

Chapitre 3

Théorie de l'estimation

3.1 Position du problème

On considère un vecteur $Z \in \mathbb{R}^N$ qui est fonction d'un autre vecteur $\theta \in \mathbb{R}^p$:

$$Z = h(\theta).$$

Le problème de l'estimation peut être énoncé de manière très générale de la manière suivante :

Si j'observe une réalisation $z \in \mathbb{R}^N$ de Z , que puis-je en tirer comme information sur θ ? De manière plus précise, on souhaite construire un estimateur de θ , noté $\hat{\theta}$.

Définition : Un estimateur $\hat{\theta}$ de θ est une fonction déterministe $g(Z)$ qui à toute réalisation z du vecteur Z fait correspondre une valeur particulière $\hat{\theta}$:

$$\hat{\theta} = g(z).$$

On peut distinguer les trois cas suivants :

1. θ et Z sont des vecteurs aléatoires ayant des densités de probabilité : l'estimateur appartiendra alors à la famille des estimateurs de Bayes.
2. θ est déterministe ; Z est un vecteur aléatoire ayant une densité de probabilité : l'estimateur appartiendra alors à la famille des estimateurs de Fisher.
3. θ est déterministe ; Z est une fonction bruitée mais non probabiliste de θ . Dans ce cas on peut encore définir un estimateur en dehors de tout cadre probabiliste.

Exemple 1

Considérons un signal utile $X(k) \in \mathbb{R}^n$ modélisé comme un processus stochastique Markovien :

$$X(k+1) = AX(k) + W(k),$$

avec $W(k)$ un bruit blanc Gaussien de moyenne nulle et de covariance Q , et avec $X(0)$ Gaussien : $X(0) \in N(X_0, P_0)$. Supposons que l'on dispose de mesures $Y(k)$ qui soient des combinaisons linéaires bruitées de $X(k)$:

$$Y(k) = CX(k) + V(k)$$

où $V(k)$ est aussi un bruit blanc Gaussien de moyenne nulle et de variance R , non corrélé avec $W(\cdot)$ ni avec $X(0)$. Par exemple, la mesure $Y(k)$ est la première composante de $X(k)$.

On peut alors poser le problème de l'estimation optimale du signal utile $X(k)$ à partir du signal mesuré $y(0), y(1), \dots, y(k)$. Dans ce cas :

$$Z \triangleq \begin{pmatrix} y(0) \\ y(1) \\ \vdots \\ y(k) \end{pmatrix} \quad \text{et} \quad \theta \triangleq X(k).$$

Remarquons que Z et θ sont tous deux des vecteurs aléatoires. Le problème posé est celui du *filtrage optimal*. La solution est fournie par le filtre de Kalman, qui produit une estimation réursive de $\hat{X}(k)$ à partir des mesures $\{y(0), y(1), \dots, y(k)\}$: voir Chapitre 4.

On pourrait aussi vouloir estimer

$$\theta = \begin{pmatrix} X(0) \\ X(1) \\ \vdots \\ X(k) \end{pmatrix}$$

à partir du même vecteur Z . Le problème est alors celui du *lissage optimal*.

Exemple 2

Supposons que l'on observe N échantillons $y(1), y(2), \dots, y(N)$ d'une variable aléatoire Gaussienne $Y \in \mathbb{R}$, de moyenne μ et de variance σ^2 . A partir de ces mesures, on souhaite construire une estimée de μ et de σ . Dans ce cas on a :

$$z = (y(1), y(2), \dots, y(N))^T, \quad \theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$$

On pourrait, par exemple, considérer les estimateurs suivants :

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^N y(k) \quad \text{et} \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^N (y(k) - \hat{\mu})^2$$

Dans cet exemple, Z est aléatoire, mais θ est déterministe.

Exemple 3

Considérons que l'on veuille estimer avec grande précision la valeur d'une résistance R , et que l'on fasse N expériences indépendantes où l'on mesure le courant $I(k)$ passant par la résistance et la tension $U(k)$ à ses bornes. Ces mesures de courants et de tensions sont toutes entachées d'erreurs, mais on peut n'avoir aucune information sur la densité de probabilité de ces erreurs. Dans ce cas on a

$$\begin{aligned} z &= (I(1), U(1), I(2), U(2), \dots, I(N), U(N)) \\ \theta &= R \end{aligned}$$

Bien que z et θ ne soient pas supposées avoir une densité de probabilité, on peut vouloir construire un estimateur $\hat{\theta} = g(z)$ qui aurait des propriétés intéressantes. Remarquons que plusieurs estimateurs différents semblent raisonnables :

$$\hat{R}_1 = \frac{\sum_{k=1}^N U(k)I(k)}{\sum_{k=1}^N I^2(k)}, \quad \hat{R}_2 = \frac{1}{N} \sum_{k=1}^N \frac{U(k)}{I(k)}, \quad \hat{R}_3 = \frac{\sum_{k=1}^N U(k)}{\sum_{k=1}^N I(k)}$$

L'estimateur \hat{R}_1 est obtenu en minimisant $\sum_{k=1}^N [U(k) - RI(k)]^2$. Ces trois estimateurs ont des propriétés très différentes, comme on le verra plus loin.

3.2 Propriétés des estimateurs

Pour juger de la qualité d'un estimateur on vérifie s'il a certaines propriétés jugées intéressantes.

3.2.1 Biais d'un estimateur

On dira qu'un estimateur $\hat{\theta} = g(Z)$ de θ est *sans biais* si

i)

$$m_{\hat{\theta}} = \mathbf{E}\{g(Z)|\theta\} = \theta \quad (3.1)$$

dans le cas où θ est un paramètre inconnu mais fixe. Dans ce cas, l'expression (3.1) signifie :

$$\mathbf{E}\{g(Z)|\theta\} = \int_z g(z)T_{Z|\theta}(z|\theta)dz = \theta \quad (3.2)$$

ii)

$$m_{\hat{\theta}} = \mathbf{E}\{g(Z)\} = \mathbf{E}\{\theta\} = m_{\theta} \quad (3.3)$$

dans le cas où θ est aléatoire de moyenne m_{θ} . Dans ce cas l'expression (3.3) signifie

$$\mathbf{E}\{g(Z)\} = \int_{\theta} \left[\int_z g(z)T_{Z|\theta}(z|\theta)dz \right] T_{\Theta}(\theta)d\theta = m_{\theta} \quad (3.4)$$

Si l'estimateur $\hat{\theta} = g(Z)$ est biaisé, on définit le biais $b_{\hat{\theta}}$ de l'estimateur :

- $b_{\hat{\theta}} = m_{\hat{\theta}} - \theta$ si θ est déterministe.
- $b_{\hat{\theta}} = m_{\hat{\theta}} - m_{\theta}$ si θ est aléatoire

3.2.2 Variance d'un estimateur

La matrice de covariance de l'estimateur $\hat{\theta} = g(Z)$ est définie comme pour n'importe quel vecteur aléatoire :

$$C_{\hat{\theta}} = \mathbf{E}\{(\hat{\theta} - m_{\hat{\theta}})(\hat{\theta} - m_{\hat{\theta}})^T\} \quad (3.5)$$

Encore une fois, l'opérateur espérance mathématique est pris par rapport à Z si θ est déterministe, et par rapport à Z et Θ si Θ est aléatoire. La matrice $C_{\hat{\theta}}$ est symétrique semi-définie positive. Si θ est scalaire, on parlera de la variance de l'estimateur.

3.2.3 Erreur quadratique moyenne d'un estimateur

L'erreur quadratique moyenne de l'estimateur $\hat{\theta} = g(Z)$ est définie comme suit :

$$EQM_{\hat{\theta}} = \mathbf{E}\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\} \quad (3.6)$$

Pour l'estimateur d'un vecteur θ déterministe on a :

$$EQM_{\hat{\theta}} = C_{\hat{\theta}} + b_{\hat{\theta}}b_{\hat{\theta}}^T \quad (3.7)$$

Dans le cas scalaire, cela s'écrit encore :

$$EQM_{\hat{\theta}} = \sigma_{\hat{\theta}}^2 + b_{\hat{\theta}}^2 \quad (3.8)$$

La formule (3.7) se démontre de la façon suivante :

$$\begin{aligned} EQM_{\hat{\theta}} &= \mathbf{E}\{(\hat{\theta} - m_{\hat{\theta}} + m_{\hat{\theta}} - \theta)(\hat{\theta} - m_{\hat{\theta}} + m_{\hat{\theta}} - \theta)^T\} \\ &= C_{\hat{\theta}} + b_{\hat{\theta}}b_{\hat{\theta}}^T + \mathbf{E}\{(\hat{\theta} - m_{\hat{\theta}})(m_{\hat{\theta}} - \theta)^T\} + \mathbf{E}\{(m_{\hat{\theta}} - \theta)(\hat{\theta} - m_{\hat{\theta}})^T\} \end{aligned}$$

Considérons le troisième terme. Puisque $m_{\hat{\theta}}$ et θ sont déterministes, on a

$$\mathbf{E}\{(\hat{\theta} - m_{\hat{\theta}})(m_{\hat{\theta}} - \theta)^T\} = \mathbf{E}\{\hat{\theta} - m_{\hat{\theta}}\}(m_{\hat{\theta}} - \theta)^T = 0$$

car $\mathbf{E}\{\hat{\theta}\} = m_{\hat{\theta}}$. Le quatrième terme est la transposée du troisième.

3.2.4 Estimateur efficace

Soit $\hat{\theta} = g(Z)$ un estimateur non biaisé de θ . On dira alors que $g(Z)$ est un estimateur efficace, si pour tout autre estimateur non-biaisé $\theta^* = f(Z)$, on a

$$C_{\hat{\theta}} \leq C_{\theta^*} \quad (3.9)$$

Autrement dit, un estimateur non-biaisé est appelé efficace si sa covariance (sa variance dans le cas scalaire) est plus petite que celle de n'importe quel autre estimateur non-biaisé. Comme les matrices de covariance sont des matrices symétriques semi-définies positives, l'inégalité (3.9) signifie que la matrice $C_{\theta^*} - C_{\hat{\theta}}$ est une matrice symétrique semi-définie positive. Un estimateur efficace est aussi appelé *estimateur à variance minimale*.

3.2.5 Matrice d'information de Fisher et borne de Cramér-Rao

Soit $\theta \in \mathbb{R}^n$ un vecteur de paramètres inconnu mais fixe, $Z \in \mathbb{R}^N$ un vecteur d'observations lié à θ , et supposons que la fonction de densité de probabilité $T_{Z|\theta}(z|\theta)$ soit différentiable par rapport à θ . On définit alors la *matrice d'information de Fisher* $I(\theta)$ comme :

$$I(\theta) = [I_{k,j}(\theta)]_{k,j=1,n} \in \mathbb{R}^{n \times n} \quad (3.10)$$

avec

$$I_{k,j}(\theta) = \mathbf{E} \left\{ \frac{\partial \ln T_{Z|\theta}(z|\theta)}{\partial \theta_k} \times \frac{\partial \ln T_{Z|\theta}(z|\theta)}{\partial \theta_j} \right\} \quad (3.11)$$

$$= -\mathbf{E} \left\{ \frac{\partial^2 \ln T_{Z|\theta}(z|\theta)}{\partial \theta_k \partial \theta_j} \right\} \quad (3.12)$$

Soit maintenant $\hat{\theta} = g(Z)$ n'importe quel estimateur *non-biaisé* d'un paramètre inconnu mais déterministe θ . Alors

$$\mathbf{E}\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\} \geq I^{-1}(\theta) \quad (3.13)$$

Ce résultat est connu sous le nom *d'inégalité de Cramér-Rao*. L'inverse de la matrice d'information de Fisher est donc une borne inférieure de la covariance de n'importe quel estimateur. Si la covariance d'un estimateur est égale à $I^{-1}(\theta)$, cet estimateur est donc efficace.

Souvent le vecteur Z est constitué de N observations indépendantes $Z = (Z_1, Z_2, \dots, Z_N)$ qui ont toutes la même densité de probabilité. Dans ce cas, on a

$$T_{Z|\theta}(z|\theta) = \prod_{i=1}^N T_{Z_i|\theta}(z_i|\theta) \quad (3.14)$$

et donc :

$$I(\theta) = N.M(\theta)$$

où

$$M_{k,j}(\theta) = \mathbf{E} \left\{ \frac{\partial \ln T_{Z_1|\theta}(z_1|\theta)}{\partial \theta_k} \times \frac{\partial \ln T_{Z_1|\theta}(z_1|\theta)}{\partial \theta_j} \right\} \quad (3.15)$$

$$= -\mathbf{E} \left\{ \frac{\partial^2 \ln T_{Z_1|\theta}(z_1|\theta)}{\partial \theta_k \partial \theta_j} \right\} \quad (3.16)$$

L'inégalité de Cramér-Rao peut alors se réécrire :

$$\mathbf{E}\{(\hat{\theta}_N - \theta)(\hat{\theta}_N - \theta)^T\} \geq I^{-1}(\theta) = \frac{1}{N} M^{-1}(\theta).$$

où $M(\theta)$ est défini en (3.15). Dans ce cas, $M(\theta)$ est une matrice fixe (indépendante du nombre de données), et on constate donc que la borne de Cramér-Rao est

inversément proportionnelle au nombre d'observations N .

Exemple : Considérons une séquence aléatoire $Z = (Z_1, Z_2, \dots, Z_N)$ où les Z_K sont tous distribués de manière identique selon une densité Gaussienne de moyenne inconnue μ et de variance 1. On veut estimer μ à partir de Z . La densité de probabilité de Z est donc :

$$T_{Z|\mu}(z|\mu) = \frac{1}{(2\pi)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2} \sum_{k=1}^N (z_k - \mu)^2 \right\}$$

D'où :

$$\frac{\partial \ln T_{Z|\mu}(z|\mu)}{\partial \mu} = \sum_{k=1}^N (z_k - \mu)$$

et

$$\frac{\partial^2 \ln T_{Z|\mu}(z|\mu)}{\partial \mu^2} = -N.$$

On sait donc que, quel que soit l'estimateur $\hat{\mu}(Z)$, on aura toujours $\text{Var}(\hat{\mu}(Z)) \geq \frac{1}{N}$. Considérons l'estimateur $\hat{\mu}(Z) = \bar{Z} = \frac{1}{N} \sum_{k=1}^N Z_K$. On vérifie facilement que $\text{Var}(\bar{Z}) = \frac{1}{N}$. Cet estimateur est donc efficace.

Les propriétés que nous venons de décrire sont définies pour les estimateurs d'un vecteur θ à partir d'un vecteur aléatoire $Z \in \mathbb{R}^N$, de dimension fixe. Dans beaucoup de situations pratiques, l'on veut estimer un vecteur θ à partir d'un nombre croissant d'observations, c'est à dire que $Z^N = (Z_1, Z_2, \dots, Z_N)$ avec $N \rightarrow \infty$. On désigne alors par $\hat{\theta}_N$ l'estimée basée sur N observations, $\hat{\theta}_N = g(Z^N)$, et on s'intéresse aux propriétés asymptotiques de l'estimateur, c'est à dire les propriétés de $\hat{\theta}_N$ quand $N \rightarrow \infty$. Pour définir les notions de biais asymptotique, covariance asymptotique, etc.. on doit d'abord introduire quelques notions de convergence.

3.3 Notions de convergence stochastique

Il existe plusieurs définitions distinctes pour la convergence de séquences aléatoires. Nous donnons ici les trois définitions les plus courantes.

3.3.1 Convergence en probabilité

Définition 1 : Soit une séquence aléatoire $X_k \in \mathbb{R}^n, k \in \mathbb{Z}_+$. On dit que X_k converge en probabilité vers X si

$$\forall \epsilon > 0 : \lim_{k \rightarrow \infty} p[\|X_k - X\| > \epsilon] = 0.$$

On utilise souvent la notation $p \lim_{k \rightarrow \infty} X_k = X$. La convergence en probabilité est la plus faible des notions de convergence stochastique.

3.3.2 Convergence en moyenne quadratique

Définition 2 : Soit une séquence aléatoire $X_k \in \mathbb{R}^n, k \in \mathbb{Z}_+$. On dit que X_k converge en moyenne quadratique vers X si

$$\lim_{k \rightarrow \infty} E\{\|X_k - X\|^2\} = 0$$

On utilise souvent la notation $l.i.m.X_k = X$ où *l.i.m.* signifie “limit in the mean”. Remarquons que si X_k est un estimateur de θ et que $X = \theta$, la convergence en moyenne quadratique de $\hat{\theta}_k$ vers θ signifie que la variance et le biais tendent tous deux vers zéro, en vertu de l’expression (3.7).

3.3.3 Convergence presque partout (ou avec probabilité 1)

Pour définir la convergence presque partout, il faut se rappeler que toute réalisation d’une variable aléatoire est liée à un événement ω dans l’espace des événements Ω . Il en est de même d’une séquence aléatoire. Pour un ω donné, $x_1(\omega), x_2(\omega), \dots, x_k(\omega)$ est donc une séquence de nombres qui peut converger (ou non) vers un nombre $x(\omega)$. Si pour tout ω , la réalisation $x_k(\omega)$ converge vers un nombre $x(\omega)$, éventuellement différent pour chaque ω , on dit que la séquence aléatoire X_k converge presque partout. Cette définition est beaucoup trop forte et on la remplace par la convergence presque partout.

Définition 3 : Soit une séquence aléatoire $X_k \in \mathbb{R}^N, k \in \mathbb{Z}_+$. On dit que X_k converge vers X presque partout (ou avec probabilité 1) si

$$p[X_k(\omega) \xrightarrow[k \rightarrow \infty]{} X(\omega)] = 1$$

De manière plus précise

$$x_k(\omega) \xrightarrow[k \rightarrow \infty]{} x(\omega) \quad \forall \omega \in A \subset \Omega, \text{ avec } P(A) = 1.$$

En anglais, on utilise les notations

$$X_k \rightarrow X \quad a.s., a.e. \text{ ou } w.p.1$$

(*a.s.* = almost surely, *a.e.* = almost everywhere, *w.p.1* = with probability one). La convergence presque partout est une notion de convergence forte ; l’estimateur est parfois appelé *fortement convergent* (strongly consistent). Enfin, la convergence presque partout et la convergence en moyenne quadratique impliquent toutes deux la convergence en probabilité. Un estimateur qui converge en probabilité est parfois appelé *faiblement convergent* (weakly consistent).

3.4 Propriétés asymptotiques des estimateurs

Ayant présenté les notions de convergence, nous pouvons maintenant décrire certaines propriétés qu’il est intéressant qu’un estimateur possède lorsque le

nombre d'observations tend vers l'infini. Ces propriétés sont appelées asymptotiques.

3.4.1 Estimateur asymptotiquement non-biaisé

Considérons un estimateur $\hat{\theta}_N = g(Z^N)$ basé sur un vecteur d'observations $Z^N = (Z_1, \dots, Z_N)$ avec N croissant. L'estimateur est dit asymptotiquement non-biaisé si

$$\text{i) } \lim_{N \rightarrow \infty} \mathbf{E}\{g(Z^N)|\theta\} = \theta \quad (3.17)$$

dans le cas où θ est un paramètre inconnu mais fixe ;

$$\text{ii) } \lim_{N \rightarrow \infty} \mathbf{E}\{g(Z^N)\} = m_\theta \quad (3.18)$$

dans le cas où θ est un vecteur aléatoire de moyenne m_θ .

3.4.2 Estimateur asymptotiquement efficace

Soit $\hat{\theta}_N = g(Z^N)$ un estimateur qui converge en moyenne quadratique vers θ . Rappelons que cela implique que le biais et la variance tendent vers zéro. On dira que cet estimateur est asymptotiquement efficace s'il existe un $N_0 > 0$ tel que pour tout autre estimateur $\theta_N^* = f(Z^N)$ et pour tout $N > N_0$, on ait :

$$\mathbf{E}\{(\hat{\theta}_N - \theta)(\hat{\theta}_N - \theta)^T\} \leq \mathbf{E}\{(\theta_N^* - \theta)(\theta_N^* - \theta)^T\}.$$

On dit aussi que $\hat{\theta}_N$ est un estimateur asymptotiquement à variance minimale de θ .

3.4.3 Estimateur asymptotiquement efficace et normal

Un estimateur $\hat{\theta}_N = g(Z^N)$ est appelé *asymptotiquement efficace et normal* (en anglais : Best Asymptotically Normal, ou BAN) si

$$\lim_{N \rightarrow \infty} T_{\hat{\theta}_N|\theta}[\hat{\theta}_N - \theta|\theta] = N(0, I^{-1}(\theta)) \quad (3.19)$$

où $I(\theta)$ est la matrice d'information de Fisher. Lorsque Z^N est constitué de N observations $Z^N = (Z_1, Z_2, \dots, Z_N)$ ayant toutes la même densité de probabilité, on a $I(\theta) = N \cdot M(\theta)$ (voir Sec. 3.2.5), et on écrira :

$$\sqrt{N}(g(Z^N) - \theta) \xrightarrow{N \rightarrow \infty} N(0, M^{-1}(\theta)) \quad (3.20)$$

3.5 Exemple : Estimation d'une résistance

Pour illustrer les propriétés statistiques des estimateurs, et aussi pour montrer que, même dans un problème très simple, on peut construire des estimateurs

ayant des propriétés fort différentes, on va reprendre l'exemple de l'estimation d'une résistance à partir d'un ensemble de mesures du courant I_k et de la tension U_k .¹

Supposons que l'on fasse passer un courant I , toujours le même, dans une résistance R , produisant la tension $U = RI$ à ses bornes. Supposons aussi que U et I soient mesurés avec du bruit, et que les mesures obtenues soient en fait :

$$I_k = I + i_k \quad \text{et} \quad U_k = U + u_k.$$

Supposons enfin que les bruits à des instants de mesures différents soient indépendants, que les bruits de mesure de courant et de tension soient indépendants, et que ces bruits soient de moyenne nulle. Donc $\forall k, j$:

$$\mathbf{E}\{i_k\} = 0, \mathbf{E}\{u_k\} = 0, \mathbf{E}\{i_k i_j\} = \sigma_i^2 \delta_{kj}, \mathbf{E}\{u_k u_j\} = \sigma_u^2 \delta_{kj}, \mathbf{E}\{u_k i_j\} = 0.$$

Examinons maintenant comment se comportent les trois estimateurs proposés à la section 1.

L'estimateur \hat{R}_1

S'il n'y avait pas de bruit de mesure, on aurait $U = RI$. Si I et U sont remplacés par les mesures U_k et I_k , cette équation n'est plus valable, et on peut écrire :

$$U_k - RI_k = e_k \quad (3.21)$$

où e_k est "l'erreur d'équation". En minimisant

$$V^{LS}(R) \triangleq \sum_{k=1}^N e_k^2 = \sum_{k=1}^N (U_k - RI_k)^2 \quad (3.22)$$

par rapport à R , on obtient l'estimateur des moindres carrés (que nous avons appelé \hat{R}_1 dans l'exemple 3 de la section 3.1) :

$$\hat{R}_1 = \hat{R}_N^{LS} = \frac{\frac{1}{N} \sum_{k=1}^N (UI + U i_k + I u_k + i_k u_k)}{\frac{1}{N} \sum_{k=1}^N (I^2 + 2I i_k + i_k^2)} \quad (3.23)$$

$$= \frac{R + \frac{1}{N} \sum_{k=1}^N (R \frac{i_k}{I} + \frac{u_k}{I} + \frac{i_k u_k}{I^2})}{1 + \frac{1}{N} \sum_{k=1}^N (2 \frac{i_k}{I} + \frac{i_k^2}{I^2})} \quad (3.24)$$

$$= \frac{R + \alpha}{1 + \beta} \simeq (R + \alpha)(1 - \beta + \beta^2 + \dots) \quad (3.25)$$

$$= R + \alpha - \beta R - \alpha \beta + \beta^2 R + \alpha \beta^2 \quad (3.26)$$

En utilisant les hypothèses statistiques sur les i_k et u_k , et en négligeant les termes d'ordre 3 ou plus, on trouve :

$$\mathbf{E}\{\alpha\} = 0, \mathbf{E}\{\beta\} = \frac{\sigma_i^2}{I^2}, \mathbf{E}\{\alpha\beta\} = \frac{2}{N} \frac{\sigma_i^2}{I^2} R, \mathbf{E}\{\beta^2\} = \frac{4}{N} \frac{\sigma_i^2}{I^2}$$

¹Cet exemple est développé dans le livre J. Schoukens et R. Pintelon, "Identification of linear systems : a practical guideline to accurate modeling", Pergamon Press, 1991.

On a donc :

$$\mathbf{E}\{\hat{R}_N^{LS}\} \simeq R\left(1 - \frac{\sigma_i^2}{I^2} + \frac{2}{N} \frac{\sigma_i^2}{I^2}\right)$$

On voit donc que l'estimateur des moindres carrés est biaisé. Il est même asymptotiquement biaisé puisque

$$\lim_{N \rightarrow \infty} \mathbf{E}\{\hat{R}_N^{LS}\} \simeq R\left(1 - \frac{\sigma_i^2}{I^2}\right)$$

Ce biais est dû uniquement au bruit de mesure sur le courant. La figure 3.1 montre l'évolution de l'estimée d'une résistance $R = 1$ en fonction du nombre de mesures pour des valeurs $U = I = 1$ et pour des erreurs de mesure ayant une distribution uniforme, centrée sur zéro et de variance $\sigma_i = \sigma_u = 0.5$.

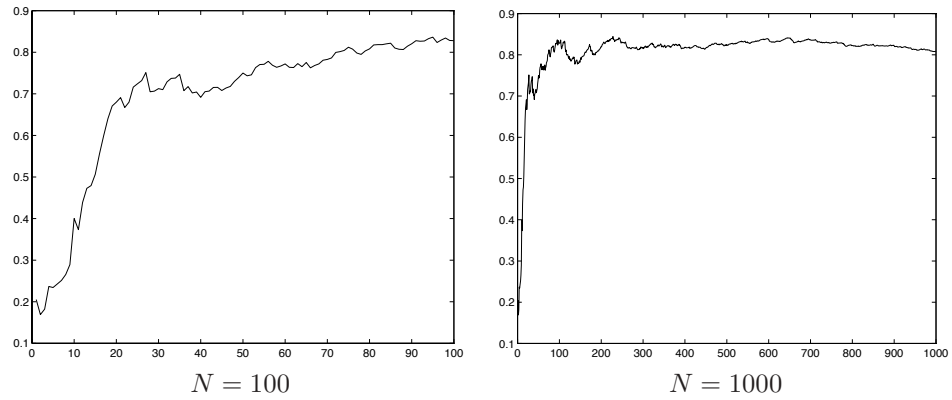


FIG. 3.1 – Estimée de la résistance par moindres carrés utilisant les 100 premières mesures (à gauche) et 1000 mesures (à droite)

les 100 premières mesures (à gauche) et 1000 mesures (à droite)

Pour la variance de l'estimateur, on trouve :

$$\sigma_{\hat{R}_N^{LS}}^2 = \mathbf{E}[\hat{R}_N^{LS} - \mathbf{E}\{\hat{R}_N^{LS}\}]^2 = \frac{R^2}{N} \left(\frac{\sigma_u^2}{U^2} + \frac{\sigma_i^2}{I^2} \right)$$

On voit donc que la variance est proportionnelle à la somme des rapports bruit/signal du courant et de la tension, et qu'elle tend vers zéro asymptotiquement comme $\frac{1}{N}$. Ceci est typique de beaucoup d'estimateurs. On constate aussi que, si le bruit de mesure de la tension n'induit aucun biais, il a par contre un impact sur la variance de l'estimateur.

L'estimateur \hat{R}_2

L'estimateur \hat{R}_2 paraît assez intuitif. Il est obtenu en estimant la résistance \hat{R}_k après chaque mesure $(U(k), I(k))$ comme le rapport de la tension au courant,

puis en faisant la moyenne de toutes ces estimées. On obtient donc

$$\hat{R}_2 = \frac{1}{N} \sum_{k=1}^N \frac{U(k)}{I(k)} \quad (3.27)$$

Le biais asymptotique de cet estimateur n'est pas bien défini car pour certaines réalisations du bruit de mesure i_k le rapport $\frac{U(k)}{I(k)}$ peut tendre vers l'infini. En pratique, on constate que l'estimateur surévalue toujours la vraie valeur de R .

L'estimateur \hat{R}_3

Enfin, plutôt que de minimiser la somme des carrés des erreurs (voir (3.22)), on peut minimiser le critère suivant :

$$V^{MV} = \sum_{k=1}^N \left(\frac{u_k^2}{\sigma_u^2} + \frac{i_k^2}{\sigma_i^2} \right) \quad (3.28)$$

$$= \sum_{k=1}^N \left[\frac{(U_k - U)^2}{\sigma_u^2} + \frac{(I_k - I)^2}{\sigma_i^2} \right] \quad (3.29)$$

à condition de connaître σ_u et σ_i . Il faut alors minimiser V^{MV} par rapport à U , I et R , en tenant compte que $U = RI$. Ce problème de minimisation sous contraintes se résout en utilisant un paramètre de Lagrange λ . On remplace V^{MV} par

$$V' = V^{MV} + \lambda(U - RI)$$

et on minimise V' par rapport à U , I , R et λ . On trouve :

$$\hat{R}_3 = \hat{R}_N^{MV} = \frac{\sum_{k=1}^N U_k}{\sum_{k=1}^N I_k}$$

On peut montrer que cet estimateur est asymptotiquement non-biaisé, $\lim_{N \rightarrow \infty} \mathbf{E}\{\hat{R}_N^{MV}\} = R$, et que, pour N suffisamment grand,

$$\sigma_{\hat{R}_N^{MV}}^2 = \frac{R^2}{N} \left(\frac{\sigma_u^2}{U^2} + \frac{\sigma_i^2}{I^2} \right).$$

La variance asymptotique de cet estimateur est la même que pour \hat{R}_N^{LS} , mais il a l'énorme avantage de converger vers la vraie valeur plutôt que vers une valeur biaisée. La figure 3.2 montre le comportement typique des trois estimateurs sur un ensemble de mesures simulées avec les variances sur les bruits indiquées plus haut.

3.6 Estimateurs de Bayes

Hypothèse : $\theta \in \mathbb{R}^N$ est un vecteur aléatoire dont la *densité a priori* $T_\Theta(\theta)$ est supposée connue.

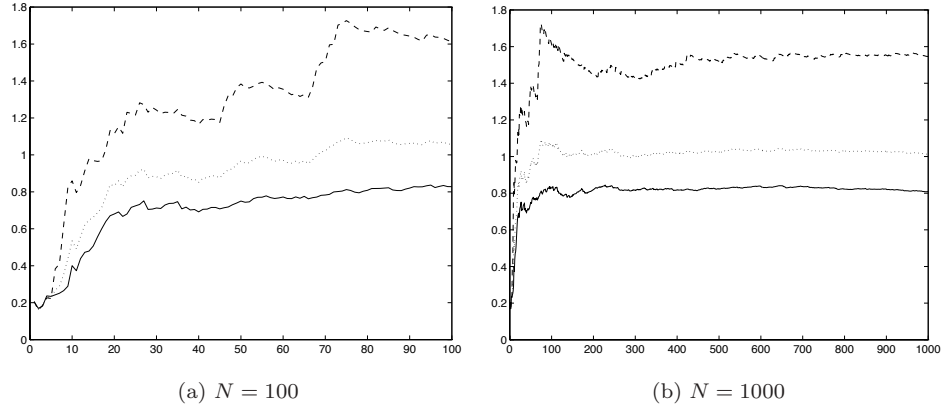


FIG. 3.2 – 3 estimées de la résistance utilisant les 100 premières mesures (à gauche) et 1000 mesures (à droite) : \hat{R}^{LS} (trait plein), \hat{R}_2 (tirets), \hat{R}^{MV} (pointillés)

Les estimateurs de Bayes sont basés sur la fonction de densité conditionnelle $T_{\Theta|Z}(\theta|z)$ et la règle de Bayes :

$$T_{\Theta|Z}(\theta|z) = \frac{T_{Z|\Theta}(z|\theta)T_{\Theta}(\theta)}{T_Z(z)} \quad (3.30)$$

Généralement $T_{\Theta}(\theta)$ est donnée, tandis que $T_{Z|\Theta}(z|\theta)$ se calcule à partir du modèle $Z = h(\theta)$ liant θ aux observations. Ceci permet ensuite de calculer $T_{\Theta|Z}(\theta|z)$ à partir de (3.30).

3.6.1 Définition

L'estimateur de Bayes $\hat{\theta}(Z)$ (parfois appelé estimateur optimal) est obtenu en minimisant un *risque de Bayes* :

$$B = \int_Z \int_{\Theta} C(\hat{\theta}(z) - \theta) T_{Z,\Theta}(z, \theta) d\theta dz \quad (3.31)$$

$$= \mathbf{E}_{Z,\Theta}\{C(\hat{\theta}(Z) - \theta)\}, \quad (3.32)$$

où $C(\hat{\theta}(Z) - \theta)$ est une fonction de coût positive que l'on choisit. On peut réécrire (3.31) sous la forme

$$B = \int_Z T_Z(z) dz \int_{\Theta} C(\hat{\theta}(z) - \theta) T_{\Theta|Z}(\theta|z) d\theta \quad (3.33)$$

$$= \int_Z T_Z(z) B(\hat{\theta}|z) dz \quad (3.34)$$

$$\text{où } B(\hat{\theta}|z) = \int_{\Theta} C(\hat{\theta}(z) - \theta) T_{\Theta|Z}(\theta|z) d\theta \quad (3.35)$$

$$(3.36)$$

Comme $T_Z(z)$ est toujours positif, minimiser le risque de Bayes B revient à minimiser le risque conditionnel $B(\hat{\theta}|z)$ pour tout z , ce qui est plus facile. En effet, $B(\hat{\theta}|z)$ ne dépend que du coût C et de la densité conditionnelle $T_{\Theta|Z}(\theta|z)$.

3.6.2 Estimateur de la moyenne conditionnelle

Si l'on prend un coût quadratique $C(\hat{\theta} - \theta) = (\hat{\theta} - \theta)^T Q (\hat{\theta} - \theta)$, où Q est une matrice symétrique semi-définie positive, on obtient l'estimateur quadratique ou de la moyenne conditionnelle :

$$\hat{\theta}^q(Z) = \mathbf{E}\{\Theta|Z\} = \int_{\Theta} \theta T_{\Theta|Z}(\theta|z) d\theta \quad (3.37)$$

On constate donc que l'espérance de θ , conditionnée sur les observations Z , minimise l'erreur quadratique moyenne. On peut montrer que cet estimateur est :

- sans biais : $\mathbf{E}\{\hat{\theta}^q(Z)\} = \mathbf{E}\{\theta\}$.
- de variance minimale.

Remarquons enfin que l'estimateur $\hat{\theta}^q(Z)$ ne dépend pas de la pondération Q ; par contre le risque optimal, B_{min} , en dépend.

3.6.3 Estimateur du maximum a posteriori (MAP)

Si l'on prend $C(\hat{\theta} - \theta) = 1 - \delta(\hat{\theta} - \theta)$, on trouve $B(\hat{\theta}|Z) = 1 - T_{\Theta|Z}(\hat{\theta}|z)$. Minimiser $B(\hat{\theta}|Z)$ par rapport à $\hat{\theta}$ revient donc à maximiser la densité de probabilité $T_{\Theta|Z}(\hat{\theta}|z)$. L'estimateur qui maximise $T_{\Theta|Z}(\hat{\theta}|z)$ est appelé *l'estimateur du maximum a posteriori* :

$$\hat{\theta}^{MAP}(Z) = \arg \max_{\theta} T_{\Theta|Z}(\theta|z) \quad (3.38)$$

Puisque $T_Z(z)$ ne dépend pas de θ , la règle de Bayes nous dit aussi que

$$\hat{\theta}^{MAP}(Z) = \arg \max_{\theta} T_{Z|\Theta}(z|\theta) T_{\Theta}(\theta) \quad (3.39)$$

On peut encore écrire, puisque la fonction logarithme est monotone, que

$$\hat{\theta}_{MAP}(Z) = \arg \max_{\theta} [\ln T_{Z|\Theta}(z|\theta) + \ln T_{\Theta}(\theta)] \quad (3.40)$$

3.6.4 Commentaires sur les estimateurs de Bayes

1. L'estimateur de Bayes est en principe le plus intéressant car il utilise l'information maximale. Mais celle-ci (en particulier la densité a priori $T_{\Theta}(\theta)$) n'est pas toujours disponible, et il faut alors recourir à des estimateurs

sous-optimaux. Remarquons en particulier que s'il n'y a pas d'information a priori sur θ et que tous les θ sont donc équiprobables (ce qui revient à dire que $T_{\Theta}(\theta)$ est uniforme), alors

$$\arg \max_{\theta} [\ln T_{Z|\Theta}(z|\theta) + \ln T_{\Theta}(\theta)] = \arg \max_{\theta} [\ln T_{Z|\Theta}(z|\theta)] \quad (3.41)$$

L'estimateur $\hat{\theta}$ qui maximise $\ln T_{Z|\Theta}(z|\theta)$ est appelé *l'estimateur du maximum de vraisemblance* (voir section suivante).

2. L'estimateur $\hat{\theta}^q(Z)$ est la moyenne de $T_{\Theta|Z}(\theta|z)$, tandis que $\hat{\theta}^{MAP}(Z)$ est le mode de $T_{\Theta|Z}(\theta|z)$ (càd. la valeur de θ qui maximise cette densité). Pour beaucoup de fonctions de densité, et notamment pour la Gaussienne, la moyenne et le mode coïncident, et ces deux estimateurs sont alors identiques.

3.7 Estimateur du Maximum de Vraisemblance (ML)

Hypothèse : $\theta \in \mathbb{R}^n$ est un vecteur déterministe mais inconnu (modèle de Fisher).

3.7.1 Définition

L'estimateur du maximum de vraisemblance $\hat{\theta}^{ML}(Z)$ est défini comme suit :

$$\hat{\theta}^{ML}(Z) = \arg \max_{\theta} T_{Z|\Theta}(z|\theta) \quad (3.42)$$

La fonction $T_{Z|\Theta}(z|\theta)$ est considérée ici comme une fonction de θ pour un vecteur d'observations $z = (z_1, z_2, \dots, z_n)$ donné. L'estimée $\hat{\theta}^{ML}(z)$ est donc la valeur de θ qui rend le vecteur z observé le plus vraisemblable. On utilise d'ailleurs souvent la notation

$$L(\theta|z) = \ln T_{Z|\Theta}(z|\theta) \quad (3.43)$$

et on appelle $L(\theta|z)$ la *fonction de vraisemblance* que l'on maximise par rapport à θ pour un vecteur z donné.

3.7.2 Propriétés de l'estimateur du max. de vraisemblance

Pour un nombre fini N de données, l'estimateur du max. de vraisemblance a typiquement une variance supérieure à celle d'un estimateur $\hat{\theta}^{MAP}(Z)$. Cependant il a de très bonnes propriétés asymptotiques :

- $\hat{\theta}_N^{ML}$ converge en probabilité vers θ
- $\hat{\theta}_N^{ML}$ est asymptotiquement normal et efficace : sa variance converge vers la borne de Cramér-Rao.

3.8 Estimateur linéaire à variance minimale

Hypothèse : $\theta \in \mathbb{R}^n$ est un vecteur aléatoire. On ne connaît pas la densité de probabilité a priori $T_\Theta(\theta)$ ni $T_{Z|\Theta}(z|\theta)$, mais seulement les moments d'ordre 1 et 2 de $T_{Z|\Theta}(z|\theta)$.

On utilise alors souvent l'estimateur linéaire non-biaisé à variance minimale, noté $\hat{\theta}^{LVM}$. En anglais, on l'appelle Best Linear Unbiased Estimator (BLUE).

3.8.1 Définition

L'estimateur linéaire non-biaisé à variance minimale est un estimateur linéaire en Z :

$$\hat{\theta}^{LVM}(Z) = \theta_0 + MZ,$$

où $\theta_0 \in \mathbb{R}^n$ et $M \in \mathbb{R}^{n \times N}$ sont choisis pour que $\hat{\theta}^{LVM}$ soit non-biaisé et que $\mathbf{E}\{|\hat{\theta}^{LVM} - \theta|^2\}$ soit minimale. On trouve :

$$\hat{\theta}^{LVM}(Z) = m_\theta + C_{\theta Z} C_{ZZ}^{-1} (Z - m_Z) \quad (3.44)$$

3.8.2 Propriétés de l'estimateur linéaire à variance minimale

La principale propriété de $\hat{\theta}^{LVM}$ est qu'il obéit au *principe d'orthogonalité* :

$$\mathbf{E}\{(\hat{\theta}^{LVM} - \theta)Z^T\} = 0 \quad (3.45)$$

L'erreur d'estimation est non-corrélée avec les observations (càd. orthogonale aux mesures). En particulier :

$$\mathbf{E}\{(\hat{\theta}^{LVM} - \theta)(\hat{\theta}^{LVM})^T\} = 0 \quad (3.46)$$

Cette dernière propriété, couplée à $\mathbf{E}\{\hat{\theta}^{LVM}\} = \theta$ (absence de biais), sont souvent utilisées pour calculer $\hat{\theta}^{LVM}$, car elles définissent cet estimateur de manière unique. Si $T_{Z|\Theta}(z|\theta)$ est Gaussienne, cette densité est entièrement déterminée par m_Z, m_θ et la covariance du vecteur $\begin{pmatrix} Z \\ \theta \end{pmatrix}$. Dans ce cas :

$$\hat{\theta}^{LVM} = \hat{\theta}^q = \hat{\theta}^{MAP}.$$

3.9 Estimateur au sens des moindres carrés

Hypothèse : On ne possède aucune information statistique sur θ et Z , mais on sait seulement que Z est une mesure bruitée de $h(\theta)$:

$$Z = h(\theta) + V \quad (3.47)$$

avec $Z \in \mathbb{R}^N, \theta \in \mathbb{R}^n$ et $V \in \mathbb{R}^N$. On peut alors utiliser l'estimateur des moindres carrés pour estimer θ .

3.9.1 Définition

Supposons que la relation entre le vecteur de paramètres θ et le vecteur d'observations soit donné par (3.47). L'estimateur des moindres carrés est alors défini comme :

$$\hat{\theta}^{LS}(Z) = \arg \min_{\theta} [(Z - h(\theta))^T W (Z - h(\theta))] \quad (3.48)$$

où W est une matrice symétrique semi-définie positive.

3.9.2 Propriétés de l'estimateur des moindres carrés

Dans le cas où la relation $h(\theta)$ est linéaire, on obtient une solution explicite pour $\hat{\theta}^{LS}(Z)$. Soit

$$Z = H\theta + V \quad (3.49)$$

avec $H \in \mathbb{R}^{N \times n}$. Alors

$$\hat{\theta}^{LS}(Z) = (H^T W H)^{-1} H^T W Z. \quad (3.50)$$

L'estimateur des moindres carrés est alors *linéaire* en Z . Aucune information statistique n'est nécessaire pour calculer l'estimateur des moindres carrés. Cependant, si on connaît certaines propriétés statistiques du bruit V , on peut en déduire des propriétés statistiques sur $\hat{\theta}^{LS}$.

3.10 Modèle linéaire et Gaussien

Dans cette section nous allons calculer différents estimateurs dans le cas particulier où le modèle $Z = h(\theta)$ est linéaire et Gaussien :

$$Z = H\Theta + V \quad (3.51)$$

avec $Z \in \mathbb{R}^N$, $V \in \mathbb{R}^N$ et $\Theta \in \mathbb{R}^n$, V et θ sont tous deux Gaussiens, indépendants l'un de l'autre, et :

$$\mathbf{E}\{\theta\} = m_{\theta}, \quad \text{Cov}\{\theta\} = Q, \quad \mathbf{E}\{V\} = 0, \quad \text{Cov}\{V\} = R. \quad (3.52)$$

On en déduit que Z est aussi Gaussien et que

$$\mathbf{E}\{Z|\theta\} = H\theta \quad \text{et} \quad \mathbf{E}\{(Z - \mathbf{E}\{Z|\theta\})(Z - \mathbf{E}\{Z|\theta\})^T | \theta\} = R \quad (3.53)$$

3.10.1 Estimateur Bayésien

Calculons d'abord l'estimateur Bayésien $\hat{\theta}^{MAP}$.

On a :

$$T_{Z|\Theta}(z|\theta) = \frac{1}{[(2\pi)^N \det R]^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(z - H\theta)^T R^{-1}(z - H\theta)\right\}$$

$$T_{\Theta}(\theta) = \frac{1}{[(2\pi)^n \det Q]^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\theta - m_{\theta})^T Q^{-1}(\theta - m_{\theta})\right\}$$

Par conséquent

$$\hat{\theta}^{MAP} = \arg \min_{\theta} [\ln T_{Z|\Theta}(z|\theta) + \ln T_{\Theta}(\theta)] \quad (3.54)$$

$$= \arg \min_{\theta} \left\{ \frac{1}{2} (z - H\theta)^T R^{-1} (z - H\theta) + \frac{1}{2} (\theta - m_{\theta})^T Q^{-1} (\theta - m_{\theta}) \right\} \quad (3.55)$$

$$\triangleq \arg \min_{\theta} J(\theta) \quad (3.56)$$

En prenant $\frac{\partial J}{\partial \theta}(\theta) = 0$ on trouve

$$\begin{aligned} \frac{\partial J}{\partial \theta}(\theta) &= 0 = Q^{-1}(\hat{\theta}^{MAP} - m_{\theta}) - H^T R^{-1}(Z - H\hat{\theta}^{MAP}) \\ \implies \hat{\theta}^{MAP} &= (I + Q H^T R^{-1} H)^{-1}(m_{\theta} + Q H^T R^{-1} Z) \quad (3.57) \\ &= (Q^{-1} + H^T R^{-1} H)^{-1}(Q^{-1}m_{\theta} + H^T R^{-1} Z) \quad (3.58) \end{aligned}$$

On voit donc que, dans ce cas de modèle linéaire avec distributions Gaussiennes, $\hat{\theta}^{MAP}$ est un estimateur linéaire en Z . L'estimateur $\hat{\theta}^{MAP}$ coïncide dans ce cas avec l'estimateur quadratique $\hat{\theta}^q$ et avec l'estimateur $\hat{\theta}^{LVM}$. Il est une combinaison linéaire de la moyenne a priori m_{θ} et du vecteur d'observations Z . Le rôle respectif de l'information a priori m_{θ} et des mesures Z apparaît très clairement si on prend le cas où les observations $Z = (Z_1, Z_2, \dots, Z_N)^T$ est constitué d'un ensemble de N observations indépendantes d'un paramètre scalaire θ . Dans ce cas $H = (1 \ 1 \ \dots \ 1)^T$, $Q = m_{\theta}$ et $R = \sigma_V^2 I$. L'expression (3.58) de $\hat{\theta}^{MAP}$ devient alors :

$$\hat{\theta}^{MAP} = \hat{\theta}^{LVM} = \frac{m_{\theta} + \frac{\sigma_{\theta}^2}{\sigma_V^2} \sum_{k=1}^N Z_k}{1 + \frac{N\sigma_{\theta}^2}{\sigma_V^2}} = \frac{\frac{m_{\theta}}{\sigma_{\theta}^2} + \frac{1}{\sigma_V^2} \sum_{k=1}^N Z_k}{\frac{1}{\sigma_{\theta}^2} + \frac{N}{\sigma_V^2}} \quad (3.59)$$

On voit ici clairement que $\hat{\theta}^{MAP}$ est combinaison linéaire de m_{θ} et des mesures Z_k . Le poids respectif donné à la moyenne a priori m_{θ} et aux mesures Z_k est inversement proportionnel à la confiance que l'on a dans ces informations, c.à.d. à leur variance (respectivement σ_{θ}^2 et σ_V^2).

Propriétés de l'estimateur $\hat{\theta}^{MAP}$

En substituant $Z = H\theta + V$ dans (3.58) et en prenant l'espérance mathématique, on voit tout de suite que $\mathbf{E}\{\hat{\theta}^{MAP}\} = m_{\theta}$. L'estimateur est donc sans biais.

Des expressions (3.51) et (3.58) on déduit que :

$$\hat{\theta}^{MAP} - \theta = (I + QH^T R^{-1} H)^{-1}(m_{\theta} - \theta + QH^T R^{-1} V).$$

Dès lors

$$Cov(\hat{\theta}^{MAP} - \theta) \triangleq \mathbf{E}\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\} \quad (3.60)$$

$$= (I + QH^T R^{-1} H)^{-1}(Q + QH^T R^{-1} H Q)(I + QH^T R^{-1} H)^{-1} \quad (3.61)$$

$$= (Q^{-1} + H^T R^{-1} H)^{-1} \quad (3.62)$$

On voit donc que $Cov(\hat{\theta}^{MAP} - \theta) \leq Q$: la covariance de l'erreur d'estimation a posteriori est inférieure à la covariance a priori. Les observations ont donc permis de réduire l'incertitude sur θ .

Dans le cas particulier des N observations indépendantes d'un paramètre scalaire on a :

$$Var\{\hat{\theta}^{MAP} - \theta\} = \frac{\sigma_V^2}{N + \frac{\sigma_V^2}{\sigma_\theta^2}}$$

La variance tend vers zéro lorsque le nombre d'observations tend vers l'infini.

3.10.2 Estimateur du maximum de vraisemblance

Supposons maintenant que la variance a priori de θ tende vers l'infini, c.à.d. $Q^{-1} \rightarrow 0$. Ceci revient à dire qu'il n'y a plus d'information a priori. L'expression (3.58) devient

$$\hat{\theta}^{ML} = (H^T R^{-1} H)^{-1} H^T R^{-1} Z. \quad (3.63)$$

Ceci est aussi l'estimateur que l'on obtient par la méthode du maximum de vraisemblance, en considérant θ comme déterministe et en maximisant $T_{Z|\Theta}(z|\theta)$: voir (3.54)-(3.55). C'est également l'estimateur linéaire à variance minimale.

En substituant (3.51) dans (3.63) on voit que $\mathbf{E}\{\hat{\theta}^{ML}\} = \theta$: l'estimateur est non-biaisé.

Enfin

$$\begin{aligned} \mathbf{E}\{(\hat{\theta}^{ML} - \theta)(\hat{\theta}^{ML} - \theta)^T\} &= (H^T R^{-1} H)^{-1} \\ &\geq (H^T R^{-1} H + Q^{-1})^{-1} = \mathbf{E}\{(\hat{\theta}^{MAP} - \theta)(\hat{\theta}^{MAP} - \theta)^T\}. \end{aligned} \quad (3.64)$$

La covariance de l'estimateur du maximum de vraisemblance est supérieure à celle de l'estimateur Bayésien.

3.10.3 Estimateur des moindres carrés

L'estimateur des moindres carrés pour le modèle linéaire (3.51) est donné par (3.50) pour le coût quadratique (3.48). Cet estimateur est calculé sans aucune hypothèse statistique. Cependant, si V est un bruit de moyenne m_V et de covariance R , on a les propriétés suivantes :

- 1) $\mathbf{E}\{\hat{\theta}^{LS}\} = \theta$ si et seulement si $m_V = 0$.
- 2) Supposons que $m_V = 0$. Alors

$$Cov(\hat{\theta}^{LS}) = (H^T W H)^{-1} H^T W R W H (H^T W H)^{-1} \quad (3.65)$$

$$= (H^T R^{-1} H)^{-1} \text{ si } W = R^{-1}. \quad (3.66)$$

Dans tous les cas, la covariance de l'erreur d'estimation $\hat{\theta}^{LS} - \theta$ est plus grande que celle de l'erreur $\hat{\theta}^{MAP} - \theta$ (voir (3.62)) :

$$\text{Cov}(\hat{\theta}^{LS} - \theta) \geq \text{Cov}(\hat{\theta}^{MAP} - \theta)$$

avec égalité si la covariance a priori est infinie ($Q^{-1} = 0$) et si $W = R^{-1}$.

