# Séminaire TEST

Andrés SÁNCHEZ PÉREZ

October 18th, 2013

## 1 Présentation du sujet

Le problème de régression non-paramétrique se pose de la façon suivante :
Supposons que l'on dispose de $n$ couples indépendantes de v.a. $(X_1, Y_1), \ldots, (X_T, Y_T)$ telles que :

$$Y_i = \theta(X_i) + \xi_i, \quad X_i \in [0, 1] \ ,$$

où $\mathbb{E}(\xi_i) = 0$ pour tout $i$ et la fonction dite de régression $\theta : [0, 1] \to \mathbb{R}$ est inconnue. On cherche à estimer $\theta$ qui appartient à un ensemble non-paramétrique (infini-dimensionnel) $\Theta$.
Ça fait partie d'une classe plus large de problèmes, celle de la estimation non paramètrique. On utilisera comme ensemble $\Theta$ la classe de Hölder $\Sigma(\beta, L)$.

**Definition 1 (Classe de Hölder)** *Pour $I \subseteq \mathbb{R}$, soient $\beta > 0, L > 0$. La classe de Hölder sur $I$ est l'ensemble de toutes les fonctions $f : I \to \mathbb{R}$ telles que la dérivée $f^{(\ell)}$, $\ell = \lfloor \beta \rfloor$ existe et vérifie*

$$\left| f^{(\ell)}(x) - f^{(\ell)}(x') \right| \leq L \left| x - x' \right|^{\beta - \ell} \ , \ \forall x, x' \in I \ .$$

Il y a trois éléments qui caracterisent un problème d'estimation non-paramétrique :
- Une classe non-paramètrique de fonctions $\Theta$ contenant la "vraie" fonction $\theta$.
- Une famille $\{\mathbb{P}_\theta, \theta \in \Theta\}$ de mesures de probabilité sur un espace mesurable $(\mathcal{X}_T, \mathcal{A}_T)$ associé à un $T-$échantillon.
- Une semi-distance $d$ sur $\Theta$ utilisée pour définir le risque.

La performance d'un estimateur $\hat{\theta}_T$ de $\theta$ est mesurée par son risque maximal sur $\theta$ :

$$r\left(\hat{\theta}_T\right) = \sup_{\theta \in \Theta} \mathbb{E}_\theta \left[ d^2\left(\hat{\theta}_T, \theta\right) \right] \ .$$

Si pour certains estimateurs on a des inégalités du type :

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta \left[ d^2\left(\hat{\theta}_T, \theta\right) \right] \leq C \psi_T^2 \ , \tag{1}$$

avec $\psi_T \to 0$ et $C < \infty$, on serait intérésés à obtenir des bornes inférieures correspondantes :

$$\forall \hat{\theta}_T, \ \sup_{\theta \in \Theta} \mathbb{E}_\theta \left[ d^2\left(\hat{\theta}_T, \theta\right) \right] \geq c \psi_T^2 \ . \tag{2}$$

**Definition 2 (Risque minimax)**

$$\mathcal{R}_T = \inf_{\hat{\theta}_T} \sup_{\theta \in \Theta} \mathbb{E}_\theta \left[ d^2\left(\hat{\theta}_T, \theta\right) \right]$$

**Definition 3 (Vitesse optimale de convergence)** *La suite $\{\psi_T\}_{T\geq 1}$ est dite vitesse optimale de convergence des estimateurs sur $(\Theta, d)$ si 1 et 2 sont verifiées. Un estimateur $\theta_T^*$ qui vérifie :*

$$\sup_{\theta\in\Theta} \mathbb{E}_\theta \left[ d^2 \left(\theta_T^*, \theta\right)\right] \leq C' \psi_T^2 \ ,$$

*est dit estimateur optimal en vitesse de convergence sur $(\Theta, d)$.*

On peut considérer un cadre plus général où le risque maximal es défini par

$$r_w\left(\hat{\theta}_T\right) = \sup_{\theta\in\Theta} \mathbb{E}_\theta \left[ w\left(\psi_T^{-1} d\left(\hat{\theta}_T, \theta\right)\right)\right] \ ,$$

avec une fonction perte $w : \mathbb{R}_+ \to \mathbb{R}_+$ monotone croissante, non idéntiquement nulle et telle que $w(0) = 0$. Exemple : $w(u) = u^p$, $p > 0$.

# 2 Rappels de théorie de la mesure

**Definition 4** *Soit $\nu$ une mesure positive sur $(\mathcal{X}, \mathcal{A})$ et soit $\rho, \bar{\rho}$ des mesures positives (resp. réelles, resp. complexes) sur $(\mathcal{X}, \mathcal{A})$.*

- *On dit que $\rho$ est absolument continue par rapport à $\nu$, et l'on note $\rho \ll \nu$, si pour tout $A \in \mathcal{A}$ tel que $\nu(A) = 0$, on a également $\rho(A) = 0$. Si $\nu$ et $\rho$ sont $\sigma-$ finies $\rho \ll \nu \Leftrightarrow \rho(A) = \int_A \frac{\mathrm{d}\rho}{\mathrm{d}\nu} \mathrm{d}\nu$. $\frac{\mathrm{d}\rho}{\mathrm{d}\nu}$ est la densité de $\rho$ par rapport à $\nu$, qui, dans ce cas est une fonction mesurable positive (resp. -intégrable réelle, resp. -intégrable complexe).*
- *On dit que $\rho$ est concentrée sur $E \in \mathcal{A}$ si pour tout $A \in \mathcal{A}$ on a $\rho(A) = \rho(A \cap E)$, ou bien encore $\rho(A \backslash E) = 0$.*
- *On dit que $\rho$ et $\bar{\rho}$ sont étrangères, et l'on note $\rho \perp \bar{\rho}$, s'il existe $E \in \mathcal{A}$ telle que $\rho$ soit portée par $E$ et $\bar{\rho}$ soit portée par $E^c$.*

**Theorem 2.1 (Radon-Nikodym-Lebesgue)** *Soient $\nu$ une mesure positive $\sigma-$ finie sur $(\mathcal{X}, \mathcal{A})$ et $\mu$ une mesure positive $\sigma-$finie (resp. réelle, resp. complexe) sur $(\mathcal{X}, \mathcal{A})$. Alors :*

*(i) Il existe un unique couple de mesures positives $\sigma-$finies (resp. réelles, resp. complexes) tel que :*
- *$\mu = \mu_1 + \mu_2$,*
- *$\mu_1 \ll \nu$,*
- *$\mu_2 \perp \nu$*

*Cette décomposition s'appelle la décomposition de Lebesgue de $\mu$.*

*(ii) Il existe une unique (à égalité -presque partout près) fonction $h$, $\nu$ -intégrable mesurable positive (resp. réelle, resp. complexe), telle que pour tout $A \in \mathcal{A}$ on ait :*

$$\mu_1(A) = \int_A h \mathrm{d}\nu = \int_X \mathbb{1}_A h \mathrm{d}\nu$$

*Cette fonction s'appelle la dérivée de Radon-Nykodym de $\mu_1$ par rapport à $\nu$.*

Supposons que $\nu$ est une mesure $\sigma-$ finie sur $(\mathcal{X}, \mathcal{A})$ telle que $P_0 \ll \nu$, $P_1 \ll \nu$, on définit $p_0 = \frac{\mathrm{d}P_0}{\mathrm{d}\nu}$, $p_1 = \frac{\mathrm{d}P_1}{\mathrm{d}\nu}$. $\nu$ existe toujours, par exemple, $\nu = P_0 + P_1$.

**Definition 5 (Divergence de Kullback-Leibler)** *Entre $P_0$ et $P_1$ est définie par :*

$$\mathcal{K}(P_0, P_1) = \begin{cases} \int \log \frac{\mathrm{d}P_0}{\mathrm{d}P_1} \mathrm{d}P_0 & \text{, si } P_0 \ll P_1 \\ +\infty & \text{, sinon.} \end{cases}$$

Si $P_0 \ll P_1$, on a $\{p_1 > 0\} \supseteq \{p_0 > 0\}$, $\{p_0 p_1 > 0\} = \{p_0 > 0\}$. De plus, l'intégrale dans la définition 5 est calculée sur l'ensemble $\{p_0 > 0\}$. On en déduit que si $P_0 \ll P_1$ :

$$\mathcal{K}(P_0, P_1) = \int_{p_0 p_1 > 0} p_0 \log \frac{p_0}{p_1} \mathrm{d}\nu \, .$$

**Lemma 1 (Inégalité de Le Cam, 1973)**

$$\int \min(p_0, p_1) \, \mathrm{d}\nu \geq \frac{1}{2} \left( \int \sqrt{p_0 p_1} \mathrm{d}\nu \right)^2$$

**Proof du Lemme 1**

$$\begin{aligned} \left( \int \sqrt{p_0 p_1} \mathrm{d}\nu \right)^2 &= \left( \int_{p_0 \geq p_1} \sqrt{p_0 p_1} \mathrm{d}\nu + \int_{p_0 < p_1} \sqrt{p_0 p_1} \mathrm{d}\nu \right)^2 \\ &\leq 2 \left[ \left( \int_{p_0 \geq p_1} \sqrt{p_0 p_1} \mathrm{d}\nu \right)^2 + \left( \int_{p_0 < p_1} \sqrt{p_0 p_1} \mathrm{d}\nu \right)^2 \right] \\ &\leq 2 \left[ \int p_0 \mathrm{d}\nu \int_{p_0 \geq p_1} p_1 \mathrm{d}\nu + \int p_1 \mathrm{d}\nu \int_{p_0 < p_1} p_0 \mathrm{d}\nu \right] \\ &= 2 \int \min(p_0, p_1) \, \mathrm{d}\nu \end{aligned}$$

■

**Lemma 2**

$$\int \min(p_0, p_1) \, \mathrm{d}\nu \geq \frac{1}{2} \exp\left(-2\mathcal{K}(P_0, P_1)\right) \, .$$

**Proof du lemme 2**
Grâce à l'inégalité de Jensen,

$$\begin{aligned} \left( \int \sqrt{p_0 p_1} \mathrm{d}\nu \right)^2 &= \exp\left( 2 \log \int_{p_0 p_1 > 0} \sqrt{p_0 p_1} \mathrm{d}\nu \right) \\ &= \exp\left( 2 \log \int_{p_0 p_1 > 0} p_0 \sqrt{\frac{p_1}{p_0}} \mathrm{d}\nu \right) \\ &= \exp\left( 2 \log \int_{p_0 p_1 > 0} \sqrt{\frac{\mathrm{d}P_1}{\mathrm{d}P_0}} \mathrm{d}P_0 \right) \end{aligned}$$

$$
\begin{aligned}
&\geq \exp\left(2 \int_{p_0 p_1 > 0} \log \sqrt{\frac{\mathrm{d}P_1}{\mathrm{d}P_0}} \mathrm{d}P_0\right) \\
&= \exp\left(-2\mathcal{K}\left(P_0, P_1\right)\right)
\end{aligned}
$$

On complète la preuve en utilisant l'Inégalité de Le Cam.

■

# 3 Schéma général de réduction

a) Réduction aux bornes en probabilité :
Soit $A$ une constante telle que $w(A) > 0$, l'inégalité de Markov conduit à :

$$
\inf_{\hat{\theta}_T} \sup_{\theta \in \Theta} \mathbb{E}_\theta \left[ w\left(\psi_T^{-1} d\left(\hat{\theta}_T, \theta\right)\right)\right] \geq w(A) \inf_{\hat{\theta}_T} \sup_{\theta \in \Theta} \mathbb{P}_\theta \left(d\left(\hat{\theta}_T, \theta\right) \geq s\right)
$$

avec $s = A\psi_T$.

b) Réduction au problème de test d'un nombre fini d'hypothèses :

$$
\inf_{\hat{\theta}_T} \sup_{\theta \in \Theta} \mathbb{P}_\theta \left(d\left(\hat{\theta}_T, \theta\right) \geq s\right) \geq \inf_{\hat{\theta}_T} \max_{\theta \in \{\theta_0, \ldots, \theta_M\}} \mathbb{P}_\theta \left(d\left(\hat{\theta}_T, \theta\right) \geq s\right) \tag{3}
$$

where $\{\theta_0, \ldots, \theta_M\} \subset \Theta$. Cet ensemble doit etre convenablement choisi.

c) Choix des hypothèses séparées d'une distance d'au moins $2s$ :
Si :

$$
d\left(\theta_j, \theta_k\right) \geq 2s, \quad k \neq j,
$$

soit $\psi^* = \underset{0 \leq k \leq M}{\arg\min}\, d\left(\hat{\theta}_T, \theta_k\right)$ le test du minimum de distance, alors

$$
\psi^* \neq j \Rightarrow d\left(\theta_{\psi^*}, \hat{\theta}_T\right) + d\left(\hat{\theta}_T, \theta_j\right) \geq d\left(\theta_{\psi^*}, \theta_j\right) \geq 2s \Rightarrow d\left(\hat{\theta}_T, \theta_j\right) \geq s \ ,
$$

et donc $\mathbb{P}_{\theta_j}\left(d\left(\hat{\theta}_T, \theta_j\right) \geq s\right) \geq \mathbb{P}_{\theta_j}\left(\psi^* \neq j\right), \forall j = 0, \ldots, M$.

$$
\inf_{\hat{\theta}_T} \max_{\theta \in \{\theta_0, \ldots, \theta_M\}} \mathbb{P}_\theta \left(d\left(\hat{\theta}_T, \theta\right) \geq s\right) \geq \inf_{\psi} \max_{0 \leq j \leq M} \mathbb{P}_{\theta_j}\left(\psi \neq j\right) \ . \tag{4}
$$

L'infimum dans le terme de droite est pris dans l'ensemble de tous les *tests*, i.e. fonctions $(Y_1, \ldots, Y_T)$ − mesurables $\psi$ à valeurs dans $\{0, \ldots, M\}$. On note :

$$
p_{e,M} = \inf_{\psi} \max_{0 \leq j \leq M} \mathbb{P}_{\theta_j}\left(\psi \neq j\right) \ .
$$

L'objectif est d'obtenir une borne inférieure pour $p_{e,M}$ qui soit indépendante de $T$. Dans la pratique 3 conditions sont cherchées afin de borner $p_{e,M}$ :

1. Assurer 3 : $\theta_j \in \Theta, \forall j = 0, \ldots, M$.
2. Assurer 4 : $d\left(\theta_j, \theta_k\right) \geq 2s > 0$, $\forall 0 \leq j < k \leq M$.
3. Assurer que $p_{e,M} \geq c > 0$ :
   – **Theorem 3.1** *Soient $P_0, P_1$ deux mesures de probabilité sur $\mathcal{A}$.
   Si $\mathcal{K}(P_0, P_1) \leq \alpha < \infty$, alors :*

$$
p_{e,1} \geq \frac{1}{2}\exp(-2\alpha)
$$

– **Theorem 3.2** *Supposons que $\Theta$ contient $M \geq 2$ éléments $\theta_0, \ldots, \theta_M$ tels que :*
*(i)* $d\left(\theta_j, \theta_k\right) \geq 2s > 0,\ \forall 0 \leq j < k \leq M,$
*(ii)* $P_j \ll P_0,\ \forall 1 \leq j < k \leq M$, *et*
*(iii)* $\frac{1}{M} \sum_{j=1}^{M} K\left(\mathbb{P}_{\theta_j}, \mathbb{P}_{\theta_0}\right) \leq \alpha \log M$, *avec* $\alpha \in \left(0, \frac{1}{4}\right)$.
*Alors,*

$$p_{e,M} \geq \frac{1}{2}\left(1 - \alpha - \sqrt{\frac{2\alpha}{\log M}}\right) > 0 \ .$$

**Proof du Théorème 3.1**

$$\begin{aligned}
p_{e,1} &= \inf_{\psi} \max_{j=0,1} P_j\left(\psi \neq j\right) \\
&\geq \inf_{\psi} \left[P_0\left(\psi \neq 0\right) + P_1\left(\psi \neq 1\right)\right] \\
&= \inf_{A \in \mathcal{A}} \int \left[\mathbb{1}_A p_0 + \mathbb{1}_{A^c} p_1\right] \mathrm{d}\nu \\
&= \int \min\left(p_0, p_1\right) \mathrm{d}\nu \\
&= P_0\left(\psi^* \neq 0\right) + P_1\left(\psi^* \neq 1\right)
\end{aligned}$$

L'ensemble $A$ où l'infimum est atteint est $A^* = \{p_0 \leq p_1\}$ et en conséquence $\psi^* = \mathbb{1}_{A^*}$. On utilise le lemme 2 et cela donne :

$$\begin{aligned}
p_{e,1} &\geq \int \min\left(p_0, p_1\right) \mathrm{d}\nu \\
&\geq \frac{1}{2} \exp\left(-2\mathcal{K}\left(P_0, P_1\right)\right) \\
&\geq \frac{1}{2} \exp\left(-2\alpha\right) \ .
\end{aligned}$$

■

# 4 Borne inférieure pour la régression en un point

Exemple d'application des bornes basées sur deux hypothèses dans le modèle de régression vérifiant les conditions suivantes.

– Le modèle statistique est celui de régression non-paramétrique :

$$Y_i = f\left(X_i\right) + \xi_i \ , \quad , i = 1, \ldots, T$$

où $f : [0,1] \to \mathbb{R}$.

– Les variables aléatoires $\xi_i$ sont i.i.d., de densité $p_\xi$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$, vérifiant :

$$\exists p_* > 0, v_0 > 0 : \quad \int p_\xi\left(u\right) \log \frac{p_\xi\left(u\right)}{p_\xi\left(u+v\right)} \mathrm{d}u \leq p_* v^2 \ ,$$

pour tout $|v| \leq v_0$.

– Les $X_i \in [0,1]$ sont déterministes.

– Il existe un $a_0 \in \mathbb{R}$ tel que, pour tout intervalle $A \subseteq [0,1]$ et tout $T \geq 1$,

$$\frac{1}{T}\sum_{i=1}^{T} \mathbb{1}_{\{X_i \in A\}} \leq a_0 \max\left(\mathrm{Leb}\,(A)\,, \frac{1}{T}\right)$$

La deuxième condition est satisfaite si, par exemple, $p_\xi$ est la densité de la loi normale $\mathcal{N}\left(0, \sigma^2\right)$, $\sigma > 0$.
Notre objectif est d'établir une borne inférieure pour le risque minimax sur $(\Theta, d)$, où $\Theta = \Sigma(\beta, L)$ est une classe de Hölder et pour la semi-distance $d$, la distance en un point fixé $x \in [0,1]$ : $d(f,g) = |f(x_0) - g(x_0)|$. La vitesse qu'on veut obtenir est $\psi_T = T^{-\frac{\beta}{2\beta+1}}$, la meme que dans les bornes supérieures des estimateurs par polynomes locaux par exemple (page 36).
Il nous suffira de prendre $M = 1$ (2 hypothèses) :

$$\theta_0(x) \equiv 0, \qquad \theta_1(x) = L h_T^\beta K\left(\frac{x - x_0}{h_T}\right), \qquad x \in [0,1] \ ,$$

où $h_T = c_0 T^{-\frac{1}{2\beta+1}}$, $c_0 > 0$, $L > 0$ et la fonction $K : \mathbb{R} \to [0, +\infty)$ vérifie :
$K \in \Sigma\left(\beta, \frac{1}{2}\right) \cap C^\infty(\mathbb{R})$, $K(u) > 0 \Leftrightarrow u \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Les fonctions $K$ vérifiants cette condition existent. Par exemple, on peut prendre, avec un $a > 0$ suffisamment petit,

$$K(u) = aK_0(2u)\,, \qquad \text{où} \qquad K_0(u) = \exp\left(-\frac{1}{1-u^2}\right)\mathbb{1}_{|u|\leq 1} \ .$$

Pour pouvoir utiliser le Théorème 3.1 on doit vérifier trois conditions :
– La condition $\theta_j \in \Sigma(\beta, L)$, $j = 0, 1$.
  Pour $\ell = \lfloor \beta \rfloor$, la dérivée $\ell-$ ème de $\theta_1$ vaut $\theta_1^{(\ell)}(x) = L h_T^{\beta-\ell} K^{(\ell)}\left(\frac{x - x_0}{h_T}\right)$ et alors :

$$\begin{aligned}
\left|\theta_1^{(\ell)}(x) - \theta_1^{(\ell)}(x')\right| &= L h_T^{\beta-\ell}\left|K^{(\ell)}\left(\frac{x - x_0}{h_T}\right) - K^{(\ell)}\left(\frac{x' - x_0}{h_T}\right)\right| \\
&\leq \frac{L h_T^{\beta-\ell}}{2}\left|\frac{x - x_0}{h_T} - \frac{x' - x_0}{h_T}\right|^{\beta-\ell} \\
&= \frac{L}{2}\left|x - x'\right|^{\beta-\ell}
\end{aligned}$$

  Donc $\theta_1 \in \Sigma\left(\beta, \frac{L}{2}\right) \subset \Sigma(\beta, L)$.
– La condition $d(\theta_0, \theta_1) \geq 2s$.

$$d(\theta_0, \theta_1) = |\theta_1(x_0)| = L h_T^\beta K(0) = L c_0^\beta K(0)\, n^{-\frac{\beta}{2\beta+1}}$$

  Et la condition est satisfaite avec $A = \frac{1}{2} L c_0^\beta K(0)$.
– La condition $\mathcal{K}(P_0, P_1) \leq \alpha$.
  Notons que $P_j$ (la loi de $Y_1, \ldots, Y_T$ pour $\theta_j$) admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^T$ de la forme

$$p_j(u_1, \ldots, u_T) = \prod_{i=1}^{T} p_\xi\left(u_i - \theta_j(X_i)\right), \quad j = 0, 1$$

Il existe un entier $T_0$ tel que pour tout $T > T_0$ on a $Th_T \geq 1$ et $Lh_T^{\beta} K_{\max} \leq v_0$ où $K_{\max} = \max_u K(u)$ et $T_0$ ne dépend que de $c_0, L, \beta, K_{\max}, v_0$. Donc :

$$
\begin{aligned}
\mathcal{K}(P_0, P_1) &= \int \log \frac{\mathrm{d}P_0}{\mathrm{d}P_1} \mathrm{d}P_0 \\
&= \int \dots \int \log \prod_{i=1}^{T} \frac{p_\xi(u_i)}{p_\xi(u_i - \theta_1(X_i))} \prod_{i=1}^{T} \left[ p_\xi(u_i)\,\mathrm{d}u_i \right] \\
&= \sum_{i=1}^{T} \int \log \frac{p_\xi(y)}{p_\xi(y - \theta_1(X_i))} p_\xi(y)\,\mathrm{d}y \\
&\leq p_* \sum_{i=1}^{T} \theta_1^2(X_i) \\
&= p_* L^2 h_T^{2\beta} \sum_{i=1}^{T} K^2\left(\frac{X_i - x_0}{h_T}\right) \\
&\leq p_* L^2 h_T^{2\beta} K_{\max}^2 \sum_{i=1}^{T} \mathbb{1}_{\left\{\left|\frac{X_i - x_0}{h_T}\right| \leq \frac{1}{2}\right\}} \\
&\leq p_* a_0 L^2 K_{\max}^2 h_T^{2\beta} \max(Th_T, 1) \\
&= p_* a_0 L^2 K_{\max}^2 T h_T^{2\beta+1}.
\end{aligned}
$$

Si on choisit $c_0 = \left(\frac{\alpha}{p_* a_0 L^2 K_{\max}^2}\right)^{\frac{1}{2\beta+1}}$, on obtient $\mathcal{K}(P_0, P_1) \leq \alpha$.