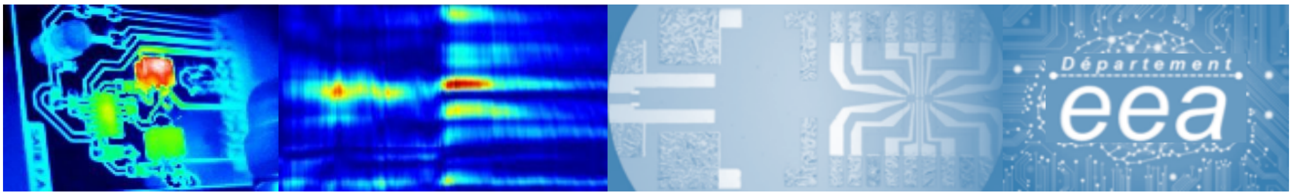
M1 E3A - Voie André Ampère

UE451

Traitement du Signal et de l'image

Un cours de:
Cécile Durieu

Rédigé par:
Pierre-Antoine Comby



Introduction

Les signaux déterministe sont utilisé pour les test ou les commandes . Ils sont cependants insuffisant pour décrire l'ensemble des phénomènes qui peuvent être aléatoire ou stochastique. Il est donc nécessaire d'introduire des "signaux aléatoires" (SA).

Exemple: *Lot de resistances "identiques"* On considère un lot de résistance de même valeur indiqué par le code couleur, et de même précision. Elles ont en réalité toutes des valeurs différentes R_i. on note $s_i(t)$, la tension au borne de R_i.

On a $s_i(t) = 0$ mais avec $A \gg 1$ on obtient $A.s_i(t) \neq 0$.

Alors : $\overline{s_i(t)} = 0$ mais il existe une fluctuation de la tension due au bruit thermique.

On note alors :

$$s_i(t) = s(t, i) = s(t, \omega)$$

C'est une réalisation particulière du SA, appelé par la suite trajectoire.

Dans le cas général on parle de fonction aléatoire $F(\theta, \omega)$,où θ est un paramètre certain (comme le temps par exemple).

Remarque: Un signal aléatoire n'est caractérisé qu'en moyenne.

- moyenne à ω donné : $s(t, \omega_0) = s_0(t)$: **moyenne temporelle**
- moyenne à t donné : $s(t_0, \omega) = s_0(\omega)$: **moyenne statistique**.

On dit aussi que $S(t, \omega) = S_t(\omega)$ est une famille de variable aléaoire indexé par le temps, qui décrit l'aspect incertain à chaque instant.

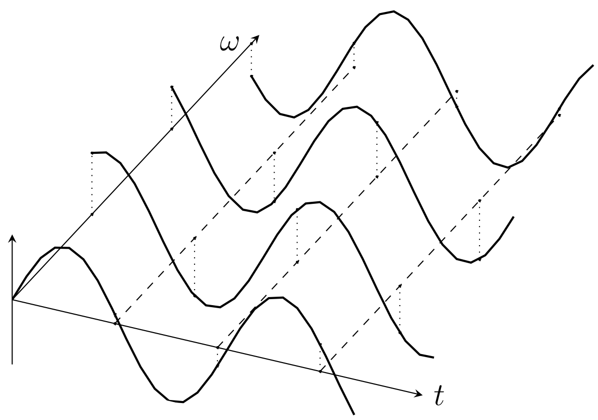


Figure 1 – Représentation des trajectoires d'un signal aléatoire

Table des matières

Chapitre 1

Rappel d'élement de probabilité et de VA

1 Probabilités

1.1 Évènement

- La réalisation d'une expérience aléatoire (on ne peux pas prédire avec certitude le résultat) est un *évènement* ω, singleton de Ω ensembles de tous les évènements.

  Exemple: *jet de dé* aux évènements "Tirer 1, ... ,6 " on associe $\Omega = \omega_1, ...\omega_6$

- $\mathcal{E}$est une tribu (ou σ-algèbre) de Ω, tel que :
  - $\Omega \in \mathcal{E}$
  - $\mathcal{E}$ est stable par union , intersection et complémentarité.

1.2 Probabilités

Définition

On appelle probabilité :

$$P : \begin{cases} \mathcal{E} & \rightarrow [0, 1] \\ E & \mapsto P(E) \end{cases}$$

tel que:

- $P(\Omega) = 1$
- $\forall E_i, i \in \mathbb{I}$, desév disjoint 2 à 2, $\Longrightarrow$ $P\left(\bigcup_{i\in\mathbb{I}} E_i\right) = \sum_{\mathbb{I}} P(E_i)$

Proposition

- $P(\bar{E}) = 1 - P(E)$
- $(P(\emptyset) = 0)$
- $A \subset B \implies P(A) \leq P(B)$
- $P(A + B) = P(A) + P(B) - P(A \cap B)$

1.3 Probabilités conditionnelles

Définition

Soit A et B deux évènements. On appelle *probabilité conditionnelle* la probabilité de A sachant que B est réalisé:

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

Proposition ***(Formule de Bayès)***

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

1.4 Indépendance

Définition

Deux évènements A et B sont dits *indépendant* si et seulement si le fait que A est réalisé n'apporte pas d'information sur la réalisaiton de B

$$\begin{aligned} & P(A|B) = P(A) \\ \iff & P(B|A) = P(B) \\ \iff & P(A \cap B) = P(A).P(B) \end{aligned}$$

Définition

Des évènements $(E_i)_{i \in \mathbb{I}}$ sont dits mutuellement indépendants (ou encore indépendants dans leur ensemble), si et seulement si:

$$P\left(\bigcap_{i \in \mathbb{I}} E_i\right) = \prod_{\mathbb{I}} P(E_i)$$

Proposition

L'indépendance dans son ensemble implique l'indépendance deux à deux.
La réciproque n'est pas forcément vraie.

2 Variable aléatoire réelle et scalaire

On se place dans un espace probabilisé Ω donné.

2.1 Généralité et exemple

Définition

On appelle *Variable aléatoire* (VA) :

$$X : \begin{cases} \Omega \to \mathbb{R} \\ \omega \mapsto X(\omega) = x \end{cases}$$

Exemple:

- Dé à n faces (discret)
- distance d'une flèche au centre de la cible.

Proposition

Pour des variables aléatoires continues,

$$P(X = x) = 0, \forall x \in \mathbb{R}$$

car x est un point de mesure nulle.

2.2 Fonction de répartition

Définition

On appelle fonction de répatition:

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(X \in]-\infty, x]) \\ &= P(\{\omega \in \Omega | X(\omega) \leq x\}) \end{aligned}$$

Proposition

- $0 \leq F_X(x) \leq 1$
- $P(a \leq X \leq b) = F_X(b) - F_X(a)$
- F_x est une fonction :
  - non décroissante
  - continue presque partout

Une variable aléatoire est complétement caractérisée par sa f.d.r

Remarque: Dans le cas d'une VAD , F_X est en marche d'escalier.

2.3 Densité de probabilité

Définition

On appelle *densité de probabilité* la fonction :

$$f_X(x) \underset{\mathcal{D}}{=} \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$$

Avec la dérivée généralisé au sens des distributions.

Proposition

- Les fonction de densité de probabilité et de répartition sont équivalentes pour décrire une variable aléatoire.
- $f_X(x) \geq 0$
- $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = 1$
- $\displaystyle\int_{-\infty}^{x} f_X(\alpha)\mathrm{d}\alpha = F_X(x)$

Remarque: Pour les variables aléatoires discrètes, la ddp est une suite d'impulsion de Dirac :

$$f_X(x) = \sum_{i \in \mathbb{I}} p_i \delta(x - x_i)$$

Exemple:

- VAC uniforme sur $[a, b]$:

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}$$

- VAC gaussienne :

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(\frac{-1}{2}\frac{(x-m_x)^2}{\sigma_X^2}\right)$$

2.4 Changement de VA

Proposition

Soit $g : \begin{cases} \mathbb{R} \to \mathbb{R} \\ X \mapsto g(X) = Y \end{cases}$ une fonction homéomorphique[1]

Alors :

$$f_Y(y) = f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right| = f_X(x) \frac{1}{\left| \frac{\mathrm{d}y}{\mathrm{d}x} \right|}$$

Dans le cas ou g n'est pas bijective :

$$f_Y(y) = \sum_{x_i | g(x_i) = y} f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|_{x = x_i}$$

2.5 Expérance, moment et fonction caractéristique

Définition

pour $g\mathbb{R} \to \mathbb{C}^p$ On appelle *espérance* d'une variable aléatoire la grandeur:

$$E(g(X)) = \int_{\mathbb{R}} g(x) f_X(x) \mathrm{d}x$$

Dans le cas discret on a:

$$E(g(X)) = \sum_{\mathbb{I}} g(x_i) P(X = x_i)$$

Proposition

L'espérance est linéaire (sous réserve d'existance) :

- $E[c] = c$
- $E[cg(x)] = cE[g(x)]$
- $E[g(x) + h(x)] = E[g(x)] + E[h(y)]$

Remarque: On note aussi $E[X] = m_X = m$ “moyenne de la variable aléatoire”. Si $m = 0$ on dit que la VA est centrée.

1. continue, bijective continue

Définition

On appelle *momemt d'ordre k*:

$$m_k = E[X^k]$$

Le *moment centré d'ordre k :*

$$m_k = E[(X - m_X)^k]$$

Le moment μ_2 est aussi appelé la *variance*

Remarque: on note $\sigma_x = \sqrt{v_x}$ l'écart type de X. Il mesure la dispersion autour de m_x. On défini la variable centrée réduite associée à X :

$$X_r = \frac{X - m_X}{\sigma_X}$$

2.6 Fonction caractéristique

Définition

On appelle fonction caractéristique:

$$\phi_X(u) = E[exp(juX)] = \int_{-\infty}^{+\infty}$$

Proposition

- $\phi_X(u)$ existe toujours $|\phi_X(u)| \leq \phi_X(0) = 1$
- Symétrie hermitienne
- $\phi_X(u)$ est continue même pour des VA discrètes
- On appelle 2ème fonction de répartition $\Psi_X(u) = \ln(\phi_X(u))$
- $$m_k = (-j)^k \left. \frac{\mathrm{d}^k \phi_X(u)}{\mathrm{d}u^k} \right|_{u=0}$$

3 Couple de variable aléatoire réelles

3.1 Généralité

Définition

Un couple de variable aléatoire est défini comme:

$$Z \begin{cases} \Omega \to \mathbb{R}^2 \\ \omega \mapsto Z(\omega) = \begin{bmatrix} X(\omega) \\ Y\omega \end{bmatrix} \end{cases}$$

On défini également:

$$Z^{-1} : \mathcal{D} \mapsto Z^{-1}(\mathcal{D}) = E_D \subset \mathcal{E}$$

3.2 Fonction de répartition

Définition

- fonction de répartition conjointe:

$$\begin{aligned} P(X < x; Y < y) &= F_{XY}(x, y) \\ &= P((x, y) \in \mathcal{D}) \\ &= F_Z(z) \end{aligned}$$

- fonction de répartition marginale

$$\begin{aligned} F_X(x) = P(X < x) &= F_{XY}(x, +\infty) \\ &= P((x, y) \in \mathcal{D}_X) \end{aligned}$$

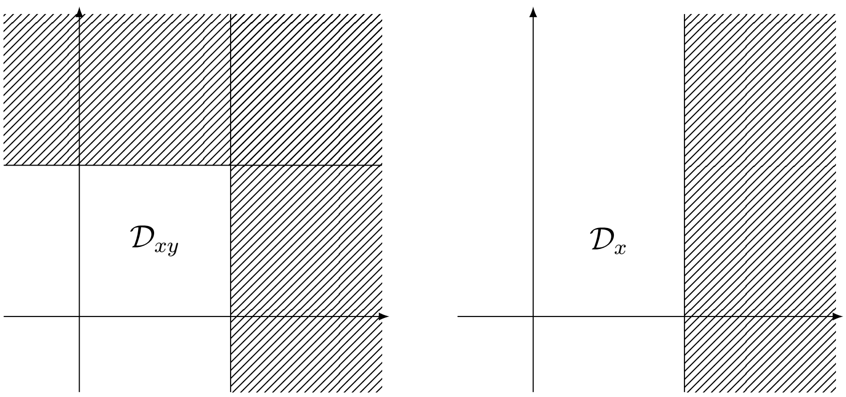


Figure 1.1 – Représentation des domaines d'existence possible pour X, Y

3.3 Densité de probabilité

Définition

on défini la densité de probabilité conjointe:

$$f_{XY} = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

Proposition

densité de probabilité conjointe et fonction de répartition sont reliées :

$$\int_{-\infty}^{x^-}\int_{-\infty}^{y^-} f_{XY}(\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\alpha = F_{XY}(x,y)$$

et :

$$\int_{-\infty}^{x}\int_{\mathbb{R}} f_{XY}(\alpha,\beta)\mathrm{d}\beta = F_{XY}(x,\infty) = F_X(x)$$

Définition

À partir de la fonction de répartion marginale on peux définir la loi marginale de X :

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = \int_{\mathbb{R}} f_{XY}(x,y)\mathrm{d}y$$

Et alors la loi conditionelle de X sachant Y:

$$f_{X|Y}(x) = \frac{f_{XY}(x,y)}{f_{Y(y)}}$$

3.4 Indépendance

Définition

On dit que X et Y sont indépendant:

$\iff \quad F_{XY}(x,y) = F_X(x)F_Y(y)$

$\iff \quad f_{XY}(x,y) = f_X(x)f_Y(y)$

3.5 Changement de VA

Proposition

On considère :

$$g \begin{cases} \mathbb{R}^2 \to \mathbb{R}^2 \\ Z = (X,Y) \mapsto W = (U,V) = g(X,Y) \end{cases}$$

Alors :

$$f_W(w) = f_Z(z)|J|$$

où :

$$J = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$$

Remarque: Il est parfois plus simple de calculer :

$$|K| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

Au quel cas on a : $f_W(w) = f_Z(z)\frac{1}{|K|}$

3.6 Espérance et moments-fonction caractéristique

Dans la suite on considère la fonction suivante :

$$g : \begin{cases} \mathbb{R}^2 \to \mathbb{C}^p \\ Z = (X,Y) \mapsto g(Z) = \begin{bmatrix} g_1(X,Y) \\ \vdots \\ g_p(X,Y) \end{bmatrix} \end{cases}$$

Théorème *(Théorème de transfert)*

On a :

$$E[g(z)] = \iint_{\mathbb{R}^2} g(X,Y) f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$

Proposition

Dans le cas de VA indépendante et pour g séparable on a : $g(X,Y) = g_X(X)g_Y(Y)$ et alors :

$$E[g(X,Y)] = E[g_X(X)]E[g_Y(Y)]$$

Définition

On peux également définir les moments d'un couple de VA:

- Moment d'ordre 1
$$E[Z] = m_Z = \begin{bmatrix} m_X \\ m_Y \end{bmatrix}$$
- Moment d'ordre 2 (Matrice de corrélation)
$$E[ZZ^T] = E\left[\begin{bmatrix} X^2 & XY \\ XY & Y^2 \end{bmatrix}\right] = \begin{bmatrix} E[X^2] & E[XY] \\ E[XY] & E[Y^2] \end{bmatrix} = C_{ZZ}$$

Remarque: C_{ZZ} est symétrique positive : $C_{ZZ} \in S_n^+(\mathbb{R})$

Définition

On appelle matrice de covaraince la matrice de corrélation des variables centrées:

$$\Sigma_{ZZ} = E[(X - m_x)(Y - m_Y)^T] = \begin{bmatrix} \sigma_x^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

où $\rho_{XY} = \frac{E[(X-m_X)(Y-m_y)^T]}{\sigma_x\sigma_y} = E[X_r Y_r]$ est le *coefficient de corrélation*

Proposition

- $\Sigma_{ZZ} \in S_n^+$
- $\rho_{XY} < 1$
- $\rho_{XY} = 1$ ssi $\exists a, b, c \neq 0, aX + bY + c = 0$. Les variables sont alignées.
- Si $\rho_{XY} = 0$ on dit que les variables sont décorélées

Théorème

L'indépendance de 2 variables aléatoires implique leur non corrélation.
La réciproque n'est vraie que dans le cas gaussien.

3.7 Espérance de loi conditionnelle

Définition

On note
$$E[X|Y=y] = \int_{\mathbb{R}} x f_{X|Y=y}(x)\mathrm{d}x = m_X(y)$$
l'espérance conditionnelle de la VA X sachant $Y = y$

Proposition

On a :
$$E[m_x(y)] = E[X]$$

Démonstration : Directement :

$$\begin{aligned}
E[m_X(y)] &= \int_{\mathbb{R}} m_X(y) f(y)\mathrm{d}y \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{XY}(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{\mathbb{R}^2} x f_{XY}(x,y)\mathrm{d}x\mathrm{d}y \\
&= E[X]
\end{aligned}$$

■

4 Variable aléatoire vectorielle et réelles

4.1 Définition

Définition

On généralise la notion de variable aléatoire et de couple de variable aléatoire :
$$X : \begin{cases} \Omega \to \mathbb{R}^n \\ \omega \mapsto X(\omega) = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \end{cases}$$

4.2 Fonction de répartition

Définition

- Fonction de répartition conjointe:(toutes les composantes jouent le même rôle)

$$F_{X_1 \ldots X_n}(x_1 \ldots x_n) = P\left(\bigcap_{i=1}^{n} X_i < x_i\right)$$

- Fonction de répartition marginale de X_i:

$$F_{X_i} = P(X_i < x_i) = P\left(X_i < x_i ; \bigcap_{j \neq i} X_j < +\infty\right)$$

Les propriétés démontrées dans le cas 2 se généralise au cas vectoriel.

4.3 Densité de Probabilité

Définition

On défini la densité de probabilité conjointe:

$$f_X(x) = \frac{\partial^n}{\partial \mathbf{x}} F_{\mathbf{X}}(\mathbf{x})$$

Et alors :

$$P(X \in \mathcal{D}) = \int_{\mathcal{D}} f_X(x) \mathrm{d}x = \iint_{\mathcal{D}} \ldots \int f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x}$$

Définition

On généralise de même les notions de ddp margianle et conditionnelle:

- ddp marginale:

$$f_{X_i}(x_i) = \frac{\mathrm{d}^n F_{x_i}(x_i)}{\mathrm{d}x_i} = \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \ldots \mathrm{d}x_n$$

- ddp conditionnelles: On considère Y et Z de VA vectorielles:

$$f_{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(\mathbf{y}) = \frac{f_{\mathbf{YZ}}(\mathbf{y}, \mathbf{z})}{f_{\mathbf{Z}(z)}}$$

4.4 Indépendance

Théorème

On donne une CNS d'indépencande dans leur ensemble des VA X_i :

$$F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} F_{X_i}(x_i) \iff f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

L'indépendance dans leur ensemble implique l'indépendance 2à2.

4.5 Changement de variable aléatoire

Proposition

Pour $g \begin{cases} \mathbb{R}^n \to \mathbb{R}^n \\ \mathbf{X} \mapsto g(\mathbf{X}) = \mathbf{Y} \end{cases}$ On peux définir le changement de variable :

$$f_{\mathbf{Y}(\mathbf{y}} = f_{\mathbf{X}}\mathbf{x}|\mathbf{J}| = f_{\mathbf{X}}(\mathbf{x})\frac{1}{|\mathbf{K}|}$$

où : $\mathbf{J} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}^T} = x_{i,j}$ et $K = \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} = y_{j,i}$

4.6 Espérance, moments et fonction caractéristique

Théorème ***(Théorème de transfert)***

$$E[g(\mathbf{X})] = \int_{\mathbb{R}^n} g(\mathbf{X}) f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x}$$

Définition

- Moment d'ordre 1:
$$\mathbf{m}_{\mathbf{x}} = E[\mathbf{X}]$$
- Moment d'ordre 2: (matrice de corrélation)
$$\mathbf{C}_{\mathbf{XX}} = E[\mathbf{X}\mathbf{X}^T] \geq 0$$
- Moment centrée d'ordre 2: (matrice de covariance)
$$\mathbf{\Sigma}_{\mathbf{XX}} = E[(\mathbf{X} - \mathbf{m}_{\mathbf{x}})(\mathbf{Y} - \mathbf{m}_{\mathbf{y}})^T]$$
- Fonction caractèristique:
$$\phi_{\mathbf{X}}(\mathbf{u}) = E[e^{j\mathbf{u}^T\mathbf{X}}]$$

4.7 Va Gaussienne et réelle

Définition

On dit que $X = \mathbf{X_1} \vdots \mathbf{X_n}$ est une VA gaussienne:
$\iff$ X_i sont gaussiens et indépendants dans leur ensembles
$\iff$ $\sum_{i=1}^{n} \alpha_i X_i$ est une gaussienne.

Proposition

- $\mathbf{X} \quad \mathcal{N} \implies X_i \quad \mathcal{N}$. La réciproque n'est pas vraie (cf ex 9/10 p 14 du fascicule)
- $\mathbf{X} \quad \mathcal{N} \implies$ loi conditionnelle gaussienne.
- $X_i \quad \mathcal{N}$ et indépendantes dans leur ensemble $\implies \mathbf{X} \quad \mathcal{N}$.
- $\mathbf{X} \quad \mathcal{N}$ et X_i indépendants 2à2 $\implies$ indépendant dans leur ensemble.
- $\mathbf{X} \quad \mathcal{N} \implies \mathbf{Y} = \mathbf{AX} + \mathbf{B} \quad \mathcal{N}$

5 Extension aux VA complexes

Définition

On généralise *encore*:

$$\mathbf{Z} \begin{cases} \Omega \to \mathbb{C}^p \\ \omega \mapsto \mathbf{Z}(\omega) = \mathbf{X} + j\mathbf{Y} \end{cases}$$

Notation : $\mathbf{Z}^\dagger = (\mathbf{Z}^*)^T$ transposé conjugué.

Proposition

- Fonction de répartition :

$$f_{\mathbf{Z}}(\mathbf{z}) = P(\mathbf{X} < \mathbf{x}; \mathbf{Y} < \mathbf{y})$$

- Matrice de corrélation :

$$\mathbf{C}_{\mathbf{zz}} = E[\mathbf{Z}\mathbf{Z}^{\dagger}]$$

- Matrice de covariance :

$$\mathbf{\Sigma}_{\mathbf{ZZ}} = E[(\mathbf{Z} - \mathbf{m}_{\mathbf{z}})(\mathbf{Z} - \mathbf{m}_{\mathbf{z}})^{\dagger}]$$

- Fonction caractéristique :

$$\phi_{\mathbf{Z}}(\mathbf{u}) = E[e^{j\mathbf{u}^{\dagger}\mathbf{Z}}]$$

- La linéarité de l'espérance donne également :

$$E[g(\mathbf{Z})^*] = E[g(\mathbf{Z})]^*$$

Chapitre 2

Signaux Aléatoire

1 Stationnarité et ergodicité

Définition

Soit $(\Omega, \mathcal{E}, P)$ un espace probabilisé. Une famille/suite de VA indexé par le temps est un *signal aléatoire* $\in \mathbb{C}^n$ noté : $X_t(\omega)$ $\forall t \in \mathbb{R}$ (ou $X_n(\omega)$ $\forall n \in \mathbb{Z}$)

En ptratique on s'interesse àdes des signaux de dimension 1.

Rappel : on appelle trajectoire la réalisation / acquisition d'un signal. il existe deux types de moyenne possible :

- temporelle, idem que celle des signaux déterministes
- Statistique, ideme que pour les VA.

Exemple: Soit le SA suivant : $X(t, \omega) = A \sin(2\pi f_0 t)$ où A est une variable aléatoire , (ici qui suit une loi uniforme). Alors une réalisation de ce SA est $x(t) = a \sin(2\pi f_0 t)$.

- $\overline{x(t, \omega)} = 0 = m_x$ et $\overline{x^2 2(t, \omega)} = \frac{a^2}{2}$
- $E[X(t, \omega)] = \sin(2\pi f_0 t) E[A] = m_X(t)$.

1.1 Moyenne temporelle

On rappelle les différentes expression des 1er et 2nd ordre (si il existe) de trajectoire particulière.

Définition

Les moments d'ordre 1 temporel sont des *moyennes temporelle*:

- Temps continu

$$\overline{x(t,\omega)} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t,\omega)\mathrm{d}t = m_x(\omega)$$

- Temps discret

$$\overline{x[n,\omega]} = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x[n,\omega] = m_x(\omega)$$

Définition

Les moments d'ordre 2 croisés définissent la fonction d'intercorrélation temporelle ((ω est fixé)

- Temps continu:

$$\overline{x(t,\omega)\cdot y^*(t-\tau,\omega)} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t,\omega)y^*(t-\tau,\omega)\mathrm{d}t = C^p_{xy}(\tau,\omega)$$

- Temps discret:

$$\overline{x[n,\omega]y^*[n-k,\omega]} = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x[n,\omega]y^*[n-k,\omega]$$

Remarque: On dit également que les Les moments temporels dépendent de la trajectoire. Si $y = x$ on parle d'autocorrélation. De plus $C^p_{xx}(0)$ est la *puissance de x*.

1.2 Ergodicité

Définition

- Un processus est *ergodique au sens stricte* si et seulement si toutes les moyennes temporelles sont indépendantes de la trajecoire considérée.
- Un processus est ergodique à l'ordre n si et seulement si tous les moments jusqu'à l'ordre de n sont indépendant de la trajectoire considéré. Les moments temporel d'un signal ergodique ne sont pas des variables aléatoires.

Remarque: Souvent $n = 2$ Pour 2 SA on parle d'ergodicité dans leur ensemble.

1.3 Moyenne statistique

On considère les signaux aléatoire à des instants particuliers, fixé.

Remarque: En fixant le temps on peux définir les fonctions de répartition et la densité de probabilité d'un signal aléatoire. Alors on peux exprimer les moments statistiques de ses signaux temporels :

Définition

On défini la moyenne statistique (moment d'ordre 1):

$$m_X(t) = E[X(t,\omega)] = \int_{\mathbb{R}} x f_X(x,t)\mathrm{d}x$$

et la fonction d'intercorrélation statistique (moment d'ordre 2):

$$\gamma_{xy}(t_1,t_2) = E[X(t_1,\omega)y^*(t_2,\omega)] = \iint xy^* f_{x,y,t_1,t_2}\mathrm{d}x\mathrm{d}y$$

Il en est de meme dans le cas discret.

1.4 Stationnarité

Définition

- Un processus aléatoire est *stationanaire au sens strict* ssi toutes ses caractéristiques statistiques sont invariantes par tout changement de l'origine des temps.

$$f_X(x,t) = f_X(x,t+\tau) = f_X(x) \quad \forall\tau$$

- Un processus aléatoire est stationnaire au sens large /au second ordre ssi ses moments d'ordre 1 et 2 sont invariants par tout changement d'origine des temps.

$$E[|X(t,\omega)|^2] = E[|X(t',\omega)|^2] < +\infty$$

1.5 Stationnarité et ergodicité

Proposition

Si un SA est à la fois stationnaire et ergodique les moyennes temporelles et statistiques sont égales.
L'ensemble des processus stochastique,stationnaire, ergodique peux être obtenu à partir d'une seule trajectoire allant de $-\infty$ à $+\infty$.

Proposition

Un SASE au second ordre est tel que :

$$m_x = E[X(t)] = \overline{x(t)} = m_x$$

et

$$\gamma_{xx}(\tau) = E[X(t)X^*(t-\tau)] = \overline{x(t)x^*(t-\tau)} = C^p_{xx}(\tau)$$

2 Corrélation et densité spectrale de puissance

Ici on s'interesse la répartition de la puissance d'un SA en fonction de la fréquence (idem que la DSE pour des signaux à énergie finie). On se restreint à des SAS du 2nd ordre.

Notation : $x(t, \omega)$ représente le SA ou une des ses réalisation
$X(f)$ représente la TF d'un signal x sous réserve d'existence.

Théorème *(Wiener-Kintchine)*

$$TF[\gamma_{xx}] = \Gamma_{XX}(f) = \text{ dsp de x(t)}$$

Proposition *(Cas du TC)*

$$\begin{aligned}
\Gamma_{xy}(f) &= \int_{\mathbb{R}} \gamma_{xy}(\tau) e^{-j2\pi f t} \mathrm{d}t \\
\gamma_{xy}(\tau) &= \int_{\mathbb{R}} \Gamma_{xy}(f) e^{j2\pi f \tau} \mathrm{d}f \\
\gamma_{xy}(0) &= \int_{\mathbb{R}} \Gamma_{xy}(f) \mathrm{d}f = \text{ puissance (statistique)}
\end{aligned}$$

Proposition *(Cas du TD)*

$$\begin{aligned}
\Gamma_{xy}(f) &= \sum \gamma_{xy}[k] e^{-j2\pi f k} \\
\gamma_{xy}[\tau] &= \int_{-1/2}^{1/2} \Gamma_{xy}(f) e^{j2\pi f k} \mathrm{d}f \\
P_x = \gamma_{xx}[0] &= \int_{-1}^{1} \Gamma_{xx}(f) df
\end{aligned}$$

Exemple: cf TP1

Proposition

- $|\gamma_{xy}(\tau)| \leq \gamma_{xx}(0) = P_x$
- $\gamma_{xx}(-\tau) = \gamma_{xx}(\tau)^* \implies \Gamma_{xx}(f) \in \mathbb{R}$
- $\Gamma_{xx}(f) > 0$

Remarque: très souvent on a $\gamma_{xx}(\tau) \xrightarrow[+\infty]{|} m_x|^2$, ce qui signifie qu'on a aps d'effet "mémoire" àl'infini. Si $m_x \neq 0$ la DSP comporte une raie à l'origine de valeur m_x.

3 Periodogramme

4 Signaux aléatoire particulier

4.1 SA indépendants

4.2 SA décorrélés

Chapitre 3

Filtrage des Signaux Aléatoires

1 Formule des moments et des interférences

On considère les filtres :

$$e \longrightarrow \boxed{\mathcal{FL}} \longrightarrow s$$

On s'interesse aux filtre linéaires :

Définition

- Un fltre linéaire conservent la linéarité des systèmes auxquels il est appliqué.
- Il est temps-invariant.
- et stationnaire.

On peux caractériser un filtre linéaire par:

- sa réponse impulsionnelle h
- sa réponse fréquentielle $H = TF[h]$
- sa fonction de transfert H_{II}.

Proposition ***(Moyenne)***

$$m_s = H(0)m_e$$

Pour deux filtres on a :

$$e_1 \longrightarrow \boxed{\mathcal{H}_1} \longrightarrow s_1$$

$$e_2 \longrightarrow \boxed{\mathcal{H}_2} \longrightarrow s_2$$

Proposition ***(Formule des interférences)***

$$\Gamma_{s_1,s_2}(f) = H_1(f) \cdot H_2(f)^* \cdot \Gamma_{e_1,e_2}(f)$$

2 Application

2.1 Blanchiement d'un signal

Pour générer un bruit blanc $s(t)$ on veux :

$$\Gamma_0 = |H(f)|^2 \Gamma_{ee}(f) \implies |H(f)|^2 = \frac{\Gamma_0}{\Gamma_{ee}(f)}$$

2.2 Identification d'un filtre linéaire

On applique en entrée un bruit blanc tel que $\Gamma_{ee}(f) = \Gamma_0$. Alors :

$$\Gamma_{se} = H(f)\Gamma_e(f) \implies H(f) = \frac{\Gamma_{se}(f)}{\Gamma_0} \propto \text{intercorrélation entrée sortie}$$

2.3 Signaux ARMA

On peux utilise un Filtre Linéaire (FL) pour définir un Signal Aléatoire. (SA). Le SA sera la sortie d'un filtre dynamique (Fonction de transfert rationnel ,stable ,causal) excité par un bruit blanc.

2.3.1 AR : autoregressif

$$\boxed{H_{II}(z) = \frac{1}{D(z)} = \frac{1}{1 - \sum_{i=1}^{q} a_i z^{-i}}}$$

Alors on aura en sortie du filtre :

$$s_k = e_k + \sum_{i=1}^{q} a_i s_{k-i}$$

on parle aussi de filtre « tout pôle »

2.3.2 MA : Moyenne ajustée

$$\boxed{H_{II}(z) = N(z) = 1 + \sum_{i=1}^{q} b_i z^{-i}}$$

Alors on aura en sortie du filtre :

$$s_k = e_k + \sum_{i=1}^{q} b_i e_{k-i}$$

2.3.3 ARMA

$$H_{II}(z) = \frac{N(z)}{D(z)}$$

On connait alors Γ_{ss} et le modèle AR. (Équation de Yule WAlker, cf TP2)

2.4 Signaux AR : illustration

pour une entrée en bruit blanc , les poles proches du cercle unités sont dominant (approche géométrique , joli dessin)

2.5 Filtre Adapté (FA)

Contexte Problème de transmission numérique (tout ou rien) d'un signal déterministe, connu avec bruit additif.

Objectif déterminer le meilleur traitement linéaire pour décider de la présence ou non d'un signal.

Exemple en TD

Méthode :Maximiser le RSB à l'instant de décision : avec $|s^f_{n_0}|^2$ puissance instantanée à l'instant de décision.

$$\boxed{\frac{|s^f_{n_0}|^2}{E[|b^f_n|^2]}}$$

Proposition ***(Application au bruit blanc)***

$$E[|b^f_n|^2] = \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^2 \Gamma_{bb}(f) df = \Gamma_0 \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^2 df$$

$$|S^f_{n_0}|^2 = \left| \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^2 S(f) e^{j2\pi n_0 f} df \right| \leq \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^2 df \int_{-\frac{1}{2}}^{\frac{1}{2}} |S(f)|^2 df$$

On a égalité si $H(f) \propto S^*(f) e^{-j2\pi n_0 f} \iff h_n \propto s^f_{n_0-n}$

LA RI du filtre est donc

- un retour temporel
- translaté autour de l'instant de décision (attention a la causalité)
- conjugué.

Remarque

- Le FA peut être non causal, la RI est alors tronqué et le filtre sous-optimal.
- Le FA est un corrélateur (d'énergie), l'objectif n'est pas de restituer le signal utile mais d'avoir le meilleur RSB à l'instant de décision.

Chapitre 4

Estimation

1 Introduction

Objectif : Présenter quelques élements de la théorue de l'estimation statistique.

1.1 Problématique

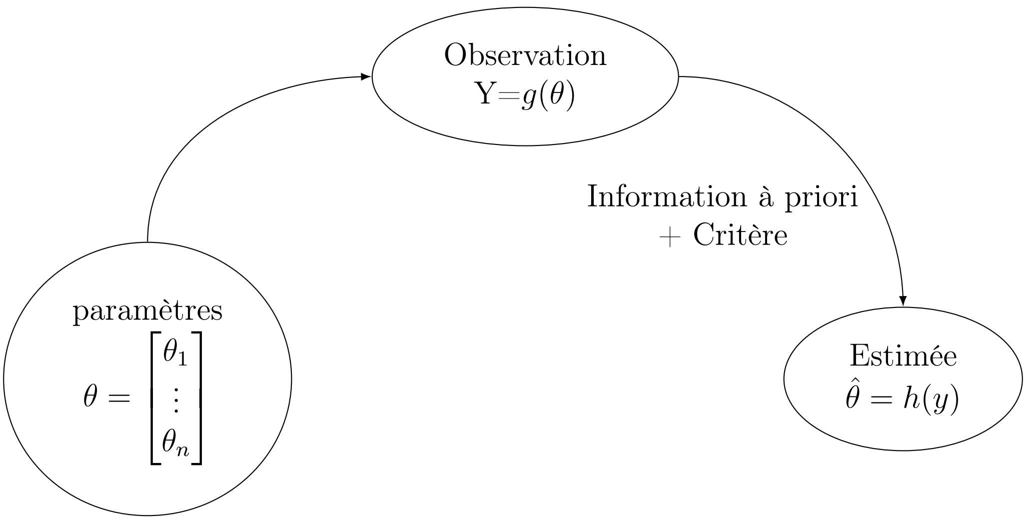


FIGURE 4.1 – Méthode d'estimation classique

Le raisonnement se transpose alors sur la figure suivante :

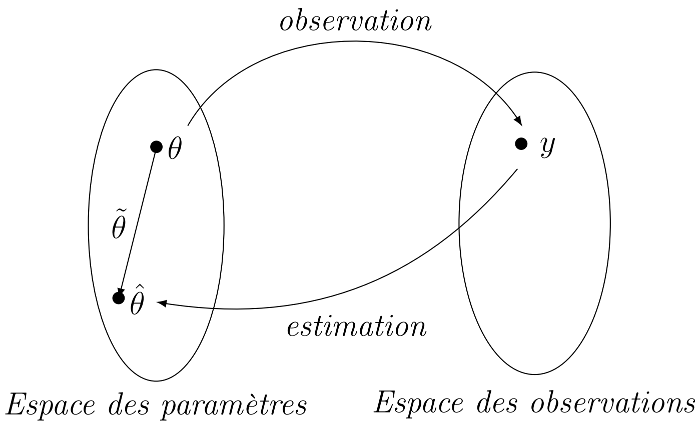


FIGURE 4.2 – Raisonnement en espace algébrique

On défini les index suivants :

m nombre d'expérience réalisée (taille de y)

n nombre de paramètres (taille de θ)

Estimateurs statistiques On observe une réalisation $y = g(\theta)$ où θ est une VA. et on détermine $\hat{\theta} = h(Y)$ estimée.

Exemple

Exemple 1 Θ tension constante.
$y(t) = \theta + b(t)$. soit $y_i = \theta + b_i$
On défini donc Y et Θ VA et on a $Y = A\Theta + B$ -> régression linéaire.

Exemple 2 filtre RC $y(t) = (1 - e^{-t/\tau})u(t) + b(t)$, $\Theta = \tau$. modèle non linéaire, traité en TD.

1.2 Performance-Qualité d'une estimation

Proposition ***(Grandeurs utiles)***

- erreur d'estimation
$$\tilde{\theta} = \hat{\theta} - \theta$$
- moment d'ordre 1 :
$$E_{Y|\Theta}[\tilde{\theta}] = E_{Y|\Theta}[\hat{\theta}] - \theta$$
- Biais moyen :
$$E[\tilde{\theta}] = E_{Y\Theta}[\tilde{\theta}] = E[\hat{\theta}] - \theta$$
- moment d'ordre 2 :
  - covariance de l'erreur d'estimation
$$C_{\tilde{\theta}\tilde{\theta}} = E[(\tilde{\theta} - m_{\tilde{\theta}})(.)^T]$$
  - Corrélation de l'erreur d'estimation
$$\Gamma_{\tilde{\theta}\tilde{\theta}} = E[\tilde{\theta}\tilde{\theta}^T]$$
  - Puissance :(Estimateur Quadratique moyen)
$$P_{\tilde{\theta}} = E[\|\tilde{\theta}\|^2] = tr(\Gamma_{\tilde{\theta}\tilde{\theta}})$$

1.3 Caractérisation des estimateurs

Définition

- Borne de Cramer Rao: borne minimale du biais de variance (qui dépend de l'estimateur choisi)
- Estimateur non biaisé: $E[\tilde{\theta}] = 0$
- Estimateur efficace: Borne de Cramer-Rao atteinte.
- Estimateur consistent: $E[\tilde{\theta}] \xrightarrow[N_{obs}\to\infty]{} 0$ et $V[\tilde{\theta}] \xrightarrow[N_{obs}\to\infty]{} 0$
- Estimateur robuste:
  Les performances de l'estimateur ne sont pas trop dégradé si on s'écarte un peu des hypothèses sous laquelle l'estimateur a été établi.
- Complexité de l'estimateur:
  sur l'o btention des connaissances et mise en oeuvre de l'estimateur.

2 Théorie classique de l'estimation

2.1 Estimateur des moindres carrés

Définition

Pour Y une VA de moyenne $m_y = m_{Y|\theta}$ on défini le critère :

$$J_{MC} = (Y - m_y)^T M (Y - m_y)$$

Avec M matrice symétrique définie positive et alors:

$$\hat{\theta}_{MC} = \arg\min_{\theta} J_{MC}(Y, \theta)$$

2.1.1 Condition nécessaire d'existance

Si $J_{MC}(y, \theta)$ est dérivable et pas de contrainte sur θ.

$$\nabla_J(\theta)|_{\hat{\theta}_{MC}} = \frac{\partial J_{MC}}{\partial \theta} = 0 \quad \text{Gradien}$$

Il faut ensuite vérifié que c'est un minimum absolu :

$$\nabla_J^2(\theta) = \frac{\partial^2 J_{MC}}{\partial \theta \partial \theta^T} > 0 \quad \text{Hessien}$$

Application $Y = A\theta + B$, avec B une VA. le critère des moindres carrés est alors :

$$J_{MC} = (Y - A\theta - m_B)^T M (Y - A\theta - m_B)$$

On a une forme quadratique positive car $A^T MA \geq 0$. (dans le cas > 0 on a une CNS sur ce qui suit)

Méthode 1

$$\nabla_J(\theta)|_{\hat{\theta}_{MC}} = 0 = -2A^T M (Y - A\theta - m_B)$$

Donc

$$A^T MA\theta = A^T M (Y - m_B)$$

Soit

$$\boxed{\hat{\theta}_{MC} = \underbrace{(A^T MA)^{-1} AM}_{D}(Y - m_B)}$$

On remarque que $DA = I_n$.

Méthode 2 Pour $A^T MA > 0$.

$$J_{MC} = \underbrace{(D(Y - m_B) - \Theta)^T A^T MA(D(Y - m_B) - \theta)}_{J_1(Y,\theta)}$$
$$+ \underbrace{(Y - m_B)^T (M - D^T A^T MAD)(Y - m_B)}_{J_2(Y)}$$

Alors $\nabla J_{MC} = 0 \implies J_1 = 0 \implies D(Y - m_B) = \hat{\theta}_{MC}$

2.1.2 Caractéristique de l'estimateur

- Estimateur non biaisé

$$\begin{aligned}\tilde{\theta}_{MC} &= \hat{\Theta} - \theta \\ &= D(Y - m_B) - \theta \\ &= D(B - m_B)\end{aligned}$$

  Donc $E[\hat{\theta_{MC}}] = 0$
- moment d'ordre 2 :

$$C_{\tilde{\theta}\tilde{\theta}} = E[(\tilde{\theta} - m_{\tilde{\theta}})(.)^T] = DE[(B - m_B)(B - m_B)^T]D^T = DC_{BB}D^T$$

  - Cas MC ordinaire ($M = I_n$)

$$C_{\tilde{\theta}\tilde{\theta}} = (A^T A)^{-1} A^T C_{BB} A(A^T A)^{-1}$$

  - Cas MC pondéré ($M = C_{BB}^{-1}$)

$$C_{\tilde{\theta}\tilde{\theta}} = (A^T C_{BB}^{-1} A)^{-1}$$

- Cas θ scalaire $Y_i = \theta + B_i$ donc :

$$C_{BB} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{bmatrix} \text{ et } A = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

  - Cas MCO : $A^T A = m$

$$\hat{\theta_{MC}} = \frac{\Sigma(y_i - m_{bi})}{m} \quad \text{et} \quad \sigma_{\tilde{\theta}}^2 = \frac{\Sigma\sigma_i^2}{m^2}$$

  - cas MCP pour $M = C_{BB}^{-1} = diag(\sigma_1^{-2}, \ldots, \sigma_m^{-2})$

$$A^T C_{BB} A = \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \quad \text{donc} \quad \hat{\theta}_{MCP} = \frac{1}{\sum \frac{1}{\sigma_i^2}} \sum \frac{Y_i - mB_i}{\sigma_i^2}$$

    - $\hat{\theta_{MCP}}$ défini un barycentre

  - Pour $\sigma_i = \sigma$ on a $M = \sigma I \implies MCO = MCP$
- Comparaison MCO et MCP (avec $M = C_{BB}$)

$$\begin{aligned} \sigma^2_{MCO} &\leq \sigma^2_{MCP} \\ \frac{1}{\sum \sigma_i^{-2}} &\leq \frac{1}{m^2} \sum \sigma_i^2 \\ m^2 &\leq \frac{1}{\sum \sigma_i^{-2}} \sum \sigma_i^2 \end{aligned}$$

2.2 Estimateur du maximum de vraisemblance

Définition

On considère $f_Y(y)$ ddp de y paramétrée par θ. On a $f_{Y|\theta}(y) = V(Y, \theta)$. on pose également $L(Y, \theta) = \ln(V(Y, \theta))$.
on défini alors:

$$\hat{\theta}_{MV} = \arg\min f_{Y|\theta}(y) = \arg\min L(Y, \theta)$$

GRAPHE

Exemple Modèle avec bruit additif gaussien.

Proposition

Dans le cas d'un brui Gaussien et pour $M = C_{BB}^{-1}$

$$\hat{\theta}_{MCP} = \hat{\theta}_{MV}$$

Remarque L'estimateur de MV n'est pas nécessairement efficace mais si un estimateur sans biais existe et est efficace c'est celui-ci.

Si $m \to \infty$ on montre que le MV est asymptotiquement efficace. (loi des grands nombres)

3 Théorie générale de l'estimation

3.1 Estimateur linéaire en moyenne quadratique (ELMQ)

Définition

Un ELMQ fourni une estimée de la forme

$$\hat{\theta} = HY + C$$

à partir de l'erreur quadratique moyenne $E[\|\tilde{\theta}\|^2] = E[\tilde{\theta}\tilde{\theta}^T] = P_{\tilde{\theta}}$

Concept H et C tel que $P_{\tilde{\theta}}$ minimal.

$$(1) \quad \frac{\partial P_{\tilde{\theta}}}{\partial H} = 0 \quad \text{et} \quad (2) \quad \frac{\partial P_{\tilde{\theta}}}{\partial C} = 0$$

Proposition

1)

$$\frac{\partial P_{\tilde{\theta}}}{\partial H} = 2E[HY + C - \theta] = 2E[\tilde{\theta}] = 0$$

L'ELMQ est un estimateur non biaisé.

et donc :

$$\begin{aligned} C &= -Hm_Y + m_\theta \\ \hat{\theta} &= H(Y - m_y) + m_\theta \\ \tilde{\theta} &= H(Y - m_y) - (\theta - m_\theta) \end{aligned}$$

Proposition

2)

$$\frac{\partial P_{\tilde{\theta}}}{\partial C} = 2E[(HY + C - \theta)Y^T] = 2E[\tilde{\theta}Y^T] = 0$$

$\tilde{\theta} \perp Y$ quand la puissance est minimale, $\tilde{\theta}$ et Y sont décorrélées, on a extrait toute l'information commune.

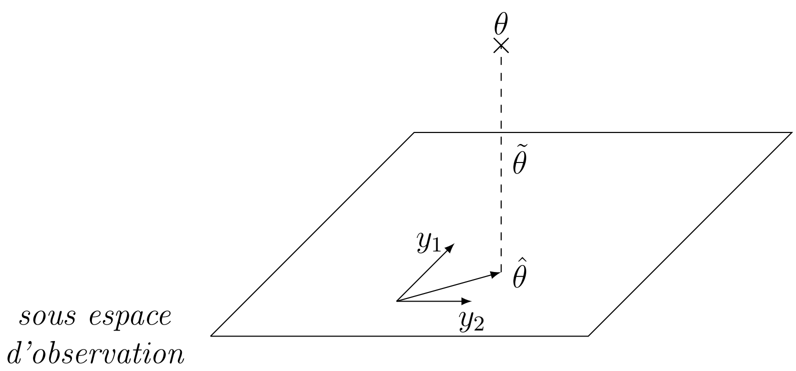


Figure 4.3 – Représentation des paramètres

De plus :

$$\begin{aligned} E[\tilde{\theta}Y^T] &= E[\tilde{\theta}(Y - m_Y)^T] \\ &= E[(H(Y - m_Y) - \theta - m_\theta)(Y - m_y)^T] \\ &= HC_{yy} - C_{\theta Y} = 0 \implies H = C_{\theta Y}C_{YY}^{-1} \end{aligned}$$

on a donc

$$\boxed{\hat{\theta} = C_{\theta Y}C_{YY}^{-1}(Y - m_Y) + m_\theta}$$

Remarque L'ELMQ nécessite des connaissances du premier et du second ordre sur θ et Y.

Proposition

$$C_{\tilde{\theta}\tilde{\theta}} = C_{\theta\theta} - C_{\theta Y} C_{YY}^{-1} C_{Y\theta}$$

La corrélation entre θ et Y permet de diminuer l'ELMQ.

3.2 Estimateur Bayésiens

3.2.1 Fonction coût/pénalité

Définition

On appelle fonction de coût ou fonction de pénalité une fonction qui mesure l'erreur entrainée par la prise de la valeur $\hat{\theta}$ pour θ.

$$C(\hat{\theta}, \theta) \geq 0 \quad \text{ou encore} \quad C(\tilde{\theta}) \geq 0$$

On prendra le plus souvent une « bonne » fonction (continue, paire , croissante ...)

Exemple de coût on représente les fonctions de coût usuelles :

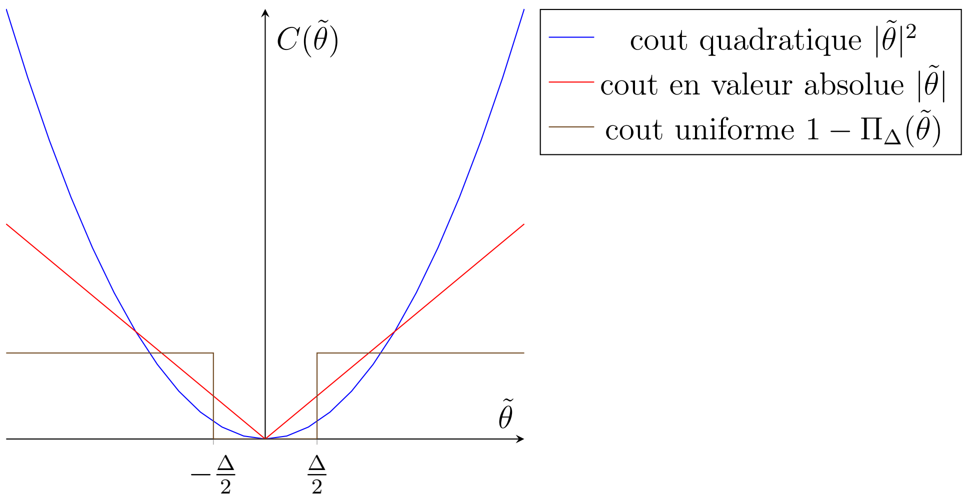


FIGURE 4.4 – Représentation des fonctions de coût classique

Définition

On appelle estimateur bayésiens l'estimateur qui minimise le coût moyen :

$$E_{\theta,Y}[C(\hat{\theta},\theta)] = \int_{\mathbb{R}^{m+n}} C(\hat{\theta},\theta) f_{\theta Y}(\theta,y) d\theta dy$$
$$= \int_{\mathbb{R}^m} \left(\underbrace{\int_{\mathbb{R}^n} C(\hat{\theta},\theta) f_{\theta|Y=y}(\theta) d\theta}_{E_{\theta|Y}[C(\hat{\theta},\theta)]} \right) f_Y(y) dy$$

On minimise donc $E_{\theta|Y}[C(\hat{\theta},\theta)]$ à coût conditionnel donné

$$\hat{\theta}_B = \arg\min_{\hat{\theta}} E_{\theta|Y}[C(\hat{\theta},\theta)]$$

3.2.2 Estimateur du maximum a posteriori (MAP)

On considère un cout uniforme.

Définition

En prenant:

$$E_{\theta|Y}[C(\hat{\theta},\theta)] = \int_{\mathbb{R}^m} (1 - \Pi_\Delta(\tilde{\theta})) f_{\theta|Y=y}(\theta) d\theta$$
$$= 1 - \int_{\hat{\theta}-\Delta/2}^{\hat{\theta}+\Delta/2} f_{\theta|Y=y}(\theta) d\theta$$
$$\simeq 1 - \Delta^n f_{\theta|Y=y}(\hat{\theta})$$

Soit

$$\hat{\theta}_{MAP} = \arg\max_{\theta} f_{\theta|Y=y}(\theta)$$

Lien MAP-MV on a $f_{\theta|Y=y}(\theta) f_Y(y) = f_{\theta Y}(\theta,y)$. Avec $f_\theta(\theta) = C^{ste}$ quand $f_{\theta Y}(\theta,y)$ à une valeur significative (ie $C_{\theta\theta}$ grand / σ_θ grand) alors :

$$\arg\max f_{\theta|Y=y}(\theta) \simeq \arg\max f_{Y|\Theta=\theta}(y)$$

On considère alors que θ est un paramètre aléatoire mais très mal connu. (ddp uniforme sur un interval tres grand, peu d'infos sur θ).

cf. TD « file d'attente »

Exemple et Application On considère θ scalaire aléatoire avec : $Y_i = \theta + B_i$ Avec :

$$\begin{cases} B \hookrightarrow \mathcal{N}(0, C_{BB}) \\ \Theta \hookrightarrow \mathcal{N}(m_\theta, \sigma_\theta^2) \\ B \perp \Theta \end{cases}$$

Rappel MC=MV avec : $\begin{cases} m_B = 0 \\ \hat{\theta}_{MV} = \hat{\theta}_{MC} = \frac{\sum_{i=1}^m Y_i}{m} \\ E[\hat{\theta}_{MV}] = E[\theta] = m_\theta \text{ et } \sigma_{\tilde{\theta}_{MV}} = \frac{\sigma_B}{m} \end{cases}$

On a donc :

$$f_{Y|\theta}(y) = f_B(Y - A\theta) = \prod_{i=1}^m f_{B_i}(Y_i - \theta) = C_1 \exp\left(-\frac{1}{2}\frac{\sum(Y_i - \theta)^2}{\sigma_B^2}\right)$$

Or

$$f_{\theta|Y=y}(\theta) = \frac{f_{Y|\theta}(y) f_\theta(\theta)}{f_Y(y)} = C_2 \exp\left(-\frac{1}{2}\underbrace{\left[\frac{\sum(Y_i - \theta)^2}{\sigma_B^2} + \frac{(\theta - m_\theta)^2}{\sigma_\theta^2}\right]}_{J_{MAP}}\right)$$

Le critère est ici une forme quadratique, donc :

$$\hat{\theta}_{MAP} = \arg\max f_{\theta|Y=y}(\theta) = \arg\min J_{MAP}(\theta, Y)$$

Alors on a la CNS :

$$\frac{\mathrm{d}J_{MAP}}{\mathrm{d}\theta} = 0 = 2\left[-\sum_{i=1}^m \frac{Y_i - \theta}{\sigma_b^2} + \frac{(\theta - m_\theta)^2}{\sigma_\theta^2}\right]$$

Soit une expression barycentrique :

$$\hat{\theta}_{MAP} = \frac{\frac{m}{\sigma_B^2}\sum\frac{Y_i}{m} + \frac{m_\theta}{\sigma_\theta^2}}{\frac{m}{\sigma_B^2} + \frac{1}{\sigma_\theta^2}}$$

Donc :

Proposition

$$E[\hat{\theta}_{MAP}] = m_\theta$$

L'estimateur est non biaisé. De plus :

$$\sigma_{\tilde{\theta}_{MAP}}^2 = \frac{1}{\frac{1}{\sigma_{MV}} + \frac{1}{\sigma_\theta^2}} < \begin{cases} \sigma_\theta^2 \\ \sigma_{MV}^2 \end{cases}$$

On a fait mieux en prenant en compte toutes les sources d'informations.

Remarque

- Si $\sigma_\theta >> \sigma_{MV}$ alors $\hat{\theta}_{MAP} \simeq \hat{\theta}_{MV}$ (ce qui arrive pour σ_B ou m grand)
- Si $\sigma_\theta << \sigma_{MV}$ et $\hat{\theta}_{MAP} \simeq m_\theta$ (l'obersavation apporte peu d'info)

3.2.3 Estimateur en moyenne quadratique (EQM)

Définition

On le cout moyen de l'EQM:

$$C(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^T M(\hat{\theta} - \theta)$$

Avec $M > 0$. On cherche a minimiser le cout moyen mais sans contrainte de linéarité avec une matrice de pondération qui peux prendre en compte des facteurs d'echelles ou des unités différentes.

Etude de l'estimateur On veut minimiser $E_{\Theta|Y}[C(\hat{\theta}, \theta)]$

$$\begin{aligned}
\nabla_{\hat{\theta}} E_{\Theta|Y}[C(\hat{\theta}, \theta)] &= 0 \\
E_{\theta|Y}[2M(\hat{\theta} - \theta)] &= 0 \\
2ME_{\theta|Y}[\underbrace{\hat{\theta}}_{h(y)}] - E_{\theta|Y}[\theta] &= 0 \\
2M(\hat{\theta} - E_{\theta|Y}[\theta]) &= 0 \\
\boxed{\hat{\theta}_{MQ} = E_{\theta|Y}[\theta]} & \\
&= \int_{\mathbb{R}^n} \theta f_{\theta|y}(\theta) d\theta = h(Y = y)
\end{aligned}$$

Par conséquent : $E[\hat{\theta}_{MQ}] = E[\theta]$. on a un estimateur non biaisé.

Remarque Si $f_{\theta|Y}$ possède un axe de symétrie (ex : gaussienne) :
FIGURE . ($\hat{\theta}_{MQ} = \hat{\theta}_{MAP}$ dans le cas gaussien. Différent avec deux bosses.)
Dans le cas général la contrainte de linéarité pour l'ELMQ conduit à une valeur plus grande qu'avec l'EQM. Dans le cas gaussien : $\hat{\theta}_{ELMQ} = \hat{\theta}_{MQ}$, mais $\hat{\theta}_{MQ}$ nécessite plus de connaissance (ddp).

3.2.4 Estimateur en valeur absolu

Définition

on s'interesse au cas $n = 1$ (un paramètre) On choisit le cout moyen :

$$C(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$$

Alors :

$$E_{\theta|Y}[C(\hat{\theta}, \theta)] = \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) f_{\theta|Y=y}(\theta) d\theta - \int_{\hat{\theta}}^{+\infty} (\hat{\theta} - \theta) f_{\theta|Y=y}(\theta) d\theta$$

Donc :

$$\begin{aligned} 0 =& \nabla_{\hat{\theta}} E_{\theta|Y}[C(\hat{\theta}, \theta)] \\ =& \dots \\ =& \int_{-\infty}^{\hat{\theta}} f_{\theta|Y=y}(\theta) d\theta - \int_{\hat{\theta}}^{+\infty} f_{\theta|Y=y}(\theta) d\theta \end{aligned}$$

Proposition

L'estimée est alors $\hat{\theta}_{VA}$ tel que :

$$\int_{-\infty}^{\hat{\theta}} f_{\theta|Y=y}(\theta) d\theta = \int_{\hat{\theta}}^{+\infty} f_{\theta|Y=y}(\theta) d\theta$$

On parle de médiane a posteriori. Le résultat se généralise pour tout n.

Remarque Dans le cas où $f_{\theta|Y=y}(\theta)$ possède un axe de symétrie (ex gaussienne) on a :

$$\hat{\theta}_{VA} = \hat{\theta}_{MV} \overset{\overset{\max}{\downarrow}}{=} \hat{\theta}_{MAP}$$

Exemple Localisation d'un véhicule / Ellipsoïde de confiance (cf poly).

4 Conclusion

- L'estimateur statistique dépend des connaissances a priori, de la complexité des calculs et de la robustesse attendue.
- Dans certains cas particuliers/ limites on retrouve des estimateurs intuitifs /empirique.
- La loi normale joue un rôle important (hypothèses qui se justifie par la loi des grands nombres) : les calculs sont simplifiés et conduisent au même résultat.