

Modélisation et Estimation pour les Signaux et Systèmes

1^{ère} partie

Hervé CARFANTAN, Herve.Carfantan@irap.omp.eu

Institut de Recherche en Astrophysique et Planétologie – IRAP

Chap. I Introduction à l'estimation

Chap. II Modélisation des Signaux et Systèmes

Chap. III Construction d'un estimateur

Chap. IV Calcul des estimateurs

2^{ème} partie

Patrick DANÈS, Patrick.Danes@laas.fr

Laboratoire d'Analyse et d'Architecture des Systèmes – LAAS

Filtrage de Kalman.

Master 1 EEA parcours SIA2 et IdS parcours IM

2024-2025

Université de Toulouse

CHAP I INTRODUCTION À L'ESTIMATION

- 1/ Introduction à la modélisation et à l'estimation pour les signaux et systèmes
- 2/ Définition d'un estimateur
- 3/ Rappels d'algèbre linéaire et de calcul matriciel
- 4/ Rappels d'outils statistiques
- 5/ Propriétés des estimateurs

I.1 INTRODUCTION

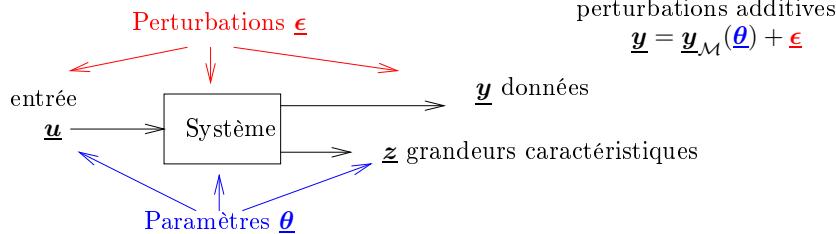
■ Cadre de travail

On dispose de données mesurées \underline{y} dépendant d'un paramètre d'intérêt $\underline{\theta}$

↓

Comment déterminer $\underline{\theta}$ à partir de \underline{y} pour une certaine utilisation ?

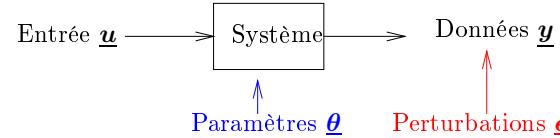
■ Modèle d'observation : $\underline{y} = \mathcal{M}(\underline{\theta}, \underline{\epsilon})$



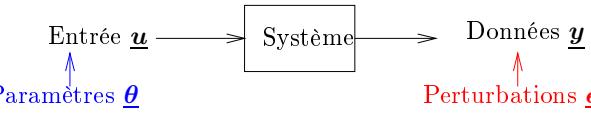
Cas le plus simple : perturbations additives
 $\underline{y} = \underline{y}_{\mathcal{M}}(\underline{\theta}) + \underline{\epsilon}$

■ Exemple : observation d'une galaxie avec un télescope

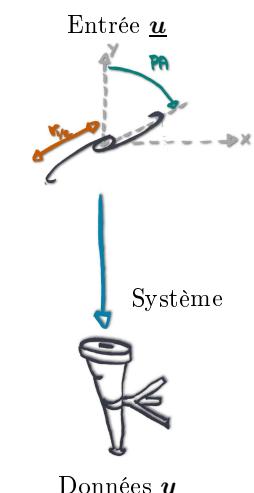
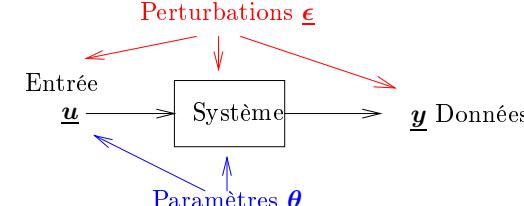
→ Calibration d'un instrument en laboratoire :



→ Mesures à partir d'un instrument calibré



→ Calibration d'un instrument en mode opérationnel



I.2 DÉFINITION D'UN ESTIMATEUR

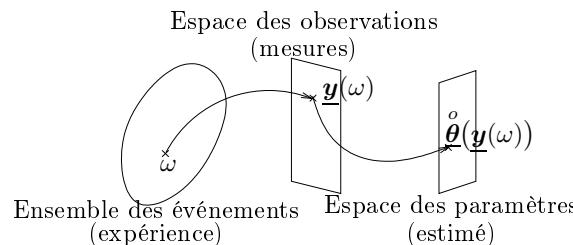
Définition

Un estimateur est une fonction de l'espace des **observations** (mesures, données) dans l'espace des **paramètres** qui pour une valeur mesurée \underline{y} associe une valeur aux paramètres $\underline{\theta}$:

$$\begin{aligned}\underline{\theta} : \mathbb{R}^N &\longrightarrow \mathbb{R}^M \\ \underline{y} &\longrightarrow \underline{\theta}(\underline{y})\end{aligned}$$

Prise en compte des perturbations :

Perturbations \Rightarrow données \underline{y} aléatoires \Rightarrow paramètres estimés aléatoires!



Rappels sur les matrices

Une matrice $M \times N$ modélise une application linéaire de \mathbb{C}^N dans \mathbb{C}^M
 $(\underline{u} \in \mathbb{C}^N \text{ et } \underline{v} \in \mathbb{C}^M)$

$$\begin{bmatrix} v_1 \\ \vdots \\ v_m \\ \vdots \\ v_M \end{bmatrix} = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M,1} & A_{M,2} & \cdots & A_{M,N} \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \text{ avec } v_m = \sum_{n=1}^N A_{m,n} u_n$$

$\Rightarrow \left\{ \begin{array}{l} \text{Nombre de ligne de } \mathbf{A} = \text{nombre de lignes de } \underline{v} \\ \text{Nombre de colonnes de } \mathbf{A} = \text{nombre de lignes de } \underline{u} \end{array} \right.$

• Transposition et application adjointe : soit \mathbf{A} une matrice $M \times N$, alors

$\mathbf{A}^H = \{A_{c,\ell}^*\}_{\ell,c}$ est une matrice $N \times M$ telle que $\langle \underline{v}, \mathbf{A} \underline{u} \rangle = \langle \mathbf{A}^H \underline{v}, \underline{u} \rangle$

$$\mathbf{A}^H = \begin{bmatrix} A_{1,1}^* & \cdots & A_{m,1}^* & \cdots & A_{M,1}^* \\ A_{1,2}^* & \cdots & A_{m,2}^* & \cdots & A_{M,2}^* \\ \vdots & & \vdots & \ddots & \vdots \\ A_{1,N}^* & \cdots & A_{m,N}^* & \cdots & A_{M,N}^* \end{bmatrix} \Rightarrow \left\{ \begin{array}{l} \text{Les lignes deviennent} \\ \text{les colonnes conjuguées} \end{array} \right.$$

$$\mathbf{A}^T = \{A_{c,\ell}\}_{\ell,c} \Rightarrow \left\{ \begin{array}{l} \text{Pour les matrices à valeurs dans } \mathbb{R} : \mathbf{A}^H = \mathbf{A}^T \\ \text{Pour les matrices à valeurs dans } \mathbb{C} : \mathbf{A}^T \text{ sans conjugaison!} \end{array} \right.$$

I.3 RAPPELS D'ALGÈBRE LINÉAIRE ET DE CALCUL MATRICIEL

Rappels sur les vecteurs

On considère ici toujours des vecteurs colonnes : $\underline{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \in \mathbb{R}^N$ ou \mathbb{C}^N

• Transposition : soit \underline{u} de dimension N , alors

$\underline{u}^T = [u_1, u_2, \dots, u_N]$ et $\underline{u}^H = [u_1^*, u_2^*, \dots, u_N^*]$ sont des vecteurs lignes

• Somme de vecteurs : \underline{u} et \underline{v} vecteurs de \mathbb{C}^N et α et β scalaires $\in \mathbb{C}$

$$\text{alors } \alpha \underline{u} + \beta \underline{v} = \begin{bmatrix} \alpha u_1 + \beta v_1 \\ \vdots \\ \alpha u_N + \beta v_N \end{bmatrix}$$

• Produits de vecteurs : soient \underline{u} et \underline{v} deux vecteurs colonnes de \mathbb{C}^N , alors :

Produit scalaire : $\langle \underline{u}, \underline{v} \rangle = \underline{u}^H \underline{v} = \sum_{n=1}^N u_n^* v_n$ est un scalaire

Norme : $\|\underline{u}\|^2 = \underline{u}^H \underline{u}$ et $\langle \underline{u}, \underline{v} \rangle = \|\underline{u}\| \cdot \|\underline{v}\| \cdot \cos(\alpha)$ (α angle entre \underline{u} et \underline{v})

$\underline{u} \underline{v}^H = \{u_\ell v_c^*\}_{\ell,c}$ est une matrice $N \times N$

Le produit $\underline{u} \underline{v}$ n'existe pas!

• Produits de matrices et composition d'applications linéaires :

→ Soient les matrices \mathbf{A} , $N \times K$ et \mathbf{B} , $K \times M$ et les vecteurs $\underline{u} \in \mathbb{C}^M$, $\underline{v} \in \mathbb{C}^K$

Alors $\mathbf{AB} = \{\sum_{k=1}^K A_{\ell,k} B_{k,c}\}_{\ell,c}$ est une matrice $N \times M$

telle que $\underline{w} = \mathbf{AB} \underline{u} = \mathbf{A} \underline{v}$ avec $\underline{v} = \mathbf{B} \underline{u}$

$$\underline{u} \in \mathbb{C}^M \xrightarrow{\mathbf{B}} \underline{v} \in \mathbb{C}^K \xrightarrow{\mathbf{A}} \underline{w} \in \mathbb{C}^N$$

Attention : Toujours vérifier que les dimensions sont cohérentes :
 nombre de lignes de \mathbf{B} = nombre de colonnes de \mathbf{A} !
 Produit non commutatif $\mathbf{AB} \neq \mathbf{BA}$

$(\mathbf{AB})^H = \mathbf{B}^H \mathbf{A}^H$ est une matrice $M \times N$

→ Soient deux vecteurs colonne $\underline{u} \in \mathbb{C}^N$ et $\underline{v} \in \mathbb{C}^M$

$\underline{u} \underline{v}^H = \{u_\ell v_c^*\}_{\ell,c}$ est une matrice $N \times M$
 (application de \mathbb{C}^M dans \mathbb{C} suivi d'une application de \mathbb{C} dans \mathbb{C}^N)

Le produit $\underline{u} \underline{v}$ n'existe pas!

- Propriétés des matrices carrées :

→ Diagonale : $\mathbf{A} = \text{diag}\{\underline{a}\} = \text{diag}\{a_1, a_2, \dots, a_N\}$ si $\begin{cases} A_{n,n} = a_n \\ A_{\ell,c} = 0, \forall \ell \neq c \end{cases}$

→ Identité : $\mathbf{I} = \text{diag}\{\underline{1}\} = \text{diag}\{1, 1, \dots, 1\}$

→ Symétrique : si $\mathbf{A}^T = \mathbf{A}$

→ Hermitienne : si $\mathbf{A}^H = \mathbf{A}$

→ Inversible : s'il existe une inverse \mathbf{A}^{-1} telle que $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$

⇒ Si \mathbf{A} est symétrique et inversible, alors \mathbf{A}^{-1} est symétrique.

⇒ Si \mathbf{A} est diagonale avec $a_n \neq 0, \forall n$ alors $\mathbf{A}^{-1} = \text{diag}\{a_1^{-1}, a_2^{-1}, \dots, a_N^{-1}\}$

→ Orthogonale : si $\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H = \mathbf{I}$ (soit $\mathbf{A}^{-1} = \mathbf{A}^H$)

■ Projection orthogonale dans un espace vectoriel

Soit \mathcal{H} un espace vectoriel réel ou complexe (c'est-à-dire tel que pour tout \underline{u} et $\underline{v} \in \mathcal{H}$ et α et β dans \mathbb{R} ou \mathbb{C} on a $\alpha\underline{u} + \beta\underline{v} \in \mathcal{H}$).

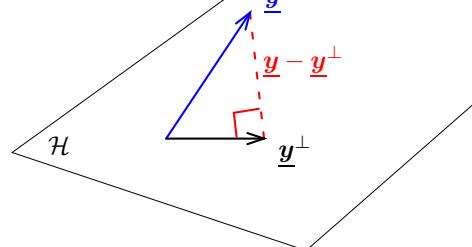
Soit $\langle \underline{u}, \underline{v} \rangle$ le produit scalaire sur \mathcal{H} et la norme associée $\|\underline{u}\|^2 = \langle \underline{u}, \underline{u} \rangle$.

On peut définir la projection orthogonale de \underline{y} sur l'espace vectoriel \mathcal{H} comme l'élément \underline{y}^\perp de \mathcal{H} minimisant la distance à \underline{y} :

$$\underline{y}^\perp = \arg \min_{\underline{u} \in \mathcal{H}} \|\underline{u} - \underline{y}\|^2 = \arg \min_{\underline{u} \in \mathcal{H}} \langle \underline{u} - \underline{y}, \underline{u} - \underline{y} \rangle$$

On a alors $\underline{y} - \underline{y}^\perp \perp \mathcal{H}$,

soit $\forall \underline{u} \in \mathcal{H}, \langle \underline{y} - \underline{y}^\perp, \underline{u} \rangle = 0$



- Propriétés des matrices carrées (suite) :

→ Valeurs et vecteurs propres : λ_n et \underline{v}_n tels que $\mathbf{A}\underline{v}_n = \lambda_n\underline{v}_n$

→ Les valeurs propres sont les racines du polynôme caractéristique :

$$P(\lambda) = \det(\mathbf{A} - \lambda I)$$

→ Diagonalisation d'une matrice symétrique :

Si les colonnes de \mathbf{V} sont les vecteurs propres de norme unité de \mathbf{A}

$$\text{alors } \begin{cases} \mathbf{V} \text{ est orthogonale : } \mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I} \\ \mathbf{A} = \mathbf{V}\text{diag}(\underline{\lambda})\mathbf{V}^T \end{cases}$$

→ Matrice définie positive : si $\forall \underline{u} \neq 0, \underline{u}^H \mathbf{A} \underline{u} > 0$

Matrice définie positive \Leftrightarrow valeurs propres positives : $\lambda_n > 0, \forall n$

→ Matrice définie non négative : si $\forall \underline{u} \neq 0, \underline{u}^H \mathbf{A} \underline{u} \geq 0$

Matrice définie positive \Leftrightarrow valeurs propres positives ou nulles : $\lambda_n \geq 0, \forall n$

I.4 RAPPELS D'OUTILS STATISTIQUES

■ Lois de probabilités pour des variables aléatoires :

→ Fonction de répartition : $F_X(x) = \text{Proba}(X \leq x) \geq 0$

$$\text{Proba}(X \in]a, b]) = F_X(b) - F_X(a)$$

$\lim_{x \rightarrow -\infty} F_X(x) = 0, \lim_{x \rightarrow +\infty} F_X(x) = 1, F_X(x)$ croissante, continue à droite

→ Densités de probabilités pour variables à valeurs continues : $f_X(x) \geq 0$

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi \Rightarrow \begin{cases} \int_{-\infty}^{\infty} f_X(x) dx = 1 \\ \text{Proba}(X \in [I]) = \int_{[I]} f_X(x) dx \end{cases}$$

$$\text{Espérance : } E_X\{g(X)\} = \int g(x) f_X(x) dx$$

→ Probabilités pour variables à val. discrètes : $P_X[k] = \text{Proba}(X=k) \geq 0, k \in \mathbb{Z}$

$$F_X(x) = \sum_{k \leq x} P_X[k] \Rightarrow \sum_{k \in \mathbb{Z}} P_X[k] = 1$$

$$\text{Espérance : } E_X\{g(X)\} = \sum_{k \in \mathbb{Z}} g(k) P_X[k]$$

■ Propriétés de l'espérance : valeur moyenne sur les réalisations

→ Interprétation : $E_X\{g(X)\}$ valeur moyenne de $g(X)$ sur les réalisations de X

→ Linéarité : pour a, b et c constantes

$$E_X\{ah(X) + bg(X) + c\} = aE_X\{h(X)\} + bE_X\{g(X)\} + c$$

■ Moments, moyenne, variance...

→ Moment d'ordre k : $E_X\{X^k\}$

→ Moment centré d'ordre k : $E_X\{(X - E_X\{X\})^k\}$

→ Moyenne (moment d'ordre 1) : $m_X = E_X\{X\}$

→ Puissance moyenne (moment d'ordre 2) : $P_X = E_X\{X^2\}$

→ Variance (moment d'ordre 2 centré) : $\sigma_X^2 = E_X\{(X - E\{X\})^2\}$

$$\sigma_X^2 = E_X\{X^2\} - E_X\{X\}^2 = P_X - m_X^2$$

$$\text{Si } Y = aX, \sigma_Y^2 = a^2\sigma_X^2$$

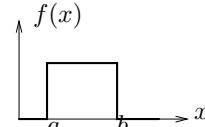
→ Écart-type : $\sigma_X = \sqrt{\sigma_X^2} = \sqrt{E_X\{(X - E\{X\})^2\}}$

■ Exemples de lois de probabilités continues : $X \in \mathbb{R}$

→ Distribution uniforme : $f(x) = Cte$ sur un intervalle et nulle en dehors !

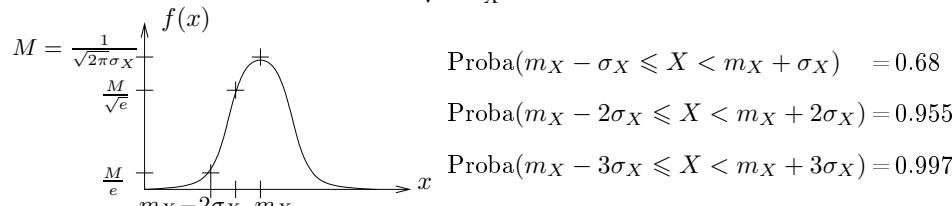
$$f(x) = \mathbb{1}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{ailleurs} \end{cases}$$

$$E\{X\} = \frac{a+b}{2} \quad \sigma_X^2 = \frac{(a-b)^2}{12}$$



→ Loi de Gauss ou loi Normale : $\mathcal{N}(m_X, \sigma_X^2)$; ($E\{X\} = m_X$, variance σ_X^2)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-m_X)^2}{2\sigma_X^2}}$$



■ Exemples de lois de probabilités discrètes : $X \in \mathbb{Z}$

→ Loi de Bernoulli :

$$\text{Proba}(X=1) = P_X[1] = \lambda, \quad \text{Proba}(X=0) = P_X[0] = 1 - \lambda$$

$$E\{X\} = \lambda, \quad \sigma_X^2 = \lambda(1 - \lambda)$$

→ Loi géométrique : (arrêt de réalisations de v. a. de Bernoulli au 1er succès)

$$P_X[k] = (1 - \lambda)^{k-1} \lambda$$

$$E\{X\} = 1/\lambda, \quad \sigma_X^2 = (1 - \lambda)/\lambda^2$$

→ Loi de Poisson :

(Comptage d'un nombre d'événements indépendants dans un intervalle)

$$P_X[k] = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots, \infty$$

$$E\{X\} = \lambda, \quad \sigma_X^2 = \lambda \quad (\text{rappel : } \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^\lambda)$$

(Pour λ grand, peut être approchée par une distribution gaussienne $\mathcal{N}(\lambda, \lambda)!$)

■ Changement de variable aléatoire dans une loi de probabilité continue

X variable aléatoire de loi $f_X(x)$ et $g : \mathbb{R} \rightarrow \mathbb{R}$

$$x \rightarrow y = g(x) \quad \text{avec } g(x) \text{ bijective}$$

$$\Downarrow \quad Y \text{ variable aléatoire de loi } f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} \cdot f_X(g^{-1}(y))$$

→ Exemple : transformation affine $y = a \cdot x + b$ avec a et b constantes ($a \neq 0$)

$$g(x) = a \cdot x + b \Rightarrow g^{-1}(y) = \frac{y-b}{a} \quad \text{et} \quad g'(x) = a$$

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

→ Cas d'une loi normale $x \sim \mathcal{N}(m_X, \sigma_X^2)$: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-m_X)^2}{2\sigma_X^2}}$

$$f_Y(y) = \frac{1}{|a|} \cdot f_X\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi}|a|\sigma_X} e^{-\frac{(\frac{y-b}{a}-m_X)^2}{2\sigma_X^2}} = \frac{1}{\sqrt{2\pi}|a|\sigma_X} e^{-\frac{(\frac{y-b}{a}-m_X)^2}{2a^2\sigma_X^2}}$$

$$\Rightarrow y \sim \mathcal{N}\left(\underbrace{a \cdot m_X + b}_{m_Y}, \underbrace{a^2 \sigma_X^2}_{\sigma_Y^2}\right)$$

■ Loi jointe pour un couple de variables aléatoires

→ v.a. continues : densité jointe $f_{X,Y}(x,y)$

$$\text{Proba}(\{X, Y\} \in [I]) = \iint_{[I]} f_{X,Y}(x,y) dx dy$$

$$\text{Espérance : } E_{X,Y}\{g(X, Y)\} = \iint g(x,y) f_{X,Y}(x,y) dx dy$$

→ v.a. discrètes : probabilité jointe $P_{X,Y}[k,\ell] = \text{Proba}((X=k) \cap (Y=\ell))$

$$\text{Espérance : } E_{X,Y}\{g(X, Y)\} = \sum_{k \in \mathbb{Z}} \sum_{\ell \in \mathbb{Z}} g(k,\ell) P_{X,Y}[k,\ell]$$

→ Covariance : $\text{Cov}(X, Y) = E_{X,Y}\{(X - E_X\{X\})(Y - E_Y\{Y\})\}$

→ Coefficient de corrélation linéaire : $r = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$

$r = 0 \Rightarrow \text{Corr}(X, Y) = 0 \Rightarrow X \text{ et } Y \text{ sont décorrélés}$

■ Couples de variables aléatoires indépendantes :

X indépendant de $Y \Rightarrow$ La loi de X sachant Y ne dépend pas de la valeur de Y

X et Y sont indépendantes si $\begin{cases} f_{X|Y}(x|y) = f_X(x) \text{ si } f_Y(y) \neq 0 \\ P_{X|Y}(k|\ell) = P_X[k] P_Y[\ell] \text{ si } P_Y[\ell] \neq 0 \end{cases}$

→ Si X et Y sont indépendantes :

$$\Rightarrow \begin{cases} f_{X,Y}(x,y) = f_{X|Y}(x|y) f_Y(y) = f_X(x) f_Y(y) \\ P_{X,Y}(k,\ell) = P_{X|Y}[k,\ell] P_Y[\ell] = P_X[k] P_Y[\ell] \end{cases}$$

$$\Rightarrow E_{X,Y}\{h(X)g(Y)\} = E_X\{h(X)\} E_Y\{g(Y)\}$$

$$\Rightarrow \text{Cov}(X, Y) = E_X\{(X - E_X\{X\})\} E_Y\{(Y - E_Y\{Y\})\} = 0$$

$\Rightarrow X$ et Y décorrélées

Réiproque fausse en général, sauf si X et Y gaussiens !

■ Lois conditionnelles

Loi de probabilité d'une variable X sachant la valeur prise par la variable Y

→ Événements aléatoires A et B : $\text{Proba}(B|A) = \frac{\text{Proba}(A \text{ et } B)}{\text{Proba}(A)}$ si $\text{Proba}(A) \neq 0$

→ Variables aléatoires à valeurs discrètes : $P_{X|Y}[k|\ell] = \frac{P_{X,Y}[k,\ell]}{P_Y[\ell]}$ si $P_Y[\ell] \neq 0$

→ Variables aléatoires à valeurs continue : $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ si $f_Y(y) \neq 0$

Ce sont des lois de probabilités !

→ Normalisation

$$\text{v.a. à valeurs continues : } \int f_{X|Y}(x|y) dx = 1$$

$$\text{v.a. à valeurs discrètes : } \sum_k P_{X|Y}[k|y] = 1$$

→ Espérance conditionnelle

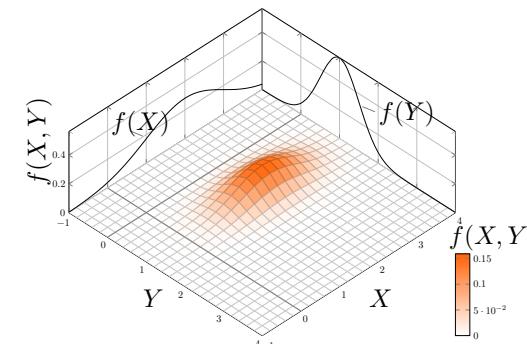
$$\text{v.a. à valeurs continues : } E_{X|Y}\{g(x,y)|y\} = \int g(x,y) f_{X|Y}(x|y) dx$$

$$\text{v.a. à valeurs discrètes : } E_{X|Y}\{g(x,y)|y\} = \sum_k g(k,y) P_{X|Y}[k|y]$$

■ Principe de marginalisation :

Comment calculer la loi de X connaissant la loi jointe de X et Y ?

$$\rightarrow \text{v.a. continues : } f_X(x) = \int f_{X,Y}(x,y) dy = f_X(x) \underbrace{\int f_{Y|X}(y|x) dy}_{=1}$$



→ Variables aléatoires à valeurs discrètes : $P_Y[\ell] = \sum_k P_{X,Y}[k,\ell]$

Règle de Bayes :

→ Événements A et B :

$$\text{Proba}(A|B) = \frac{\text{Proba}(B|A)\text{Proba}(A)}{\text{Proba}(B)} \quad \text{si } \text{Proba}(B) \neq 0$$

→ v.a. discrètes : $(P_{X,Y}[k,\ell] = P_{X|Y}[k|\ell]P_Y[\ell] = P_{Y|X}[\ell|k]P_X[k])$

$$\Rightarrow P_{Y|X}[\ell|k] = \frac{P_{X|Y}[k|\ell]P_Y[\ell]}{P_X[k]} \quad \text{si } P_X[k] \neq 0$$

→ v.a. continues : $(f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x))$

$$\Rightarrow f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)} \quad \text{si } f_X(x) \neq 0$$

Transformation affine d'un vecteur aléatoire :

Soit \underline{x} de moyenne \underline{m}_X et de matrice de covariance Γ_X

Alors le vecteur $\underline{y} = \mathbf{A}\underline{x} + \underline{b}$ (\mathbf{A} matrice et \underline{b} vecteur constants)

a pour moyenne $\underline{m}_Y = \mathbf{A}\underline{m}_X + \underline{b}$ et pour matrice de covariance $\Gamma_Y = \mathbf{A}\Gamma_X\mathbf{A}^T$

Changement de variables dans une loi de probabilité vectorielle

$$\begin{aligned} \underline{x} \text{ vecteur aléatoire de loi } f_X(\underline{x}) \text{ et } \underline{g} : \mathbb{R}^N &\rightarrow \mathbb{R}^N \\ \underline{x} &\rightarrow \underline{y} = \underline{g}(\underline{x}), \text{ bijective} \\ \Downarrow & \end{aligned}$$

\underline{y} vecteur aléatoire de loi $f_Y(\underline{y}) = |\det(J_{\underline{g}^{-1}}(\underline{y}))| \cdot f_X(\underline{g}^{-1}(\underline{y}))$

$$\det(J_{\underline{g}^{-1}}(\underline{y})) \text{ déterminant de la matrice } \left\{ \frac{\partial \underline{g}_{\ell}^{-1}(\underline{y})}{\partial y_c} \right\}_{\ell,c}$$

• Exemple : $\underline{y} = \mathbf{A}\underline{x} + \underline{b}$ avec \mathbf{A} matrice inversible $\Rightarrow \underline{g}^{-1}(\underline{y}) = \mathbf{A}^{-1}(\underline{y} - \underline{b})$

$$\Rightarrow \det(J_{\underline{g}^{-1}}(\underline{y})) = \frac{1}{\det(\mathbf{A})} \Rightarrow f_Y(\underline{y}) = \frac{1}{|\det(\mathbf{A})|} \cdot f_X(\mathbf{A}^{-1}(\underline{y} - \underline{b}))$$

Vecteur aléatoire, vecteur moyen, matrice de covariance

→ Vecteur aléatoire (vecteur colonne) : $\underline{x} = [x_1, x_2, \dots, x_N]^T$

$F_{\underline{X}}(\underline{x}) = \text{Proba}(\underline{X} \leq \underline{x}) \Rightarrow \text{Loi jointe sur des } x_n$

$f_{\underline{X}}(\underline{x}) = f_{X_1, X_2, \dots, X_N}(x_1, x_2, \dots, x_N)$, parfois notée $f_X(\{x_k\}_{k=1\dots N})$

$$\rightarrow \text{Vecteur moyen : } \underline{m}_X = E\{\underline{x}\} = \begin{bmatrix} E\{x_1\} \\ E\{x_2\} \\ \vdots \\ E\{x_N\} \end{bmatrix}$$

→ Matrice de (variance-)covariance : $\Gamma_X = E\{(\underline{x} - E\underline{x})(\underline{x} - E\underline{x})^T\}$

$$\Gamma_X = \{\text{Cov}(x_\ell, x_c)\}_{\ell,c} = \begin{bmatrix} \sigma_{x_1}^2 & \text{Cov}(x_1, x_2) & \cdots & \text{Cov}(x_1, x_N) \\ \text{Cov}(x_2, x_1) & \sigma_{x_2}^2 & \cdots & \text{Cov}(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(x_N, x_1) & \text{Cov}(x_N, x_2) & \cdots & \sigma_{x_N}^2 \end{bmatrix}$$

Matrice Γ_X symétrique si $\underline{x} \in \mathbb{R}^N$: $\Gamma_X^T = \Gamma_X$, hermitienne si $\underline{x} \in \mathbb{C}^N$: $\Gamma_X^H = \Gamma_X$

Matrice Γ_X définie non négative : $\forall \underline{u}, \underline{u}^T \Gamma_X \underline{u} \geq 0$

Loi Gaussienne (Normale) d'un vecteur (dimension N) : $\mathcal{N}(\underline{m}_X, \Gamma_X)$

$$E\{\underline{x}\} = \underline{m}_X, \quad \Gamma_X = E\{(\underline{x} - \underline{m}_X)(\underline{x} - \underline{m}_X)^T\}$$

$$f_X(\underline{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} (\det \Gamma_X)^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{x} - \underline{m}_X)^T \Gamma_X^{-1} (\underline{x} - \underline{m}_X)}$$

→ Transformation affine d'un vecteur aléatoire gaussien :

(formule de changement de variables)

Soit $\underline{x} \sim \mathcal{N}(\underline{m}_X, \Gamma_X)$, et $\underline{y} = \mathbf{A}\underline{x} + \underline{b}$ alors $\underline{y} \sim \mathcal{N}(\overbrace{\mathbf{A}\underline{m}_X + \underline{b}}^{\underline{m}_Y}, \overbrace{\mathbf{A}\Gamma_X\mathbf{A}^T}^{\Gamma_Y})$,

→ Exemples :

- \underline{x} variables i.i.d. Gaussiennes $x_n \sim \mathcal{N}(m_x, \sigma_x^2) \Rightarrow \underline{x} \sim \mathcal{N}(m_x \mathbf{1}, \sigma_x^2 \mathbf{I})$
(indépendants et identiquement distribués)

- \underline{x} indépendant Gaussien $x_n \sim \mathcal{N}(0, \sigma_n^2) \Rightarrow \underline{x} \sim \mathcal{N}(\mathbf{0}, \text{diag}\{\sigma_n^2\}_n)$

- $\underline{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \Rightarrow \underline{y} = \mathbf{A}\underline{x} + \underline{b} \sim \mathcal{N}(\underline{b}, \mathbf{A}\mathbf{A}^T)$

- x_n indép. $\sim \mathcal{N}(0, \sigma_n^2) \Rightarrow y = \sum_n x_n \sim \mathcal{N}(0, \sigma_y^2)$ avec $\sigma_y^2 = \sum_n \sigma_n^2$

I.5 PROPRIÉTÉS DES ESTIMATEURS

Biais : $b_{\hat{\theta}}(\theta) = E_{Y|\theta}\{\hat{\theta} | \theta\} - \theta \Rightarrow$ Erreur moyenne de l'estimateur

Si $\hat{\theta}$ est une v.a. : **biais moyen** $b_{\hat{\theta}} = E_{\theta}\{b_{\hat{\theta}}(\theta)\} = E_{Y,\theta}\{\hat{\theta}\} - E_{\theta}\{\theta\}$

Matrice de covariance :

$$V_{\hat{\theta}}(\theta) = E_{Y|\theta}\{(\hat{\theta} - E_{Y|\theta}\{\hat{\theta} | \theta\})(\hat{\theta} - E_{Y|\theta}\{\hat{\theta} | \theta\})^T | \theta\}$$

\Rightarrow Variation de l'estimateur autour de sa valeur moyenne

Si $\hat{\theta}$ est une v.a. : **covariance moyenne** $V_{\hat{\theta}} = E_{\theta}\{V_{\hat{\theta}}(\theta)\}$

Erreur quadratique moyenne : Compromis biais/variance

\Rightarrow Variation de l'estimateur autour de la vraie valeur

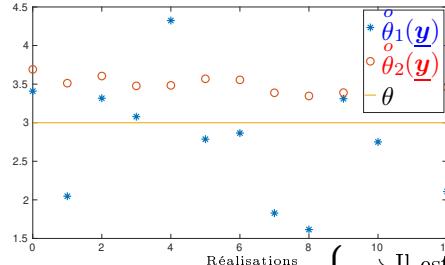
$$\text{EQM}_{\hat{\theta}}(\theta) = E_{Y|\theta}\{(\hat{\theta} - \theta)^T (\hat{\theta} - \theta) | \theta\} = \text{Tr}\{V_{\hat{\theta}}(\theta) + b_{\hat{\theta}}(\theta)b_{\hat{\theta}}(\theta)^T\}$$

Si $\hat{\theta}$ est une v.a. :

$$\text{EQM}_{\hat{\theta}} = E_{Y,\theta}\{(\hat{\theta} - \theta)^T (\hat{\theta} - \theta)\}$$

Exemple : Soient $\left\{ Y \text{ une variable aléatoire de moyenne } E_{Y|\theta}\{Y|\theta\} = \theta, \text{ les données } \{y_n\}_{n=1 \dots N} \text{ sont } N \text{ réalisations de } Y. \right.$

Étudions deux estimateurs de θ à partir de 13 réalisations : $\hat{\theta}_1(\underline{y})$ and $\hat{\theta}_2(\underline{y})$



\rightarrow Quel estimateur préférez-vous ?

\rightarrow Lequel a le biais le plus faible ?

\rightarrow Lequel a la variance la plus faible ?

\rightarrow Lequel a l'EQM la plus faible ?

\rightarrow Convergent-ils quand $N \rightarrow \infty$?

Soit $\hat{\theta}_1(\underline{y}) = \frac{1}{N} \sum_{n=1}^N y_n$: $\left\{ \begin{array}{l} \rightarrow \text{Il est non biaisé car } E_{Y|\theta}\{\hat{\theta}_1(\underline{y}) | \theta\} = \theta \\ \rightarrow \text{Si les données } \{y_n\}_{n=1 \dots N} \text{ sont indépendantes :} \\ \Rightarrow \text{sa variance est } \sigma_{\hat{\theta}_1}^2(\theta) = \frac{\sigma_Y^2}{N} \\ \Rightarrow \text{il converge en moyenne quadratique vers } \theta \\ \Rightarrow \text{il converge en loi vers } \mathcal{N}(\theta, \frac{\sigma_Y^2}{N}) \\ \rightarrow \text{Si } Y \sim \mathcal{N}(\theta, \sigma_Y^2) : \sigma_{\hat{\theta}_1}^2(\theta) = F(\theta)^{-1} \end{array} \right.$

Convergence : quand $N \rightarrow \infty$ $\overset{o}{\hat{\theta}}(\underline{y})$ avec \underline{y} de dimension N

- Convergence en loi ($\hat{\theta}$ v.a. de loi $f_{\hat{\theta}}$, F fonction de répartition)

- Convergence en probabilité

$$\forall \epsilon > 0, \lim_{N \rightarrow \infty} \text{Proba}(\|\overset{o}{\hat{\theta}}(\underline{y}) - \theta\| \geq \epsilon) = 0$$

- Convergence en moyenne quadratique

$$\lim_{N \rightarrow \infty} E\{(\overset{o}{\hat{\theta}}(\underline{y}) - \theta)^T (\overset{o}{\hat{\theta}}(\underline{y}) - \theta)\} = 0$$

Inégalité de Cramer-Rao : (inégalité matricielle : $\mathbf{A} \geq \mathbf{0} \Leftrightarrow \forall \underline{u}, \underline{u}^T \mathbf{A} \underline{u} \geq 0$)

$$V_{\hat{\theta}}(\theta) \geq \left(\mathbf{I} + \frac{\partial b_{\hat{\theta}}}{\partial \theta^T}(\theta) \right)^T \mathbf{F}(\theta)^{-1} \left(\mathbf{I} + \frac{\partial b_{\hat{\theta}}}{\partial \theta^T}(\theta) \right)$$

$$\left. \begin{array}{l} \mathbf{F}(\theta) \text{ matrice d'Information de Fisher} \\ \mathbf{F}(\theta) = -E \left\{ \frac{d^2 \log f_{Y|\theta}(\underline{y} | \theta)}{d\theta d\theta^T} \right\} \end{array} \right\} \text{matrice de covariance du vecteur } \frac{d \log f_{Y|\theta}(\underline{y} | \theta)}{d\theta}$$

Cas 1D, estimateur non biaisé : borne sur la variance $V_{\hat{\theta}}(\theta) \geq F(\theta)^{-1}$

Exemple (suite) : propriétés de l'estimateur $\hat{\theta}_1(\underline{y}) = \frac{1}{N} \sum_{n=1}^N y_n$

\rightarrow Il est non biaisé car :

$$E_{Y|\theta}\{\hat{\theta}_1(\underline{y}) | \theta\} = E_{Y|\theta}\left\{ \frac{1}{N} \sum_{n=1}^N y_n | \theta \right\} \stackrel{\text{linéarité de l'espérance}}{\downarrow} \frac{1}{N} \sum_{n=1}^N E_{Y|\theta}\{y_n | \theta\} = \frac{1}{N} \sum_{n=1}^N \theta = \theta$$

\rightarrow Variance : $\sigma_{\hat{\theta}_1}^2(\theta) = E_{Y|\theta}\left\{ (\hat{\theta}_1(\underline{y}) - E_{Y|\theta}\{\hat{\theta}_1(\underline{y}) | \theta\})^2 | \theta \right\}$

$$\sigma_{\hat{\theta}_1}^2(\theta) = E_{Y|\theta}\left\{ \left(\frac{1}{N} \sum_{n=1}^N y_n - \theta \right)^2 | \theta \right\} = E_{Y|\theta}\left\{ \frac{1}{N^2} \left(\sum_{n=1}^N (y_n - \theta) \right)^2 | \theta \right\}$$

$$= \frac{1}{N^2} E_{Y|\theta}\left\{ \left(\sum_{n=1}^N (y_n - \theta) \right) \left(\sum_{m=1}^N (y_m - \theta) \right) | \theta \right\} = \frac{1}{N^2} \sum_{n=1}^N \sum_{m=1}^N \underbrace{E_{Y|\theta}\{(y_n - \theta)(y_m - \theta) | \theta\}}$$

Si les données $\{y_n\}_{n=1 \dots N}$ sont indépendantes : pour $n \neq m \uparrow = 0$

$$\Rightarrow \text{Pour des données } \{y_n\}_{n=1 \dots N} \text{ indépendantes : } \sigma_{\hat{\theta}_1}^2(\theta) = \frac{\sigma_Y^2}{N}$$

\Rightarrow Il converge en moyenne quadratique vers θ car $\lim_{N \rightarrow \infty} \sigma_{\hat{\theta}_1}^2(\theta) = 0$

→ Si $\{y_n\}_{n=1 \dots N}$ sont indépendantes et Y est gaussien $\sim \mathcal{N}(\theta, \sigma_Y^2)$:

Indépendantes Gaussiennes

$$\begin{aligned} f(\underline{y}|\theta) &= \prod_{n=1}^N f(y_n|\theta) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left\{-\frac{(y_n - \theta)^2}{2\sigma_Y^2}\right\} = \frac{1}{\sqrt{2\pi}^N \sigma_Y^N} \exp\left\{-\frac{1}{2\sigma_Y^2} \sum_{n=1}^N (y_n - \theta)^2\right\} \\ &\Rightarrow \log(f(\underline{y}|\theta)) = -\frac{1}{2\sigma_Y^2} \sum_{n=1}^N (y_n - \theta)^2 + \text{constante} \\ &\Rightarrow \frac{d \log(f(\underline{y}|\theta))}{d\theta} = \frac{1}{\sigma_Y^2} \sum_{n=1}^N (y_n - \theta) \quad \text{et} \quad \frac{d^2 \log(f(\underline{y}|\theta))}{d\theta^2} = \frac{-N}{\sigma_Y^2} \\ &\Rightarrow \text{Information de Fisher : } F(\theta) = -E\left\{\frac{d^2 \log(f(\underline{y}|\theta))}{d\theta^2}\right\} = \frac{N}{\sigma_Y^2} \end{aligned}$$

On a donc : $\sigma_{\hat{\theta}}^2(\theta) = F(\theta)^{-1}$ c'est l'estimateur non biaisé de variance minimale

→ Question : Peux-tu formaliser cela un peu plus au niveau mathématique ?

Bien sûr, je vais formaliser davantage la notion d'estimateur en utilisant des notations mathématiques. Soit X une variable aléatoire représentant la caractéristique que l'on souhaite estimer dans une population. On suppose que l'on a un échantillon de taille n , noté $\{X_1, X_2, \dots, X_n\}$, tiré de cette population.

Un estimateur, noté $\hat{\theta}$, est une fonction de l'échantillon qui donne une estimation de la valeur du paramètre θ associé à la variable aléatoire X . Mathématiquement, cela peut être représenté comme suit : $\hat{\theta} = g(X_1, X_2, \dots, X_n)$ où g est une fonction qui prend en entrée les valeurs de l'échantillon et renvoie une estimation de θ .

Un bon estimateur possède deux propriétés importantes :

- Impartialité : L'estimateur est dit impartial s'il a une espérance égale à la vraie valeur du paramètre. Formellement, cela signifie que $E\{\hat{\theta}\} = \theta$.
- Efficacité : L'estimateur est dit efficace s'il a une variance aussi faible que possible parmi tous les estimateurs impartiaux. En d'autres termes, il minimise la dispersion de ses estimations autour de la vraie valeur du paramètre.

■ Qu'en penses ChatPGT ?

→ Question : Quelle est la définition d'un estimateur ?

En statistiques, un estimateur est une règle ou une formule mathématique utilisée pour estimer une caractéristique d'une population à partir d'un échantillon de cette population. L'objectif d'un estimateur est de fournir une estimation aussi précise que possible de la valeur réelle de la caractéristique recherchée.

Plus formellement, si l'on considère une variable aléatoire X qui représente la caractéristique de la population que l'on souhaite estimer, un estimateur est une fonction de l'échantillon observé, notée $T(X)$, qui donne une approximation de la valeur de X .

Il est important que les estimateurs soient impartiaux (leur espérance est égale à la valeur réelle du paramètre que l'on cherche à estimer) et qu'ils aient une variance aussi faible que possible pour être considérés comme efficaces. Il existe différents types d'estimateurs, tels que les estimateurs ponctuels et les intervalles de confiance, qui permettent de donner une plage probable dans laquelle se trouve la véritable valeur du paramètre.

→ Question : Es-tu sûr du terme "impartialité" ?

Je m'excuse pour la confusion. Le terme correct utilisé en statistiques est "biais" plutôt que "impartialité". Permettez-moi de clarifier la terminologie :

1. Biais : Un estimateur est dit non biaisé si l'espérance de l'estimateur est égale à la vraie valeur du paramètre que l'on souhaite estimer. Mathématiquement, un estimateur $\hat{\theta}$ est non biaisé si $E\{\hat{\theta}\} = \theta$.
2. Efficacité : L'estimateur est dit efficace s'il a une variance aussi faible que possible parmi tous les estimateurs non biaisés.

Veuillez remplacer le terme "impartialité" par "non biaisé" dans le contexte de la définition d'un bon estimateur. Merci de me l'avoir signalé, et je suis là pour toute clarification supplémentaire.

→ Question : Je ne comprends pas : on cherche à estimer θ et pas X !

Vous avez raison, et je m'excuse pour la confusion dans mes précédentes réponses. Permettez-moi de clarifier cela.

Supposons que nous voulions estimer un paramètre θ associé à une distribution de probabilité représentée par une variable aléatoire X . Alors, un estimateur, noté $\hat{\theta}$, est une fonction de l'échantillon qui donne une estimation de ce paramètre. Mathématiquement, cela peut être formulé comme suit : $\hat{\theta} = g(X_1, X_2, \dots, X_n)$ où X_1, X_2, \dots, X_n est l'échantillon observé et g est une fonction qui donne une estimation de θ basée sur cet échantillon.

L'objectif est d'avoir un estimateur $\hat{\theta}$ qui soit non biaisé, c'est-à-dire que son espérance soit égale à la vraie valeur du paramètre θ : $E\{\hat{\theta}\} = \theta$, et idéalement, cet estimateur devrait être efficace, c'est-à-dire qu'il devrait avoir une variance aussi faible que possible parmi tous les estimateurs non biaisés de θ .

II.1 NOTION DE MODÈLES ET CLASSIFICATION

■ Modèle de connaissance / représentation

	Connaissance	Représentation
Paramètres	sens physique	pas de sens physique
Simulation	souvent délicate	aisée
Information <i>a priori</i>	prise en compte	délicate à prendre en compte
Domaine de validité	étendu	souvent restreint

■ Modèle à temps continu / temps discret

	Temps continu	Temps discret
Paramètres	indépendant de discréte.	dépendant discréttisation
Simulation	souvent délicate	aisée
Information <i>a priori</i>	prise en compte	délicate à prendre en compte
Instants de mesure	quelconques	imposés

CHAP II MODÉLISATION DES SIGNAUX ET SYSTÈMES

- 1/ Notion de modèles et classification
- 2/ Modèles de systèmes
- 3/ Modèles de signaux
- 4/ Propriétés structurelles des modèles

■ Modèle déterministe / aléatoire

- Déterministes :

On sait calculer la valeur du signal ou la sortie du système de façon exacte connaissant les paramètres.
- Aléatoires (stochastiques) :

Le modèle ne caractérise que certaines propriétés statistiques du signal ou du système.

Modèles aléatoires utilisés pour modéliser des incertitudes

⇒ information *a priori*

II.2 MODÈLES DE SYSTÈMES

■ Filtres : réponse impulsionnelle h , réponse en fréquence $\hat{h}(f)$

Relation entrée sortie convolutionnelle $s = e * h \Leftrightarrow \hat{s} = \hat{e} \cdot \hat{h}$

■ Équations différentielles/aux différences : fonction de transfert H

- Temps continu : (conditions initiales nulles)

$$\frac{d^N s}{dt^N}(t) = - \sum_{n=0}^{N-1} a_n \frac{d^n s}{dt^n}(t) + \sum_{m=0}^M b_m \frac{d^m e}{dt^m}(t);$$

$$H(p) = \frac{S(p)}{E(p)} = \frac{b_M p^M + b_{M-1} p^{M-1} + \dots + b_1 p + b_0}{p^N + a_{N-1} p^{N-1} + \dots + a_1 p + a_0}$$

- Temps discret : (conditions initiales nulles)

$$s[k] = - \sum_{n=0}^N a_n s[k-n] + \sum_{m=0}^M b_m e[k-m]$$

$$H(z) = \frac{S(z)}{E(z)} = \frac{b_0 + b_1 z^{-1} + \dots + b_M z^{-M}}{1 + a_1 z^{-1} + \dots + a_N z^{-N}}.$$

II.3 MODÈLES DE SIGNAUX

■ Modèles analytiques

Équation permettant de calculer à chaque instant la valeur du signal pour des paramètres connus.

Exemple : $x(t) = \sum_{k=1}^K A_k \sin(2\pi f_k t + \phi_k)$, $\underline{\theta} = (K, \{A_k\}_k, \{f_k\}_k, \{\phi_k\}_k)$

$$x(t) = \sum_{k=0}^K P_k t^k, \quad \underline{\theta} = \{P_k\}_k$$

■ Modèles de type lois de probabilité

Connaissance de caractéristiques statistiques de la valeur du signal

- Bruit blanc** : $\epsilon[k]$ indépendants identiquement distribués (i.i.d.) gaussien, binaire...
- Chaines de Markov :

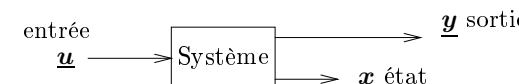
Proba($X[k] = \ell | \{X[n] = \ell_n\}_{n=0 \dots k-1}\right) = \text{Proba}(X[k] = \ell | \{X[k-1] = \ell_{k-1}\})$

Exemple : file d'attente $x[k]$ nb personnes présent dans la file à l'instant k .

■ Représentation d'état d'un système

(voir filtre de Kalman)

État \underline{x} non accessible à la mesure caractérisant le passé du système



- Temps continu : $\begin{cases} \frac{d\underline{x}}{dt}(t) = f(\underline{x}(t), \underline{u}(t), t) \\ \underline{y}(t) = h(\underline{x}(t), \underline{u}(t), t) \end{cases}$ équation d'état
équation d'observation

- Temps discret : $\begin{cases} \underline{x}[k+1] = f(\underline{x}[k], \underline{u}[k], k) \\ \underline{y}[k] = h(\underline{x}[k], \underline{u}[k], k) \end{cases}$ équation d'état
équation d'observation

- Cas particulier des systèmes linéaires invariants :

$$\begin{cases} \underline{x}[k+1] = \mathbf{A}\underline{x}[k] + \mathbf{B}\underline{u}[k] \\ \underline{y}[k] = \mathbf{C}\underline{x}[k] + \mathbf{D}\underline{u}[k] \end{cases} \quad \begin{cases} \frac{d\underline{x}}{dt}(t) = \mathbf{A}\underline{x}(t) + \mathbf{B}\underline{u}(t) \\ \underline{y}(t) = \mathbf{C}\underline{x}(t) + \mathbf{D}\underline{u}(t) \end{cases}$$

A matrice de dynamique, **B** matrice de commande

C matrice d'observation, **D** matrice de transmission directe

- Lois de probabilité de v.a. à valeur discrète : binomiale, multinomiale, bernoulli, hyper-géométrique, Poisson...

Loi de poisson : $P_X[k] = \frac{(m_X)^k}{k!} e^{-m_X}$

- Lois de probabilité de v.a. à valeur continue : uniforme, gaussienne, exponentielle, exponentielle symétrique...

Loi gaussienne $\mathcal{N}(m_X, \sigma_X^2) : f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp(-\frac{1}{2} \frac{(x - m_X)^2}{\sigma_X^2})$

Loi exponentielle symétrique $m_X = 0$, $\sigma_X^2 = \frac{2}{\lambda^2} : f_X(x) = \frac{\lambda}{2} \exp(-\lambda|x|)$

Loi exponentielle positive $m_X = \frac{1}{\lambda}$, $\sigma_X^2 = \frac{1}{\lambda^2} : f_X(x) = \lambda \exp(-\lambda x)$, si $x > 0$

- Lois de probabilité de vecteurs aléatoires (corrélation)

$\mathcal{N}(\underline{m}_X, \Gamma_X) : f_X(\underline{x}) = \frac{1}{\sqrt{2\pi^N} \sqrt{|\det(\Gamma_X)|}} \exp(-\frac{1}{2} (\underline{x} - \underline{m}_X)^T \Gamma_X^{-1} (\underline{x} - \underline{m}_X))$

Champs de Gibbs $f_X(\underline{x}) \propto \exp(-\frac{1}{T} \sum_{k \sim \ell} \phi(x_k - x_\ell))$ (~ notion de voisinage)

■ Filtrage des signaux aléatoires



- Modèle ARMA (*Auto-Regressive Moving Average*)

$$x[k] = - \sum_{n=0}^N a_n x[k-n] + \sum_{m=0}^M b_m \epsilon[k-m]$$

- Modèle AR (*Auto-Regressive*)

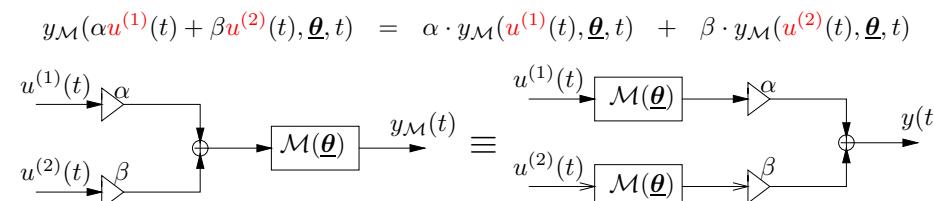
$$x[k] = - \sum_{n=0}^N a_n x[k-n] + \epsilon[k]$$

- Modèle MA (*Moving Average*)

$$s[k] = \sum_{m=0}^M b_m \epsilon[k-m]$$

■ (Structure de) modèle linéaire / non linéaire

- Linéarité par rapport aux entrées (LE) si $\forall \alpha, \beta, \forall \underline{u}^{(1)}, \underline{u}^{(2)}, \forall \underline{\theta}, \forall t :$



⇒ Sens classique de la linéarité pour les signaux et systèmes !

II.4 PROPRIÉTÉS STRUCTURELLES DES MODÈLES

■ Structure de modèle

- Modèle $\mathcal{M}(\underline{\theta})$ (données idéales $y_{\mathcal{M}}(u(t), \underline{\theta}, t)$ ou $\underline{y}_{\mathcal{M}}(\underline{\theta})$)

- Structure de modèle $\mathcal{M}(\cdot)$:

Ensemble des modèles possibles suivant les valeurs de $\underline{\theta}$

■ Complexité d'une structure de modèle

Par simplicité :

Complexité d'une structure = nombre de paramètres, degrés de liberté

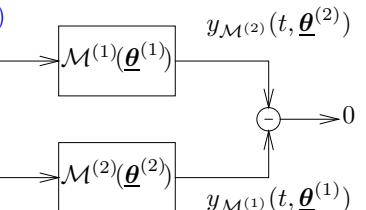
■ Modèles équivalents $\mathcal{M}^{(1)}(\underline{\theta}^{(1)}) \equiv \mathcal{M}^{(2)}(\underline{\theta}^{(2)})$

Cadre idéal :

$\forall \underline{\theta}^{(1)}$ on peut régler $\underline{\theta}^{(2)}$ tel que :

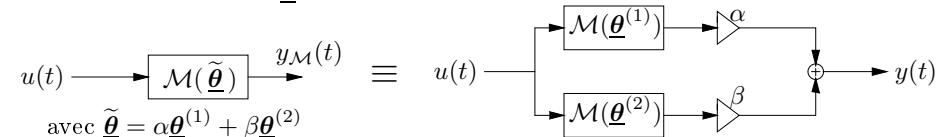
$\forall u(t), t, y_{\mathcal{M}^{(1)}}(t; \underline{\theta}^{(1)}) = y_{\mathcal{M}^{(2)}}(t; \underline{\theta}^{(2)})$

et réciproquement



- Linéarité par rapport aux paramètres (LP) si $\forall \alpha, \beta, \forall \underline{\theta}^{(1)}, \underline{\theta}^{(2)}, \forall u, \forall t :$

$$y_{\mathcal{M}}(\underline{u}, \underbrace{\alpha \underline{\theta}^{(1)} + \beta \underline{\theta}^{(2)}}_{\underline{\theta}}, t) = \alpha y_{\mathcal{M}}(\underline{u}, \underline{\theta}^{(1)}, t) + \beta y_{\mathcal{M}}(\underline{u}, \underline{\theta}^{(2)}, t).$$



⇒ Modèles LP particulièrement avantageux pour l'estimation !

→ Le modèle s'écrit comme une **combinaison linéaire des paramètres** :

$$y_{\mathcal{M}}(n, \underline{\theta}) = \underline{r}_n^T \underline{\theta} = \sum_{m=1}^M r_{n,m} \theta_m$$

$\underline{r}_n^T = [r_{n,1}, r_{n,2}, \dots, r_{n,M}]$ vecteur de régression (connu)

Écriture matricielle : $\underline{y}_{\mathcal{M}} = \mathbf{R}\underline{\theta}$ avec la matrice $\mathbf{R} = \{r_{n,m}\}_{n,m}$

■ Identifiabilité Si la structure est la bonne, les paramètres seront-ils les bons ?

- Le paramètre θ_k est dit structurellement globalement identifiable (*s.g.i.*) si pour presque tout $\underline{\theta}^{(1)} : \mathcal{M}(\underline{\theta}^{(1)}) \equiv \mathcal{M}(\underline{\theta}^{(2)}) \Rightarrow \theta_k^{(1)} = \theta_k^{(2)}$ (presque tout : sauf sur un ensemble de mesure nulle)
- \mathcal{M} est dit structurellement globalement identifiable si tous ses paramètres le sont.
- On parle d'identifiabilité structurelle locale si cette propriété n'est vrai que pour un voisinage de $\underline{\theta}^{(1)}$.

■ Discernabilité Si la structure est mauvaise, puis-je le détecter ?

- On dit que la structure de modèle $\mathcal{M}^{(1)}(\cdot)$ est structurellement discernable (*s.d.*) de la structure de modèle $\mathcal{M}^{(2)}(\cdot)$ si pour presque tout $\underline{\theta}^{(1)}$ il n'existe aucun $\underline{\theta}^{(2)}$ tel que $\mathcal{M}^{(2)}(\underline{\theta}^{(2)}) \equiv \mathcal{M}^{(1)}(\underline{\theta}^{(1)})$
- Si $\mathcal{M}^{(1)}(\cdot)$ est *s.d.* de $\mathcal{M}^{(2)}(\cdot)$ et $\mathcal{M}^{(2)}(\cdot)$ est *s.d.* de $\mathcal{M}^{(1)}(\cdot)$ on dira que $\mathcal{M}^{(1)}(\cdot)$ et $\mathcal{M}^{(2)}(\cdot)$ sont structurellement discernables.

III.1 ESTIMATEURS EMPIRIQUES

Estimateurs construits d'après des considérations intuitives, de façon artisanale...

Pas de règles systématiques, mais quelques principes exploitables...

■ Méthode des moments

Soient $\{y_n\}_{n=1 \dots N}$ des réalisations indépendantes d'une même variable aléatoire Y

- Loi de la variable Y : $\begin{cases} f_{Y|\theta}(y|\theta_1, \theta_2, \dots, \theta_M) & (Y \text{ à valeur continue}) \\ P_{Y|\theta}[y|\theta_1, \theta_2, \dots, \theta_M] & (Y \text{ à valeur discrète}) \end{cases}$
- Calcul des moments de $f_{Y|\theta}$ en fonction de $\underline{\theta}$: $\mu_Y^{(k)}(\theta_1, \theta_2, \dots, \theta_M) = E_{Y|\theta}\{y^k | \underline{\theta}\}$ (centrés ou non centrés?)
- Estimation empirique de ces moments à partir des données : $\hat{\mu}_Y^{(k)} = \frac{1}{N} \sum_{n=1}^N y_n^k$
- Résolution du système (généralement non linéaire en $\underline{\theta}$) :

$$\mu_Y^{(k)}(\theta_1, \theta_2, \dots, \theta_M) = \hat{\mu}_Y^{(k)}, \forall k = 1 \dots M$$

CHAP III CONSTRUCTION D'UN ESTIMATEUR

- 1/ Estimateurs empiriques
- 2/ Estimateur du maximum de vraisemblance
- 3/ Estimateurs bayésiens

■ Démarche pour le calcul d'un estimateur pas méthode des moments :

- ① A partir des hypothèses du problème, déduire la loi de la v.a. $Y : y|\underline{\theta}$
- ② Choisir M moments (centrés, non centrés, d'ordre 1, 2, ...) de la loi de $y|\underline{\theta}$ simples à calculer en fonction des paramètres $(\theta_1, \theta_2, \dots, \theta_M)$
- ③ Calculer les estimateurs empiriques de ces moments à partir des données \underline{y}
- ④ Calculer la solution $\underline{\theta}$ du système de M équations à M inconnues déduites des égalités entre moment théoriques et estimateurs empiriques des moments