Fiche de révision Méthodes d'estimation et Théorie moderne du Codage

Alexis GRACIAS

7 décembre 2024

Table des matières

Introduction

Les objectifs du cours sont de présenter les bases de la théorie de la décision :
- **L'estimation pure**
- **La détection**

Plan du cours :

- Notions d'estimation
  - Introduction
  - Estimateurs bayésiens (espaces de Hilbert, projection orthogonale, estimation moyenne quadratique avec contrainte linéaire)
  - Estimateurs non bayésiens (Inégalité de Cramer-Rao, maximum de vraisemblance)
- Estimation d'un signal dans un bruit auditif
- Analyse spectrale non paramétrique
- Détection
  - Test des hypothèses (théorie bayésienne, stratégie de Nayman-Person, courbes COR)
  - Application à la détection du signal dans un bruit (décomposition de Karhunen-Loève, détection d"un signal déterminisme dans un bruit gaussien)
- Détection
- Filtrage linéaire statistique
  - Introduction
  - Filtrage de Wiener
  - Filtrage de Wiener avec filtrage linéaire
- Prédiction à un pas et passé infini
  - Cas d'un signal dont la densité spectrale de puissance est bornée et admet une factorisation forte
  - Cas général, décomposition de Wold
- Interpolation d'un signal stationnaire
- Prédiction à un pas passé et infini
- Primitives de la théorie de l'information
  - Introduction (source d'information discrète, canal discret, message)
  - Deux problèmes clés de codage (codage canal, codage source distribuée)
  - Théorèmes fondamentaux (codage aléatoire (random coding), compartimentage aléatoire (random binning))
- Exercices de la théorie de l'information (2H de TD)

Prérequis :

— Cours de CIP
— Cours de Traitement du Signal
— Cours de Statistique et Apprentissage
— COurs de SIP

Chapitre 1

Notions d'estimation

1.1 Rappels des cours de 1A

1.1.1 Espace de Hilbert

Definition 1: Espaces de Hilbert

Un espace de Hilbert est un **espace préhilbertien complet**, c'est à dire un **espace de Banach** dont la norme $||.||$ découle d'un **produit scalaire** ou **hermitien** par la formule suivante :

$$||x|| = \sqrt{\langle x, x\rangle} \tag{1.1}$$

Definition 2: Espaces de Banach

Un espace de Banach est un **espace vectoriel complet et normé** sur un sous corps $\mathbb{K}$ de $\mathbb{C}$ ($\mathbb{R}$ ou $\mathbb{C}$), peut importe la norme.

En d'autres termes, un espace de Hilbert est un espace vectoriel complet muni d'une norme définie par un produit scalaire.

1.1.2 Espace Lp

Definition 3: Espaces $L^p[a, b]$

Espace vectoriel des fonctions p intégrables au sens de **Lebesgue** sur $[a, b]$. i.e.

$$\int_a^b |f(t)|^p dt \tag{1.2}$$

converge, avec la norme L^p :

$$||f||_p = \sqrt[\frac{1}{p}]{\int_a^b |f(t)|^p dt} \tag{1.3}$$

1.1.3 Précisions

Definition 4: Espace vectoriel normé

Un K-espace vectoriel E est dit normé si il est muni d'une norme, c'est-à-dire d'une application $\mathcal{N} : E \to \mathbb{R}^+$ qui satisfait les conditions suivantes :

- **Séparation**
$$\forall x \in E, \mathcal{N}(x) = 0 \implies x = 0_E \tag{1.4}$$
- **Homogénéité**
$$\forall (x, \lambda) \in E \times \mathbb{R}, \mathcal{N}(\lambda x) = |\lambda| \mathcal{N}(x) \tag{1.5}$$
- **Sous-additivité (inégalité triangulaire)**
$$\forall (x, y) \in E^2, \mathcal{N}(x + y) < \mathcal{N}(x)\mathcal{N}(y) \tag{1.6}$$

Definition 5: Espace métrique complet

Un espace métrique (E, d) est dit complet si toute suite de Cauchy [a] converge dans ce même espace, c'est-à-dire :
$$\forall n \in \mathbb{N}, (x_n)_{n\in\mathbb{N}} \in E, x_n \longrightarrow l \in E \tag{1.8}$$

[a] Suite qui vérifie le critère de Cauchy, c'est-à-dire que les éléments de la suite se rapprochent uniformément entre-eux à l'infini, i.e.
$$\forall \epsilon > 0, \exists (n_0, p_0) \in \mathbb{N}^2, \forall n > n_0, \forall p > p_0, d(x_n, x_p) < \epsilon \tag{1.7}$$

Definition 6: Espace métrique

On note (E, d) un espace métrique (E ensemble et d la distance définie pour tout éléments de E). C'est un espace vectoriel au sein duquel la notion de distance est bien définie pour tout éléments de E. L'application d satisfait les conditions suivantes :

- **Symétrie**
$$\forall (x, y) \in E^2, d(x, y) = d(y, x) \tag{1.9}$$
- **Séparation**
$$\forall (x, y) \in E^2, d(x, y) = 0 \Longleftrightarrow x = y \tag{1.10}$$
- **Inégalité triangulaire**
$$\forall (x, y, z) \in E^3, d(x, y) < d(x, z) + d(z, y) \tag{1.11}$$

1.1.4 Probabilités

Definition 7: Espaces probabilisé

Un espace probabilisé est constitué d'un **espace probabilisable** et d'une **mesure de probabilité**, noté $(\Omega, \mathcal{A}, \mathbb{P})$ avec :

- Ω : l'univers, l'espace des observations ou espace des évènements élémentaires
- $\mathcal{A}$ est une **tribu** sur Ω
- $\mathbb{P}$: mesure de probabilité

tel que $\mathbb{P}(\Omega) = 1$ et $\forall A \in \mathcal{A}$, $\mathbb{P}(A)$ est appelé *probabilité de l'évènement A*

Definition 8: Espaces probabilisables

Un espace probabilisable est noté $(\Omega, \mathcal{A})$, il est constitué de **l'univers** et de la **tribu** de cet univers.

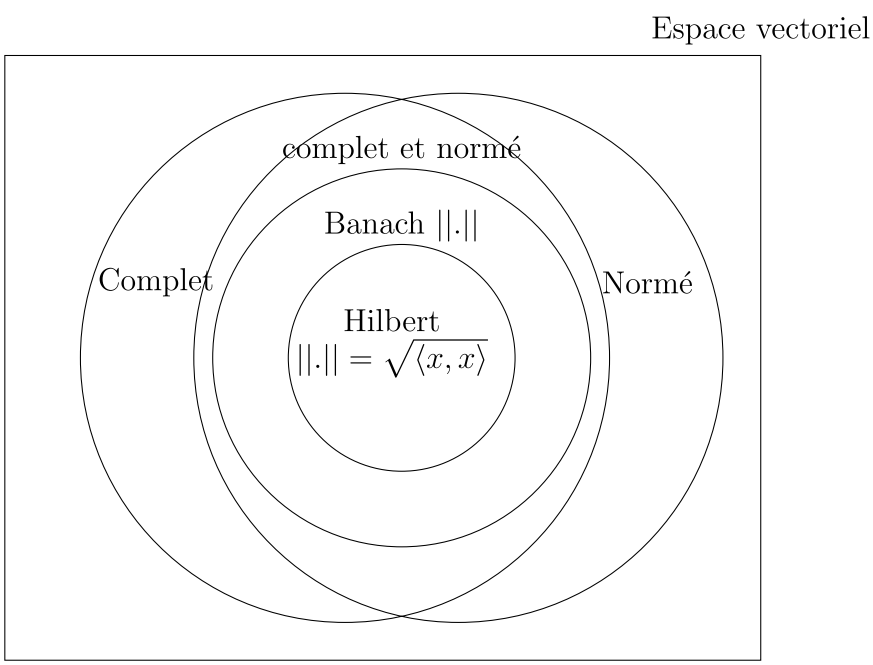
Definition 9: Mesure de probabilité

La mesure de probabilité est définie par l'application $\mathbb{P} : \Omega \longrightarrow [0; 1]$ telle que :

- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}(\{\}) = 0$
- σ**-additivité** : $\forall$ collection dénombrable $\{A_i\}$ d'ensemble disjoints :

$$\mathbb{P}(\bigcup_{I \in \mathcal{I}} A_i) = \sum_{I \in \mathcal{I}} \mathbb{P}(A_i) \tag{1.12}$$

1.1.5 Schématiquement



1.2 Estimation

1.2.1 Notions et définitions de base

On cherche à reconstituer un signal $\underline{x}$ ou $\underline{\theta}$ à partir de son observation, le vecteur $\underline{y}$. On cherche alors une fonction $\widehat{\theta}$ tel que $x \circ y = \widehat{\theta}(\underline{y})$ soit la meilleure estimation de $\underline{\theta}$.

Mathématiquement, on a :

$$\underline{\theta} = \begin{bmatrix}\theta_0 & \theta_1 & \dots & \theta_n\end{bmatrix}, (\theta_i)_{i\in\mathbb{N}} \in G$$
$$\underline{y} = \begin{bmatrix}y_0 & y_1 & \dots & y_n\end{bmatrix}, (y_i)_{i\in\mathbb{N}} \in F$$

Avec :

- $\underline{\theta}$: **paramètre décisionnel**
- $\underline{y}$: **vecteur d'observation**
- $(\theta_i)_{i\in\mathbb{N}}$, $(y_i)_{i\in\mathbb{N}}$: **variables aléatoires** de G et F
- $(G, \mathcal{G})$ et $(F, \mathcal{F})$ sont des espaces probabilisables

Estimateur
Pour un **estimateur** $\widehat{\theta}$, le meilleur estimateur (ou filtre) de θ est donc l'application $\widehat{\theta}$:

$$\boxed{\widehat{\theta} : F \longrightarrow G, \widehat{\theta} \circ y = \widehat{\theta}(\underline{y}) \in \mathcal{M}(F, G)}$$

Avec $\mathcal{M}(F, G)$ l'ensemble des applications de F dans G.

Densité de probabilité conditionnelle
(Ω, ϵ, P) est un espace probabilisé qui modélise l'expérience aléatoire (signaux aléatoires rencontrés) $\underline{y}$ est le vecteur d'observation qui admet une **densité** par rapport à F, de **mesure canonique**[1] ν, tandis que $\widehat{\theta}(\underline{y})$ est une VA à valeurs dans G, alors la **densité de probabilité conditionnelle de $\underline{y}$ sachant θ** s'exprime sous la forme :

$$\boxed{f_{\underline{y}|\theta}(\underline{\nu}|\theta), \forall \nu \in F}$$

Fonction de perte ou de coût : application L telle que :

$$L : G \times G \longrightarrow \mathbb{R}, (\theta_1, \theta_2) \mapsto L[\theta_1, \theta_2]$$

Pour un estimateur $\widehat{\theta}$ et $\nu \in F$ donnés, $L[\widehat{\theta}(\underline{\nu}), \theta]$ représente le **coût de la décision** $\widehat{\theta}(\nu)$ quand la vraie valeur du paramètre décisionnel est θ. Ce coût est généralement arbitraire.
Fonction de risque

$$\mathcal{R}(\widehat{\theta}, \theta) = E[L(\widehat{\theta}(\underline{y}), \theta)|\theta] = \int_F L[\widehat{\theta}(\underline{\nu}), \theta] f_{\underline{y}|\theta}(\underline{\nu}|\theta) d\underline{\nu}$$

1. Mesure canonique que l'on prend comme mesure de Lebesgue

1.2.2 Point de vu Bayésien et non Bayésien

- **Non Bayésien** : θ est supposé déterministe, c'est à dire qu'il obéit à des lois non probabilistes

- **Bayésien** : θ est supposé aléatoire et de loi connue

- **Paramétrique** : θ est aléatoire et de loi connue (i.e. Bayésien sans supposition)

- **Non Paramétrique** : θ est aléatoire et de loi inconnue.

Le point de vu **fréquentiste**, i.e. non paramétrique ne sera étudié dans le cadre de ce cours. Le point de vu Bayésien permet d'introduire d'autres notions telles que :

Risque moyen
pour une **densité de probabilité a fortiori** $f_\theta(u)$ du paramètre décisionnel, le risque moyen s'exprime sous la forme :

$$\overline{R}(\widehat{\theta}) = E[R(\widehat{\theta}), \theta] = \int_G R(\widehat{\theta}, u) f_\theta(u) du = \int_F H_c(\underline{\nu}) f_y(\underline{\nu}) d\underline{\nu} \tag{1.13}$$

Avec

$$H_c(\underline{\nu}) = \int_G L[\widehat{\theta}(\underline{\nu}), u] f_{\theta|y}(u|\underline{\nu}) du = E\{L[\widehat{\theta}(\underline{y}), \theta] | \underline{y} = \underline{\nu}\} \tag{1.14}$$

La ddp conditionnelle $f_{\theta|y}(u|\underline{\nu})$ devient la **densité de probabilité a posteriori** et la loi de $\underline{y}$ est **l'evidence**