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

Alexis GRACIAS

15 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**

$$\forall (x, y) \in E^2, d(x, y) = 0 \iff x = y \tag{1.10}$$

— **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.

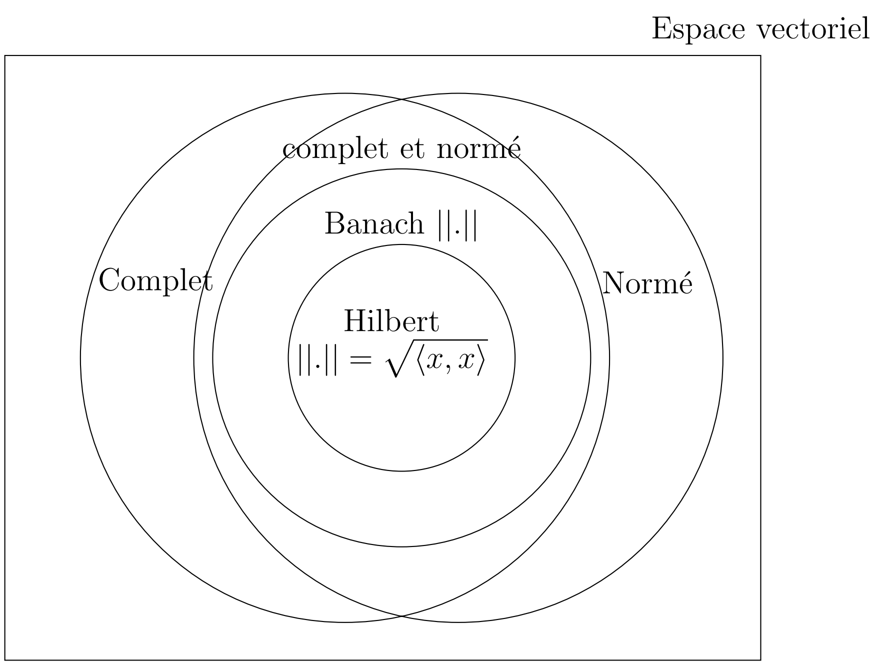
Definition 9: Mesure de probabilité

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 Transformée de Fourrier continue

1.3 Transformée de Fourrier discrète

1.4 Estimation

1.4.1 Notions et définitions de base

On cherche à reconstituer un signal aléatoire $\underline{x}$ ou $\underline{\theta}$ à partir de son observation, le vecteur $\underline{y}$. Par conséquent, les éléments y_i du vecteur d'observation $\underline{y}$ sont des **VA** sur un espace probabilisé. On cherche alors une fonction $\widehat{\theta}$ tel que $x \circ y = \widehat{x}(\underline{y})$ soit la meilleure estimation de $\underline{x}$.

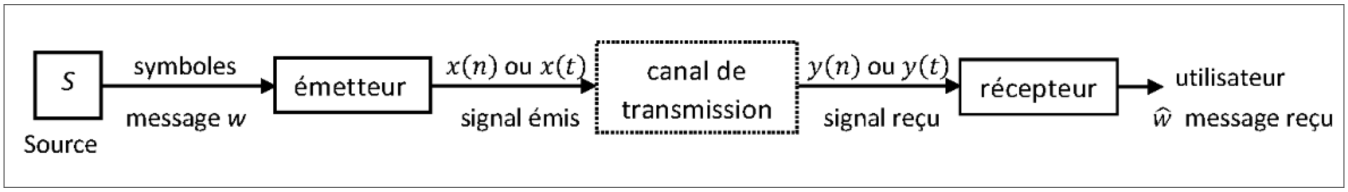


FIGURE 1.1 – Modèle de transmission de Shanon

Mathématiquement, on a :

$$\underline{x} = \begin{bmatrix} x0 & x_1 & \dots & x_n \end{bmatrix} \in G$$
$$\underline{y} = \begin{bmatrix} y_0 & y_1 & \dots & y_n \end{bmatrix} \in F$$

Avec :

— $\underline{x}$: **paramètre décisionnel** (aussi noté $\underline{\theta}$) : c'est le paramètre sur lequel est pris la décision. Il est peut être déterministe ou aléatoire.

— $\underline{y}$: **vecteur d'observation**. (Ω, ϵ, P) est un espace probabilisé qui modélise l'expérience aléatoire permettant d'observer les signaux aléatoires rencontrés (i.e. $\underline{x}$). Par conséquent, $\underline{y} : \Omega \longrightarrow F$, soit $\underline{y} \in \mathcal{M}(\Omega, F)$ [1]

— $(x_i)_{i \in \mathbb{N}}$ et $(y_i)_{i \in \mathbb{N}}$: **variables aléatoires** de $\underline{x}$ et de $\underline{y}$

— $(G, \mathcal{G})$ et $(F, \mathcal{F})$ sont des espaces probabilisables. F est **l'espace des observations** et G **l'esapce des paramètres**

1.4.2 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 **mesurables** de F dans G.

1. i. e. $\underline{y}$ est une **VA** $\Longleftrightarrow$ $\underline{y}$ est une fonction qui à tout évènement $\omega \in \Omega$ associe une valeur $y_i \in F$

1.4.3 Densité de probabilité conditionnelle - vraisemblance

(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**[2], 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 :

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

On supposera dans le cours que la vraisemblance est connue (i.e. la loi de la VA est connue)

1.4.4 Fonction de perte ou de coût

C'est une application L telle que :

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

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

1.4.5 Fonction de risque

Fonction déterministe qui dépend de $\widehat{\theta}$ et θ. C'est l'espérance de la perte entre la valeur du paramètre θ estimée et sa vraie valeur, sachant que θ vaut tant.

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

Ici, l'objectif est de trouver une fonction $\widehat{\theta}$ qui minimise le risque, c'est-à-dire une fonction $\widehat{\theta}$ qui se rapproche le plus possible de sa vraie valeur.

1.5 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

2. Mesure canonique que l'on prend comme mesure de Lebesgue.
On a, pour $\underline{y} \in \mathcal{M}(\Omega, F)$ de mesure $d\underline{v} = d\mu(\underline{v})$, que $\forall y_i \in F, \mu(y_i) = P(\underline{y}^{-1}(y_i))$. De même, on note $d\underline{u} = dv(\underline{u})$ pour la mesure de l'espace décisionnel $\mathcal{G}$

— **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 :

1.5.1 Risque moyen - θ Bayésien

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

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

Avec

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

La ddp conditionnelle $f_{\theta|y}(u|\underline{v})$ devient la **densité de probabilité a posteriori** et la loi de $\underline{y}$ est **l'evidence**. Le but est ici de trouver le $\underline{v} \in F$ qui minimise $H(\underline{v})$

Exemple - coût quadratique
Supposons $G = \mathbb{R}, L[\widehat{\theta}, \theta] = (\widehat{\theta} - \theta)^2$. Pour $\underline{v} \in F$, calculer le risque conditionnel :

$$\begin{aligned} H(\underline{v}) &= E[(\widehat{\theta} - \theta)^2 | \underline{y} = \underline{v}] \\ &= \widehat{\theta}^2(\underline{v}) - 2\widehat{\theta}(\underline{v}) E[\theta | \underline{y} = \underline{v}] + E[\theta^2 | \underline{y} = \underline{v}] \\ &= (\widehat{\theta}(\underline{v}) - E[\theta | \underline{y} = \underline{v}])^2 + E[(\theta - E[\theta|\underline{y}])^2 | \underline{y} = \underline{v}] \end{aligned}$$

On obtient une équation quadraitque en $\widehat{\theta}(\underline{y})$ à résoudre. Le minimum de H est atteint en

$$\widehat{\theta}(\underline{v}) = E[\theta | \underline{y} = \underline{v}] = \int_G u f_{\theta|\underline{y}}(u|\underline{v} du)$$

1.5.2 Risque moyen - θ non Bayésien

D'un **point de vue non Bayésien**, on utilise plutôt le maximum de vraisemblance :

$$\widehat{\theta}(\underline{y}) = arg(max f_{\underline{y}|\theta}(\underline{v}|\theta))$$

En effet, pour θ déterministe, on définit :
— **Le biais** : $B(\widehat{\theta}) = E[\widehat{\theta}] - \widehat{\theta}$
— **La variance** : $V(\widehat{\theta}) = E[\widehat{\theta}^2] - E[\widehat{\theta}]^2$

1.6 Estimation en moyenne quadratique (EMQ)

1.7 Estimation linéaire en moyenne quadratique (ELMQ)

D'un point de vue Bayésien, on prend un coût quadratique et on suppose $\underline{x}, \underline{y} \in L^2(P) \times L^2(P)^n$ avec :

$$\underline{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^H$$

[3]. Le problème est de trouver l'estimateur $a^H : \mathbb{C}^n \longrightarrow \mathbb{C}$ tel que :

$$\widehat{x}_L = a^H \underline{y}$$

soit la meilleure estimation de $\underline{x}$ et a^H désigne la transposée de la matrice a dans le **sous-espace des observations filtrées avec contraintes linéaires** $H_L[\underline{y}]$ tel que :

$$\boxed{H_L[\underline{y}] := \{z \in L^2(P), \exists \underline{\omega} \in \mathbb{C}^n, z = \underline{\omega}^H \underline{y}\}}$$

$\widehat{x_L} \in H_L[\underline{y}]$ est l'élément qui minimise la distance à $\underline{x}$.

3. H est ici l'opérateur transpose définit sur H_L, le sous-espace des observations filtrées avec contrainte, qui sera définit juste après.

1.8 Estimation affine en moyenne quadratique

1.9 Estimations en moyenne quadratique avec contraintes linéaires

1.10 Estimation non-Bayésienne