Introduction au cours.

Contenu et organisation.

- Méthodologies de construction d'estimateurs de paramètres.
- Méthodologies d'évaluation des performances des estimateurs.
- Applications aux communications numériques.

3 intervenants : J.P. Barbot (ENS Cachan), W. Hachem (Supelec), Ph. Loubaton (UMLV).

Exemples: synchronisation

Signal transmis $x(t) = \sum_{n \in \mathbb{Z}} a_n g(t - nT)$.

Dans le cas d'un canal mono-trajet, et présence d'un offset de fréquence

Signal reçu.

$$
\begin{aligned}
y(t) &= \lambda x(t-\tau) e^{2i\pi \delta f_0 t} + b(t) \\
&= \mu e^{i\phi_0} \Big(\sum_{n \in \mathbb{Z}} a_n g(t - \tau - nT)\Big) e^{2i\pi \delta f_0 t} + b(t)
\end{aligned}
$$

Exemples: synchronisation

$$y(t) = \mu e^{i\phi_0} (\sum_{n \in \mathbb{Z}} a_n g(t - \tau - nT)) e^{2i\pi \delta f_0 t} + b(t)$$

A partir d'échantillons prévelés à la période T ou $\frac{T}{2}$, estimer $\phi_0, \delta f_0, \tau$.

- Contexte entrainé : les a_n connus sur la durée d'observation (séquence d'apprentissage).
- Contexte aveugle : les a_n inconnus sur la durée d'observation.
- Contexte semi-aveugle : une partie des a_n est connue sur la durée d'observation.

Exemples: estimation de canal.

Signal transmis $x(t) = \sum_{n \in \mathbb{Z}} a_n g(t - nT)$.

Dans le cas d'un canal multi-trajets-trajet, sans offset de fréquence

Signal reçu : $y(t) = \sum_k \lambda_k x(t - \tau_k) + b(t)$

A partir d'échantillons prévelés à la période T ou $\frac{T}{2}$, estimer les (λ_k, τ_k).

Autre formulation.

Signal reçu : $y(t) = \sum_n a_n h(t - nT) + b(t)$

$$h(t) = \sum_k \lambda_k g(t - \tau_k)$$

Si $y_n = y(nT)$, $y_n = \sum_l h_l a_{n-l} + b_n$

Estimer $h_0, \ldots, h_L$ à partir de $(y_n)_{n=0,\ldots,N-1}$.

Le formalisme de l'estimation statistique.

L'observation: $\mathbf{y} = (y_0, y_1, \ldots, y_N)^T$ vecteur aléatoire.

La densité de probabilité de $\mathbf{y}$ est connue à un nombre fini de paramètres près

Notations: θ le vecteur des paramètres inconnus, de dimension K, $p_\theta(\mathbf{x})$ la densité de probabilité de $\mathbf{y}$.

Le problème: on observe $\mathbf{y} = (y_0, y_1, \ldots, y_N)^T$, et il s'agit d'estimer θ.

Estimateur $\hat{\theta}$ correspondance qui à $\mathbf{y}$ associé un vecteur de dimension K noté $\hat{\theta}(\mathbf{y})$.

Exemple: estimation entrainée de canal.

L'observation $y_n = \sum_{l=0}^{L} h_l a_{n-l} + b_n$, pour $n = 0, \ldots, N-1$.

Estimation entrainée: les symboles $a_{-L}, \ldots, a_0, \ldots, a_{N-1}$ sont connus.

$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{b}$.

- $\mathbf{h} = (h_0, \ldots, h_L)^T$, $\mathbf{b} = (b_0, \ldots, b_{N-1})^T$
- $\mathbf{A} = \begin{pmatrix} a_0 & a_1 & \ldots & a_{-L} \\ a_1 & a_0 & \ldots & a_{-(L-1)} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ a_{N-1} & a_{N-2} & \ldots & a_{N-1-L} \end{pmatrix}$

Exemple: estimation entrainée de canal.

Le paramètre $\theta = \mathbf{h}$.

On suppose que $\mathbf{b}$ est blanc Gaussien complexe de variance σ^2.

$p_{\mathbf{h}}(\mathbf{x}) = \frac{1}{\pi^N \sigma^{2N}} \exp - \frac{\|\mathbf{x}-\mathbf{A}\mathbf{h}\|^2}{\sigma^2}$

Estimateurs "raisonnables"

- Moindre carrés : $\min_{\mathbf{h}} \|\mathbf{y} - \mathbf{A}\mathbf{h}\|^2$, $\hat{\mathbf{h}} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{y}$.
- Si les colonnes de $\mathbf{A}$ sont proches d'être orthogonales, $\hat{\mathbf{h}} = \frac{1}{N} \mathbf{A}^H \mathbf{y}$

Exemple: synchronisation en fréquence.

L'observation: $y_n = \lambda a_n e^{2i\pi n \delta f_0} + b_n$, estimer δf_0.

Estimation entrainée: les symboles $a_L, \ldots, a_0, \ldots, a_N$ sont connus, et de module 1.

λ inconnu, δf_0 inconnu, $\theta = (\lambda, \delta f_0)$

$p_\theta(\mathbf{x}) = \frac{1}{\pi^N \sigma^{2N}} \exp - \frac{\sum_{n=0}^{N-1} |y_n - \lambda a_n e^{2i\pi n \delta f_0}|^2}{\sigma^2}$

Estimateurs "raisonnables"

- $\hat{\delta f_0} = \mathrm{Arg}(\frac{1}{N} \sum_{n=0}^{N-1} y_n y_{n-1}^* a_n^* a_{n-1})$
- $\hat{\delta f_0}$ maximise $\left| \frac{1}{N} \sum_{n=0}^{N-1} y_n a_n^* e^{-2i\pi \delta f} \right|^2$

Comparaison des estimateurs.

Biais.

$$\mathbf{b}_{\hat{\theta}}(\theta) = \mathrm{E}_\theta\left((\hat{\theta}(\mathbf{y}) - \theta\right)$$

Covariance et erreur quadratique moyenne

$$\mathbf{\Gamma}_{\hat{\theta}}(\theta) = \mathrm{E}_\theta\left((\hat{\theta}(\mathbf{y}) - \theta)(\hat{\theta}(\mathbf{y}) - \theta)^H\right)$$

$$R_{\hat{\theta}}(\theta) = \mathrm{E}_\theta\left(\|\hat{\theta}(\mathbf{y}) - \theta\|^2\right) = \mathrm{Trace}(\Gamma_{\hat{\theta}}(\theta))$$

Ces quantités dépendent de θ.

Inégalité de Cramer Rao.

Information de Fisher : $\mathbf{I}(\theta) = \mathrm{E}\left(\frac{\partial}{\partial\theta} \log p_\theta(\mathbf{y}) \frac{\partial}{\partial\theta} \log p_\theta(\mathbf{y})^H\right)$.

Aussi égal à $\mathbf{I}(\theta) = -\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2} \log p_\theta(\mathbf{y})\right)$.

Théorème 1 *Si $\hat{\theta}(\mathbf{y})$ est un estimateur sans biais, alors pour tout θ,*

$$\mathbf{\Gamma}_{\hat{\theta}}(\theta) \geq \mathbf{I}^{-1}(\theta)$$

Exemple: estimation de canal.

$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{b}$.

Information de Fisher: $\mathbf{I}(\mathbf{h}) = \frac{\mathbf{A}^H \mathbf{A}}{\sigma^2}$.

Borne de Cramer Rao : $\sigma^2 \left(\mathbf{A}^H \mathbf{A}\right)^{-1}$.

Coincide avec la variance de l'estimateur des moindres carrés.

Exemple: synchronisation.

$y_n = \lambda a_n e^{2i\pi n \delta f_0} + b_n$, $n = 0, \ldots, N-1$

- Les $(a_n)_{n=0,\ldots,N-1}$ sont connus et de modules 1.
- λ est réel et positif.

Se ramène au modèle: $y_n = \lambda e^{2i\pi n \delta f_0} + b_n$, $n = 0, \ldots, N-1$, $\theta = (\lambda, \delta f_0)^T$.

$$(\mathbf{I}(\theta))^{-1} = \begin{pmatrix} \frac{\sigma^2}{2N} & 0 \\ 0 & \frac{\sigma^2}{\lambda} \frac{3}{4\pi^2 N(N+1)(2N+1)} \end{pmatrix}$$

Estimateur du maximum de vraisemblance.

$\hat{\theta}_{ml}$ maximise en θ $\log p_\theta(y)$.

Exemples.

- Estimation de canal :$\hat{\mathbf{h}}_{ml} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{y}$.
- Synchronisation, λ réel :

$$\begin{aligned}
\delta \hat{f}_{0_{ml}} &= \text{Argmax} \left| \frac{1}{N} \sum_{n=0}^{N-1} y_n e^{-2i\pi n \delta f} \right|^2 \\
\hat{\lambda}_{ml} &= \left| \frac{1}{N} \sum_{n=0}^{N-1} y_n e^{-2i\pi n \delta \hat{f}_{0_{ml}}} \right|
\end{aligned}$$

Propriétés asymptotiques des estimateurs.

Comportement des estimateurs quand $N \to +\infty$.

Définition 1 $\hat{\theta}_N(\mathbf{y})$ *consistant si* $\hat{\theta}_N(\mathbf{y}) \to \theta$ *presque surement.*

Tous les estimateurs "raisonnables" sont consistants.

Définition 2 $\hat{\theta}_N(\mathbf{y})$ *est asymptotiquement gaussien si il existe* $\alpha > 0$ *tel que* $N^{\alpha}\left(\hat{\theta}_N(\mathbf{y}) - \theta\right) \to \mathcal{N}(0, \mathbf{\Delta}(\theta))$.

$\mathcal{N}(0, \mathbf{\Delta}(\theta))$ gaussienne centrée de covariance $\mathbf{\Delta}(\theta)$.

Propriétés asymptotiques de l'estimateur du maximum de vraisemblance.

Les $(y_n)_{n=0,\ldots,N-1}$ indépendantes de même loi $p_\theta(x)$.

$\log p_\theta(\mathbf{y}) = \sum_{n=0}^{N-1} \log p_\theta(y_n)$

$$(\hat{\theta}_N - \theta_0) \simeq - \frac{\left(\frac{\partial}{\partial\theta} \log p_\theta(\mathbf{y})\right)_{\theta_0}}{\left(\frac{\partial^2}{\partial\theta^2} \log p_\theta(\mathbf{y})\right)_{\theta_0}}$$

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \simeq - \frac{\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \left(\frac{\partial}{\partial\theta} \log p_\theta(y_n)\right)_{\theta_0}}{\frac{1}{N} \sum_{n=0}^{N-1} \left(\frac{\partial^2}{\partial\theta^2} \log p_\theta(y_n)\right)_{\theta_0}}$$

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \simeq \mathcal{N}\left(0, \mathbf{I}^{-1}(\theta_0)\right)$$

$\hat{\theta}_N$ **est asymptotiquement efficace.**

Méthode des moments généralisés.

On suppose que $\mathrm{E}_\theta(\mathbf{g}(y_n)) = \mathbf{h}(\theta)$.

On peut alors estimer θ par $\hat{\theta}$ défini par

$$\hat{\theta} = \mathrm{Argmin} V_N(\mathbf{y}, \theta)$$

où

$$V_N(\mathbf{y}, \theta) = \left\| \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{g}(y_n) - \mathbf{h}(\theta) \right\|$$

Méthode des moments généralisés: exemple.

synchronisation aveugle sans résidu de porteuse et avec symboles binaires.

$y_n = \mu e^{i\phi} a_n + b_n$ avec $a_n = \pm 1$ inconnus, σ^2 connu. Estimer $\theta = (\mu, \phi)$.

Utiliser que $\mathrm{E}_\theta(y_n^2) = \mu^2 e^{2i\phi}$ et $\mathrm{E}_\theta(|y_n|^2) = \mu^2 + \sigma^2$.

Estimer $\mathrm{E}_\theta(y_n^2)$ par $\frac{1}{N}\sum_{n=0}^{N-1} y_n^2$

Estimer $\mathrm{E}_\theta(|y_n|^2)$ par $\frac{1}{N}\sum_{n=0}^{N-1} |y_n|^2$

Puis ajuster μ et ϕ.

$$\begin{aligned}
\hat{\phi} &= \frac{1}{2}\mathrm{Arg}\left(\frac{1}{N}\sum_{n=0}^{N-1} y_n^2\right) \\
\hat{\mu}^2 &= \frac{1}{2}\left(\left|\frac{1}{N}\sum_{n=0}^{N-1} y_n^2\right| + \frac{1}{N}\sum_{n=0}^{N-1} |y_n|^2\right)
\end{aligned}$$

Propriétés asymptotiques de la méthode des moments I.

Faire un développement limité au second de $V_N(\mathbf{y}, \theta)$ au voisinage du vrai paramètre θ_0.

$$\hat{\theta}_N - \theta_0 \simeq - \left(\frac{\partial^2}{\partial \theta^2} V_N(\mathbf{y}, \theta) \right)^{-1}_{\theta=\theta_0} \left(\frac{\partial}{\partial \theta} V_N(\mathbf{y}, \theta) \right)_{\theta=\theta_0}$$

Etudier les comportements des 2 termes.

Propriétés asymptotiques de la méthode des moments II.

La dérivée seconde:

$$\frac{\partial^2 V_N}{\partial \theta_k \theta_l} = 2\frac{\partial \mathbf{h}^T}{\partial \theta_k}\frac{\partial \mathbf{h}}{\partial \theta_l} - 2\frac{\partial^2 \mathbf{h}^T}{\partial \theta_k \theta_l}\left(\frac{1}{N}\sum_{n=1}^{N} g(y_n) - \mathbf{h}(\theta)\right)$$

Or, $\frac{1}{N}\sum_{n=1}^{N} g(y_n) - \mathbf{h}(\theta_0) \rightarrow 0$**.**

Le terme de dérivée seconde converge donc vers $2\left(\frac{\partial \mathbf{h}^T}{\partial \theta}\frac{\partial \mathbf{h}}{\partial \theta}\right)_{\theta=\theta_0}$

Propriétés asymptotiques de la méthode des moments II.

La dérivée première:

$$\frac{\partial V_N}{\partial \theta} = -2 \left(\frac{\partial \mathbf{h}^T}{\partial \theta} \right) \left(\frac{1}{N} \sum_{n=1}^{N} g(y_n) - \mathbf{h}(\theta) \right)$$

Or, dans la plupart des cas, $\sqrt{N} \left[\frac{1}{N} \sum_{n=1}^{N} g(y_n) - \mathbf{h}(\theta_0) \right] \rightarrow \mathcal{N}(0, \Gamma(\theta_0))$.

Propriétés asymptotiques de la méthode des moments III.

Conclusion:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \to \mathcal{N}(0, \Delta(\theta_0))$$

avec

$$\Delta(\theta_0) = \left(\frac{\partial \mathbf{h}^T}{\partial \theta}\frac{\partial \mathbf{h}}{\partial \theta}\right)_{\theta=\theta_0} \left((\frac{\partial \mathbf{h}^T}{\partial \theta})_{\theta_0}\Gamma(\theta_0)(\frac{\partial \mathbf{h}}{\partial \theta})_{\theta_0}\right) \left(\frac{\partial \mathbf{h}^T}{\partial \theta}\frac{\partial \mathbf{h}}{\partial \theta}\right)_{\theta=\theta_0}$$

Estimation de canal.

Signal transmis $x(t) = \sum_{n \in \mathbb{Z}} a_n g(t - nT)$.

Signal reçu : $y(t) = \sum_{k=1}^{K} \lambda_k x(t - \tau_k) + b(t)$

Signal reçu : $y(t) = \sum_n a_n h(t - nT) + b(t)$

$$h(t) = \sum_{k=1}^{K} \lambda_k g(t - \tau_k)$$

Si $y_n = y(nT)$ et $h_l = h(lT) = \sum_{k=1}^{K} \lambda_k g(lT - \tau_k)$, alors $y_n = \sum_l h_l a_{n-l} + b_n$

Estimation entrainée standard.

L'observation $y_n = \sum_{l=0}^{L} h_l a_{n-l} + b_n$, pour $n = 0, \ldots, N-1$.

Estimation entrainée: les symboles $a_{-L}, \ldots, a_0, \ldots, a_{N-1}$ **sont connus.**

$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{b}$.

- $\mathbf{h} = (h_0, \ldots, h_L)^T$, $\mathbf{b} = (b_0, \ldots, b_{N-1})^T$
- $\mathbf{A} = \begin{pmatrix} a_0 & a_1 & \ldots & a_{-L} \\ a_1 & a_0 & \ldots & a_{-(L-1)} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ a_{N-1} & a_{N-2} & \ldots & a_{N-1-L} \end{pmatrix}$

$\hat{\mathbf{h}} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{y}$.

Estimation entrainée multi-slots I.

Estimer le canal sur chaque slot à partir d'une séquence d'apprentissage.

Sur le slot m, $\mathbf{h}_m = \sum_{k=1}^{K} \lambda_{k,m} \mathbf{g}(\tau_k)$, avec :

- les $(\lambda_{k,m})_{m=1,\ldots,M}$ variables aléatoires centrées de même loi
- $K < (L+1)$
- $\mathbf{g}(\tau_k)$ fixe sur un nombre de slots raisonnables.

Profiter de ce modèle pour améliorer l'estimation de $\mathbf{h}_m$ pour tout m.

Estimation entrainée multi-slots II.

$\Gamma = \mathbb{E}(\mathbf{h}_m \mathbf{h}_m^H)$ matrice de rang $K < L$.

Si Γ était connue : $\Gamma = \mathbf{U}\Delta\mathbf{U}^H$ la décomposition en elts propres.

$\mathbf{h}_m = \mathbf{U}\mathbf{f}_m$ avec $\mathbf{U}$ orthogonale, et $\mathbf{f}_m$ de dimension K à composantes indépendantes.

Estimer $\mathbf{h}_m$ équivaut à estimer $\mathbf{f}_m$: moins de paramètres, estimation meilleure.

En pratique, il faut estimer Γ à partir des $(\hat{\mathbf{h}}_m)_{m=1,\dots,M}$.

Estimation entrainée multi-slots III.

Estimation de Γ: on suppose que $\mathbf{A}^H\mathbf{A} = N\mathbf{I}_N$.

$$\hat{\mathbf{h}}_m = \mathbf{h}_m + \epsilon_m \text{ avec } \mathbb{E}(\epsilon_m \epsilon_m^H) = \frac{\sigma^2}{N}\mathbf{I}_{L+1}$$

$$\mathbb{E}(\hat{\mathbf{h}}_m \hat{\mathbf{h}}_m^H = \Gamma + \tfrac{\sigma^2}{N}\mathbf{I}_{L+1}.$$

Estimer Γ par

$$\hat{\Gamma} = \frac{1}{M}\sum_{m=1}^{M} \hat{\mathbf{h}}_m \hat{\mathbf{h}}_m^H - \frac{\sigma^2}{N}\mathbf{I}_{L+1}$$

Estimation aveugle / semi-aveugle.

estimer **h** à partir de la séquence d'apprentissage **et** et des observations correspondant aux symboles inconnus.

Quelques notations.

- $\mathbf{h} = \theta$
- $(a_n, \ldots, a_{n-L})^T = \mathbf{z}_n$
- $\mathbf{y} = (y_0, \ldots, y_{N-1})^T$
- $\mathbf{z} = (\mathbf{z}_0^T, \ldots, \mathbf{z}_{N-1})^T$

Estimer θ à partir de $\mathbf{y}$, $y_n = \theta^T \mathbf{z}_n + b_n$

Estimation au sens du maximum de vraisemblance.

Maximiser $p(\mathbf{y}/\theta)$.

$$p(\mathbf{y}/\theta) = \sum_{\mathbf{z}} p(\mathbf{y}, \mathbf{z}/\theta)$$

$$p(\mathbf{y}, \mathbf{z}/\theta) = p(\mathbf{y}/\mathbf{z}, \theta) p(\mathbf{z})$$

$$p(\mathbf{y}/\mathbf{z}, \theta) = \frac{1}{(2\pi)^{N/2}} \exp - \frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} |y_n - \theta^T \mathbf{z}_n|^2 \right)$$

Impossible de calculer $p(\mathbf{y}/\theta)$.

Algorithme EM.

Algorithme itératif : supposons que $\theta^{[k]}$ soit disponible.

On considère $p(\mathbf{z}/\mathbf{y}, \theta^{[k]})$ la loi conditionnelle de $\mathbf{z}$ sachant $\mathbf{y}$ si le paramètre était $\theta^{[k]}$.

$$Q(\theta, \theta^{[k]}) = \int \log p(\mathbf{y}, \mathbf{z}/\theta) \, p(\mathbf{z}/\mathbf{y}, \theta^{[k]}) \, d\mathbf{z}$$

$\theta^{[k+1]}$ est la valeur de θ qui maximise $Q(\theta, \theta^{[k]})$.

Théorème 2 *Au cours de l'algorithme, la vraisemblance augmente, c'est-à-dire que*

$$\log p(\mathbf{y}/\theta^{[k+1]}) \geq \log p(\mathbf{y}/\theta^{[k]})$$