Estimation en traitement du signal

Olivier Besson

Institut Supérieur de l'Aéronautique et de l'Espace
SUPAERO

Sommaire

Problématique

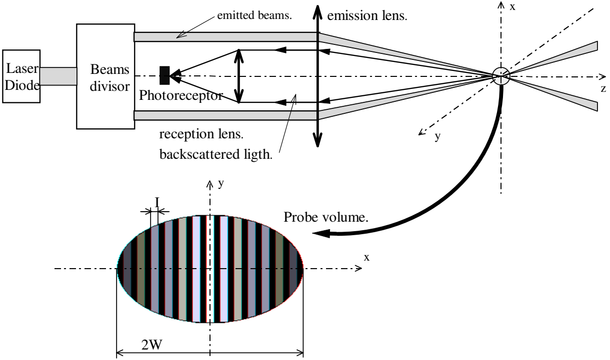
Objectif

Mettre en œuvre un système permettant d'accéder à une information pertinente, par exemple vitesse et/ou position d'un objet en mouvement.

Questions

- Quel est le problème ; qu'est-ce que je cherche ?
- Comment choisir le capteur pour résoudre le problème ?
- Quel est le principe de fonctionnement du système ?
- De quelle façon le capteur donne t-il les informations recherchées ?
- Comment vais-je utiliser les données qu'il me fournit ?

Exemple : anémométrie laser



Le signal reçu s'écrit

$$x(t) = A \exp\left\{-2\alpha^2 f_d^2 t^2\right\} \cos(2\pi f_d t) + b(t).$$

$\Rightarrow$ estimation de $v = I f_d$ à partir de $x(t)$.

Exemple : effet Doppler

On considère une antenne émettant un signal $s(t) = e^{i\omega_c t}$ et rétro-diffusé par une cible en mouvement uniforme à la vitesse radiale v. Le signal reçu s'écrit

$$r(t) = As\left(t - 2\tau(t)\right) = As\left(t - 2\frac{d_0 - vt}{c}\right) = Ae^{i\omega_c t}e^{-i\omega_c\frac{2d_0}{c}}e^{i\frac{2\omega_c v}{c}t}$$

Après démodulation, on obtient

$$x(t) = Ae^{i\phi}e^{i2\pi\frac{2v}{\lambda}t} + b(t)$$

et donc la vitesse de la cible se déduit de la fréquence du signal utile.

Modèle-Mesures-Estimation

Signal
$\{x(n)\}_{n=1}^{N}$

Estimateur

$\hat{\boldsymbol{\theta}} = \boldsymbol{g}\left(\{x(n)\}_{n=1}^{N}\right)$

$\hat{\boldsymbol{\theta}}$ est t-il un **bon** estimateur?

Comment trouver $\hat{\boldsymbol{\theta}}$?

Modèle $\boldsymbol{\theta}$

dépend de ce que l'on sait sur le modèle

Contexte et objectif

Ce que l'on sait (croit savoir)

- Un capteur reçoit des mesures $x(1), \cdots, x(N)$.
- Le signal contient une information via un vecteur paramètre $\boldsymbol{\theta}$, par exemple $x(n) = s(n; \boldsymbol{\theta}) + b(n)$ et on connaît (ou pas) la loi de $\boldsymbol{b} = \begin{bmatrix} b(1) & \cdots & b(N) \end{bmatrix}^T$.

Ce que l'on veut (voudrait) savoir

- **Comment** trouver un estimateur $\hat{\boldsymbol{\theta}} = g(x(1), \cdots, x(N))$ de $\boldsymbol{\theta}$?
- Est-ce que $\hat{\boldsymbol{\theta}}$ est un bon estimateur ?

Contraintes

- métrologie, performance, coût calculatoire, robustesse, ...

Approches pour l'estimation

Approche déterministe

On suppose que $\boldsymbol{\theta}$ est une quantité déterministe inconnue et on considère alors la loi $p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$ du vecteur $\boldsymbol{X} = \begin{bmatrix} X(1) & X(2) & \cdots & X(N) \end{bmatrix}^T$ qui est paramétrée par $\boldsymbol{\theta}$.

Approche Bayésienne

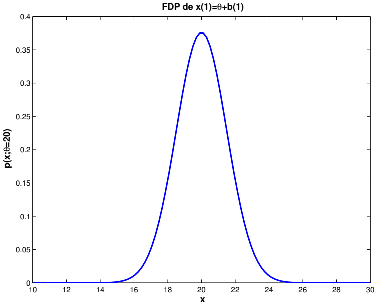
On suppose que $\boldsymbol{\theta}$ est une réalisation d'un vecteur aléatoire $\boldsymbol{\Theta}$ et on considère alors la loi jointe $p_{\boldsymbol{X},\boldsymbol{\Theta}}(\boldsymbol{x},\boldsymbol{\theta}) = p_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}|\boldsymbol{\theta})\, \pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, où $\pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ désigne la loi a priori.

Sommaire

Densité de probabilité et fonction de vraisemblance

- La loi $p(\boldsymbol{x};\boldsymbol{\theta})$ donne toute l'information sur le signal $\boldsymbol{x}$. Par exemple, si $x(n) = \theta + b(n)$ où $b(n)$ sont i.i.d. $b(n) \sim \mathcal{N}\left(0, \sigma^2\right)$,

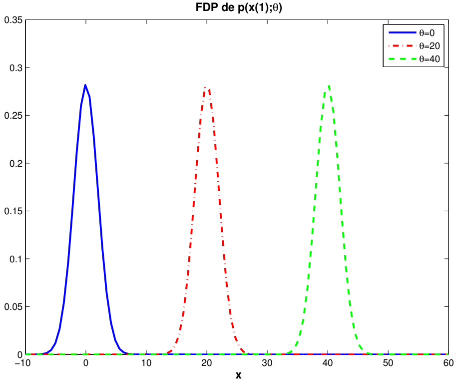
$$p\left(\boldsymbol{x};\theta\right) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(x(n)-\theta\right)^2\right\}.$$



- $p(\boldsymbol{x};\boldsymbol{\theta})$, vue comme une fonction de $\boldsymbol{\theta}$, est dénommée vraisemblance.

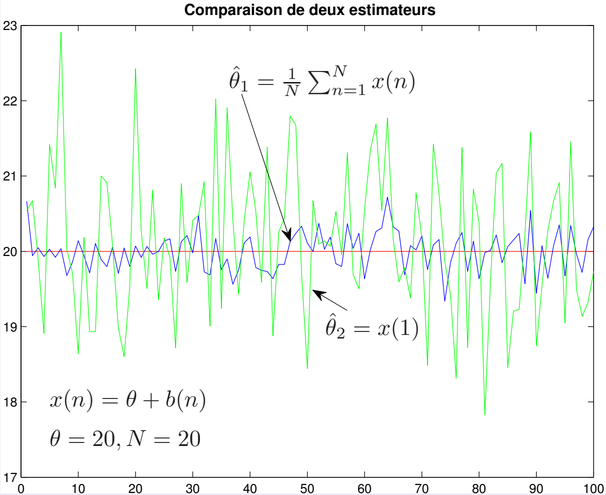
Vraisemblance et estimation

- Toute l'information sur $\boldsymbol{x}$ est contenue dans $p(\boldsymbol{x};\boldsymbol{\theta})$.
- $p(\boldsymbol{x};\boldsymbol{\theta})$ dépend de $\boldsymbol{\theta} \Rightarrow$ on peut inférer $\boldsymbol{\theta}$ à partir de $\boldsymbol{x}$.



- $\hat{\boldsymbol{\theta}}$ est une fonction de $\boldsymbol{x}$: $\boldsymbol{x}$ étant aléatoire, $\hat{\boldsymbol{\theta}}$ est *aléatoire*.

Critères de qualité d'un estimateur ?



Caractérisation d'un estimateur

Biais

$$\mathrm{b}(\hat{\boldsymbol{\theta}}) = \mathrm{E}\{\hat{\boldsymbol{\theta}}\} - \boldsymbol{\theta}$$

Matrice de covariance

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) = \mathrm{E}\{(\hat{\boldsymbol{\theta}} - \mathrm{E}\{\hat{\boldsymbol{\theta}}\})(\hat{\boldsymbol{\theta}} - \mathrm{E}\{\hat{\boldsymbol{\theta}}\})^T\}$$

Erreur quadratique moyenne

$$\mathrm{eqm}(\hat{\boldsymbol{\theta}}) = \mathrm{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\} = \mathrm{b}(\hat{\boldsymbol{\theta}})\mathrm{b}(\hat{\boldsymbol{\theta}})^T + \mathrm{cov}(\hat{\boldsymbol{\theta}})$$

Loi

La loi $p(\hat{\boldsymbol{\theta}})$ constitue la caractérisation complète du vecteur aléatoire $\hat{\boldsymbol{\theta}}$.

Exemple de caractérisation d'un estimateur

Exemple $x(n) = \theta + b(n)$ avec $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$

On suppose que les variables $b(n)$ sont i.i.d. $b(n) \sim \mathcal{N}\left(0, \sigma^2\right)$, i.e. $\boldsymbol{b} = \begin{bmatrix} b(1) & \cdots & b(N) \end{bmatrix}^T \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$. Soit $\hat{\theta} = \frac{1}{N}\sum_{n=1}^{N} x(n)$. On a

$$\mathrm{E}\{\hat{\theta}\} = \mathrm{E}\{\frac{1}{N}\sum_{n=1}^{N} x(n)\} = \frac{1}{N}\sum_{n=1}^{N} \mathrm{E}\{x(n)\} = \theta$$

$$\begin{aligned}
\mathrm{cov}(\hat{\theta}) &= \frac{1}{N^2}\mathrm{E}\{\sum_{n,m=1}^{N} \left(x(n) - \theta\right)\left(x(m) - \theta\right)\} \\
&= \frac{1}{N^2}\mathrm{E}\{\sum_{n,m=1}^{N} b(n)b(m)\} = \frac{1}{N^2}\mathrm{E}\{\sum_{n=1}^{N} b(n)^2\} = \frac{\sigma^2}{N}.
\end{aligned}$$

Dans ce cas, on a la loi de l'estimateur : $\hat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2/N\right)$.

Convergence d'un estimateur

Estimateur consistant

Soit $\hat{\boldsymbol{\theta}}_N$ un estimateur de $\boldsymbol{\theta}$ à partir de N échantillons $x(1), \cdots, x(N)$.
Alors $\hat{\boldsymbol{\theta}}_N$ est un estimateur consistant (au sens large) si

$$\lim_{N \to \infty} \Pr \left\{ \left\| \hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta} \right\| < \delta \right\} = 1 \qquad \forall \delta > 0 \quad \forall \boldsymbol{\theta}.$$

Il est dit consistant en moyenne quadratique si

$$\lim_{N \to \infty} \mathrm{E}\{(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})^T\} = \mathbf{0} \qquad \forall \boldsymbol{\theta}.$$

“Meilleur” estimateur

Estimation optimale ?

- quel critère dois-je adopter pour mesurer la performance d’un estimateur ?
- en supposant que ce critère existe, y a t-il des estimateurs qui produisent la performance optimale ?
- si oui, existe t-il une démarche systématique (et laquelle) pour calculer cet estimateur optimal ?

Approches

1. soit on considère que l’estimateur doit être non biaisé et on cherche donc l’estimateur à minimum de variance ;
2. soit on autorise un biais et on cherche alors à minimiser l’erreur quadratique moyenne.

Minimisation de l'erreur quadratique moyenne : exemple

Minimisation de l'eqm pour $x(n) = \theta + b(n)$

- On autorise un biais dont on espère qu'il sera contre-balancé par une diminution de la variance.
- Par exemple, on cherche un estimateur du type $\hat{\theta} = a\frac{1}{N}\sum_{n=1}^{N} x(n)$ dont l'eqm $\mathrm{eqm}(\hat{\theta}) = (a-1)^2\,\theta^2 + \frac{a^2\sigma^2}{N}$ est la plus faible possible.
- La valeur optimale de a est alors $a_{\text{opt}} = \left[\theta^2 + N^{-1}\sigma^2\right]^{-1}\theta^2$ *qui dépend de θ inconnu*. On a, avec ce choix de a

$$\mathrm{b}(\hat{\theta}) = -\left[\theta^2 + N^{-1}\sigma^2\right]^{-1}\theta N^{-1}\sigma^2 \neq 0$$

$$\mathrm{cov}(\hat{\theta}) = \left[\theta^2 + N^{-1}\sigma^2\right]^{-2}\theta^4 N^{-1}\sigma^2 < N^{-1}\sigma^2 = \mathrm{cov}(\frac{1}{N}\sum_{n=1}^{N} x(n))$$

$$\mathrm{eqm}(\hat{\theta}) = \left[\theta^2 + N^{-1}\sigma^2\right]^{-1}\theta^2 N^{-1}\sigma^2 < N^{-1}\sigma^2 = \mathrm{eqm}(\frac{1}{N}\sum_{n=1}^{N} x(n)).$$

Estimation non biaisée à variance minimale

Estimateur MVU

On considère comme "**référence**" l'estimateur non biaisé, à minimum de variance :

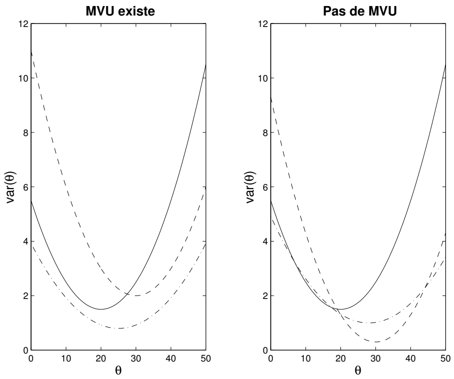
$$\mathrm{E}\{\hat{\boldsymbol{\theta}}^{MVU}\} = \boldsymbol{\theta}$$
$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} - \boldsymbol{C}^{MVU} \geq \mathbf{0} \qquad \forall \hat{\boldsymbol{\theta}}, \forall \boldsymbol{\theta}$$

où $\boldsymbol{C}^{MVU} = \mathrm{cov}(\hat{\boldsymbol{\theta}}^{MVU})$ et $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} = \mathrm{cov}(\hat{\boldsymbol{\theta}})$.
Si cet estimateur existe (et si on peut l'implémenter), il est considéré comme l'estimateur optimal.

Estimation non biaisée à variance minimale

Questions

- Cet estimateur existe-t-il ? Si oui comment le trouver ?
- Existe-t-il une borne inférieure à $\mathrm{cov}(\hat{\boldsymbol{\theta}})$?



Borne minimale de la variance d'un estimateur non biaisé

Bornes de Cramér-Rao

La borne de Cramér-Rao (BCR) donne l'expression de la covariance minimale que l'on peut atteindre en estimant (de manière non biaisée) le paramètre $\boldsymbol{\theta}$:

- si un estimateur atteint les bornes, et ce quel que soit $\boldsymbol{\theta}$, alors c'est l'estimateur MVU ;
- cette borne est une référence à laquelle on peut comparer tous les estimateurs ;
- la théorie permet de déterminer s'il existe un estimateur qui atteindra la borne.

Borne minimale de la variance d'un estimateur non biaisé

Théorème (Borne de Cramér-Rao)

Si $p\left(\boldsymbol{x};\boldsymbol{\theta}\right)$ vérifie la condition de régularité $\mathrm{E}\{\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\} = \mathbf{0}$ alors, quel que soit l'estimateur non biaisé $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$, sa matrice de covariance vérifie $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} - \boldsymbol{I}^{-1}\left(\boldsymbol{\theta}\right) \geq \mathbf{0}$ où

$$\boldsymbol{I}\left(\boldsymbol{\theta}\right) = -\mathrm{E}\{\frac{\partial^2 \ln p\left(\boldsymbol{x};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\} = \mathrm{E}\{\frac{\partial \ln p\left(\boldsymbol{x};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} \frac{\partial \ln p\left(\boldsymbol{x};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^T}\}$$

est la **matrice d'information de Fisher**. De plus, on peut trouver un estimateur qui atteint la borne i.e. $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{I}^{-1}\left(\boldsymbol{\theta}\right)$ **si et seulement si**

$$\frac{\partial \ln p\left(\boldsymbol{x};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} = \boldsymbol{I}\left(\boldsymbol{\theta}\right)\left(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{\theta}\right).$$

Cet estimateur, **qui est l'estimateur MVU**, est donné par $\hat{\boldsymbol{\theta}} = \boldsymbol{f}(\boldsymbol{x})$ et sa matrice de covariance est alors $\boldsymbol{I}^{-1}\left(\boldsymbol{\theta}\right)$: il est dit **efficace**.

Démonstration

$\hat{\boldsymbol{\theta}}$ étant un estimateur non biaisé, on a $\int \hat{\boldsymbol{\theta}} p(\boldsymbol{x};\boldsymbol{\theta}) dx = \boldsymbol{\theta}$. En différentiant, on obtient

$$\int \hat{\boldsymbol{\theta}} \frac{\partial p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} d\boldsymbol{x} = \int \hat{\boldsymbol{\theta}} \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} p(\boldsymbol{x};\boldsymbol{\theta}) d\boldsymbol{x} = \boldsymbol{I}$$

et ainsi

$$\mathrm{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\} = \boldsymbol{I}.$$

Par conséquent, la matrice de covariance de $\begin{bmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T & \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \end{bmatrix}^T$ s'écrit

$$\mathrm{cov}(\begin{bmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ \partial \ln p(\boldsymbol{x};\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \end{bmatrix}) = \begin{bmatrix} \boldsymbol{C}_{\hat{\boldsymbol{\theta}}} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{I}(\boldsymbol{\theta}) \end{bmatrix}.$$

Borne de Cramér-Rao

Démonstration

Cette matrice étant définie positive, on a nécessairement

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} - \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \geq \mathbf{0}.$$

D'autre part

$$\begin{aligned}
&\int \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p(\boldsymbol{x};\boldsymbol{\theta}) d\boldsymbol{x} = \mathbf{0} \\
&\Rightarrow \int \frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} p(\boldsymbol{x};\boldsymbol{\theta}) d\boldsymbol{x} + \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} p(\boldsymbol{x};\boldsymbol{\theta}) d\boldsymbol{x} = \mathbf{0} \\
&\Rightarrow \boldsymbol{I}(\boldsymbol{\theta}) = \mathrm{E}\{\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\} = -\mathrm{E}\{\frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\}.
\end{aligned}$$

Démonstration

Si de plus

$$\frac{\partial \ln p\,(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{I}\,(\boldsymbol{\theta})\,(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{\theta})$$

alors, pour $\hat{\boldsymbol{\theta}} = \boldsymbol{f}(\boldsymbol{x})$, on a

$$\begin{aligned}
\mathrm{E}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\} &= \int (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{\theta})\, p\,(\boldsymbol{x};\boldsymbol{\theta})\, d\boldsymbol{x} \\
&= \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \int \frac{\partial \ln p\,(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p(\boldsymbol{x};\boldsymbol{\theta}) d\boldsymbol{x} = \mathbf{0} \\
\mathrm{cov}(\hat{\boldsymbol{\theta}}) &= \mathrm{E}\{\boldsymbol{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \ln p\,(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p\,(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \boldsymbol{I}^{-1}(\boldsymbol{\theta})\} \\
&= \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \mathrm{E}\{\frac{\partial \ln p\,(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p\,(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\} \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \\
&= \boldsymbol{I}^{-1}(\boldsymbol{\theta}).
\end{aligned}$$

BCR : un exemple

Exemple $x(n) = \theta + b(n)$, $\boldsymbol{b} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}\right)$

On a dans ce cas

$$p\left(\boldsymbol{x};\theta\right) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left\{\frac{-1}{2\sigma^2}\sum_{n=1}^{N}\left(x(n)-\theta\right)^2\right\}$$

$$\ln p\left(\boldsymbol{x};\theta\right) = -\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(x(n)-\theta\right)^2 .$$

Il s'ensuit que

$$\frac{\partial \ln p\left(\boldsymbol{x};\theta\right)}{\partial\theta} = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left(x(n)-\theta\right)$$

$$\frac{\partial^2 \ln p\left(\boldsymbol{x};\theta\right)}{\partial\theta^2} = -\frac{N}{\sigma^2} = -I(\theta).$$

BCR : un exemple

Exemple $x(n) = \theta + b(n)$, $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$

La BCR est donc simplement

$$\boxed{BCR(\theta) = \frac{\sigma^2}{N}.}$$

De plus,

$$\frac{\partial \ln p\left(\boldsymbol{x};\theta\right)}{\partial \theta} = \frac{N}{\sigma^2}\left[\frac{1}{N}\sum_{n=1}^{N} x(n) - \theta\right].$$

Ainsi il existe un estimateur efficace (donc forcément MVU), à savoir

$$\boxed{\hat{\theta}^{MVU} = \frac{1}{N}\sum_{n=1}^{N} x(n).}$$

BCR : cas signal plus bruit

Exemple $x(n) = s(n;\boldsymbol{\theta}) + b(n)$, $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$

La loi s'écrit maintenant

$$\ln p\left(\boldsymbol{x};\boldsymbol{\theta}\right) = -\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}\left[x(n) - s(n;\boldsymbol{\theta})\right]^2 .$$

Les dérivées deviennent alors

$$\frac{\partial \ln p\left(\boldsymbol{x};\boldsymbol{\theta}\right)}{\partial \theta_k} = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left[x(n) - s(n;\boldsymbol{\theta})\right]\frac{\partial s(n;\boldsymbol{\theta})}{\partial \theta_k} = \frac{1}{\sigma^2}\left[\frac{\partial \boldsymbol{s}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\left[\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta})\right]\right]_k .$$

La matrice d'information de Fisher $\boldsymbol{I}(\boldsymbol{\theta})$ s'écrit donc simplement

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\theta}) &= \frac{1}{\sigma^4}\frac{\partial \boldsymbol{s}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathrm{E}\{\left[\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta})\right]\left[\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta})\right]^T\}\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \\ &= \frac{1}{\sigma^2}\frac{\partial \boldsymbol{s}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}.\end{aligned}$$

BCR : cas signal plus bruit

Exemple $x(n) = s(n;\boldsymbol{\theta}) + b(n)$, $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$

- Considérons maintenant **l'existence** d'un estimateur efficace. On doit avoir la factorisation

$$\begin{aligned}\frac{\partial \ln p\left(\boldsymbol{x};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} &= \frac{1}{\sigma^2}\frac{\partial \boldsymbol{s}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\left[\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta})\right] \\ &= \boldsymbol{I}\left(\boldsymbol{\theta}\right)\left[\boldsymbol{f}\left(\boldsymbol{x}\right) - \boldsymbol{\theta}\right] = \frac{1}{\sigma^2}\frac{\partial \boldsymbol{s}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\left[\boldsymbol{f}\left(\boldsymbol{x}\right) - \boldsymbol{\theta}\right].\end{aligned}$$

- Cette factorisation n'est possible que si $\boldsymbol{s}(\boldsymbol{\theta})$ est une fonction **affine** de $\boldsymbol{\theta}$, soit $\boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{c}$, auquel cas $\boldsymbol{I}(\boldsymbol{\theta}) = \sigma^{-2}\boldsymbol{H}^T\boldsymbol{H}$ et

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T(\boldsymbol{x} - \boldsymbol{c})$$

est l'estimateur efficace.

BCR : cas signal plus bruit

Synthèse $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{s}(\boldsymbol{\theta}), \sigma^2 \boldsymbol{I}\right)$ avec σ^2 connu

- si $\boldsymbol{s}(\boldsymbol{\theta})$ est une fonction non linéaire de $\boldsymbol{\theta}$, il n'existe pas d'estimateur efficace ;
- si $\boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{H}\boldsymbol{\theta}$, il existe un estimateur efficace $\hat{\boldsymbol{\theta}} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{x}$.

Cas σ^2 inconnu

Alors, la matrice de Fisher s'écrit

$$\boldsymbol{I}\left(\boldsymbol{\theta}, \sigma^2\right) = \frac{1}{\sigma^2}\begin{bmatrix} \frac{\partial \boldsymbol{s}(\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} & \mathbf{0} \\ \mathbf{0} & \frac{N}{2\sigma^2} \end{bmatrix}.$$

Il n'y a pas d'estimateur efficace du vecteur $\begin{bmatrix}\boldsymbol{\theta} & \sigma^2\end{bmatrix}^T$. Par contre, si $\boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{H}\boldsymbol{\theta}$,$\hat{\boldsymbol{\theta}} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{x}$ reste efficace.

Modèle linéaire $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{H}\boldsymbol{\theta}, \boldsymbol{C}\right)$, $\boldsymbol{C}$ connue

- Le signal utile s'écrit comme une combinaison linéaire, à coefficients inconnus, de signaux connus :

$$s(n;\boldsymbol{\theta}) = \sum_{k=1}^{p} \theta_k h_k(n) \Leftrightarrow \boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{H}\boldsymbol{\theta}$$

  et les colonnes de $\boldsymbol{H}$ contiennent les signaux de la base sur laquelle on décompose $s(n;\boldsymbol{\theta})$.

- Le bruit additif est Gaussien, coloré, de matrice de covariance $\boldsymbol{C}$ connue.
- La vraisemblance s'écrit

$$p(\boldsymbol{x};\boldsymbol{\theta}) = (2\pi)^{-N/2} \left|\boldsymbol{C}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\right)^T \boldsymbol{C}^{-1} \left(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\right)\right\}$$

$$\ln p(\boldsymbol{x};\boldsymbol{\theta}) = \text{const.} - \frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\right)^T \boldsymbol{C}^{-1} \left(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\right).$$

Modèle linéaire $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{H}\boldsymbol{\theta}, \boldsymbol{C}\right)$, $\boldsymbol{C}$ connue

- Les dérivées de la log-vraisemblance s'écrivent

$$\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{H}^T\boldsymbol{C}^{-1}\left(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\right)$$
$$\frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} = -\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H} = -\boldsymbol{I}(\boldsymbol{\theta})$$

- On a donc la factorisation

$$\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{I}(\boldsymbol{\theta})\left[\left(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{x} - \boldsymbol{\theta}\right]$$

- L'estimateur MVU efficace s'en déduit comme

$$\hat{\boldsymbol{\theta}}^{MVU} = \left(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{x}$$

et cet estimateur est une fonction **linéaire** de $\boldsymbol{x}$.

Modèle linéaire $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{H\theta}, \boldsymbol{C})$, $\boldsymbol{C}$ connue

Remarques

- L'estimateur MVU est solution du problème des moindres carrés pondérés suivant :

$$\hat{\boldsymbol{\theta}}^{MVU} = \arg\min_{\boldsymbol{\theta}} (\boldsymbol{x} - \boldsymbol{H\theta})^T \boldsymbol{C}^{-1} (\boldsymbol{x} - \boldsymbol{H\theta}) .$$

- Dans le cas où $\boldsymbol{x} = \boldsymbol{H\theta} + \boldsymbol{b}$ où $\boldsymbol{b}$ est de moyenne nulle, de matrice de covariance $\boldsymbol{C}$ (mais $\boldsymbol{b}$ n'est pas nécessairement Gaussien),

$$\hat{\boldsymbol{\theta}}^{BLUE} = \left(\boldsymbol{H}^T \boldsymbol{C}^{-1} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^T \boldsymbol{C}^{-1} \boldsymbol{x}$$

est l'estimateur **BLUE** (best linear unbiased estimate), i.e.

$$\forall \hat{\boldsymbol{\theta}} = \boldsymbol{F}\boldsymbol{x} \text{ tel que } \mathrm{E}\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta}, \quad \mathrm{cov}(\hat{\boldsymbol{\theta}}) - \mathrm{cov}(\hat{\boldsymbol{\theta}}^{BLUE}) \geq \mathbf{0}.$$

BCR pour $f(\boldsymbol{\theta})$

Transformation de paramètres

- Si l'on veut estimer $\boldsymbol{\beta} = f(\boldsymbol{\theta})$ alors

$$\boxed{\boldsymbol{I}^{-1}(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}^T} \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\beta}^T}{\partial \boldsymbol{\theta}}.}$$

- Si $\hat{\boldsymbol{\theta}}$ est un estimateur efficace et si $f(.)$ est linéaire, alors $\hat{\boldsymbol{\beta}} = f(\hat{\boldsymbol{\theta}})$ est efficace. Si $f(.)$ n'est pas linéaire, alors $\hat{\boldsymbol{\beta}} = f(\hat{\boldsymbol{\theta}})$ n'est plus efficace. Par contre, en général on a l'efficacité asympotique car

$$\hat{\boldsymbol{\beta}} = f\left(\hat{\boldsymbol{\theta}}\right) \simeq f(\boldsymbol{\theta}) + \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}^T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

et donc

$$\boldsymbol{C}_{\hat{\boldsymbol{\beta}}} \simeq \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}^T} \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\beta}^T}{\partial \boldsymbol{\theta}} = \boldsymbol{I}^{-1}(\boldsymbol{\beta}).$$

BCR pour signaux Gaussiens (formule de Slepian-Bangs)

Cas de signaux Gaussiens réels $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{C}(\boldsymbol{\theta})\right)$

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{k,\ell} = \frac{\partial \boldsymbol{\mu}^T(\boldsymbol{\theta})}{\partial \theta_k} \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_\ell} + \frac{1}{2} \mathrm{Tr}\left\{ \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{C}(\boldsymbol{\theta})}{\partial \theta_k} \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{C}(\boldsymbol{\theta})}{\partial \theta_\ell} \right\}.$$

Cas de signaux Gaussiens complexes $\boldsymbol{x} \sim \mathcal{CN}\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{C}(\boldsymbol{\theta})\right)$

Pour des signaux complexes circulaires dont la loi est

$$p(\boldsymbol{x}; \boldsymbol{\theta}) = \pi^{-N} \left|\boldsymbol{C}(\boldsymbol{\theta})\right|^{-1} \exp\left\{ -\left(\boldsymbol{x} - \boldsymbol{\mu}(\boldsymbol{\theta})\right)^H \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \left(\boldsymbol{x} - \boldsymbol{\mu}(\boldsymbol{\theta})\right) \right\}$$

la matrice d'information de Fisher devient

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{k,\ell} = 2\mathrm{Re}\left[\frac{\partial \boldsymbol{\mu}^H(\boldsymbol{\theta})}{\partial \theta_k} \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_\ell} \right] + \mathrm{Tr}\left\{ \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{C}(\boldsymbol{\theta})}{\partial \theta_k} \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{C}(\boldsymbol{\theta})}{\partial \theta_\ell} \right\}.$$

BCR : cas de l'exponentielle complexe bruitée

Exponentielle complexe

- Soit le signal $x(n) = Ae^{i(n\omega_0+\phi)} + b(n)$, $n = 0, \cdots, N-1$, où $\boldsymbol{b} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$ et A, ϕ et ω_0 sont des paramètres déterministes inconnus. Si $\boldsymbol{\theta} = \begin{bmatrix} A & \omega_0 & \phi & \sigma^2 \end{bmatrix}^T$ on obtient

$$\boldsymbol{I}\left(\boldsymbol{\theta}\right) = \frac{2}{\sigma^2} \begin{pmatrix} N & 0 & 0 & 0 \\ 0 & A^2Q & A^2P & 0 \\ 0 & A^2P & NA^2 & \\ 0 & 0 & 0 & \frac{N}{2\sigma^2} \end{pmatrix}; P = \sum_{n=0}^{N-1} n, Q = \sum_{n=0}^{N-1} n^2$$

- On en déduit les bornes de Cramér-Rao :

$$BCR(A) = \frac{\sigma^2}{2N} \qquad BCR(\sigma^2) = \frac{\sigma^4}{N}$$

$$BCR(\phi) = \frac{\sigma^2(2N-1)}{A^2N(N+1)} \qquad BCR(\omega_0) = \frac{6\sigma^2}{A^2N(N^2-1)}.$$

MVU et statistique exhaustive

Question

Existe-t-il une fonction des données qui concentre à elle seule toute l'information sur $\boldsymbol{\theta}$? Par exemple, pour $x(n) = \theta + b(n)$, $\hat{\theta}^{MVU} = N^{-1} \sum_{n=1}^{N} x(n)$ et il suffit d'observer $t = \sum_{n=1}^{N} x(n)$ pour obtenir l'estimateur MVU.

Statistique exhaustive

$\boldsymbol{t}$ est une statistique exhaustive pour l'estimation de $\boldsymbol{\theta}$ si, une fois que l'on a observé $\boldsymbol{t}$, rien de plus ne peut être inféré sur $\boldsymbol{\theta}$ à partir des données : $p\left(\boldsymbol{x}|\boldsymbol{t}\right)$ ne dépend plus de $\boldsymbol{\theta}$.

Exemple $x(n) = \theta + b(n)$, $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$

Puisque $\boldsymbol{x} \sim \mathcal{N}\left(\theta\,\mathbf{1}, \sigma^2 \boldsymbol{I}\right)$ et $t = \boldsymbol{x}^T \mathbf{1} \sim \mathcal{N}\left(N\theta, N\sigma^2\right)$, on a

$$p(\boldsymbol{x}, t; \theta) = \left(2\pi\sigma^2\right)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{x}^T\boldsymbol{x} - 2t\theta + N\theta^2\right)\right\} \delta(t - \boldsymbol{x}^T\mathbf{1})$$

$$p(t; \theta) = \left(2\pi N\sigma^2\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{t^2}{N} - 2t\theta + N\theta^2\right)\right\}.$$

Par conséquent la loi conditionnelle de $\boldsymbol{x}|t$ s'écrit

$$p(\boldsymbol{x}|t) = \frac{\left(2\pi N\sigma^2\right)^{1/2}}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{x}^T\boldsymbol{x} - \frac{t^2}{N}\right)\right\} \delta(t - \boldsymbol{x}^T\mathbf{1})$$

et ne dépend plus de θ.

MVU et statistique exhaustive

Statistique exhaustive

- $\boldsymbol{t}$ est une statistique exhaustive pour l'estimation de $\boldsymbol{\theta}$ ssi $p\left(\boldsymbol{x}|\boldsymbol{t}\right)$ ne dépend plus de $\boldsymbol{\theta}$.
- L'estimateur MVU doit être une fonction d'une statistique exhaustive.

Recherche de l'estimateur MVU

- $p\left(\boldsymbol{x};\boldsymbol{\theta}\right) = g\left(\boldsymbol{t},\boldsymbol{\theta}\right) h\left(\boldsymbol{x}\right) \Leftrightarrow \boldsymbol{t}$ statistique exhaustive.
- $\boldsymbol{t}$ est **complète** s'il n'existe qu'une **seule** fonction f telle que $f\left(\boldsymbol{t}\right)$ soit un estimateur non biaisé de $\boldsymbol{\theta}$: c'est justement l'estimateur MVU.
- Si $\breve{\boldsymbol{\theta}}$ est un estimateur non biaisé de $\boldsymbol{\theta}$ et si $\boldsymbol{t}$ est une statistique exhaustive alors $\hat{\boldsymbol{\theta}} = \mathrm{E}\{\breve{\boldsymbol{\theta}}|\boldsymbol{t}\}$ est un estimateur non biaisé de $\boldsymbol{\theta}$ de variance inférieure ou égale à celle de $\breve{\boldsymbol{\theta}}$. Si si $\boldsymbol{t}$ est complète, il s'agit de l'estimateur MVU.

Modèle $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{b}$ avec $\boldsymbol{b} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{C}\right)$

On a

$$p(\boldsymbol{x};\boldsymbol{\theta}) = (2\pi)^{-N/2} \left|\boldsymbol{C}\right|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x}\right\}$$
$$\times \exp\left\{\boldsymbol{\theta}^T\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H}\boldsymbol{\theta}\right\}$$

et donc $\boldsymbol{t} = \boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{x}$ est une statistique exhaustive. En supposant qu'elle est complète, on cherche l'estimateur non biaisé de $\boldsymbol{\theta}$ utilisant $\boldsymbol{t}$. Or

$$\mathrm{E}\{\boldsymbol{t}\} = \boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H}\boldsymbol{\theta} \Rightarrow \hat{\boldsymbol{\theta}}^{MVU} = \left(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{t}.$$

Statistique exhaustive du modèle non linéaire

Modèle $\boldsymbol{x} = \boldsymbol{s}(\boldsymbol{\theta}) + \boldsymbol{b}$ avec $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C})$

On a

$$\begin{aligned} p(\boldsymbol{x}; \boldsymbol{\theta}) &= (2\pi)^{-N/2} |\boldsymbol{C}|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{x}^T \boldsymbol{C}^{-1} \boldsymbol{x}\right\} \\ &\times \exp\left\{\boldsymbol{s}^T(\boldsymbol{\theta})\boldsymbol{C}^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{s}^T(\boldsymbol{\theta})\boldsymbol{C}^{-1}\boldsymbol{s}(\boldsymbol{\theta})\right\} \end{aligned}$$

et donc il n'existe pas de statistique exhaustive puisque $\boldsymbol{s}^T(\boldsymbol{\theta})\boldsymbol{C}^{-1}\boldsymbol{x} = \sum_{n=1}^{N} s(n; \boldsymbol{\theta}) \left[\boldsymbol{C}^{-1}\boldsymbol{x}\right](n)$ n'est pas observable car ne dépendant pas des données $\boldsymbol{x}$ uniquement.

Ecart-type d'une loi uniforme

On considère N données indépendantes $x(n) \sim \mathcal{U}([-\beta, \beta])$ et on cherche à estimer β. La loi de $\boldsymbol{x}$ s'écrit, avec $u(.)$ la fonction échelon

$$
\begin{aligned}
p(\boldsymbol{x}; \beta) &= \frac{1}{(2\beta)^N} \prod_{n=0}^{N-1} \left[u(x(n) + \beta) - u(x(n) - \beta) \right] \\
&= \begin{cases} \frac{1}{(2\beta)^N} & -\beta < x(n) < \beta \quad n = 0, \cdots, N-1 \\ 0 & \text{sinon} \end{cases} \\
&= \begin{cases} \frac{1}{(2\beta)^N} & \max |x(n)| < \beta \\ 0 & \text{sinon} \end{cases}
\end{aligned}
$$

et donc $t = \max |x(n)|$ est une statistique exhaustive. En supposant qu'elle est complète on cherche un estimateur non biaisé de β à partir de t. Pour cela, on cherche la loi de t.

Ecart-type d'une loi uniforme

$$\Pr\left[t \leq \xi\right] = \prod_{n=0}^{N-1} \Pr\left[|x(n)| \leq \xi\right] = \Pr\left[|x(n)| \leq \xi\right]^N$$

$$= \begin{cases} 0 & \xi \leq 0 \\ \left(\frac{\xi}{\beta}\right)^N & 0 \leq \xi \leq \beta \\ 1 & \xi \geq \beta \end{cases}$$

ce qui implique que la FDP de t est

$$p_t(\xi) = \frac{\partial \Pr\left[t \leq \xi\right]}{\partial \xi} = \begin{cases} 0 & \xi \leq 0 \\ \frac{N}{\beta}\left(\frac{\xi}{\beta}\right)^{N-1} & 0 \leq \xi \leq \beta \\ 1 & \xi \geq \beta \end{cases}$$

MVU : exemple

Ecart-type d'une loi uniforme

On a

$$\mathrm{E}\{t\} = \int_0^\beta \frac{N}{\beta}\xi\left(\frac{\xi}{\beta}\right)^{N-1} d\,\xi = \frac{N}{N+1}\beta \Rightarrow \hat{\beta}^{MVU} = \frac{N+1}{N}\max|x(n)|.$$

Sa variance peut se calculer :

$$\begin{aligned}
\operatorname{cov}(\hat{\beta}^{MVU}) &= \left(\frac{N+1}{N}\right)^2 \operatorname{cov}(t) \\
&= \left(\frac{N+1}{N}\right)^2 \left\{\int_0^\beta \frac{N}{\beta}\xi^2\left(\frac{\xi}{\beta}\right)^{N-1} d\,\xi - \left(\frac{N\beta}{N+1}\right)^2\right\} \\
&= \frac{\beta^2}{N(N+2)}.
\end{aligned}$$

MVU : exemple

Exemple $x(n) = n\theta + b(n)$ avec $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \theta\sigma^2 \boldsymbol{I}\right)$

- Dans ce cas on a

$$
\begin{aligned}
p(\boldsymbol{x};\boldsymbol{\theta}) &= \left(2\pi\theta\sigma^2\right)^{-N/2} e^{\frac{-1}{2\theta\sigma^2}\sum_{n=1}^{N}(x(n)-n\theta)^2} \\
&= \left(2\pi\theta\sigma^2\right)^{-N/2} e^{\frac{-\sum_{n=1}^{N} x^2(n)}{2\theta\sigma^2} - \frac{\theta\sum_{n=1}^{N} n^2}{2\sigma^2}} e^{\frac{\sum_{n=1}^{N} n x(n)}{\sigma^2}}.
\end{aligned}
$$

- Par conséquent $t = \sum_{n=1}^{N} x^2(n)$ est une statistique exhaustive pour l'estimation de θ.
- Cependant, $\mathrm{E}\{t\} = \theta^2 \sum_{n=1}^{N} n^2 + N\theta\sigma^2$ et il n'est donc pas évident de trouver un estimateur non biaisé de θ à partir de t.

Estimateur du maximum de vraisemblance

Définition

L'estimateur ML consiste à maximiser la fonction de vraisemblance :

$$\hat{\boldsymbol{\theta}}^{ML} = \arg\max_{\boldsymbol{\theta}} p\left(\boldsymbol{x};\boldsymbol{\theta}\right).$$

Propriétés

- **Efficacité asymptotique** : $\hat{\boldsymbol{\theta}}^{ML} \overset{\text{as}}{\sim} \mathcal{N}\left(\boldsymbol{\theta}, \boldsymbol{I}^{-1}\left(\boldsymbol{\theta}\right)\right)$.
- Si un estimateur efficace existe, l'estimateur ML le produit :

$$\frac{\partial \ln p\left(\boldsymbol{x};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} = \boldsymbol{I}\left(\boldsymbol{\theta}\right)\left[\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{\theta}\right] \Rightarrow \hat{\boldsymbol{\theta}}^{ML} = \boldsymbol{f}(\boldsymbol{x}).$$

- Si une statistique exhaustive existe, i.e., $p\left(\boldsymbol{x};\boldsymbol{\theta}\right) = g\left(\boldsymbol{t},\boldsymbol{\theta}\right) h\left(\boldsymbol{x}\right)$, alors $\hat{\boldsymbol{\theta}}^{ML}$ est une fonction de $\boldsymbol{t}$ uniquement.
- **Propriété d'invariance** : $\boldsymbol{\beta} = u\left(\boldsymbol{\theta}\right) \Rightarrow \hat{\boldsymbol{\beta}}^{ML} = u\left(\hat{\boldsymbol{\theta}}^{ML}\right)$.

Estimateur ML : cas signal plus bruit

Exemple $\boldsymbol{x} \sim \mathcal{CN}\left(\boldsymbol{s}(\boldsymbol{\theta}), \sigma^2 \boldsymbol{I}\right)$, σ^2 inconnu

- La fonction log-vraisemblance s'écrit

$$\Lambda(\boldsymbol{x}; \boldsymbol{\theta}, \sigma^2) = -N \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \left\| \boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta}) \right\|^2 .$$

- On maximise d'abord par rapport à σ^2 :

$$\begin{aligned} \frac{\partial \Lambda(\boldsymbol{x}; \boldsymbol{\theta}, \sigma^2)}{\partial \sigma^2} &= -\frac{N}{\sigma^2} + \frac{1}{\sigma^4} \left\| \boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta}) \right\|^2 \\ \hat{\sigma}^{2^{ML}} &= N^{-1} \left\| \boldsymbol{x} - \boldsymbol{s}(\hat{\boldsymbol{\theta}}^{ML}) \right\|^2 . \end{aligned}$$

- L'estimateur ML de $\boldsymbol{\theta}$ s'obtient comme

$$\hat{\boldsymbol{\theta}}^{ML} = \arg\min_{\boldsymbol{\theta}} \left\| \boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta}) \right\|^2$$

qui est un problème des **moindres carrés non linéaires**.

Estimateur ML : cas signal non linéaire plus bruit

Exemple $\boldsymbol{x} \sim \mathcal{CN}\left(A\boldsymbol{s}(\omega), \sigma^2 \boldsymbol{I}\right)$

- L'estimateur ML de $\boldsymbol{\theta} = \begin{bmatrix} \omega & A \end{bmatrix}^T$ avec $A \in \mathbb{C}$ s'écrit

$$\hat{\boldsymbol{\theta}}^{ML} = \arg\min_{\boldsymbol{\theta}} \left\| \boldsymbol{x} - A\boldsymbol{s}(\omega) \right\|^2.$$

- On minimise d'abord par rapport à A :

$$\begin{aligned}
\left\| \boldsymbol{x} - A\boldsymbol{s}(\omega) \right\|^2 &= \boldsymbol{x}^H \boldsymbol{x} - A \boldsymbol{x}^H \boldsymbol{s}(\omega) - A^* \boldsymbol{s}^H(\omega) \boldsymbol{x} + |A|^2 \boldsymbol{s}^H(\omega) \boldsymbol{s}(\omega) \\
&= \boldsymbol{s}^H(\omega)\boldsymbol{s}(\omega) \left| A - \frac{\boldsymbol{s}^H(\omega)\boldsymbol{x}}{\boldsymbol{s}^H(\omega)\boldsymbol{s}(\omega)} \right|^2 + \boldsymbol{x}^H\boldsymbol{x} - \frac{\left|\boldsymbol{s}^H(\omega)\boldsymbol{x}\right|^2}{\boldsymbol{s}^H(\omega)\boldsymbol{s}(\omega)}
\end{aligned}$$

et l'estimateur ML de A est alors

$$\hat{A}^{ML} = \frac{\boldsymbol{s}^H(\hat{\omega}^{ML})\boldsymbol{x}}{\boldsymbol{s}^H(\hat{\omega}^{ML})\boldsymbol{s}(\hat{\omega}^{ML})}.$$

Estimateur ML : cas signal non linéaire plus bruit

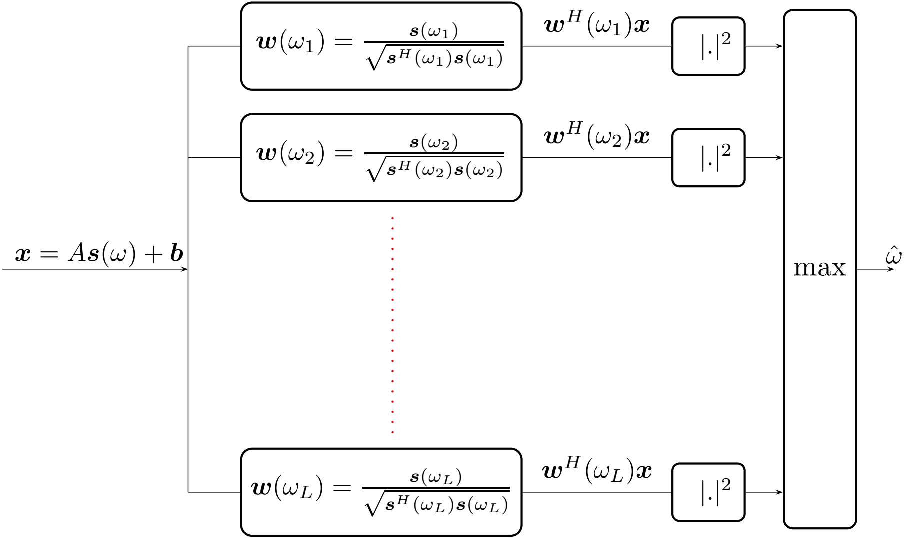
Exemple $\boldsymbol{x} \sim \mathcal{CN}\left(A\boldsymbol{s}(\omega), \sigma^2 \boldsymbol{I}\right)$

- L'estimateur ML de ω est donc obtenu comme

$$\begin{aligned}
\hat{\omega}^{ML} &= \arg\max_{\omega} \frac{|\boldsymbol{s}^H(\omega)\boldsymbol{x}|^2}{\boldsymbol{s}^H(\omega)\boldsymbol{s}(\omega)} \\
&= \arg\max_{\omega} \boldsymbol{x}^H \boldsymbol{P}_{\boldsymbol{s}(\omega)} \boldsymbol{x}; \qquad \boldsymbol{P}_{\boldsymbol{s}(\omega)} = \frac{\boldsymbol{s}(\omega)\boldsymbol{s}^H(\omega)}{\boldsymbol{s}^H(\omega)\boldsymbol{s}(\omega)} \\
&= \arg\max_{\omega} \left|\boldsymbol{w}^H(\omega)\boldsymbol{x}\right|^2; \qquad \boldsymbol{w}(\omega) = \frac{\boldsymbol{s}(\omega)}{\sqrt{\boldsymbol{s}^H(\omega)\boldsymbol{s}(\omega)}}
\end{aligned}$$

- Il peut s'interpréter comme
  1. la valeur de ω qui maximise la norme de la projection de $\boldsymbol{x}$ sur la variété $\boldsymbol{s}(\omega)$;
  2. la valeur de ω qui maximise la puissance en sortie du **filtre adapté** $\boldsymbol{w}(\omega)$.

Filtre adapté (bruit blanc Gaussien)



Estimateur ML : cas signal non linéaire plus bruit

Exemple $\boldsymbol{x} \sim \mathcal{CN}\left(A\boldsymbol{s}(\omega), \sigma^2 \boldsymbol{I}\right)$

- Le rapport signal sur bruit en entrée du filtre adapté s'écrit

$$RSB_{\text{in}} = \frac{|A|^2 \boldsymbol{s}^H(\omega)\boldsymbol{s}(\omega)}{\mathrm{E}\{\boldsymbol{b}^H\boldsymbol{b}\}} = \frac{|A|^2 \boldsymbol{s}^H(\omega)\boldsymbol{s}(\omega)}{N\sigma^2}.$$

- Le signal en sortie de filtre adapté s'écrit $A\boldsymbol{w}^H(\omega)\boldsymbol{s}(\omega) + \boldsymbol{w}^H(\omega)\boldsymbol{b}$, ce qui conduit à un rapport signal sur bruit

$$\begin{aligned} RSB_{\text{out}} &= \frac{|A\boldsymbol{w}^H(\omega)\boldsymbol{s}(\omega)|^2}{\mathrm{E}\{|\boldsymbol{w}^H(\omega)\boldsymbol{b}|^2\}} = \frac{|A\boldsymbol{w}^H(\omega)\boldsymbol{s}(\omega)|^2}{\sigma^2\boldsymbol{w}^H(\omega)\boldsymbol{w}(\omega)} \\ &= \frac{|A|^2 \boldsymbol{s}^H(\omega)\boldsymbol{s}(\omega)}{\sigma^2} = N \times RSB_{\text{in}}. \end{aligned}$$

Le filtre adapté permet un gain de traitement (gain en RSB) de N.

Estimateur ML : cas signal non linéaire plus bruit

Exemple $\boldsymbol{x} \sim \mathcal{CN}\left(A\boldsymbol{s}(\omega), \sigma^2 \boldsymbol{I}\right)$

- **Cas d'un retard :** si $x(t) = As(t-\tau) + b(t)$ et si la forme d'onde $s(t)$ a une énergie constante

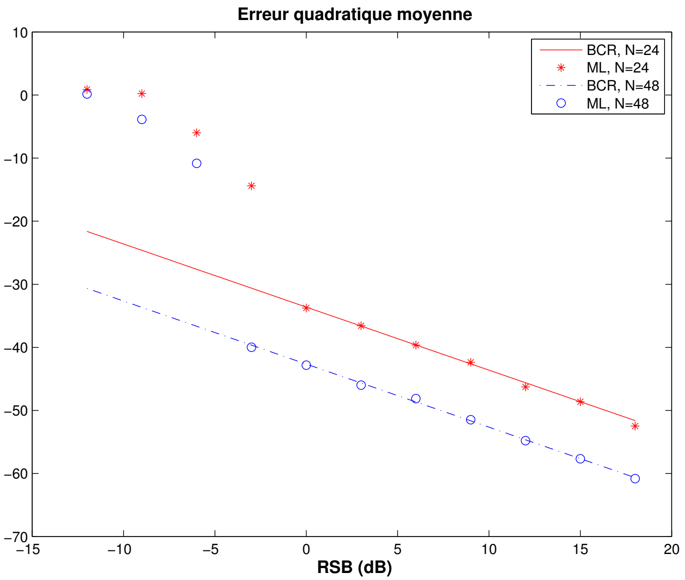
$$\hat{\tau}^{ML} = \arg\max_{\tau} \left| \sum_t x(t) s^*(t-\tau) \right|^2.$$

- **Exponentielle complexe :** si $x(n) = Ae^{in\omega} + b(n)$ alors

$$\hat{\omega}^{ML} = \arg\max_{\omega} \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-in\omega} \right|^2$$

qui n'est autre que le maximum du **périodogramme** du signal, i.e. le module carré de sa transformée de Fourier.

ML et BCR pour une exponentielle complexe



Estimateur ML et moindres carrés

Remarques

- Sous hypothèse Gaussienne $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{s}(\boldsymbol{\theta}), \boldsymbol{C}\right)$ avec $\boldsymbol{C}$ connue, l'estimateur ML revient à résoudre un problème des moindres carrés :
    - moindres carrés linéaires si $\boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{H}\boldsymbol{\theta}$: dans ce cas on a une solution analytique ;
    - moindres carrés non linéaires si $\boldsymbol{s}(\boldsymbol{\theta})$ n'est pas linéaire en $\boldsymbol{\theta}$: dans ce cas, on a recours à des méthodes itératives -utilisant gradient et éventuellement Hessien de $p(\boldsymbol{x}; \boldsymbol{\theta})$- pour trouver l'estimateur.
- Dans le cas où $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_{\text{lin}} & \boldsymbol{\theta}_{\text{nonlin}} \end{bmatrix}^T$ et où $\boldsymbol{s}(\boldsymbol{\theta})$ est linéaire par rapport à $\boldsymbol{\theta}_{\text{lin}}$ et non linéaire par rapport à $\boldsymbol{\theta}_{\text{nonlin}}$, on minimise explicitement $p(\boldsymbol{x}; \boldsymbol{\theta})$ par rapport à $\boldsymbol{\theta}_{\text{lin}}$. On reporte dans la fonction de vraisemblance, et il reste un problème non linéaire par rapport à $\boldsymbol{\theta}_{\text{nonlin}}$.
- Les méthodes des moindres carrés sont également souvent utilisées lorsque la loi n'est plus Gaussienne : dans ce cas, on privilégie la facilité d'implémentation au détriment de l'optimalité.

Estimateur ML : synthèse

Synthèse

- De part ses propriétés d'optimalité asymptotique et compte tenu de sa formulation (maximisation d'une fonction), l'estimateur du maximum de vraisemblance est très souvent utilisé.
- Les problèmes auxquels il se heurte :
  - difficulté à résoudre le problème de maximisation, par exemple due à la présence de nombreux maxima locaux
  - coût calculatoire parfois élevé dû à une maximisation compliquée, par exemple

$$p(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}\left|\boldsymbol{C}(\boldsymbol{\theta})\right|^{1/2}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\theta})\right]^T \boldsymbol{C}^{-1}(\boldsymbol{\theta})\left[\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\theta})\right]\right\}$$

  - estimation peu précise à faible RSB ou faible N (mais dans ce cas peu d'estimateurs feront mieux que le ML)

Méthode des moments

Principe

Exploiter les propriétés des moments du signal (notamment la fonction de corrélation) et leur relation avec $\boldsymbol{\theta}$: par exemple, si $\boldsymbol{\theta} = \boldsymbol{g}(r_{xx}(0), \cdots, r_{xx}(M-1))$ où $r_{xx}(m) = \mathrm{E}\{x^*(n)x(n+m)\}$ est la fonction de corrélation de $x(n)$, on peut envisager l'estimateur

$$\hat{\boldsymbol{\theta}} = \boldsymbol{g}(\hat{r}_{xx}(0), \cdots, \hat{r}_{xx}(M-1))$$

où $\hat{r}_{xx}(m)$ désigne une estimée de $r_{xx}(m)$.

Avantages et inconvénients

- ☺ Estimateurs souvent simples à mettre en œuvre et ne nécessitant pas d'hypothèses sur la loi de $x(n)$.
- ☹ Pas d'optimalité (analyse statistique au cas par cas).

Méthode des moments : exemple

Exponentielle complexe

Soit le signal $x(n) = Ae^{i(n\omega_0+\phi)} + b(n)$, $n = 0, \cdots, N-1$, où $b(n)$ est un bruit blanc complexe de moyenne nulle. On a alors

$$r_{xx}(m) = \mathrm{E}\{x^*(n)x(n+m)\} = A^2 e^{im\omega_0} + \sigma^2\delta(m)$$

et, par exemple,

$$\omega_0 = \text{angle}\left[r_{xx}(1)\right]$$

ce qui suggère une estimation possible de ω_0 comme

$$\hat{\omega}_0 = \text{angle}\left[\hat{r}_{xx}(1)\right] = \text{angle}\left[\frac{1}{N-1}\sum_{n=0}^{N-2} x^*(n)x(n+1)\right].$$

Méthode des moments : analyse de performances

Analyse statistique

Soit $\hat{\boldsymbol{\theta}} = \boldsymbol{f}(\hat{\boldsymbol{t}})$ où $\hat{\boldsymbol{t}}$ est un vecteur contenant des estimées de certains moments du signal, par exemple $\hat{\boldsymbol{t}} = \begin{bmatrix} \hat{r}_{xx}(0) & \cdots & \hat{r}_{xx}(M-1) \end{bmatrix}^T$. En supposant que $\hat{\boldsymbol{t}}$ est un estimateur consistant de $\boldsymbol{t}$, on peut écrire un développement de Taylor au voisinage de $\boldsymbol{t}$:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{f}(\hat{\boldsymbol{t}}) \simeq \boldsymbol{f}(\boldsymbol{t}) + \left[\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{t}^T}\right]_{\boldsymbol{t}} (\hat{\boldsymbol{t}} - \boldsymbol{t}).$$

On en déduit alors $\mathrm{E}\{\hat{\boldsymbol{\theta}}\} \simeq \boldsymbol{f}(\boldsymbol{t})$ et

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \simeq \left[\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{t}^T}\right]_{\boldsymbol{t}} \mathrm{E}\{(\hat{\boldsymbol{t}} - \boldsymbol{t})(\hat{\boldsymbol{t}} - \boldsymbol{t})^T\} \left[\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{t}^T}\right]_{\boldsymbol{t}}^T.$$

$\mathrm{E}\{(\hat{\boldsymbol{t}} - \boldsymbol{t})(\hat{\boldsymbol{t}} - \boldsymbol{t})^T\}$ est souvent connu, du moins pour une grande classe de processus stationnaires.

Méthode des moments : analyse de performances

Analyse à fort rapport signal sur bruit (RSB)

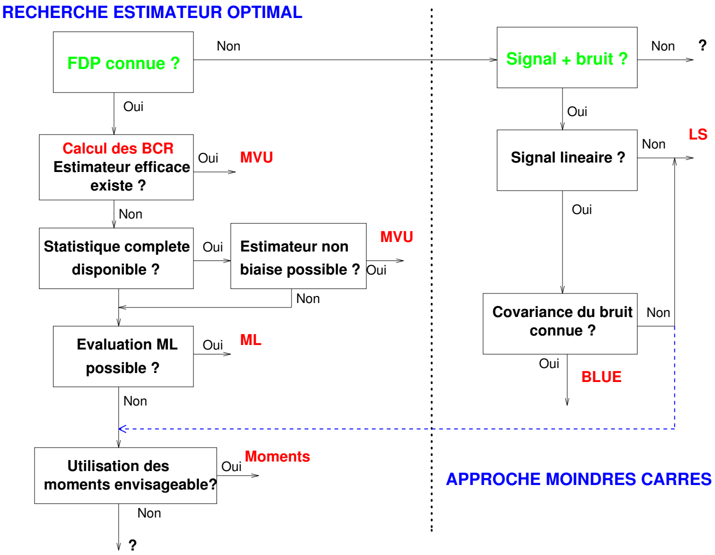
Dans le cas d'un signal déterministe noyé dans un bruit blanc $x(n) = s(n; \boldsymbol{\theta}) + b(n)$, pour lequel $\boldsymbol{\theta} = \boldsymbol{f}(\boldsymbol{s}(\boldsymbol{\theta}))$, on fait souvent l'analyse à fort RSB en considérant $\boldsymbol{t} = \boldsymbol{x}$. On a alors

$$\hat{\boldsymbol{\theta}} = \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{s}(\boldsymbol{\theta}) + \boldsymbol{b}) \simeq \boldsymbol{f}(\boldsymbol{s}(\boldsymbol{\theta})) + \left[\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}^T}\right]_{\boldsymbol{s}(\boldsymbol{\theta})} \boldsymbol{b}.$$

La moyenne et la matrice de covariance asymptotique (i.e. à fort RSB) s'en déduisent

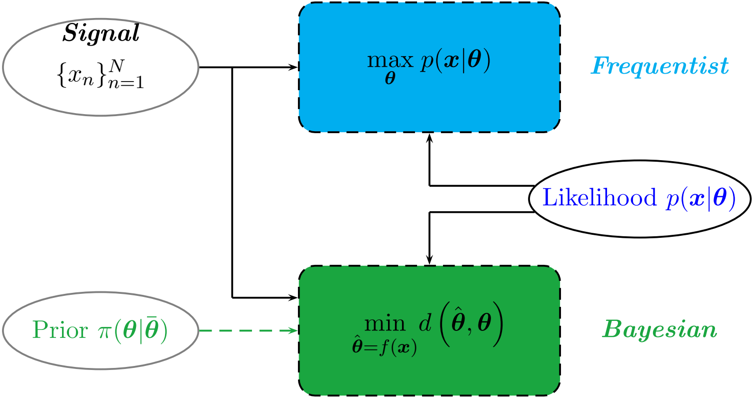
$$\begin{aligned}
\mathrm{E}\{\hat{\boldsymbol{\theta}}\} &\simeq \boldsymbol{f}(\boldsymbol{s}(\boldsymbol{\theta})) = \boldsymbol{\theta} \\
\mathrm{cov}(\hat{\boldsymbol{\theta}}) &\simeq \left[\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}^T}\right]_{\boldsymbol{s}(\boldsymbol{\theta})} \mathrm{E}\{\boldsymbol{b}\boldsymbol{b}^T\} \left[\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}^T}\right]_{\boldsymbol{s}(\boldsymbol{\theta})}^T .
\end{aligned}$$

Comment chercher un estimateur ?



Sommaire

Approche Bayésienne vs approche déterministe



Approche Bayésienne

Principe

Considérer le vecteur paramètre $\boldsymbol{\theta}$ à estimer comme **aléatoire**, avec une **loi a priori** $\pi\left(\boldsymbol{\theta}\right)$: nécessite d'avoir une information a priori sur $\boldsymbol{\theta}$.

Caractérisation

On considère alors la loi jointe $p\left(\boldsymbol{x},\boldsymbol{\theta}\right) = p\left(\boldsymbol{x}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)$.

Loi a posteriori

La **loi a posteriori**

$$p\left(\boldsymbol{\theta}|\boldsymbol{x}\right) = \frac{p\left(\boldsymbol{x},\boldsymbol{\theta}\right)}{p\left(\boldsymbol{x}\right)} = \frac{p\left(\boldsymbol{x}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)}{\int p\left(\boldsymbol{x},\boldsymbol{\theta}\right)\,d\boldsymbol{\theta}}$$

synthétise l'information sur $\boldsymbol{\theta}$ apportée par l'a priori et par les données.

Lois conjuguées

Des lois sont dîtes **conjuguées** si elles appartiennent à la même famille. Par exemple, si

$$p\left(\boldsymbol{x}|\mu,\sigma^2\right) \propto \left(\sigma^2\right)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{x}-\mu\boldsymbol{s}\right)^T\left(\boldsymbol{x}-\mu\boldsymbol{s}\right)\right\}$$

$$\propto \left(\sigma^2\right)^{-N/2} \exp\left\{-\frac{\sigma^{-2}}{2}\left[\boldsymbol{x}^T\boldsymbol{x}-\frac{(\boldsymbol{s}^T\boldsymbol{x})^2}{\boldsymbol{s}^T\boldsymbol{s}}+(\boldsymbol{s}^T\boldsymbol{s})\left(\mu-\frac{\boldsymbol{s}^T\boldsymbol{x}}{\boldsymbol{s}^T\boldsymbol{s}}\right)^2\right]\right\}$$

alors la loi conjuguée pour μ est $\mathcal{N}\left(\bar{\mu},\sigma_\mu^2\right)$ et celle pour σ^2 est $\mathcal{IG}\left(a,b\right)$:

$$\pi\left(\mu\right) \propto \exp\left\{-\frac{1}{2\sigma_\mu^2}\left(\mu-\bar{\mu}\right)^2\right\}$$

$$\pi\left(\sigma^2\right) \propto \left(\sigma^2\right)^{-(a+1)} \exp\left\{-b\sigma^{-2}\right\}.$$

Estimateur à erreur quadratique moyenne minimale

Estimateur MMSE

L'estimateur qui minimise l'erreur quadratique moyenne est donné par

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{\mathsf{mmse}} &= \arg \min_{\hat{\boldsymbol{\theta}}=f(\boldsymbol{x})} \mathrm{E}_{\boldsymbol{x},\boldsymbol{\theta}}\{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\} \\ &= \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}\{\boldsymbol{\theta}\} = \int \boldsymbol{\theta} p\left(\boldsymbol{\theta}|\boldsymbol{x}\right) \, d\boldsymbol{\theta}.\end{aligned}$$

Démonstration MMSE

On note tout d'abord que l'eqm s'écrit

$$\mathrm{eqm}(\hat{\boldsymbol{\theta}}) = \int \left[\int (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T p(\boldsymbol{\theta}|\boldsymbol{x})\, d\boldsymbol{\theta} \right] p(\boldsymbol{x})\, d\boldsymbol{x}.$$

L'intégrale interne s'écrit

$$\begin{aligned}
&\int (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T p(\boldsymbol{\theta}|\boldsymbol{x})\, d\boldsymbol{\theta} \\
&= \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^T - \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}\{\boldsymbol{\theta}\}\hat{\boldsymbol{\theta}}^T - \hat{\boldsymbol{\theta}}\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}\{\boldsymbol{\theta}\}^T + \int \boldsymbol{\theta}\boldsymbol{\theta}^T p\left(\boldsymbol{\theta}|\boldsymbol{x}\right)\, d\boldsymbol{\theta} \\
&= (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{\text{mmse}})(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{\text{mmse}})^T + \int (\hat{\boldsymbol{\theta}}^{\text{mmse}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{\text{mmse}} - \boldsymbol{\theta})^T p(\boldsymbol{\theta}|\boldsymbol{x})\, d\boldsymbol{\theta} \\
&\geq \int (\hat{\boldsymbol{\theta}}^{\text{mmse}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{\text{mmse}} - \boldsymbol{\theta})^T p(\boldsymbol{\theta}|\boldsymbol{x})\, d\boldsymbol{\theta}
\end{aligned}$$

ce qui démontre le résultat.

Propriétés de l'estimateur MMSE

Remarques

- L'estimateur MMSE correspond à la moyenne a posteriori de $\boldsymbol{\theta}$, conditionnellement à $\boldsymbol{x}$: nécessité de calculer cette loi.
- L'estimateur MMSE est non biaisé car

$$\mathrm{E}_{\boldsymbol{x},\boldsymbol{\theta}}\{\hat{\boldsymbol{\theta}}^{\text{mmse}} - \boldsymbol{\theta}\} = \mathrm{E}_{\boldsymbol{x}}\{\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}\{\hat{\boldsymbol{\theta}}^{\text{mmse}} - \boldsymbol{\theta}\}\} = \mathrm{E}_{\boldsymbol{x}}\{\mathbf{0}\} = \mathbf{0}.$$

- Sa matrice de covariance s'écrit

$$\mathrm{E}_{\boldsymbol{x},\boldsymbol{\theta}}\{(\hat{\boldsymbol{\theta}}^{\text{mmse}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{\text{mmse}} - \boldsymbol{\theta})^T\} = \mathrm{E}_{\boldsymbol{x}}\{\boldsymbol{C}_{\boldsymbol{\theta}|\boldsymbol{x}}\}$$

  où $\boldsymbol{C}_{\boldsymbol{\theta}|\boldsymbol{x}} = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}\{(\boldsymbol{\theta} - \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}\{\boldsymbol{\theta}\})(\boldsymbol{\theta} - \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}\{\boldsymbol{\theta}\})^T\}$ est la matrice de covariance a posteriori de $\boldsymbol{\theta}$.

Estimateur du maximum a posteriori

Estimateur MAP

$$\hat{\boldsymbol{\theta}}^{\text{map}} = \arg\max_{\boldsymbol{\theta}} p\left(\boldsymbol{\theta}|\boldsymbol{x}\right) = \arg\max_{\boldsymbol{\theta}} \Lambda\left(\boldsymbol{\theta}|\boldsymbol{x}\right)$$
$$= \arg\max_{\boldsymbol{\theta}} \left[\Lambda\left(\boldsymbol{x}|\boldsymbol{\theta}\right) + \Lambda\left(\boldsymbol{\theta}\right)\right].$$

Lien avec l'estimateur ML

Si $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{s}(\boldsymbol{\theta}), \boldsymbol{C}\right)$ et $\boldsymbol{\theta} \sim \mathcal{N}\left(\bar{\boldsymbol{\theta}}, \boldsymbol{C}_{\boldsymbol{\theta}}\right)$, alors

$$\hat{\boldsymbol{\theta}}^{\text{map}} = \arg\min_{\boldsymbol{\theta}} \left\{\left[\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta})\right]^T \boldsymbol{C}^{-1} \left[\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta})\right] + \left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)^T \boldsymbol{C}_{\boldsymbol{\theta}}^{-1} \left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)^T\right\}$$

et l'estimateur MAP peut être vu comme un **estimateur ML régularisé**.

Borne minimale de la variance en estimation Bayésienne

Borne Bayésienne

L'eqm est bornée inférieurement par $\boldsymbol{I}_B^{-1}(\boldsymbol{\theta})$ où

$$\begin{aligned}
\boldsymbol{I}_B(\boldsymbol{\theta}) &= \mathrm{E}_{\boldsymbol{x},\boldsymbol{\theta}}\{-\frac{\partial^2 \Lambda\left(\boldsymbol{x},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\} \\
&= \mathrm{E}_{\boldsymbol{\theta}}\{\mathrm{E}_{\boldsymbol{x}|\boldsymbol{\theta}}\{-\frac{\partial^2 \Lambda\left(\boldsymbol{x}|\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} - \frac{\partial^2 \Lambda\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\}\} \\
&= \mathrm{E}_{\boldsymbol{\theta}}\{\mathrm{E}_{\boldsymbol{x}|\boldsymbol{\theta}}\{-\frac{\partial^2 \Lambda\left(\boldsymbol{x}|\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\} - \frac{\partial^2 \Lambda\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\} \\
&= \mathrm{E}_{\boldsymbol{\theta}}\{\boldsymbol{I}(\boldsymbol{\theta})\} + \mathrm{E}_{\boldsymbol{\theta}}\{-\frac{\partial^2 \Lambda\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\}.
\end{aligned}$$

Exemple $x(n) = \theta + b(n)$ avec $\boldsymbol{b} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}\right)$ et $\theta \sim \mathcal{N}\left(\bar{\theta}, \sigma_\theta^2\right)$

$$I_B(\theta) = \frac{N}{\sigma^2} + \frac{1}{\sigma_\theta^2} \geq \frac{N}{\sigma^2} = I(\theta).$$

Estimation Bayésienne : exemple

Exemple $x(n) = \theta + b(n)$ avec $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$ et $\theta \sim \mathcal{N}\left(\bar{\theta}, \sigma_\theta^2\right)$

On a alors

$$p\left(\boldsymbol{x}|\theta\right) \propto \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{x} - \theta\mathbf{1}\right)^T\left(\boldsymbol{x} - \theta\mathbf{1}\right)\right\}$$

$$\pi\left(\theta\right) \propto \exp\left\{-\frac{1}{2\sigma_\theta^2}\left(\theta - \bar{\theta}\right)^2\right\}.$$

La loi a posteriori s'écrit donc

$$p\left(\theta|\boldsymbol{x}\right) \propto \exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_\theta^2}\right) - 2\theta\left(\frac{\boldsymbol{x}^T\mathbf{1}}{\sigma^2} + \frac{\bar{\theta}}{\sigma_\theta^2}\right)\right]\right\}.$$

Estimation Bayésienne : exemple

Exemple $x(n) = \theta + b(n)$ avec $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$ et $\theta \sim \mathcal{N}\left(\bar{\theta}, \sigma_\theta^2\right)$

Par conséquent, $\theta | \boldsymbol{x} \sim \mathcal{N}\left(\mu_{\theta|\boldsymbol{x}}, \sigma^2_{\theta|\boldsymbol{x}}\right)$ avec

$$\mu_{\theta|\boldsymbol{x}} = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_\theta^2}\right)^{-1} \left(\frac{\boldsymbol{x}^T \mathbf{1}}{\sigma^2} + \frac{\bar{\theta}}{\sigma_\theta^2}\right); \quad \sigma^2_{\theta|\boldsymbol{x}} = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_\theta^2}\right)^{-1}.$$

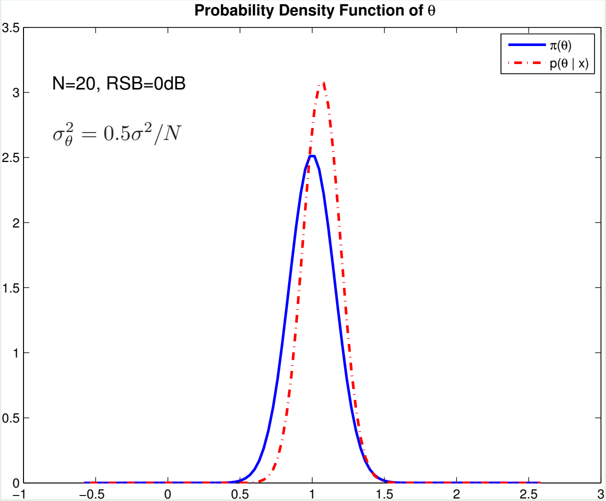
L'estimateur MMSE s'écrit donc

$$\begin{aligned}
\hat{\theta}^{\mathsf{mmse}} &= \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_\theta^2}\right)^{-1} \left(\frac{\boldsymbol{x}^T \mathbf{1}}{\sigma^2} + \frac{\bar{\theta}}{\sigma_\theta^2}\right) \\
&= \frac{N\sigma_\theta^2}{\sigma^2 + N\sigma_\theta^2} \frac{\boldsymbol{x}^T \mathbf{1}}{N} + \frac{\sigma^2}{\sigma^2 + N\sigma_\theta^2} \bar{\theta} \\
&= \alpha \hat{\theta}^{\mathsf{ml}} + (1 - \alpha)\bar{\theta}.
\end{aligned}$$

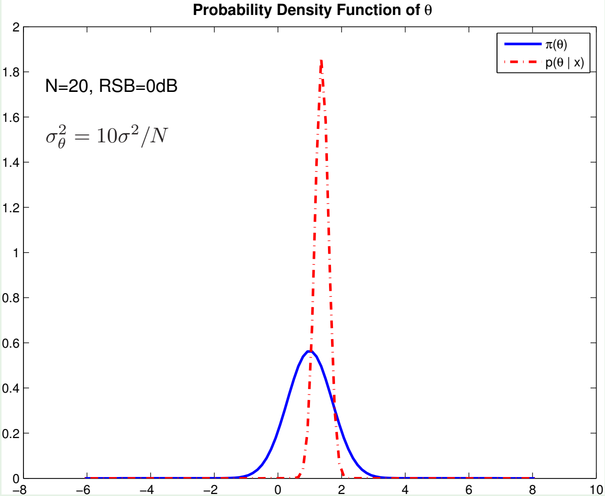
C'est une combinaison de l'estimateur ML et de l'information a priori $\bar{\theta}$.

Estimation Bayésienne : exemple

Exemple $x(n) = \theta + b(n)$ avec $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$ et $\theta \sim \mathcal{N}\left(\bar{\theta}, \sigma_\theta^2\right)$

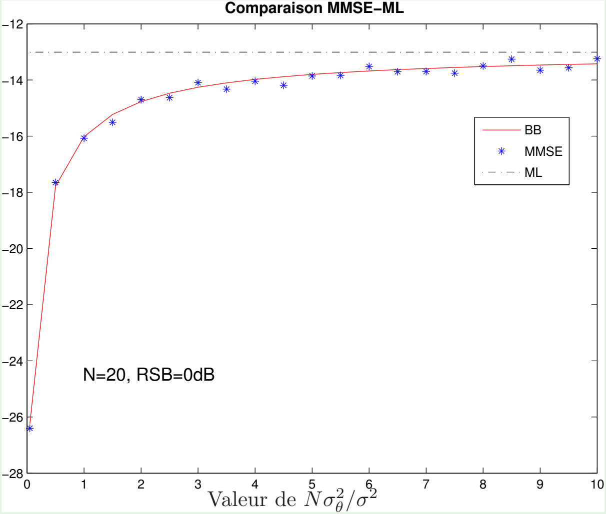


Exemple $x(n) = \theta + b(n)$ avec $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$ et $\theta \sim \mathcal{N}\left(\bar{\theta}, \sigma_\theta^2\right)$



Estimation Bayésienne : exemple

Exemple $x(n) = \theta + b(n)$ avec $\boldsymbol{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$ et $\theta \sim \mathcal{N}\left(\bar{\theta}, \sigma_\theta^2\right)$



Modèle linéaire

Soit le modèle linéaire $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{b}$ avec $\boldsymbol{b} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{C})$ et $\boldsymbol{\theta} \sim \mathcal{N}\left(\bar{\boldsymbol{\theta}}, \boldsymbol{C}_{\boldsymbol{\theta}}\right)$. On a alors les lois suivantes

$$
\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{\theta}) &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta})^T \boldsymbol{C}^{-1} (\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta})\right\} \\
\pi(\boldsymbol{\theta}) &\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)^T \boldsymbol{C}_{\boldsymbol{\theta}}^{-1} \left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)\right\}.
\end{aligned}
$$

La loi a posteriori est dans ce cas

$$
\begin{aligned}
p(\boldsymbol{\theta}|\boldsymbol{x}) \propto &\exp\left\{-\frac{1}{2}\boldsymbol{\theta}^T \left(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H} + \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}\right)\boldsymbol{\theta}\right\} \\
&\times \exp\left\{\boldsymbol{\theta}^T \left(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{x} + \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}\bar{\boldsymbol{\theta}}\right)\right\}.
\end{aligned}
$$

Estimation Bayésienne : modèle linéaire

Modèle linéaire

$\boldsymbol{\theta}|\boldsymbol{x}$ est donc distribuée selon une loi normale, de moyenne et de matrice de covariance

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{\theta}|\boldsymbol{x}} &= \left(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H} + \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}\right)^{-1}\left(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{x} + \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}\bar{\boldsymbol{\theta}}\right) \\ \boldsymbol{C}_{\boldsymbol{\theta}|\boldsymbol{x}} &= \left(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H} + \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}\right)^{-1}.\end{aligned}$$

L'estimateur MMSE est donc

$$\boxed{\hat{\boldsymbol{\theta}}^{\mathsf{mmse}} = \left(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{H} + \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}\right)^{-1}\left(\boldsymbol{H}^T\boldsymbol{C}^{-1}\boldsymbol{x} + \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}\bar{\boldsymbol{\theta}}\right).}$$

Puisque $\boldsymbol{C}_{\boldsymbol{\theta}|\boldsymbol{x}}$ ne dépend pas de $\boldsymbol{x}$, la matrice de covariance de $\hat{\boldsymbol{\theta}}^{\mathsf{mmse}}$ est $\mathrm{E}_{\boldsymbol{x}}\{\boldsymbol{C}_{\boldsymbol{\theta}|\boldsymbol{x}}\} = \boldsymbol{C}_{\boldsymbol{\theta}|\boldsymbol{x}}$.

Estimation Bayésienne : observations

- Dans de nombreux cas, la loi a posteriori $p(\boldsymbol{\theta}|\boldsymbol{x})$ n'appartient pas nécessairement à une famille connue et on n'a pas d'expression analytique de $\int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{x})\, d\boldsymbol{\theta}$. Dans ce cas, une solution consiste à utiliser des méthodes de simulation qui génèrent des variables aléatoires $\boldsymbol{\theta}_n$ distribuées selon $p(\boldsymbol{\theta}|\boldsymbol{x})$. On approxime alors l'intégrale par
$$\hat{\boldsymbol{\theta}}^{\text{mmse}} \simeq \frac{1}{N_r} \sum_{n=1}^{N_r} \boldsymbol{\theta}_n.$$
- L'approche Bayésienne peut être utilisée pour l'estimation de paramètres déterministes : dans ce cas, on choisit en général des lois a priori $\pi(\boldsymbol{\theta})$ peu informatives.
- L'utilisation d'une information a priori permet "d'aider" l'estimation dans des cas difficiles tels que faible RSB ou faible N.

Références

1. Steven M. Kay, *Fundamentals of Statistical Signal Processing : Estimation Theory*, Prentice Hall, Englewood Cliffs, NJ, 1993
2. Louis L. Scharf, *Statistical Signal Processing : Detection, Estimation and Time Series Analysis*, Addison Wesley, Reading, MA, 1991
3. Harry L. Van Trees, *Detection, estimation and modulation theory, Part I*, John Wiley, 2004
4. Harry L. Van Trees, Kristine L. Bell *Detection, estimation and modulation theory, 2nd Edition, Part I, Detection, Estimation, and Filtering Theory*, John Wiley, 2013
5. Christian Robert, *The Bayesian Choice - From Decision-Theoretic Foundations to Computational Implementation*, Springer Verlag, 2007
6. Christian Robert, George Casella *Monte-Carlo statistical methods, 2nd edition*, Springer Verlag, 2004