes sont donc

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 3 \\ 3 & 0 & 1 \end{pmatrix} \text{ et } N = \begin{pmatrix} 0 & 0 \\ 0 & -1 \\ -1 & 0 \end{pmatrix}$$

Le calcul donne $(C_N^T - C_M^T M^{-1} N)^T = (-\frac{3}{4}, \frac{5}{4})$; c'est donc la variable v qui doit entrer en base. L'accroissement de cette variable est limité par les conditions $X_M^* - (M^{-1} a_2)v \geq 0$ qui donnent

$$\begin{cases} 15 - \frac{1}{8}v & \geq & 0 \\ 20 - \frac{1}{4}v & \geq & 0 \\ 5 + \frac{3}{8}v & \geq & 0 \end{cases}$$

On prend donc $v = 80$ et c'est la variable u qui doit sortir de la base. La nouvelle solution admissible est donc

$$X_M^* = \begin{pmatrix} x & \rightarrow & 15 - \frac{1}{8} \times 80 = 5 \\ v & \rightarrow & 80 \\ y & \rightarrow & 5 + \frac{3}{8} \times 80 = 35 \end{pmatrix}$$

Pour les variables hors-base, w reste nulle et u passe à 0. Quant à la fonction objectif ψ, elle passe de 0 à $0 + \frac{5}{4} \times 80 = 100$. On se trouve donc au point C du polyèdre des contraintes (cf. figure (5.1)).

Etape 3

Les variables en base sont x, v et y et w et u constituent les variables hors-base. Les matrices M et N correspondantes sont donc

$$M = \begin{pmatrix} 1 & 0 & 1 \\ 1 & -1 & 3 \\ 3 & 0 & 1 \end{pmatrix} \text{ et } N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ -1 & 0 \end{pmatrix}$$

Le calcul donne $(C_N^T - C_M^T M^{-1} N)^T = (-2, -5)$: l'algorithme est terminé et le point $C = (5, 35)$ est solution.

5.2 Le simplexe modifié : principe et exemple

Le simplexe modifié est un simplexe optimisé.

5.2.1 Principe

Il existe une autre façon d'utiliser l'algorithme du simplexe précédent. Elle repose sur l'utilisation de tableaux appelés dictionnaires ([3]). C'est sur cette formulation que repose l'algorithme du simplexe modifié.

On se place dans le contexte du modèle linéaire des moindres carrés. On cherche donc à résoudre le système $AX = B$ mais, cette fois-ci, avec un critère en norme $\mathcal{L}_1$. On ne cherche donc plus à minimiser la quantité quadratique $\sum_i (\sum_j a_{ij}x_j - b_i)^2$ mais $\sum_i |\sum_j a_{ij}x_j - b_i|$. Pour nous placer dans le contexte d'utilisation de l'algorithme du simplexe précédent, il nous faut définir une fonction objectif ainsi qu'un ensemble de contraintes. Pour ce faire, on définit les variables ω_i par $\forall i,\ \omega_i \geq |\sum_j a_{ij}x_j - b_i|$. Ainsi, on peut prendre comme fonction coût la fonction $\psi(X) = \sum_i \omega_i$ et le problème se formule de la manière suivante :

$$\begin{cases} \text{Minimiser } \psi(X) = \sum_i \omega_i \\ \quad \text{sous les contraintes } \omega_i \geq \sum_j a_{ij}x_j - b_i \\ \quad \text{et } \omega_i \geq b_i - \sum_j a_{ij}x_j \end{cases}$$

L'algorithme utilisé ici est l'algorithme 552 de I. Barrodale et F. D. K. Roberts.

Cet algorithme général permet de minimiser $\|AX - B\|_1$ sous les contraintes (non obligatoires) $CX = D$ et $EX \leq F$. Il utilise une formulation par dictionnaires et repose sur le principe suivant. A chaque itération, la variable entrante est celle qui correspond à la valeur maximale de $\sum_i a_{ij}$ de sorte que AX approche au mieux et le plus vite possible le vecteur B. Cette méthode permet le saut de points intermédiaires sur la surface du polyèdre des contraintes. Pour plus de détails, consulter [3].

5.2.2 Exemple numérique

On reprend ici le précédent exemple de la droite de régression. On considère donc un nuage de points au travers duquel on veut faire passer "au mieux" une droite. On effectue deux estimations : l'une pour un nuage de points sans valeur aberrante et une pour un nuage de points dans lequel on introduit une faute de mesure.

La figure (5.2) montre bien la robustesse de l'estimation par l'algorithme du simplexe modifié : la détermination de la droite de régression est

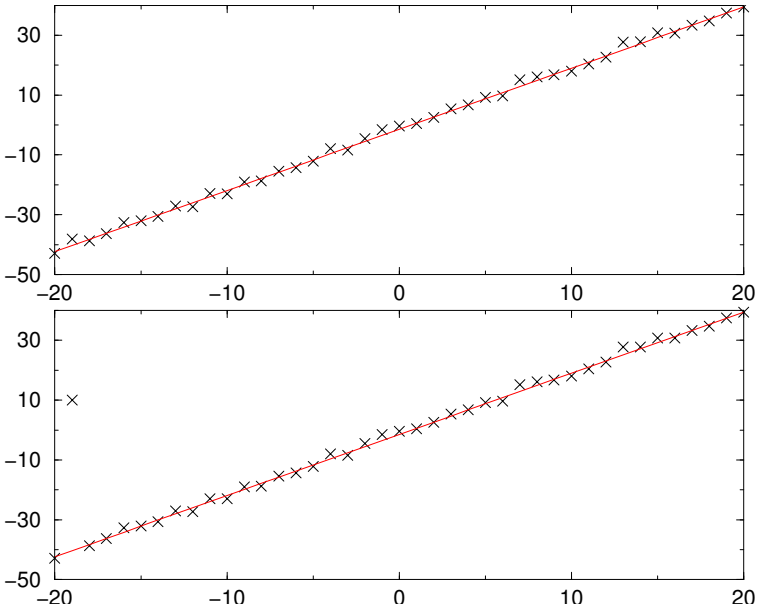

FIGURE 5.2 – *En haut* : nuage sans erreur. *En bas* : nuage avec erreur.

totalement insensible à la présence de la faute de mesure.

Ceci est confirmé par l'analyse des valeurs absolues des résidus de l'estimation présentées en figure (5.3).

Seul le résidu correspondant à l'observation erronée est modifié entre les deux estimations.

5.2.3 Remarques

L'exemple précédent est très parlant. En effet, du fait de la robustesse de la méthode, la droite "estimée" n'est pas modifiée par l'observation erronée et donc, seul le résidu d'estimation correspondant à cette valeur aberrante a une valeur importante. L'estimation par l'algorithme du simplexe modifié peut ainsi constituer un outil efficace de détection des fautes de mesure.

C'est également une méthode d'estimation efficace mais, reposant sur un processus itératif, elle peut s'avérer gourmande en temps de calcul et, d'autre part, elle ne fournit pas de grandeurs statistiques permettant de juger de la qualité des estimations comme c'est le cas pour le critère des moindres carrés.

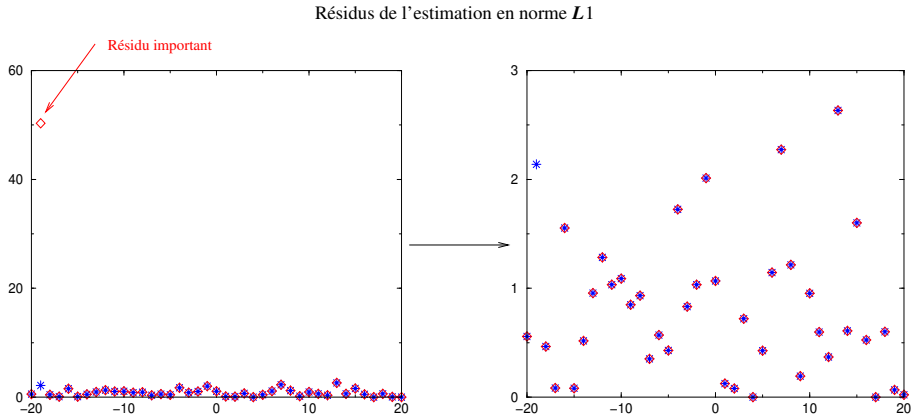

FIGURE 5.3 – *Etoiles* : estimation pour le nuage sans erreur. *Losanges* : estimation pour le nuage avec erreur.

Chapitre 6

Estimation stochastique : exemple des algorithmes génétiques

Il existe différentes méthodes d'optimisation stochastique (cf. [2] pour une liste non exhaustive) et, parmi ces méthodes, j'ai choisi de présenter ici les algorithmes génétiques.

La force des algorithmes génétiques par rapport à la méthode des moindres carrés est de pouvoir considérer des fonctions f non différentiables, présentant des minima locaux et ne possédant pas obligatoirement de forme analytique.

Ces algorithmes reposent sur les lois de la génétique et utilisent ces lois (sélection, croisement et mutation) pour améliorer la population des solutions éventuelles, population recherchée dans un espace dans lequel se trouve la solution [1] du problème d'optimisation. Cette technique est quelque peu élitiste puisque l'on cherche sans cesse l'amélioration de la population existante.

6.1 Présentation générale

Les algorithmes génétiques ont un déroulement relativement simple reposant sur trois étapes principales : l'étape d'évaluation de la population courante, l'étape de sélection et l'étape de croisement/mutation dont découle la population suivante (cf. figure (6.1)).

1. voire les solutions dans le cas d'une fonction présentant des minima locaux...

6.1.1 Etapes d'un algorithme génétique

Les algorithmes génétiques utilisent un codage du vecteur des paramètres x. Chaque composante est codée (le codage binaire est le plus couramment utilisé) et les codes ainsi obtenus sont concaténés en un seul vecteur appelé individu, génotype ou chromosome (voir figure (6.2)).

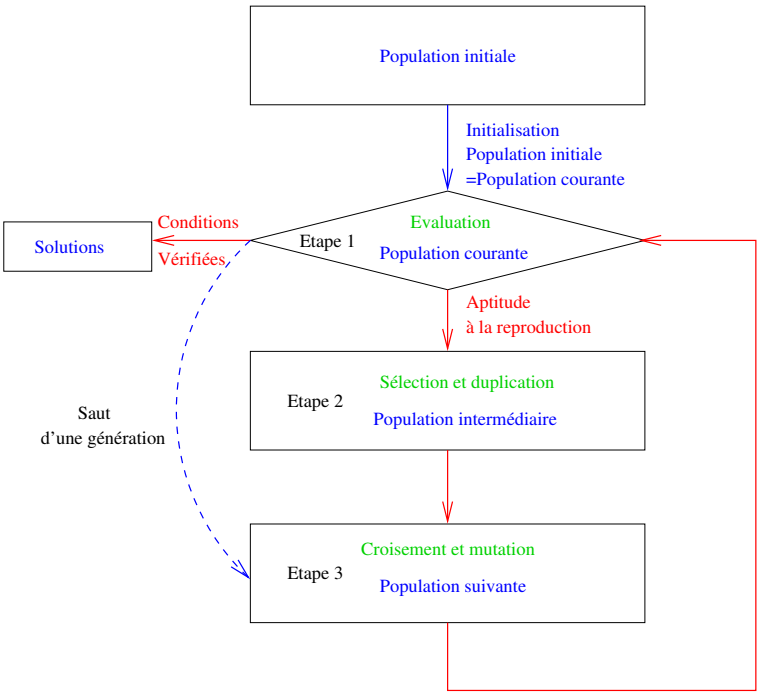

FIGURE 6.1 – Déroulement d'un algorithme génétique.

Les algorithmes génétiques travaillent ainsi sur des populations de chromosomes. Pour démarrer l'algorithme, il faut générer une population initiale, population qui constitue la population courante pour la première itération.

Dans une première étape, la population courante est évaluée à l'aide d'une fonction d'évaluation qui donne la performance de chaque individu de la population. Une fonction d'aptitude transforme alors ces performances individuelles en aptitudes à la reproduction. Intervient alors la deuxième étape (la phase de sélection) au cours de laquelle les meilleurs individus de la population courante sont recopiés en un certain nombre d'exemplaires dans la population intermédiaire. C'est sur cette dernière qu'opère la troisième et dernière étape, celle du croisement. Elle s'effectue à partir de deux chromosomes de la population intermédiaire qui vont "donner naissance" à deux enfants. La population issue du croisement peut ensuite subir des mutations

(modifications de certains gènes). De là résulte la population suivante qui constitue la nouvelle génération. Les itérations s'achèvent lorsque la population courante satisfait aux conditions d'arrêt.

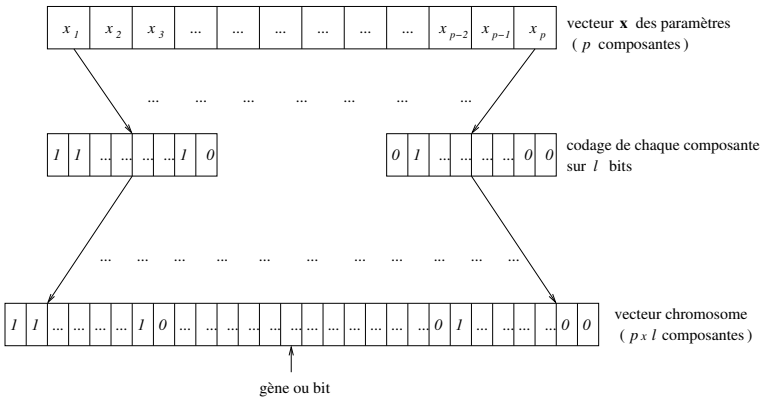

FIGURE 6.2 – Codage du vecteur des paramètres (exemple du codage binaire).

Il n'existe pas de démonstration mathématique rigoureuse de l'efficacité des algorithmes génétiques. Néanmoins, une interprétation par partitions de l'espace de recherche permet de comprendre intuitivement pourquoi ces algorithmes donnent de bons résultats et, entre autres, pour quelle raison on peut affirmer qu'ils "scannent" l'espace de recherche des paramètres de façon "intelligente".

6.1.2 Partitions par schémas de l'espace de recherche

[17] donne une explication de l'efficacité des algorithmes génétiques. Plaçons-nous dans le cas d'un codage binaire sur l bits (ou gènes). On peut, à l'aide de ce codage, représenter 2^l individus. A partir de ce codage, on définit les schémas comme des individus comportant des bits pouvant valoir indifféremment 0 ou 1 (on leur attribue la valeur *) [2]. L'ordre d'un schéma est le nombre de bits "non libres" du schéma [3]. Les schémas de même ordre et ayant les mêmes gènes fixés génèrent une partition de l'espace de recherche [4].

Un chromosome appartient à 2^l schémas différents (l'espace de recherche compris) et il existe 3^l schémas différents dans tout l'espace (cet

2. Par exemple, dans le cas d'un codage sur trois bits, le chromosome 011 appartient aux schémas ***, 0**, 01*, 0*1, *11, **1 et *1*.

3. Reprenant l'exemple précédent, le schéma *** est d'ordre 0, les schémas 01* et 0*1 sont d'ordre 2.

4. Pour le codage sur trois gènes, {0**,1**} et {00*,11*,10*,01*} sont des exemples de partitions de l'espace de recherche.

espace compris). On comprend ainsi l'intérêt de raisonner non pas sur un chromosome isolé mais sur une population de chromosomes afin de mieux couvrir l'ensemble des partitions de l'espace de recherche. De plus, on peut constater que le nombre de schémas représentés dans une population est nettement plus important que le nombre d'individus de cette population. Ainsi, lorsqu'un seul individu d'une population est évalué, on évalue déjà de nombreux schémas différents qui peuvent être concurrents. C'est ce parallélisme implicite des algorithmes génétiques qui les rend efficaces quant à la méthode de parcours de l'espace de recherche des paramètres.

Les opérateurs génétiques que sont la sélection, le croisement et la mutation permettent également d'optimiser ce parcours de l'espace de recherche des paramètres. En effet, ce sont eux qui permettent de répondre au dilemme "exploitation contre exploration" [14] : il est impératif d'exploiter les schémas prometteurs découverts tout en continuant d'explorer l'espace de recherche pour découvrir d'autres schémas.

6.1.3 Opérateurs génétiques

Sont décrits dans cette partie les points clefs des algorithmes génétiques qui en constituent ainsi également des points délicats.

Initialisation de l'algorithme

Pour alimenter la première boucle de l'algorithme génétique, il faut fournir une population initiale. Cette dernière peut être générée par tirage aléatoire des gènes des chromosomes ou avec des chromosomes solutions d'un problème correspondant à une fonction d'évaluation simplifiée [14].

Fonction d'évaluation

La fonction d'évaluation permet de déterminer la performance d'un individu, donc d'un vecteur de paramètres x donné. On peut imaginer utiliser une fonction d'évaluation issue d'autres méthodes d'optimisation. Par exemple, on peut considérer $\sum_{i=1}^{n}(f_i(x)-y_i)^2$ si l'on veut se rapprocher de la méthode des moindres carrés ou encore $\sum_{i=1}^{n}|f_i(x)-y_i|$ pour se rapprocher d'une estimation en norme $\mathcal{L}_1$. [7] propose une fonction d'évaluation rendant l'estimation plus robuste : $C - \text{médiane}((f_i(x)-y_i)^2)$, C étant une constante arbitraire qui permet de passer d'une minimisation à une maximisation.

C'est à l'aide de la fonction d'évaluation que fonctionne le premier opérateur d'évolution : la sélection.

Sélection

La sélection est l'opérateur déterminant la population intermédiaire (individus susceptibles de se reproduire) à partir de la population courante. Pour ce faire, elle utilise le plus souvent une fonction d'aptitude. Cette fonction transforme les performances individuelles en qualité des individus de la population courante les uns par rapport aux autres. Une façon simple de procéder est d'utiliser une fonction [5] d'aptitude a du type

$$a(\text{individu}_i) = \frac{e(\text{individu}_i)}{\bar{e}}$$

$\bar{e}$ désignant la moyenne des évaluations de tous les individus de la population courante. Il existe alors plusieurs méthodes de sélection.

On peut tout d'abord utiliser un tirage à la roulette ([14], [17]). On considère un disque duquel chaque individu de la population se voit attribuer une portion proportionnelle à son aptitude. Il suffit alors de procéder à des tirages aléatoires successifs avec replacement. On peut à ce titre utiliser le tirage aléatoire universel reposant sur une roulette externe avec des pointeurs également espacés qu'on ne fait tourner qu'une seule fois (cf. figure (6.3)).

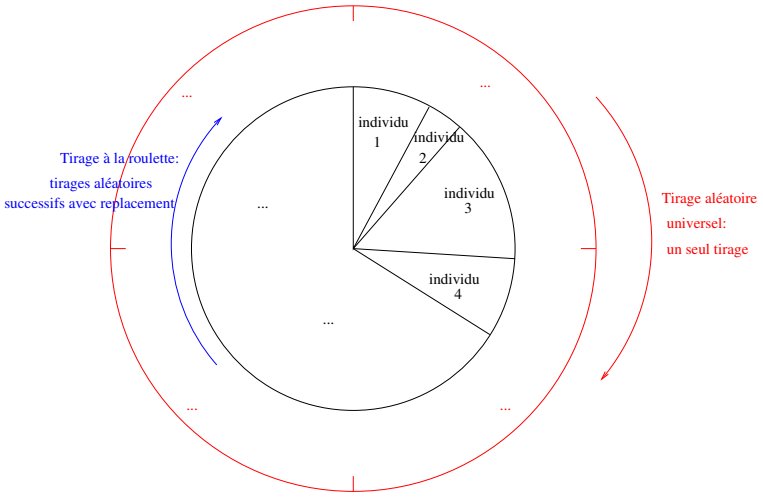

FIGURE 6.3 – Sélection par tirages à la roulette.

On peut également utiliser un tirage aléatoire avec le reste [17]. Si pour l'individu 1, $a = 1,58$, on place d'office un individu 1 dans la population intermédiaire et cet individu peut être copié une deuxième fois avec une probabilité de 0,58. Si pour l'individu 2, $a = 0,14$, il a une probabilité de

5. On note e la fonction d'évaluation.

0,14 d'être copié dans la population intermédiaire.

Enfin, on peut aussi effectuer une sélection par tournois [14]. On tire alors aléatoirement un nombre donné d'individus de la population courante et le meilleur est copié dans la population intermédiaire.

On peut, lors de la sélection, adopter une stratégie élitiste, considérant alors que tout individu de la population courante doit être remplacé, dans la population intermédiaire, par un individu meilleur.

La sélection est un opérateur d'évolution orienté vers l'exploitation. En effet, seuls sont *a priori* sélectionnés les meilleurs individus : la sélection tend à faire rester l'algorithme dans les schémas les plus prometteurs. Les opérateurs de croisement et de mutation permettent de poursuivre l'exploration de schémas encore inconnus.

Croisement

Des individus α et β de la population intermédiaire ont une probablité p_c d'être croisés et de donner lieu à deux enfants (voir figure (6.4)).

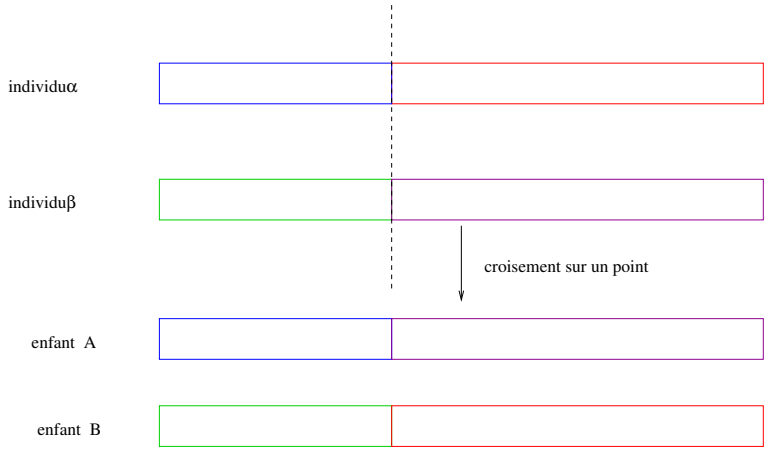

FIGURE 6.4 – Exemple de croisement sur un point.

On peut effectuer plusieurs types de croisement [17]. Le croisement sur un point est illustré par la figure (6.4). Pour le croisement sur deux points, les deux points de recombinaison sont choisis aléatoirement et les deux chromosomes échangent le segment situé entre ces deux points. C'est la forme de croisement la moins destructrice pour les schémas [6]. Enfin, on peut utiliser un croisement uniforme pour lequel les gènes hérités des individus

6. Ce croisement est alors plus orienté "exploitation" qu'"exploration".

parents sont tirés aléatoirement.

D'après [17], pour de petites populations, des opérateurs de croisement destructeurs (type croisement uniforme ou croisement sur plus de deux points) permettent de surmonter le faible niveau d'information contenu dans ces populations [7].

Une fois le croisement effectué, des mutations peuvent se produire.

Mutation

Chaque gène de chaque individu de la population "recombinée" peut muter avec une probabilité p_m. Il y a deux types de mutations [17] : soit un nouveau bit est généré aléatoirement, soit le bit mutant est changé d'office.

Cet opérateur est clairement orienté vers l'exploration et son principal intérêt est de préserver la diversité génétique. En effet, du fait d'une sélection trop stricte, l'algorithme peut converger prématurément vers une solution qui n'est pas satisfaisante. Dans ce cas, la mutation peut faire évoluer de nouveau la recherche vers des schémas inexplorés contenant une solution plus satisfaisante.

6.2 Exemple de traitement

L'exemple de traitement présenté ici est issu des travaux de recherche effectués par P. Berio, ancien ingénieur de recherche de l'équipe GMC (anciennement OCA/CERGA), et moi-même, au début des années 2000.

6.2.1 Présentation

Le cadre de ces travaux est la détermination, par une méthode de calcul géométrique, de séries temporelles de positions de stations laser à partir de mesures de télémétrie laser [4]. Dans ce contexte, on souhaite trouver une méthode d'estimation permettant de diminuer le pas d'échantillonnage de ces séries, limité pour le moment (en utilisant les satellites LAGEOS 1 et 2) à 30 jours, en gardant une bonne précision. On peut bien entendu utiliser la méthode des moindres carrés, qui donnent des résultats tout à fait satisfaisants quoique quelque peu bruités parfois (voir [4], figure (B14.4)).

On a voulu tester l'utilisation des algorithmes génétiques pour ce type de détermination. Pour ce faire, on a utilisé trois fonctions objectifs différentes : la fonction coût de la méthode des moindres carrés, celle de la méthode du

7. L'exploration prend alors le pas sur l'exploitation.

simplexe et, enfin, la médiane des carrés des composantes du vecteur des résidus. Les résultats sont fournis figure (6.5).

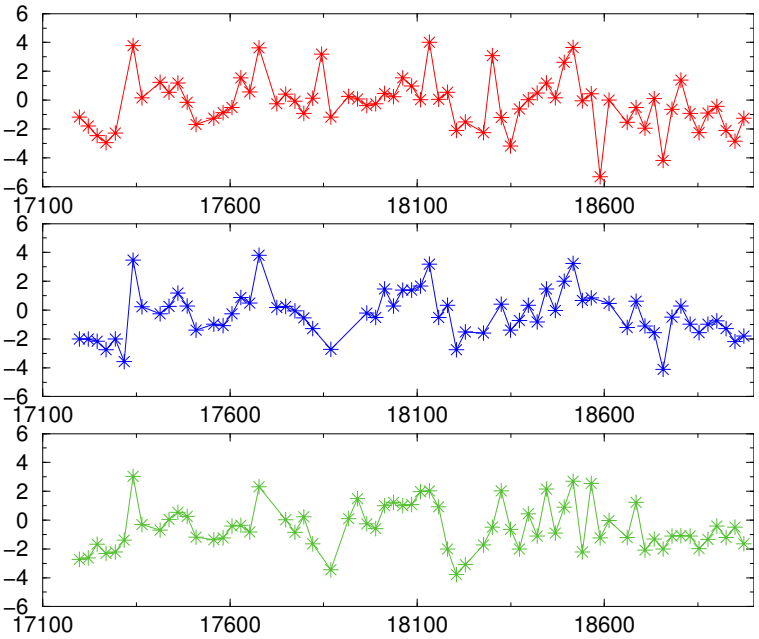

FIGURE 6.5 – Séries temporelles pour les trois fonctions coûts : "moindres carrés", "norme $\mathcal{L}_1$" et "médiane".

Les séries présentées ici sont des écarts verticaux (en cm) pour la station laser-satellites de Grasse par rapport à l'ITRF 2000, après correction de la tectonique des plaques, des marées terrestres ainsi que de la marée polaire selon les standards IERS 96. L'échantillonnage est de 24 jours. Les satellites géodésiques utilisés sont les deux LAGEOS.

Ces séries montrent un bon accord et sont également cohérentes avec les séries déterminées par moindres carrés [4]. Les points discordants sont des points pour lesquels le nombre de mesures laser disponibles est faible. Dans ce cas, toute méthode d'optimisation est quelque peu mise en défaut.

6.2.2 Perspectives

Pour diminuer encore l'échantillonnage temporel des séries en conservant une bonne précision, on peut imaginer cumuler les mesures de plusieurs satellites.

Les algorithmes génétiques, tout comme l'estimation par le simplexe,

ne permettent pas d'obtenir des grandeurs statistiques caractéristiques de la qualité des estimations. On peut ainsi envisager l'utilisation d'une méthode combinée dans laquelle les algorithmes génétiques permettraient de déterminer une solution approchée des paramètres ; le simplexe permettrait de détecter les mesures aberrantes et la méthode des moindres carrés, quant à elle, permettrait de déterminer des critères de qualité statistiques.

On peut aussi envisager l'utilisation d'autres méthodes d'optimisation globale telles que la méthode du recuit simulé ou la méthode Aliénor [2].

Troisième partie

Annexes

Annexe A

Rappels d'algèbre linéaire

Dans toute cette partie, $\mathbb{K}$ désigne un corps quelconque (en général, le corps des réels ou celui des complexes).

A.1 Projecteurs

A.1.1 Sous-espaces vectoriels supplémentaires

Soient E un $\mathbb{K}$ espace vectoriel de dimension finie, F et G deux sous-espaces vectoriels de E. F et G sont dits supplémentaires (et on note $E = F \oplus G$) ssi

$$\forall u \in E,\ \exists!(f,g) \in F \times G\ /u = f + g$$

Théorème A.1 *$E = F \oplus G \Leftrightarrow E = F + G$ et $F \cap G = \{0_E\}$*

Démonstration

Si $E = F \oplus G$,

$$\forall x \in E,\ \exists(f,g) \in F \times G\ /x = f + g$$

donc $E \subset F + G$. Or $F + G \subset E$ donc $E = F + G$. De plus, $\forall x \in F \cap G$,

$$x = \underbrace{x}_{\in F} + \underbrace{0_E}_{\in G} = \underbrace{0_E}_{\in F} + \underbrace{x}_{\in G}$$

La décomposition de x étant unique, $x = 0_E$ donc $F \cap G \subset \{0_E\}$ donc $F \cap G = \{0_E\}$.

Réciproquement, $E = F + G$ donc tout vecteur x de E se décompose en somme d'un vecteur de F et d'un vecteur de G. Si x admet deux telles décompositions

$$x = x_1 + y_1 = x_2 + y_2$$

alors,

$$\underbrace{x_1 - x_2}_{\in F} = \underbrace{y_2 - y_1}_{\in G} = 0_E$$

car $F \cap G = \{0_E\}$.

Théorème A.2 *Tout sous-espace vectoriel de E possède un supplémentaire.*

Démonstration

Le sous espace nul possède un supplémentaire : E. Si F est un sous-espace vectoriel non-nul de E alors toute base de F est un système libre de E et peut donc être complétée en une base de E. La base qui complète est une base d'un sous-espace supplémentaire de F.

A.1.2 Projections

Soient E un $\mathbb{K}$ espace vectoriel de dimension finie, F et G deux sous-espaces vectoriels supplémentaires de E. Soient $x \in E$ et $f + g$ son unique décomposition ($f \in F$ et $g \in G$). La projection sur F parallèlement à G est l'endomorphisme p de E ($p \in \mathcal{L}(E)$) défini par

$$\begin{cases} p & : & E \to F \\ & & x \mapsto f \end{cases}$$

$F = Im(p)$ et $G = \ker(p)$. On introduit également le projecteur q sur G parallèlement à F.

On a $p + q = Id_E$, $p \circ q = q \circ p = 0_{\mathcal{L}(E)}$ et $E = Im(p) \oplus Im(q)$.

Théorème A.3 *Soit* $p \in \mathcal{L}(E)$. *p projection* $\Longleftrightarrow p \circ p = p$.

Démonstration

La condition nécessaire est évidente. Démontrons la condition suffisante. Introduisons l'endomorphisme $q = Id_E - p$. $\forall x \in E$, *on peut écrire* $x = p(x) + q(x)$ *donc* $x \in Im(p) + Im(q)$ *et donc* $E = Im(p) + Im(q)$.

De plus, si $x \in Im(p) \cap Im(q)$,

$$\exists (y, z) \in Im(p) \times Im(q) \ / x = p(y) = z - p(z)$$

soit

$$p(x) = p(p(y)) = p(y) = x = p(z) - p(p(z)) = p(z) - p(z) = 0_E$$

d'où $Im(p) \cap Im(q) = \{0_E\}$. *Donc* $E = Im(p) \oplus Im(q)$. *Enfin,*

$$\forall x \in Im(q),\ \exists y\ / x = y - p(y) \Rightarrow p(x) = 0_E$$

et le théorème du rang donne

$$\dim E = rg(p) + \dim(ker(p)) = rg(p) + rg(q)$$

soit $Im(q) = ker(p)$ *et donc* $E = Im(p) \oplus ker(p)$. p *est donc bien une projection.*

A.1.3 Projections orthogonales

Soit E un $\mathbb{K}$ espace vectoriel euclidien de dimension n muni du produit scalaire $<,>$. Soit F un sous-espace de E. L'orthogonal de F est

$$F^{\perp} = \{u \in E/\ \forall f \in F < u, f >= 0\}$$

Si F est muni d'une base $\mathcal{B}$, $F^{\perp}$ est l'ensemble des vecteurs de E orthogonaux à tous les vecteurs de $\mathcal{B}$.

On note $\|.\|$ la norme induite sur E par $<,>$. Une base $(e_1, \cdots, e_n)$ de E est dite orthonormale ssi $\forall i \neq j,\ < e_i, e_j >= 0$ et $\forall i\ \|e_i\| = 1$.

Théorème A.4 *Soit* F *un sous-espace vectoriel de* E. *On a* $E = F \oplus F^{\perp}$. *La projection* p *sur* F *parallèlement à* $F^{\perp}$ *est appelée projection orthogonale sur* F. *Si* $(e_1, \cdots, e_m)$ *est une base orthonormale de* F,

$$\forall x \in E,\ p(x) = \sum_{i=1}^{m} < e_i, x > e_i$$

Démonstration

Il est clair que $F \cap F^{\perp} = \{0_E\}$.

Soit $x \in E$. *Soit* $(e_1, \cdots, e_m)$ *une base orthonormale de* F. *On cherche à écrire* $x = f + g$ *avec* $f \in F$ *et* $g \in F^{\perp}$. f *peut s'écrire sous la forme* $\sum_{i=1}^{m} a_i e_i$ *de sorte que* $g = x - \sum_{i=1}^{m} a_i e_i$. *Or*

$$g \in F^{\perp} \Leftrightarrow \forall i \in \{1, ..., m\},\ < e_i, g >= 0$$

ainsi $\forall i \in \{1, ..., m\}$, $a_i =< e_i, x >$. *On peut donc écrire*

$$x = \underbrace{\sum_{i=1}^{m} < e_i, x > e_i}_{\in F} + \underbrace{x - \sum_{i=1}^{m} < e_i, x > e_i}_{\in F^\perp}$$

et $E = F + F^\perp$ *soit finalement* $E = F \oplus F^\perp$. *Si* p *est la projection orthogonale sur* F,

$$p(x) = \sum_{i=1}^{m} < e_i, x > e_i$$

A.1.4 Distance à un sous-espace

On se place dans l'espace euclidien E précédent. Soient F un sous-espace de E et $x \in E$.On définit la distance de x au sous-espace F par

$$d(x, F) = \inf_{\{f \in F\}} (\|x - f\|)$$

Cette distance est atteinte pour $f = p(x)$ où p est la projection orthogonale sur F :

$$d(x, F) = \|x - p(x)\|$$

Démonstration

On sait que $E = F \oplus F^\perp$. *On décompose donc* x *selon* $x = x_F + x_{F^\perp}$. *Ainsi*

$$d(x, F) = \inf_{\{f \in F\}} (\|f - x\|) = \inf_{\{f \in F\}} (\|f - x_F - x_{F^\perp}\|) = \inf_{\{f \in F\}} (\|f - x_{F^\perp}\|)$$

car, quand f *décrit* F, $f - x_F$ *décrit également* F.

Comme $x_{F^\perp} \in F^\perp$, $\forall f \in F$,

$$\|f - x_{F^\perp}\| = \sqrt{\|f\|^2 + \|x_{F^\perp}\|^2} \geq \|x_{F^\perp}\|$$

De plus, pour $f = 0_E$, $\|f - x_{F^\perp}\| = \|x_{F^\perp}\|$ *donc*

$$d(x, F) = \|x_{F^\perp}\|.$$

Si $(e_1, \cdots, e_m)$ est une base orthonormale de F,

$$d(x, F) = \sqrt{\|x\|^2 - \sum_{i=1}^{m} < e_i, x >^2}.$$

A.2 Inégalité de Cauchy-Schwartz

Dans cette partie, on suppose que $\mathbb{K}$ est soit le corps des réels, soit le corps des complexes. Soit E un espace vectoriel préhilbertien sur $\mathbb{K}$ muni du produit scalaire $<,>$.

Théorème A.5 (Inégalité de Cauchy-Schwartz)

$$\forall (u,v) \in E^2, \ | < u,v > |^2 \leq < u,u > . < v,v >$$

L'inégalité devient une égalité si les vecteurs u et v sont colinéaires.

Démonstration

$\forall (u,v) \in E^2$, $\forall \lambda \in \mathbb{K}$ [1],

$$(u + \lambda v)^2 = u^2 + |\lambda|^2 v^2 + 2Re(\lambda < u,v >)$$

On suppose v non nul et on pose $< u,v >= re^{i\theta}$ ($r = | < u,v > |$ et $\theta \in \mathbb{R}$ [2]*).*

Soit $x \in \mathbb{R}$. On prend $\lambda = xe^{-i\theta}$. On a

$$P(x) = (u + xe^{-i\theta}v)^2 = u^2 + x^2 v^2 + 2xr$$

P est une fonction polynôme du second degré qui ne prend que des valeurs positives sur $\mathbb{R}$ donc son discriminant réduit Δ' est négatif ou nul soit

$$r^2 - u^2 v^2 \leq 0$$

On constate que cette inégalité reste valable pour $v = 0_E$.

On a égalité lorsque $v = 0_E$ ou, si $v \neq 0_E$, $\Delta' = 0$. Dans ce dernier cas, P admet une racine réelle double x_0 et

$$(u + x_0 e^{-i\theta}v)^2 = 0 \Rightarrow u = -x_0 e^{-i\theta}v$$

donc u et v sont colinéaires. Réciproquement, si u et v sont colinéaires, $u = \mu v$ avec $\mu \in \mathbb{K}$.

$$| < u,v > |^2 = |\bar{\mu} < v,v > |^2 = |\mu|^2 < v,v >^2 = < u,u > . < v,v >$$

1. *Re* désigne la partie réelle.
2. Si $\mathbb{K} = \mathbb{R}$, $\theta \in \{0, \pi\}$.

Exemple

On prend $\mathbb{K} = \mathbb{R}$ et $E = \mathbb{R}^n$. On munit $\mathbb{R}^n$ de son produit scalaire canonique. D'après l'inégalité de Cauchy-Schwartz, on peut écrire,

$$\forall x = (x_1, \cdots, x_n) \in \mathbb{R}^n, \forall y = (y_1, \cdots, y_n) \in \mathbb{R}^n,$$
$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right) . \left(\sum_{i=1}^{n} y_i^2\right)$$

En particulier, si y est le vecteur dont toutes les composantes sont égales à 1,

$$\frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)$$

A.3 Matrices

On se place dans le $\mathbb{K}$ espace vectoriel $\mathcal{M}_{(n,n)}(\mathbb{K})$ des matrices carrées d'ordre n.

A.3.1 Propriétés de la trace

Soit $A \in \mathcal{M}_{(n,n)}(\mathbb{K})$, $A = (a_{ij})$, rappelons que la trace de A ($tr(A)$) est définie comme la somme des éléments diagonaux de A :

$$tr(A) = \sum_{i=1}^{n} a_{ii}$$

Propriété A.1 *L'application trace est une forme linéaire.*

Démonstration

Soit $A \in \mathcal{M}_{(n,n)}(\mathbb{K})$, on a clairement $tr(A) \in \mathbb{K}$.
$\forall A \in \mathcal{M}_{(n,n)}(\mathbb{K})$ $(A = (a_{ij}))$, $\forall B \in \mathcal{M}_{(n,n)}(\mathbb{K})$ $(B = (b_{ij}))$, $\forall (\lambda, \mu) \in \mathbb{K}^2$, $(\lambda A + \mu B) \in \mathcal{M}_{(n,n)}(\mathbb{K})$ *et* $\lambda A + \mu B = (\lambda a_{ij} + \mu b_{ij})$ *d'où*

$$tr(\lambda A + \mu B) = \sum_{i=1}^{n} (\lambda a_{ii} + \mu b_{ii}) = \lambda tr(A) + \mu tr(B)$$

Propriété A.2 *Lorsque AB et BA existent, tr(AB)=tr(BA)*

Démonstration

Soient $A \in \mathcal{M}_{(p,n)}(\mathbb{K})$ *et* $B \in \mathcal{M}_{(n,p)}(\mathbb{K})$.
$AB \in \mathcal{M}_{(p,p)}(\mathbb{K})$ *et* $BA \in \mathcal{M}_{(n,n)}(\mathbb{K})$.

$$tr(AB) = \sum_{i=1}^{p}\sum_{k=1}^{n} a_{ik}b_{ki}$$

et

$$tr(BA) = \sum_{j=1}^{n}\sum_{l=1}^{p} b_{jl}a_{lj} = \sum_{j=1}^{n}\sum_{i=1}^{p} a_{ij}b_{ji} = tr(AB)$$

A.3.2 Diagonalisation

Valeurs propres, vecteurs propres et sous-espaces propres

Soit E un $\mathbb{K}$ espace vectoriel de dimension finie, $f \in \mathcal{L}(E)$.

$\lambda \in \mathbb{K}$ est une valeur propre de f ssi

$$\exists u \in E,\ u \neq 0_E\ / f(u) = \lambda u \Longleftrightarrow ker(f - \lambda Id_E) \neq \{0_E\}$$

$u \in E$ est un vecteur propre de f si

$$\exists \lambda \in \mathbb{K}\ / f(u) = \lambda u$$

u est alors dit vecteur propre de f pour la valeur propre λ.

Si λ est une valeur propre de f, l'espace propre attaché à cette valeur propre E_λ est constitué du vecteur nul et des vecteurs propres de f pour la valeur propre λ. Autrement dit,

$$E_\lambda = ker(f - \lambda Id_E) = \{u \in E\ / f(u) = \lambda u\}$$

L'ensemble des valeurs propres de f est le spectre de f.

Endomorphisme diagonalisable

Un endomorphisme f de E est diagonalisable s'il existe une base de E composée uniquement de vecteurs propres de f. Une telle base est appelée base de diagonalisation de f.

Si $\mathcal{B} = (e_1, ..., e_n)$ est une telle base pour l'endomorphisme diagonalisable f avec $\forall i\ f(e_i) = \lambda_i e_i$, la matrice de f dans $\mathcal{B}$ est diagonale :

$$\begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$

Ainsi, $f \in \mathcal{L}(E)$ est diagonalisable ssi il existe une base $\mathcal{B}$ de E dans laquelle la matrice de f est diagonale.

A.4 Endomorphisme symétrique

Soit E un $\mathbb{R}$ espace vectoriel euclidien muni du produit scalaire $<,>$.

A.4.1 Définition et propriété

f endomorphisme de E est dit symétrique [3] ssi

$$\forall (u,v) \in E^2, \ < u, f(v) > = < f(u), v >$$

Propriété A.3 *f endomorphisme de E espace vectoriel euclidien est symétrique ssi sa matrice relativement à une base orthonormale est symétrique.*

Démonstration

C.N. *: On suppose que f est symétrique. Soit $A = (a_{ij})$ sa matrice relativement à la base orthonormale $\mathcal{B} = (e_1, ..., e_n)$ de E. a_{ij} est la coordonnée de $f(e_j)$ sur e_i soit*

$$a_{ij} = < e_i, f(e_j) > = < e_j, f(e_i) > = a_{ji}$$

donc $A = A^T$.

C.S. *: On suppose que la matrice $A = (a_{ij})$ de f relativement à la base orthonormale $\mathcal{B} = (e_1, ..., e_n)$ de E est symétrique. Soient $u = \sum_i u_i e_i$ et $v = \sum_j v_j e_j$ deux vecteurs de E.*

$$\begin{aligned} < u, f(v) > &= \sum_i \sum_j u_i v_j < e_i, f(e_j) > = \sum_i \sum_j a_{ij} u_i v_j \\ < u, f(v) > &= \sum_i \sum_j a_{ji} u_i v_j = \sum_i \sum_j u_i v_j < e_j, f(e_i) > \\ < u, f(v) > &= < f(u), v > \end{aligned}$$

f est donc symétrique.

A.4.2 Diagonalisation

Théorème A.6 *Un endomorphisme de E espace vectoriel euclidien est symétrique ssi il est diagonalisable dans une base orthonormale.*

3. On parle aussi d'endomorphisme autoadjoint.

Démonstration

La condition suffisante est évidente. Démontrons la condition nécessaire.

Considérons donc un endomorphisme symétrique f de E. Montrons par récurrence sur n, dimension de E, que f est diagonalisable dans une base orthonormale.

Si $n = 1$, tout endomorphisme est diagonalisable et, a fortiori, tout endomorphisme symétrique l'est.

Supposons maintenant que $n > 1$ et que le théorème soit démontré pour toutes les dimensions strictement inférieures à n. Soit λ une valeur propre de f[4]*. E_λ, l'espace propre correspondant, est stable par f*[5] *et il en est de même pour $E_\lambda^\perp$. En effet,*

$$\forall v \in E_\lambda^\perp,\ \forall u \in E_\lambda,\ < u, v >= 0$$

et

$$< f(v), u >=< v, f(u) >=< v, \lambda u >= 0 \Rightarrow f(v) \in E_\lambda^\perp$$

Si $E_\lambda = E$, f est bien diagonalisable dans une base orthonormale de E. Sinon, on sait que $E_\lambda^\perp$ est stable par f et que $\dim E_\lambda^\perp < n$. La restriction de f à $E_\lambda^\perp$ restant symétrique, par hypothèse de récurrence, on peut munir $E_\lambda^\perp$ d'une base orthonormale de diagonalisation de f. En prenant une base orthonormale de E_λ et en la joignant à la précédente, on obtient bien une base orthonormale de diagonalisation pour f dans E.

A.5 Formes quadratiques

Dans toute cette partie, on se place dans E, $\mathbb{R}$ espace euclidien de dimension n.

A.5.1 Formes bilinéaires symétriques

Une forme bilinéaire symétrique est une application $B : E \times E \to \mathbb{R}$ bilinéaire : $\forall (u, v, w) \in E^3$, $\forall (\lambda, \mu) \in \mathbb{R}^2$,

$$\begin{aligned} B(\lambda u + \mu v, w) &= \lambda B(u, w) + \mu B(v, w)\ (\textit{linéarité à gauche}) \\ B(u, \lambda v + \mu w) &= \lambda B(u, v) + \mu B(u, w)\ (\textit{linéarité à droite}) \end{aligned}$$

4. On est sûr que f admet au moins une valeur propre car le corps de base étant $\mathbb{R}$, le polynôme $\det(Mat(f) - \lambda I_n)$ est scindé.

5. Un sous-espace F de E est stable par f ssi $\forall x \in F,\ f(x) \in F$.

et symétrique : $\forall (u,v) \in E^2$, $B(u,v) = B(v,u)$.

On munit E d'une base $\mathcal{B} = (e_1, \cdots, e_n)$. Soit $(x,y) \in E^2$ avec

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ et } y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$B(x,y) = B(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j B(e_i, e_j)$$

La matrice de B relativement à la base $\mathcal{B}$ est la matrice A de $\mathcal{M}_{(n,n)}(\mathbb{R})$ de terme général $(B(e_i,e_j))$. De sorte que

$$B(x,y) = X^T A Y$$

A.5.2 Formes quadratiques

Une application Q de E dans $\mathbb{R}$ est une forme quadratique s'il existe une forme bilinéaire symétrique B sur E telle que $\forall x \in E$, $Q(x) = B(x,x)$. B est alors appelée forme bilinéaire symétrique associée à Q ou encore forme polaire de Q.

B a pour matrice $A = (a_{ij})$ relativement à la base $\mathcal{B}$. Si $x \in E$ avec $x = \sum_{i=1}^{n} x_i e_i$,

$$Q(x) = X^T A X = \sum_{i=1}^{n} a_{ii} x_i^2 + 2 \sum_{1 \le i < j \le n} a_{ij} x_i x_j$$

L'application [6]

$$\begin{cases} \varphi & : & E \to E^* \\ & & x \mapsto B(x,.) \end{cases}$$

est l'application linéaire associée à Q. La forme quadratique Q est dite non dégénérée si φ est bijective.

La matrice de φ relativement aux bases $\mathcal{B}$ et $\mathcal{B}^*$ est la matrice A de la forme polaire de Q.

Démonstration

6. E^* désigne le dual de E, i.e. le $\mathbb{R}$ espace vectoriel des formes linéaires sur E.

On pose $\mathcal{B} = (e_1, ..., e_n)$ *et* $\mathcal{B}^* = (e_1^*, ..., e_n^*)$ *avec*[7] $\forall (i,j),\ e_j^*(e_i) = \delta_i^j$.

Soit $A = (a_{ij})$ *la matrice de* Q *et* B *relativement à* $\mathcal{B}$. *Si* $u = \sum_i u_i e_i$ *et* $v = \sum_j v_j e_j$,

$$B(u,v) = \sum_j (\sum_i a_{ij} u_i) v_j = \sum_j (\sum_i a_{ij} u_i) e_j^*(v)$$

donc

$$\varphi(u) = \sum_j (\sum_i a_{ij} u_i) e_j^*$$

et la matrice de φ *relativement aux bases* $\mathcal{B}$ *et* $\mathcal{B}^*$ *est* $A^T = A$.

Ainsi, Q est non dégénérée ssi A est inversible.

Q est définie positive (resp. négative) sur E ssi $\forall x \in E$, $Q(x) \geq 0$(resp. $Q(x) \leq 0$) et $Q(x) = 0 \Rightarrow x = 0_E$.

Comme toute matrice symétrique, A est diagonalisable sur une base orthonormale de E. Si, de plus, Q est définie positive (resp. négative), toutes les valeurs propres de A sont strictement positives (resp. négatives).

Démonstration

Soit $\mathcal{C}$ *une base orthonormale de diagonalisation de l'endomorphisme symétrique associé à la matrice* A *de la forme quadratique* Q. *Dans la base* $\mathcal{C}$, *la matrice de cet endomorphisme est diagonale*

$$\begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$

et donc, si le vecteur u *a pour coordonnées* $(u_1, ..., u_n)$ *dans* $\mathcal{C}$,

$$Q(u) = \lambda_1 u_1^2 + \cdots + \lambda_n u_n^2$$

A.6 Relation d'ordre pour les matrices symétriques

Soient $\mathcal{S}_{(n,n)}(\mathbb{R})$ l'ensemble des matrices réelles symétriques de dimension n et $\mathcal{S}^+_{(n,n)}(\mathbb{R})$ l'ensemble des matrices réelles symétriques positives[8] de dimension n. Sur l'espace vectoriel $\mathcal{S}_{(n,n)}(\mathbb{R})$, on définit la relation $\succeq$ par

7. δ désigne le symbole de Kronecker.
8. La forme quadratique dont la matrice est un élément de $\mathcal{S}^+_{(n,n)}(\mathbb{R})$ est positive.

$$A \succeq B \Longleftrightarrow (A - B) \in \mathcal{S}^{+}_{(n,n)}(\mathbb{R})$$

Cette relation est une relation d'ordre non total sur $\mathcal{S}_{(n,n)}(\mathbb{R})$.

Démonstration

$\forall A \in \mathcal{S}_{(n,n)}(\mathbb{R})$, $A - A$ *est la matrice nulle dont les valeurs propres sont nulles donc positives.* $\succeq$ *est réflexive.*

On suppose que $A \succeq B$ *et* $B \succeq A$. $A - B$ *est symétrique donc l'endomorphisme symétrique qui lui est associé est diagonalisable dans une base orthornormale de* $\mathbb{R}^n$. *De plus, ses valeurs propres* λ_i *vérifient* $\forall i$, $\lambda_i \geq 0$ *et* $\lambda_i \leq 0$ *donc* $\lambda_i = 0$. *Cet endomorphisme est donc l'endomorphisme nul et, par conséquent,* $A = B$; $\succeq$ *est antisymétrique.*

Si $A \succeq B$ *et* $B \succeq C$, $\mathcal{S}^{+}_{(n,n)}(\mathbb{R})$ *étant stable pour l'addition,*

$$A - C = (A - B) + (B - C) \Rightarrow (A - C) \in \mathcal{S}^{+}_{(n,n)}(\mathbb{R})$$

$\succeq$ *est donc transitive.*

Enfin, en dimension 2, les deux matrices suivantes ne sont pas comparables :

$$\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{pmatrix} \quad \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 3 \end{pmatrix}$$

Annexe B

Rappels d'analyse

B.1 Quelques rappels de topologie

Dans cette partie, E désigne un espace vectoriel normé sur le corps $\mathbb{K}$ des réels ou des complexes. On note $\|.\|$ la norme de E.

B.1.1 Boule ouverte

Soit $x_0 \in E$. La boule ouverte de centre x_0 et de rayon $\rho \geq 0$ est l'ensemble $\mathcal{B}_o(x_0, \rho)$ défini par

$$\mathcal{B}_o(x_0, \rho) = \{x \in E \,/ \|x - x_0\| < \rho\}$$

B.1.2 Voisinage

Soit $x_0 \in E$. Une partie V de E est un voisinage de x_0 si elle contient une boule ouverte de x_0 et de rayon non nul soit

$$V \text{ voisinage de } x_0 \Longleftrightarrow \exists \rho > 0 \,/ \mathcal{B}_o(x_0, \rho) \subset V$$

B.1.3 Ouvert et fermé

Une partie O de E est un ouvert de E ssi elle est voisinage de tous ses points. Autrement dit,

$$O \text{ ouvert de } E \Longleftrightarrow \forall x \in O,\ \exists \rho > 0 \,/ \mathcal{B}_o(x, \rho) \subset O$$

Une partie F de E est un fermé de E ssi son complémentaire dans E est un ouvert de E soit

$$F \text{ fermé de } E \Longleftrightarrow \forall x \notin F,\ \exists \rho > 0 \,/ \mathcal{B}_o(x, \rho) \subset F^c$$

Remarques

(i) $\emptyset$ et E sont des ouverts et des fermés de E.

(ii) Si $(O_i)_{i\in I}$ est une famille d'ouverts de E, $\bigcup_{i\in I} O_i$ est un ouvert de E.

En effet,

$$\forall x \in \bigcup_{i\in I} O_i,\ \exists i_0\ / x \in O_{i_0}$$

et

$$\exists \rho > 0\ / \mathcal{B}_o(x,\rho) \subset O_{i_0} \Rightarrow \mathcal{B}_o(x,\rho) \subset \bigcup_{i\in I} O_i$$

(iii) Si $(F_i)_{i\in I}$ est une famille de fermés de E, $\bigcap_{i\in I} F_i$ est un fermé de E.

En effet, $(\bigcap_{i\in I} F_i)^c = \bigcup_{i\in I} F_i^c$.

(iv) Si $O_1, \cdots, O_p$ sont des ouverts de E, $O_1 \cap \cdots \cap O_p$ est un ouvert de E.

En effet,

$$\forall x \in O_1 \cap \cdots \cap O_p,\ \forall i\ \exists \rho_i\ / \mathcal{B}_o(x,\rho_i) \subset O_i$$

et donc

$$\mathcal{B}_o(x, \inf_{\{1\le i\le p\}} \rho_i) \subset \bigcap_{1\le i\le p} O_i$$

(v) Si $F_1, \cdots, F_p$ sont des fermés de E, $F_1 \cup \cdots \cup F_p$ est un fermé de E.

En effet, $(\bigcap_{1\le i\le p} F_i^c = \bigcup_{1\le i\le p} F_i)^c$.

(vi) Les ouverts de E sont les boules ouvertes et les réunions de boules ouvertes de E.

Les boules ouvertes de E sont des ouverts de E. En effet, si $x \in \mathcal{B}_o(x_0,\rho)$, $\mathcal{B}_o(x, \rho - \|x - x_0\|) \subset \mathcal{B}_o(x_0,\rho)$.

De plus, si O est un ouvert de E,

$$\forall x \in O,\ \exists \rho_x\ / \mathcal{B}_o(x,\rho_x) \subset O$$

et donc $O = \bigcup_{x\in O} \mathcal{B}_o(x,\rho_x)$: O est réunion de boules ouvertes.

L'ensemble des ouverts de E est appelé topologie de E. E muni de sa topologie est un espace topologique.

B.2 Calcul différentiel

Dans cette partie, E et F désignent deux $\mathbb{R}$ espaces vectoriels normés munis de leurs normes respectives $\|.\|_E$ et $\|.\|_F$. Soit f une application de Ω ouvert de E dans F.

B.2.1 Différentiabilité

Soit $a \in \Omega$. f est différentiable en a s'il existe une application linéaire L de E dans F telle que

$$\lim_{h \to 0_E} \frac{1}{\|h\|_E}(f(a+h) - f(a) - L(h)) = 0_F$$

Autrement dit,

$$f(a+h) = f(a) + L(h) + \|h\|_E \varepsilon(h) = f(a) + L(h) + o(\|h\|_E)$$

ε étant définie au voisinage de 0_E et telle que $\lim_{h \to 0_E} \varepsilon(h) = 0_F$.

Théorème B.1 *L est unique.*

Démonstration

Supposons donc qu'il existe L et M, éléments de $\mathcal{L}(E,F)$, telles que

$$\lim_{h \to 0_E} \frac{1}{\|h\|_E}(f(a+h) - f(a) - L(h)) = 0_F$$

et

$$\lim_{h \to 0_E} \frac{1}{\|h\|_E}(f(a+h) - f(a) - M(h)) = 0_F$$

On a ainsi

$$\lim_{h \to 0_E} \frac{L(h) - M(h)}{\|h\|_E} = 0_F$$

Soit $u \in E$, $u \neq 0_E$. $\lim_{x \to 0^+} xu = 0_E$ et donc

$$\lim_{x \to 0^+} \frac{L(xu) - M(xu)}{\|xu\|_E} = \frac{L(u) - M(u)}{\|u\|_E} = 0_F$$

Donc $\forall u \in E\backslash\{0_E\}$, $L(u) = M(u)$. *Or* $L(0_E) = M(0_E) = 0_F$ *donc* $L = M$.

L est appelée différentielle de f au point a ou application linéaire tangente à f en a et est notée $f'(a)$.

B.2.2 Dérivées partielles

Dérivée suivant un vecteur

Soient $a \in \Omega$ et v un vecteur non nul de E. f est dérivable en a suivant le vecteur v ssi la fonction

$$(t \in \mathbb{R} \mapsto f(a + tv))$$

définie sur un voisinage de 0 est dérivable en 0. Le vecteur

$$\left[\frac{d}{dt}(f(a + tv))\right]_{t=0}$$

est la dérivée de f suivant v au point a. Il est noté $\frac{\partial f}{\partial v}(a)$.

Propriété B.1 *Si f est différentiable en a, f admet alors une dérivée suivant tous les vecteurs v non nuls au point a et*

$$\frac{\partial f}{\partial v}(a) = f'(a)(v)$$

Démonstration

Le développement limité à l'ordre 1 de f au voisinage de a donne

$$f(a + tv) = f(a) + f'(a)(tv) + |t|.\|v\|_E \varepsilon(tv)$$

soit

$$\frac{f(a + tv) - f(a)}{t} = f'(a)(v) + sgn(t).\|v\|_E \varepsilon(tv)$$

Dérivées partielles

On suppose que E est de dimension n et on le munit d'une base $\mathcal{B}_E$: $(e_1, \cdots, e_n)$. Soit $(x_1, \cdots, x_n)$ un système de coordonnées génériques suivant $\mathcal{B}_E$. Soit $a \in \Omega$.

On dit que f admet une dérivée partielle suivant la variable x_i en a si elle admet une dérivée suivant le vecteur e_i en a. On note

$$\frac{\partial f}{\partial x_i}(a) = \frac{\partial f}{\partial e_i}(a)$$

f admet une dérivée partielle suivant la variable x_i sur Ω si elle admet une dérivée partielle $\frac{\partial f}{\partial x_i}(a)$ suivant la variable x_i en tout point a de Ω.

B.2.3 Fonctions de classe $\mathcal{C}^k$

Dans cette partie, $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$.

Les dérivées partielles permettent de caractériser la classe d'une fonction. En effet, f est dite de classe $\mathcal{C}^k$, $k \in \mathbb{N}$, ssi pour toute suite $i_0, i_1, \cdots, i_k$ d'entiers compris entre 1 et n, il existe une suite $g_0, g_1, \cdots, g_k$ de fonctions définies sur Ω telle que $g_0 = f$, $\forall j \in \{0, \cdots, k-1\}$, g_j admet une dérivée partielle suivant x_{i_j} sur Ω : $\frac{\partial g_j}{\partial x_{i_j}} = g_{j+1}$ et g_k est continue sur Ω. Dans ce cas, on note

$$g_k = \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}$$

f est de classe $\mathcal{C}^\infty$ sur Ω si elle est de classe $\mathcal{C}^k$ sur Ω pour tout entier k.

B.2.4 Matrice jacobienne

On suppose que f est différentiable en $a \in \Omega$ et que E et F sont respectivement de dimensions n et p. On munit E et F de leurs bases respectives $\mathcal{B}_E$ et $\mathcal{B}_F$. La matrice jacobienne de f en a suivant les bases $\mathcal{B}_E$ et $\mathcal{B}_F$ est la matrice de $f'(a)$ relativement à ces deux bases :

$$Mat_{\mathcal{B}_E, \mathcal{B}_F}(f'(a)) = \left(\frac{\partial f_i}{\partial x_j}(a) \right)_{1 \leq i \leq p, 1 \leq j \leq n}$$

Lorsque $n = p$, le déterminant de la matrice jacobienne de f en a est la jacobien de f en a.

B.2.5 Théorème de Schwartz

Soit f de classe $\mathcal{C}^k$ sur Ω, ouvert de $\mathbb{R}^n$, à valeurs dans $\mathbb{R}$. Pour toute permutation σ de $\{1, ..., k\}$, [8]

$$\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}} = \frac{\partial^k f}{\partial x_{i_{\sigma(k)}} \cdots \partial x_{i_{\sigma(1)}}}$$

B.2.6 Développement de Taylor-Young d'ordre 2

Dans cette partie, $E = \mathbb{R}^n$ et $F = \mathbb{R}$.

On suppose f de classe $\mathcal{C}^2$ sur Ω. Soit $a \in \Omega$. Le développement de Taylor-Young d'ordre 2 de f en a s'écrit [8]

$$f(a+h) = f(a) + f'(a)(h) + \frac{1}{2}Q_{f,a}(h) + o(\|h\|^2)$$

$Q_{f,a}$ est la forme quadratique fondamentale de f en a de matrice

$$\left(\frac{\partial^2 f}{\partial x_i x_j}(a)\right)_{1\leq i,j\leq n}$$

relativement à la base canonique de $\mathbb{R}^n$.

Si $h = (h_1, \cdots, h_n)^T$,

$$\begin{aligned} f(a+h) &= f(a) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(a) \\ &+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i x_j}(a) + o(\|h\|^2) \end{aligned}$$

B.2.7 Notation différentielle

Cette notation correspond à une approche du calcul différentiel par variables liées. Si $f(x) = y$, l'écriture $dy = f'(x).dx$ n'est autre que l'écriture de l'application linéaire tangente à f en x.

Cette notation a un sens intuitif puissant. En effet, si x devient $x + \Delta x$, y devient $y + \Delta y$ et $\Delta y \simeq f'(x).\Delta x$. Plus Δx est petit, plus cette approximation se vérifie.

Enfin, la notation différentielle est formellement très simple à utiliser [1]. Si E est de dimension n et F de dimension p. Avec $x = (x_1, ..., x_n)$ et $y = (y_1, ..., y_p)$, $y_i = f_i(x_1, ..., x_n)$, on écrit généralement

$$dy_i = \frac{\partial f_i}{\partial x_1}dx_1 + \cdots + \frac{\partial f_i}{\partial x_n}dx_n = \frac{\partial y_i}{\partial x_1}dx_1 + \cdots + \frac{\partial y_i}{\partial x_n}dx_n$$

B.2.8 Extrema d'une fonction numérique

Dans cette dernière section, on suppose que $E = \mathbb{R}^n$ et $F = \mathbb{R}$ et que la fonction numérique f est de classe $\mathcal{C}^2$ sur Ω.

1. Elle est d'ailleurs abondamment utilisée en physique !

Point critique

On appelle point critique de f un point a de $\mathbb{R}^n$ où $f'(a) = 0$. Autrement dit,

$$a \text{ point critique de } f \Longleftrightarrow \forall i \in \mathbb{N},\ \frac{\partial f}{\partial x_i}(a) = 0$$

Si a est un point critique de f, a est dit critique non dégénéré ssi [2]

$$det\left(\frac{\partial^2 f}{\partial x_i x_j}(a)\right) \neq 0$$

Extrema locaux

On dit que f admet un maximum (resp. un minimum) local en $a \in \Omega$ s'il existe un voisinage relatif V de a dans Ω tel que

$$\forall x \in V,\ f(x) \leq f(a) (\text{resp. } f(x) \geq f(a))$$

On dit que f admet un extremum local en a ssi f admet un minimum ou un maximum local en a.

Théorème B.2 *Si f fonction numérique de classe $\mathcal{C}^1$ sur Ω ouvert de $\mathbb{R}^n$ admet un extremum local en $a \in \Omega$ alors a est un point critique de f.*

Démonstration

Soit $v \in \mathbb{R}^n \backslash \{0\}$. Soit la fonction $g : t \mapsto f(a+tv)$. g est une fonction de $\mathbb{R}$ dans $\mathbb{R}$, définie sur un voisinage de 0 et admettant un extremum local en 0.

g est dérivable en 0 et $g'(0) = 0$. Or $g'(0) = f'(a)(v)$ donc [3] *$f'(a) = 0$.*

Théorème B.3 *Si f numérique de classe $\mathcal{C}^2$ sur Ω ouvert de $\mathbb{R}^n$ admet un maximum (resp. minimum) local en $a \in \Omega$ alors a est un point critique de f et $Q_{f,a}$ est négative (resp. positive).*

Démonstration

On effectue la démonstration dans le cas d'un minimum local. On suppose donc que f admet un minimum local en a. On sait déjà que a est un point critique de f.

2. $Q_{f,a}$ est alors non dégénérée.
3. comme $f'(a)(0) = 0$ par linéarité de $f'(a)$.

Soit $v \in \mathbb{R}^n$. On définit la fonction $g : t \mapsto g(t) = f(a+tv)$. g admet un minimum local en 0. Le développement de Taylor-Young d'ordre 2 de f permet d'écrire

$$f(a+tv) = f(a) + f'(a)(tv) + \frac{1}{2}Q_{f,a}(tv) + \|tv\|^2\varepsilon(tv)$$

Or $f'(a) = 0$ et $Q_{f,a}$ est homogène de degré 2 donc, sur un voisinage de 0,

$$\frac{g(t) - g(0)}{t^2} \sim \frac{1}{2}Q_{f,a}(v)$$

Sur ce voisinage, $g(t) - g(0) \geq 0$ d'où $Q_{f,a}(v) \geq 0$.

Théorème B.4 *Soit a un point critique non dégénéré de f numérique de classe $\mathcal{C}^2$ sur Ω ouvert de $\mathbb{R}^n$.*
(i) Si $Q_{f,a}$ est définie positive, f admet un minimum local strict en a.
(ii) Si $Q_{f,a}$ est définie négative, f admet un maximum local strict en a.

Démonstration

On effectue la démonstration dans le cas où $Q_{f,a}$ est définie positive. On munit $\mathbb{R}^n$ de sa norme euclidienne canonique.

On diagonalise l'endomorphisme symétrique φ associé à $Q_{f,a}$ dans une base orthonormale $(e_1, \cdots, e_n)$ de $\mathbb{R}^n$ avec $\varphi(e_i) = \lambda_i e_i$. $Q_{f,a}$ étant définie positive, $\forall i\ \lambda_i > 0$ et on impose même $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Ainsi, si $h = (h_1, \cdots, h_n)^T$,

$$Q_{f,a}(h) = \sum_{i=1}^{n} \lambda_i h_i^2 \geq \lambda_n \|h\|^2$$

Le développement limité d'ordre 2 de f au voisinage de a donne

$$f(a+h) = f(a) + \frac{1}{2}Q_{f,a}(h) + \|h\|^2\varepsilon(h)$$

Comme $\lim_{h\to 0} \varepsilon(h) = 0$,

$$\exists \eta > 0 \ / \ \|h\| < \eta \Rightarrow |\varepsilon(h)| < \frac{\lambda_n}{4}$$

Pour $\|h\| < \eta$, on a donc

$$f(a+h) - f(a) \geq \frac{\lambda_n}{4}\|h\|^2$$

B.3 Théorie de la mesure

Dans cette partie, on expose les bases de la théorie de la mesure, bases utiles dans le domaine des probabilités, entre autres.

Soit un ensemble Ω. On note $\mathcal{P}(\Omega)$ l'ensemble de ses parties.

B.3.1 Algèbre et σ-algèbre

Algèbre

Une partie $\mathcal{A}$ de $\mathcal{P}(\Omega)$ (i.e. une famille de parties de Ω) est une algèbre si

(i) $\emptyset \in \mathcal{A}$

(ii) $\Omega \in \mathcal{A}$

et si [4] $\forall (A, B) \in \mathcal{A}^2$,

(iii) $A \cup B \in \mathcal{A}$

(iv) $A \cap B \in \mathcal{A}$

(v) $A^c = \Omega \backslash A \in \mathcal{A}$

Tribu

Une σ-algèbre (ou tribu) est une algèbre stable par réunion dénombrable croissante [5], c'est-à-dire

$$\text{si } A_1 \subset A_2 \subset \cdots \text{ avec } \forall n \; A_n \in \mathcal{A} \text{ alors } \bigcup_n A_n \in \mathcal{A}$$

Une tribu est ainsi stable par réunion et intersection dénombrables.

En effet, si $A_1, A_2, \ldots$ est une famille dénombrable non croissante, on peut créer une famille dénombrable croissante $A'_1, A'_2, \ldots$ avec $A'_1 = A_1$, $A'_2 = A_1 \cup A_2$,...

4. Certaines de ces conditions se recoupent...
5. **Rappel** : un ensemble est dénombrable s'il peut être mis en bijection avec $\mathbb{N}$.

Tribu engendrée

Soit $\mathcal{M} \subset \mathcal{P}(\Omega)$. L'intersection de toutes les tribus contenant $\mathcal{M}$ est une tribu. Elle est notée $\sigma(\mathcal{M})$: c'est la tribu engendrée par $\mathcal{M}$ et c'est la plus petite tribu contenant $\mathcal{M}$.

Tribu borélienne

Si Ω est un espace topologique, la tribu borélienne est la tribu engendrée par les parties ouvertes de Ω. On la note $\mathcal{B}(\Omega)$. Toute partie de Ω appartenant à $\mathcal{B}(\Omega)$ est un borélien.

Si $\Omega = \mathbb{R}^n$, la tribu borélienne est engendrée indifféremment par les familles suivantes :

(i) les parties ouvertes (c'est la définition)

(ii) les parties fermées.

En effet, si une tribu contient les parties ouvertes, par passage au complémentaire, elle contient les parties fermées et réciproquement.

B.3.2 Mesure positive

Soit $\mathcal{A}$ une tribu de $\mathcal{P}(\Omega)$. $(\Omega, \mathcal{A})$ est appelé espace mesurable.

Une mesure positive est une fonction $\mu : \mathcal{A} \to [0, +\infty[$ vérifiant :

(i)$\mu(\emptyset) = 0$

(ii) la propriété de σ-additivité : pour toute famille dénombrable de parties A_n de Ω, deux à deux disjointes et appartenant à $\mathcal{A}$,

$$\mu\left(\bigcup_1^\infty A_n\right) = \sum_1^\infty \mu(A_n)$$

Le triplet $(\Omega, \mathcal{A}, \mu)$ est appelé espace mesuré.

B.3.3 Application mesurable

Soit $(\Omega, \mathcal{A}, \mu)$ et $(\Lambda, \mathcal{B}, \nu)$ deux espaces mesurés.

Une application $f : \Omega \to \Lambda$ est mesurable si

$$\forall B \in \mathcal{B},\ f^{-1}(B) \in \mathcal{A}$$

Annexe C

Rappels de statistiques

Cette annexe rappelle brièvement les concepts de la théorie des probabilités ainsi que les bases des statistiques et, plus particulièrement, de l'estimation statistique.

C.1 Rappels de probabilités

C.1.1 Définitions

On considère une expérience aléatoire ε, c'est-à-dire un phénomène naturel étudié dont le déroulement, *a priori* inconnu, peut avoir lieu de différentes façons. *Par exemple, le lancer d'un dé à six faces constitue une expérience aléatoire.*

Ω désigne l'ensemble des résultats de l'expérience aléatoire ε ou univers des possibles. *Pour le lancer d'un dé à six faces,* $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Un événement ω est un fait observable qui peut se produire ou non selon le déroulement de l'expérience ε. *Dans l'exemple, "le dé sort un chiffre pair" est un événement.*

Si Ω est fini ou dénombrable, $\mathcal{P}(\Omega)$ est l'ensemble des événements. Dans le cas général, cet ensemble est une tribu $\mathcal{A}$ de $\mathcal{P}(\Omega)$. Le couple $(\Omega, \mathcal{A})$ est alors un espace probabilisable.

Le tableau (C.1) effectue un parallèle entre les terminologies ensembliste et probabiliste. Il est inspiré de [11].

Une loi de probabilité P sur l'espace probabilisable $(\Omega, \mathcal{A})$ est une mesure positive normalisée ($P(\Omega) = 1$) de $\mathcal{A}$ dans $[0, 1]$. Elle est le reflet du degré d'occurrence d'un événement. Le triplet $(\Omega, \mathcal{A}, P)$ est un espace probabilisé. $P($*"le dé sort un deux"*$) = \frac{1}{6}$ *pour un dé non pipé.*

Notation	***Terminologie ensembliste***	***Terminologie probabiliste***
$\omega \in \Omega$	élément	événement élémentaire
$A \subset \Omega$	partie	événement
Ω	ensemble	événement certain
$\emptyset$	ensemble vide	événement impossible
$A \cap B$	intersection de A et B	A et B réalisés simultanément
$A \cup B$	réunion de A et B	au moins un événement est réalisé
$\bar{A} = \mathcal{C}_\Omega A$	complémentaire de A	événement contraire de A
$A \subset B$	A inclus dans B	la réalisation de B implique celle de A
$A \cap B = \emptyset$	A et B disjoints	A et B incompatibles

TABLE C.1 – Terminologies ensembliste et probabiliste.

C.1.2 Propriétés

P vérifie les propriétés élémentaires suivantes :

(i) $P(\emptyset) = 0$

(ii) $\forall A \in \mathcal{A}$, $P(\bar{A}) = 1 - P(A)$

En effet, $\Omega = A \cup \bar{A}$ *avec* $A \cap \bar{A} = \emptyset$ *donc, par* σ*-additivité,* $P(\Omega) = P(A) + P(\bar{A}) = 1$.

(iii) Si $B \subset A$, $P(A) \geq P(B)$

En effet, $A = B \cup \mathcal{C}_A B$ *avec* $B \cap \mathcal{C}_A B = \emptyset$ *donc, par* σ*-additivité,* $P(A) = P(B) + P(\mathcal{C}_A B)$ *et* $P(\mathcal{C}_A B) \geq 0$.

(iv) $\forall (A, B) \in \mathcal{A}^2$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

En effet, d'après la figure (C.1), $A = (A \cap \bar{B}) \cup (A \cap B)$, $A \cap \bar{B}$ *et* $A \cap B$ *étant incompatibles, donc* $P(A) = P(A \cap \bar{B}) + P(A \cap B)$. *De même,* $P(B) = P(B \cap \bar{A}) + P(A \cap B)$.

Enfin, $A \cup B = (A \cap \bar{B}) \cup (B \cap \bar{A}) \cup (A \cap B)$, *les trois événements étant deux à deux incompatibles. Combinant toutes les relations, on obtient bien* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

En particulier, si A *et* B *sont incompatibles,*

$$P(A \cup B) = P(A) + P(B)$$

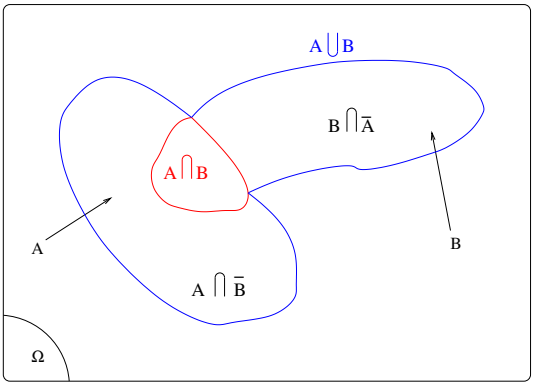

FIGURE C.1 – Ensembles A et B.

Le théorème suivant est, en quelque sorte, une généralisation de cette propriété.

Théorème C.1 (Théorème des probabilités totales) *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé. Si* $(A_i)_{i\in I}$ *constitue une partition dénombrable de* Ω*, les* A_i *étant deux à deux incompatibles,*

$$\forall A \in \mathcal{A},\ P(A) = \sum_{i\in I} P(A \cap A_i)$$

Démonstration

Par hypothèse, $\Omega = \bigcup_{i\in I} A_i$*. Ainsi,*

$$\forall A \in \mathcal{A},\ A = A \cap \Omega = \bigcup_{i\in I} A \cap A_i$$

Les événements $A \cap A_i$ *sont deux à deux incompatibles : la* σ*-additivité donne donc*

$$P(A) = P\left(\bigcup_{i\in I} A \cap A_i\right) = \sum_{i\in I} P(A \cap A_i)$$

C.1.3 Probabilités conditionnelles et indépendance

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. On s'intéresse à la probabilité d'un événement A lorsqu'un certain événement B a eu lieu. On note A/B l'événement à caractère conditionnel ainsi considéré (on dit "A si B" ou "A sachant B") et on définit la probabilité conditionnelle de A/B par [1]

1. On montre ([16]) que cette relation définit bien une mesure de probabilité.

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

A partir de cette notion, on peut définir celle d'indépendance. En effet, les événements A et B sont dits indépendants ssi

$$P(A/B) = P(A) \Longleftrightarrow P(A \cap B) = P(A)P(B)$$

Cette notion est intuitive : deux événements sont indépendants si la réalisation d'un des événements n'influe pas sur la réalisation de l'autre...

A l'aide de cette formulation conditionnelle, on peut reformuler le précédent théorème des probabilités totales.

Théorème C.2 (Théorème des probabilités totales (bis)) *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Si $(A_i)_{i \in I}$ constitue une partition dénombrable de Ω, les A_i étant deux à deux incompatibles,*

$$\forall A \in \mathcal{A},\ P(A) = \sum_{i \in I} P(A/A_i)P(A_i)$$

Les deux théorèmes qui suivent (dus à Bayes) sont utiles pour les calculs de probabilités conditionnelles.

Théorème C.3 (Formule de Bayes) *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. $\forall (A, B) \in \mathcal{A}^2$,*

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)}$$

Démonstration

Par définition, $P(B/A) = \frac{P(A \cap B)}{P(A)}$ et $P(A/B) = \frac{P(A \cap B)}{P(B)}$ d'où

$$\frac{P(B/A)}{P(A/B)} = \frac{P(B)}{P(A)}$$

Théorème C.4 (Théorème sur la probabilité des causes) *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Si les $(B_i)_{i \in I}$ forment une partition dénombrable de Ω formée d'événements deux à deux incompatibles, $\forall A \in \mathcal{A}$,*

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{\sum_{j \in I} P(A/B_j)P(B_j)}$$

Démonstration

D'après la formule de Bayes,

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{P(A)}$$

et, d'après le théorème des probabilités totales,

$$P(A) = \sum_{j \in I} P(A/B_j)P(B_j)$$

C.2 Variable aléatoire

Les variables aléatoires sont des entités fondamentales en statistiques et, plus particulièrement, en théorie de l'estimation.

C.2.1 Variable aléatoire réelle

Dans toute cette partie, on considère un espace probabilisé $(\Omega, \mathcal{A}, P)$ et l'espace probabilisable [2] $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Définition

Une variable aléatoire réelle est une application X mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Dans le précédent exemple du lancer de dé, on peut définir X comme étant le résultat du tirage.

Les valeurs prises par la variable aléatoire X suite au déroulement d'une expérience aléatoire ε sont des réalisations de cette variable aléatoire.

Loi de probabilité

La loi de probabilité P_X de la variable aléatoire X est définie à l'aide de la mesure de probabilité P par

$$\forall B \in \mathcal{B}(\mathbb{R}),\ P_X(B) = P(X^{-1}(B))$$

Densité de probabilité

La densité de probabilité de la variable aléatoire X est l'application positive p_X définie par (si elle existe)

2. $\mathcal{B}(\mathbb{R})$ est la tribu borélienne de $\mathbb{R}$.

$$P_X([a,b]) = P(a \leq X \leq b) = \int_a^b p_X(x)dx$$

Fonction de répartition

La fonction de répartition de la variable aléatoire X est la fonction F_X définie par

$$\begin{cases} F_X & : & \mathbb{R} \to [0,1] \\ & & x \mapsto F_X(x) = P_X(]-\infty, x]) = P(X \leq x) = \int_{-\infty}^{x} p_X(u)du \end{cases}$$

F_X vérifie

(i) $p_X(x) = \frac{dF_X}{dx}(x)$

(ii) $P_X([a,b]) = F_X(b) - F_X(a)$

(iii) $\lim_{x \to -\infty} F_X(x) = 0$

(iv) $\lim_{x \to +\infty} F_X(x) = 1$

Le fractile d'ordre α de F_X est le réel x_α tel que

$$F_X(x_\alpha) = P(X \leq x_\alpha) = \alpha$$

Moments

Le moment d'ordre i $\mathbb{E}X^i$ de la variable aléatoire X est défini par

$$\mathbb{E}X^i = \int_{-\infty}^{+\infty} x^i p_X(x)dx$$

Un moment important en statistiques est l'espérance (ou moyenne) de la variable aléatoire X : c'est le moment d'ordre 1

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x p_X(x)dx$$

Ce moment mesure la valeur prise en moyenne par la variable aléatoire X, compte tenu de sa densité de probabilité p_X.

Moments centrés

Le moment centré d'ordre i de la variable aléatoire X est défini à l'aide de son espérance $\mathbb{E}(X)$ par

$$\mathbb{E}[(X - \mathbb{E}(X))^i] = \int_{-\infty}^{+\infty} [x - \mathbb{E}(X)]^i p_X(x) dx$$

Le moment centré d'ordre 2 revêt une importance particulière en statistiques puisqu'il mesure la dispersion des réalisations de la variable aléatoire X autour de sa moyenne. On l'appelle variance de X et il a pour expression

$$var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \int_{-\infty}^{+\infty} [x - \mathbb{E}(X)]^2 p_X(x) dx$$

$$var(X) = \int_{-\infty}^{+\infty} x^2 p_X(x) dx - 2\mathbb{E}(X) \int_{-\infty}^{+\infty} x p_X(x) dx + \mathbb{E}(X)^2 \int_{-\infty}^{+\infty} p_X(x) dx$$

$$var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

La racine carrée de $var(X)$ définit l'écart-type $\sigma(X)$ de la variable aléatoire X.

La variable aléatoire centrée X^c déduite de X est $X^c = X - \mathbb{E}(X)$. La variable aléatoire réduite X^r déduite de X est $X^r = \frac{X}{\sigma(X)}$.

Théorème C.5 (Inégalité de Bien-Aymé Tchebicheff) *Soit X une variable aléatoire réelle sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$.*

$$\forall t \in \mathbb{R}_+^*,\ P(|X - \mathbb{E}(X)| > t) < \frac{\sigma(X)^2}{t^2}$$

Démonstration

Par définition,

$$P(|X| > t) = P(\{X > t\} \cap \{X < -t\}) = \int_t^{+\infty} p_X(x) dx + \int_{-\infty}^{-t} p_X(x) dx$$

Or, $|X| > t \Rightarrow |\frac{X}{t}|^2 > 1$ et donc

$$P(|X| > t) < \int_t^{+\infty} |\frac{x}{t}|^2 p_X(x) dx + \int_{-\infty}^{-t} |\frac{x}{t}|^2 p_X(x) dx < \underbrace{\int_{-\infty}^{+\infty} |\frac{x}{t}|^2 p_X(x) dx}_{\frac{\mathbb{E}(X^2)}{t^2}}$$

d'où, pour la variable aléatoire centrée,

$$P(|X - \mathbb{E}(X)| > t) < \frac{1}{t^2} \underbrace{\mathbb{E}[(X - \mathbb{E}(X))^2]}_{\sigma^2(X)}$$

Ce théorème traduit le fait que la probabilité de trouver des réalisations de la variable aléatoire X s'écartant fortement de son espérance $\mathbb{E}(X)$ est faible.

Cas de deux variables aléatoires réelles

On considère deux variables aléatoires réelles X et Y sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$.

Fonctions conjointes On définit la densité de probabilité conjointe p_{XY} (fonction positive de $\mathbb{R}^2$ dans $\mathbb{R}$) par la relation

$$P(\{x < X < x + dx\} \cap \{y < Y < y + dy\}) = p_{XY}(x, y)dxdy$$

La fonction de répartition conjointe F_{XY} est définie par

$$F_{XY}(x, y) = P(\{X < x\} \cap \{Y < y\}) = \int_{-\infty}^{x} \int_{-\infty}^{y} p_{XY}(u, v)dudv$$

Fonctions et moments marginaux On définit également les deux densités marginales p_X et p_Y par

$$p_X(x) = \int_{-\infty}^{+\infty} p(x, y)dy \text{ et } p_Y(y) = \int_{-\infty}^{+\infty} p(x, y)dx$$

Ces densités permettent de définir les moments et moments centrés marginaux et, en particulier, les espérances et variances marginales [3]

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x p_X(x)dx \text{ et } \sigma(X)^2 = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 p_X(x)dx$$

Corrélation et indépendance La covariance des variables aléatoires X et Y est

$$\begin{aligned} cov(X, Y) &= \mathbb{E}[(X - \mathbb{E}(X)).(Y - \mathbb{E}(Y))] \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mathbb{E}(X)).(y - \mathbb{E}(Y)) p_{XY}(x, y)dxdy \end{aligned}$$

3. On a les relations équivalentes pour la variable aléatoire Y.

A partir de cette covariance, on définit le facteur de corrélation des deux variables X et Y :

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma(X).\sigma(Y)}$$

Ce facteur est un indicateur du degré de similitude entre les variables aléatoires X et Y. On a de plus

$$\begin{aligned} cov(X,Y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyp_{XY}(x,y)dxdy + \mathbb{E}(X)\mathbb{E}(Y)\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{XY}(x,y)dxdy \\ &- \mathbb{E}(X)\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yp_{XY}(x,y)dxdy - \mathbb{E}(Y)\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xp_{XY}(x,y)dxdy \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

Ainsi, si les variables aléatoires X et Y sont décorrélées, $corr(X,Y) = 0$ et donc $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Les variables aléatoires X et Y sont indépendantes si leur densité de probabilité conjointe est égale au produit de leurs densités marginales, soit

$$p_{XY}(x,y) = p_X(x).p_Y(y)$$

La covariance de deux variables indépendantes est nulle donc, en particulier, deux variables aléatoires indépendantes sont non corrélées. ***La réciproque est fausse !***

En effet, si X et Y sont indépendantes, $p_{XY}(x,y) = p_X(x).p_Y(y)$ et

$$\mathbb{E}(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyp_{XY}(x,y)dxdy = \int_{-\infty}^{+\infty} xp_X(x)dx\int_{-\infty}^{+\infty} yp_Y(y)dy = \mathbb{E}(X)\mathbb{E}(Y)$$

A titre de contre-exemple, on montre ([16]) que deux variables aléatoires X et Y dont la densité de probabilité conjointe est

$$p_{XY}(x,y) = \frac{2}{(2\pi)^{\frac{3}{2}}}\frac{e^{-\frac{x^2+y^2}{2}}}{\sqrt{x^2+y^2}}$$

sont décorrélées mais ne sont pas indépendantes.

Densité conditionnelle La densité de probabilité conditionnelle de X/Y est la fonction

$$p_{XY}(x/y) = \frac{p_{XY}(x,y)}{p_Y(y)}$$

On a alors

$$p_X(x) = \int_{-\infty}^{+\infty} p_{XY}(x/y)p_Y(y)dy$$

Cette relation est l'équivalent du théorème des probabilités totales précédemment démontré. On a également le pendant en densités de probabilités de la formule de Bayes :

$$p_{XY}(x/y) = \frac{p_{XY}(y/x)p_X(x)}{p_Y(y)}$$

C.2.2 Variable aléatoire vectorielle

Tous les concepts précédemment vus pour deux variables aléatoires réelles peuvent être généralisés pour un vecteur de n variables aléatoires.

On considère donc le vecteur aléatoire $X = (X_1, \cdots, X_n)$ où les X_i sont n variables aléatoires réelles.

On définit alors la densité de probabilité conjointe comme étant (si elle existe) la fonction positive $p_{X_1,\cdots,X_n}$ de $\mathbb{R}^n$ dans $\mathbb{R}$ telle que

$$P(x_1 < X_1 < x_1+dx_1 \cdots x_n < X_n < x_n+dx_n) = p_{X_1,\cdots,X_n}(x_1, \cdots, x_n)dx_1 \cdots dx_n$$

On définit également les densités de probabilité marginales. Elles sont au nombre de n et la densité de probabilité marginale relative à la variable aléatoire X_i est

$$p_{X_i}(x_i) = \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n-1} p_{X_1,\cdots,X_n}(x_1, \cdots, x_i, \cdots, x_n)dx_1 \cdots dx_{i-1}dx_{i+1} \cdots dx_n$$

On définit de nouveau l'espérance de la variable aléatoire X, vecteur des espérances marginales des variables aléatoires X_i :

$$\mathbb{E}\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{pmatrix} = \begin{pmatrix} \int_{-\infty}^{+\infty} x_1 p_{X_1}(x_1)dx_1 \\ \vdots \\ \int_{-\infty}^{+\infty} x_n p_{X_n}(x_n)dx_n \end{pmatrix}$$

L'opérateur espérance étant linéaire, quelle que soit la matrice A, on a $\mathbb{E}(AX) = A\mathbb{E}(X)$. De plus, l'espérance d'une matrice aléatoire est la matrice des espérances de chaque élément : $\mathbb{E}((a_{ij})) = (\mathbb{E}(a_{ij}))$ et, en particulier, quelle que soit la matrice A, $tr(\mathbb{E}(A)) = \mathbb{E}(tr(A))$.

L'équivalent de la variance est désormais la matrice de variance-covariance du vecteur X définie par

$$var(X) = \mathbb{E}[(X - \mathbb{E}(X)).(X - \mathbb{E}(X))^T]$$

Les éléments diagonaux de cette matrice symétrique sont les variances marginales des variables aléatoires X_i et les autres éléments sont les covariances des variables aléatoires X_i, X_j. En particulier, si les variables aléatoires X_i sont indépendantes deux à deux, la matrice $var(X)$ est diagonale.

Enfin, quelle que soit la matrice A,

$$\begin{aligned} var(AX) &= \mathbb{E}[(AX - \mathbb{E}(AX)).(AX - \mathbb{E}(AX))^T] \\ var(AX) &= \mathbb{E}[A(X - \mathbb{E}(X)).(X - \mathbb{E}(X))^T A^T] = A.var(X).A^T \end{aligned}$$

C.2.3 Quelques lois de probabilité usuelles

Pour de plus amples précisions sur ces lois de probabilité, le lecteur est invité à se reporter aux ouvrages donnés en références.

Loi normale

Une variable aléatoire réelle X d'espérance μ et d'écart-type σ suit une loi normale (ou loi de Laplace-Gauss ou loi gaussienne) $\mathcal{N}(\mu, \sigma)$ lorsque sa densité de probabilité est

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

La loi normale standard (ou loi normale centrée réduite) est la loi normale $\mathcal{N}(0, 1)$.

Une variable aléatoire vectorielle X de dimension n, d'espérance vectorielle μ et de matrice de variance-covariance Σ supposée inversible suit une loi normale multidimensionnelle si sa densité de probabilité conjointe est donnée par

$$p_X(x) = \frac{1}{(\sqrt{2\pi})^n \det \Sigma} e^{-\frac{1}{2}(x-\mu)^T.\Sigma^{-1}.(x-\mu)}$$

On montre ([6]) que toute combinaison linéaire de variables aléatoires suivant des lois normales est une variable aléatoire suivant également une loi normale [4].

4. Ceci est valable pour les variables aléatoires réelles et vectorielles.

Loi du χ^2

Soient n variables aléatoires réelles X_i suivant une loi normale standard, deux à deux indépendantes. La loi suivie par la variable aléatoire $\chi_n^2 = \sum_{i=1}^n X_i^2$ est appelée loi du χ^2 à n degrés de liberté $\chi^2(n)$.

Loi de Student

Soient X et χ_n^2 deux variables aléatoires indépendantes suivant respectivement les lois $\mathcal{N}(0,1)$ et $\chi^2(n)$. La variable aléatoire $\frac{X}{\sqrt{\frac{\chi^2(n)}{n}}}$ suit alors une loi de Student à n degrés de liberté $T(n)$.

C.3 Erreurs de mesure

On suppose généralement que les erreurs de mesure suivent une loi normale. Outre le fait que cette hypothèse se trouve justifiée par l'empirisme, elle est également validée par les théorèmes de la limite centrale, théorèmes de tendance vers la loi normale.

Théorème C.6 (Théorème de la limite centrale) *Si X_1, X_2, ..., X_n sont des variables aléatoires indépendantes obéissant toutes à une même loi de moyenne μ et d'écart-type σ, la variable aléatoire centrée réduite*

$$\frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$$

suit asymptotiquement la loi normale standard.

La répétition des épreuves permet donc de substituer à une loi de probabilité inconnue une loi sensiblement normale. Ainsi, dans un calcul de compensation de mesures surabondantes, on peut supposer (avec une bonne approximation) que les erreurs de mesure sont gaussiennes quand chaque mesure compensée est moyenne de quelques observations élémentaires [6].

Il existe une autre version de ce théorème [15] :

Théorème C.7 (Théorème de la limite centrale (bis)) *Soient n variables aléatoires indépendantes Y_1, Y_2, ..., Y_n et X_1, X_2, ..., X_p p variables aléatoires déduites des premières par les relations*

$$\begin{aligned}
\forall j \in \{1,...,n\} \quad &, \quad \mathbb{E}(Y_j) = \mu_j,\ var(Y_j) = \sigma_j^2 \\
\forall i \in \{1,...,p\} \quad &, \quad \forall j \in \{1,...,n\},\ a_{ij} \in \mathbb{R} \\
\forall i \in \{1,...,p\} \quad &, \quad X_i = \frac{\sum_{j=1}^n a_{ij}(Y_j - \mu_j)}{\sqrt{\sum_{j=1}^n a_{ij}^2 \sigma_j^2}}
\end{aligned}$$

Les variables aléatoires X_i suivent asymptotiquement la loi normale standard.

C.4 Estimation

Sont présentées dans cette partie quelques notions concernant l'estimation statistique, dont la méthode des moindres carrés est un exemple.

C.4.1 Notions sur les échantillons

Un échantillon de taille n est un n-uplet de variables aléatoires réelles $(X_1, \cdots, X_n)$. Un ensemble $(x_1, \cdots, x_n)$ de n réalisations des variables aléatoires X_i est une réalisation ou une observation de l'échantillon. Les lois de probabilité suivies par les variables aléatoires X_i sont les populations parentes de l'échantillon. Toute fonction $f(X_1, \cdots, X_n)$ des variables aléatoires X_i est une statistique.

Un échantillon est donc un ensemble de variables aléatoires réelles, ces variables pouvant être indépendantes ou non, issues de la même population parente ou de populations parentes différentes. Un échantillon de taille n peut aussi être considéré comme un vecteur aléatoire de dimension n suivant une loi de probabilité à n dimensions telle que la loi normale multidimensionnelle. Un échantillon abondamment étudié est l'échantillon "Indépendant et Identiquement Distribué" (échantillon IID), i.e. un n-uplet de variables aléatoires issues de la même population parente et deux à deux indépendantes.

C.4.2 Information et vraisemblance

Dans toute cette partie, on considère un échantillon $(X_1, \cdots, X_n)$.

Fonction de vraisemblance

On suppose que la densité de probabilité conjointe de l'échantillon dépend d'un paramètre θ : $p_{X_1,\cdots,X_n}(x_1, \cdots, x_n, \theta)$. On appelle alors fonction de vraisemblance la fonction

$$L(x_1, \cdots, x_n, \theta) = p_{X_1,\cdots,X_n}(x_1, \cdots, x_n, \theta)$$

D'un point de vue probabiliste, la fonction $p_{X_1,\cdots,X_n}$, θ étant connu, permet de déduire des informations sur les réalisations des variables aléatoires $(X_1, \cdots, X_n)$. Du point de vue de l'estimation, la situation est inversée : on connait des réalisations $(x_1, \cdots, x_n)$ des variables aléatoires et on veut en déduire des informations sur θ.

Si l'échantillon est IID de population parente f, la fonction de vraisemblance a l'expression simplifiée

$$L(x_1, \cdots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

Information de Fisher

A partir de la fonction de vraisemblance précédente, on définit la fonction log-vraisemblance $\mathcal{L}$ par

$$\mathcal{L}(x_1, \cdots, x_n, \theta) = \ln[L(x_1, \cdots, x_n, \theta)]$$

On appelle fonction score la fonction $\mathcal{S}$ définie par

$$\mathcal{S}(x_1, \cdots, x_n, \theta) = \frac{\partial}{\partial \theta}\mathcal{L}(x_1, \cdots, x_n, \theta)$$

Enfin, l'information de Fisher apportée par l'échantillon $(X_1, \cdots, X_n)$ sur le paramètre θ est la grandeur

$$I(\theta) = \mathbb{E}[\mathcal{S}(X_1, \cdots, X_n, \theta)^2]$$

Cette quantité est utile pour le calcul de la borne de Fréchet [16] (voir plus loin).

C.4.3 Notions sur les estimateurs

On considère un échantillon de taille n $(X_1, \cdots, X_n)$. Les statistiques $f(X_1, \cdots, X_n)$ servent généralement à estimer des paramètres auxquels est lié l'échantillon [5]. De telles statistiques sont appelées estimateurs.

Une statistique étant une variable aléatoire, un estimateur est donc lui-même une variable aléatoire, fonction de variables aléatoires. Une réalisation d'un estimateur est une estimation. Les estimateurs peuvent satisfaire un certain nombre de propriétés. Dans la suite, on considère un estimateur $\hat{T}_n$, fonction de l'échantillon $(X_1, \cdots, X_n)$ destinée à estimer le paramètre θ.

Convergence

$\hat{T}_n$ est convergent si

$$\forall \varepsilon > 0, \lim_{n \to \infty} P(\{|\hat{T}_n - \theta| \geq \varepsilon\}) = 0$$

Autrement dit, lorque la taille de l'échantillon sur lequel se fonde l'estimation tend vers l'infini, l'estimateur converge en probabilité vers la valeur qu'il estime.

5. Bien souvent, ce sont des paramètres dont dépendent les populations parentes.

Absence de biais

Le biais de $\hat{T}_n$ est la quantité

$$B(\hat{T}_n) = \mathbb{E}(\hat{T}_n) - \theta$$

On dit d'un estimateur qu'il est non-biaisé si $B(\hat{T}_n) = 0$. $\hat{T}_n$ est dit asymptotiquement sans biais si $\lim_{n\to\infty} B(\hat{T}_n) = 0$. ***Un estimateur convergent est toujours asymptotiquement sans biais.*** Par contre, convergence et absence de biais sont deux propriétés complètement indépendantes.

L'écart-type $\sigma(\hat{T}_n)$ de l'estimateur $\hat{T}_n$ mesure la précision de ce dernier, c'est-à-dire la dispersion des réalisations de $\hat{T}_n$ autour de sa valeur moyenne [6] :

$$\sigma(\hat{T}_n) = \sqrt{\mathbb{E}[(\hat{T}_n - \mathbb{E}(\hat{T}_n))^2]}$$

Une autre grandeur intéressante est l'erreur moyenne quadratique :

$$emq(\hat{T}_n) = \sqrt{\mathbb{E}[(\hat{T}_n - \theta)^2]}$$

Cette grandeur, caractéristique de la dispersion des réalisations de $\hat{T}_n$ autour de la valeur qu'il estime, mesure l'exactitude de ce dernier, exactitude directement reliée au biais $B(\hat{T}_n)$ (cf. figure C.2). En effet,

$$\begin{aligned} emq(\hat{T}_n)^2 &= \mathbb{E}[(\hat{T}_n - \theta)^2] \\ &= \mathbb{E}[(\hat{T}_n - \mathbb{E}(\hat{T}_n)) + (\mathbb{E}(\hat{T}_n) - \theta))^2] \\ &= \mathbb{E}[(\hat{T}_n - \mathbb{E}(\hat{T}_n))^2] + \mathbb{E}[(\hat{T}_n - \theta)^2] \\ &+ \mathbb{E}[(\hat{T}_n - \mathbb{E}(\hat{T}_n)).(\hat{T}_n - \theta)] \\ emq(\hat{T}_n)^2 &= \sigma(\hat{T}_n)^2 + B(\hat{T}_n)^2 \end{aligned}$$

Ainsi, lorsque $\hat{T}_n$ est non-biaisé, $emq(\hat{T}_n)$ et $\sigma(\hat{T}_n)$ (donc, par conséquent, précision et exactitude) se confondent.

Très souvent, il existe plusieurs estimateurs d'un même paramètre θ. Il est alors légitime de rechercher, parmi tous ces estimateurs, celui qui a l'erreur moyenne quadratique minimale. Si les estimateurs considérés sont non-biaisés, cela revient ainsi à trouver l'estimateur de variance minimale. C'est l'objet des propriétés qui suivent.

6. On se place implicitement dans le cas où $\hat{T}_n$ est une variable aléatoire réelle ; les notions données ici peuvent facilement être généralisées à un vecteur aléatoire.

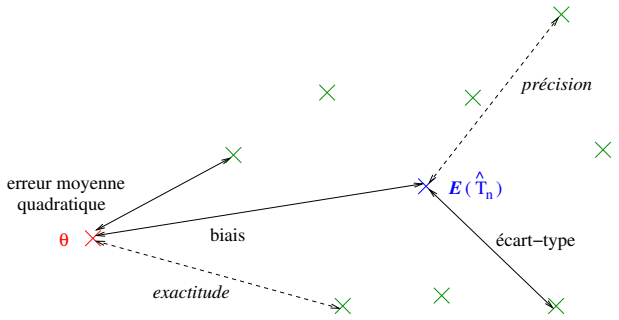

FIGURE C.2 – Biais, précision et exactitude.

Estimateur optimal

Un estimateur non-biaisé $\hat{T}_n$ du paramètre θ est dit optimal si, quelque soit l'estimateur $\hat{T}'_n$ de θ,

$$var(\hat{T}_n) \leq var(\hat{T}'_n)$$

On parle également d'estimateur de variance minimale.

Inégalité de Fréchet

Ayant défini les estimateurs optimaux, on peut maintenant se demander si la variance des estimateurs sans biais du paramètre θ admet une borne inférieure.

On considère donc un estimateur sans biais $\hat{T}_n$ d'une fonction dérivable f du paramètre θ que l'on cherche à estimer. On montre ([11]) que la variance de $\hat{T}_n$ ne peut pas être plus petite que la quantité

$$\frac{f'(\theta)}{I(\theta)}$$

Cette quantité est la borne de Fréchet $\mathcal{B}_F$.

Estimateur efficace

Soit $\hat{T}_n$ un estimateur sans biais d'une fonction dérivable f du paramètre θ. On mesure l'efficacité de $\hat{T}_n$ par le quotient

$$eff(\hat{T}_n) = \frac{\mathcal{B}_F}{var(\hat{T}_n)}$$

$\hat{T}_n$ est dit efficace si $eff(\hat{T}_n) = 1$: un estimateur efficace atteint donc la borne de Fréchet.

Un estimateur efficace est optimal. Un estimateur optimal n'est pas nécessairement efficace.

Dans le cas où θ est un paramètre multidimensionnel, on peut généraliser les notions précédentes (à l'aide de la relation d'ordre $\succeq$ précédente) et, plus particulièrement, l'inégalité de Fréchet trouve son équivalent dans les inégalités de Bhattacharyya. Le lecteur intéressé peut se reporter à [11] pour de plus amples détails.

C.4.4 Exemples d'estimation

Pour finir cette partie sur l'estimation, on aborde deux exemples de méthodes d'estimation.

Estimation Bayesienne

Dans le cadre de cette méthode d'estimation, on suppose que le paramètre θ recherché est une réalisation d'une variable aléatoire Θ dont on connait la densité de probabilité p_Θ. La densité de probabilité conjointe de l'échantillon $(X_1, \cdots, X_n)$ est alors considérée comme une densité de probabilité conditionnelle $p_X(x/\theta)$.

Exemple : si l'échantillon est de taille 1, on choisit comme estimateur l'espérance conditionnelle de Θ/X et le problème d'estimation se résume alors à la relation [16]

$$\hat{T}_1 = \mathbb{E}(\Theta/X) = \frac{1}{\int_{-\infty}^{+\infty} p_X(x/\theta) p_\Theta(\theta) d\theta} \int_{-\infty}^{+\infty} \theta p_X(x/\theta) p_\Theta(\theta) d\theta$$

Estimation par maximum de vraisemblance

La méthode d'estimation par le maximum de vraisemblance propose de choisir comme estimateur $\hat{T}_n$ du paramètre θ la valeur qui rend maximale la fonction de vraisemblance $L(x_1, \cdots, x_n, \theta)$. On peut également, en pratique, chercher plutôt à rendre maximale la fonction log-vraisemblance.

Dans le cas où l'échantillon $(X_1, \cdots, X_n)$ est IID, s'il existe un estimateur efficace, la méthode par maximum de vraisemblance fournit cet estimateur et, en général, l'estimateur du maximum de vraisemblance est convergent et donc asymptotiquement sans biais [11].

La méthode peut être généralisée pour un paramètre θ multidimensionnel (voir [16], par exemple). Dans ce cas, la fonction de vraisemblance s'écrit sous la forme

$$L(x_1, \cdots, x_n, \theta_1, \cdots, \theta_m) = p_{X_1,\cdots,X_n}(x_1, \cdots, x_n, \theta_1, \cdots, \theta_m)$$

Rendre maximale L équivaut à rendre maximale la fonction log-vraisemblance associée $\mathcal{L}$. L'estimateur du maximum de vraisemblance est donc le vecteur θ solution des équations

$$\frac{\partial \mathcal{L}}{\partial \theta_i},\ i \in \{1, ..., m\}$$

la forme quadratique de matrice $(\frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j})$ étant définie négative.

C.5 Tests statistiques

Dans cette dernière partie, quelques notions de base sur les tests statistiques sont présentées. Pour plus d'informations, consulter [6] et [16].

Le principe général est de tester une hypothèse sur la base d'un échantillon dont le comportement général dépend de la vérification (ou non) de cette hypothèse.

On considère donc un échantillon $(X_1, ..., X_n)$ de densité de probabilité notée $p_X(x_1, ..., x_n, \theta)$. p_X est supposée faible dans un ensemble de régions Ω_c de l'univers des possibles et prend des valeurs significatives dans $\bar{\Omega}_c$. On se donne une hypothèse H_0. On rejette H_0 lorsqu'une réalisation de $(X_1, ..., X_n)$ se trouve dans Ω_c et on accepte H_0 lorsque cette réalisation se trouve dans $\bar{\Omega}_c$.

H_0 est l'hypothèse nulle et sa négation H_1 est l'hypothèse alternative. Sous l'hypothèse H_1, la densité de probabilité de l'échantillon devient généralement $p_X(x_1, ..., x_n, \theta')$ avec $\theta' \neq \theta$. Ω_c est la région critique du test et son complémentaire $\bar{\Omega}_c$ est la région d'acceptation du test. Les risques de première et seconde espèce α et β sont définis dans le tableau (C.2).

vérité / décision	H_0 vraie	H_1 vraie
accepter H_0	bonne décision	β
accepter H_1	α	bonne décision

TABLE C.2 – Risques de première et de seconde espèce.

α est donc la probabilité de refuser à tort H_0 et β celle d'accepter à tort H_0. La quantité $\pi = 1 - \beta$ est la puissance du test.
Exemple : un test basé sur l'échantillon IID $(X_1, ..., X_n)$ peut être le suivant :

$$\begin{cases} H_0 & : & X \hookrightarrow \mathcal{N}(\theta, 1) \\ H_1 & : & X \hookrightarrow \mathcal{N}(\theta', 1),\ \theta' \neq \theta \end{cases}$$

On se fixe alors un risque de première espèce α. Si t est le fractile d'ordre $1-\alpha$ de la loi normale $\mathcal{N}(\theta, 1)$, les risques de première et seconde espèces sont représentés sur la figure (C.3).

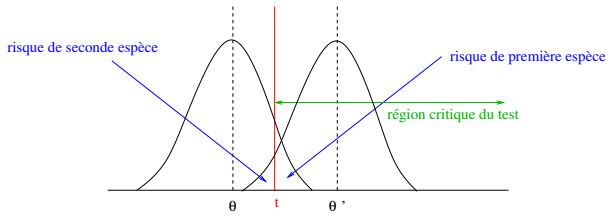

FIGURE C.3 – Risques de première et de seconde espèce.

Bibliographie

[1] Bicking, F., C. Fonteix, J.P. Corriou, I. Marc. Global optimization by artificial life : a new technique using genetic population evolution. *Operations Research*, vol. 28, 1 : 23-36, 1994.

[2] Berthiau, G., P. Siarry. Etat de l'art des méthodes d'"optimisation globale". *RAIRO Operations Research*, 35 : 329-365, 2001.

[3] Bonnel, T. Combinaison de jeux de coordonnées de stations terrestres en norme L1. Publications du LAREG, MS11, juin 1999.

[4] Coulot, D., J. Nicolas, P. Exertier. *Contribution de la télémétrie laser à la nouvelle représentation du système de référence terrestre et à la géodynamique.* Bulletin d'information de l'IGN n°73, mars 2002.

[5] Hottier, P. *Analyse numérique : moindres carrés.* ENSG, 1990.

[6] Hottier, P. *Précis de statistiques.* ENSG, 1990.

[7] Karr, C.L., B. Weck, D.L. Massart, P. Vankeerberghen. Least median squares curve fitting using a genetic algorithm. *Eng. Applic. Artif. Intell.* vol. 8, 2 : 177-189, 1995.

[8] Laudenbach, F. *Calcul différentiel et intégral.* Editions de l'école polytechnique, Ellipses, 2000.

[9] Malliavin, M.P. *Formes quadratiques et groupes classiques.* Cours de maîtrise. Université Pierre et Marie Curie, Paris.

[10] Papoulis, A. *Probability, random variables and stochastic processes.* Mc Graw-Hill, 1991.

[11] Pelat, D. *Bruits et signaux.* Cours de l'école doctorale d'Ile-de-France, Astronomie-Astrophysique.

[12] Persson, C.G. Adjustment, weight-testing and detection of outliers in mixed SFF-SFS models *Manuscripta Geodetica*, 7 :299-323, 1982.

[13] Sahin, M., P.A. Cross, P.C. Sellers. Variance component estimation applied to satellite laser ranging. *Bulletin Géodésique*, 66 :284-295, 1992.

[14] Sebag, M., Schoenauer M. Contrôle d'un algorithme génétique. *Revue française d'intelligence artificielle.* 1996

[15] Sillard, P. *Estimation par moindres carrés.* Collection ENSG-IGN, Hermes Sciences Publications, Lavoisier, 2001.

[16] Sillard, P. *Probabilités et statistiques.* ENSG, 2000.

[17] Withley, D. *A Genetic Algorithm Tutorial.* Technical report Cs-93-103. Department of Computer Science, Colorado State University, Fort Collins, 1993.