Bornes inférieures minimax

Eddie Aamari
13/01/16

Références : . Bin Yu - Assouad, Fano, Lecam - 1997
. A. Tsybakov - Introduction to Nonparametric Estimation - 2009
. I. Johnstone - Gaussian Estimation : Sequence & Wavelet models - (draft)
. L. Gibbs, F. Su - On Choosing & Bounding Probability Metrics - 2002

<u>But</u> : Donner des techniques classiques pour minorer des risques minimax du type,

$$\inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \, l(\hat{\theta}, \theta) \geq \cdots > 0 .$$

<u>Plan</u> : 1/ Risque minimax, optimalité

2/ Risques minimax fréquentiste et bayésien

3/ Théorèmes usuels de bornes inférieures

a) Lemme de Lecam

a') Proximités stochastiques

b) Comparaison de plusieurs hypothèses

c) Lemme d'Assouad

d) Autres méthodes

4/ Exemples : regression sur les classes de Hölder

a) Perte ponctuelle
b) Perte quadratique
c) Perte uniforme

1/ Risque minimax, optimalité

Soit $\mathcal{P}$ un modèle. Pour tout $P \in \mathcal{P}$, le paramètre d'intérêt est $\theta(P) \in \Theta$. On s'intéresse à l'estimation de θ uniformément sur $\mathcal{P}$ pour la perte $l(\cdot,\cdot) : \Theta \times \Theta \to \mathbb{R}$. L'uniformité est à prendre au sens du risque minimax. Si $\boxed{r(\hat\theta, \theta(P)) = \mathbb{E}_{\theta(P)}\, l(\hat\theta, \theta)}$ est le risque de l'estimateur $\hat\theta$ en P (ou par abus, en $\theta = \theta(P)$), le risque minimax s'écrit

$$\boxed{R(\mathcal{P}) = \underset{\substack{\hat\theta \\ \text{estimateur}}}{\operatorname{Inf}} \underbrace{\sup_{P \in \mathcal{P}} r(\hat\theta, \theta(P))}_{R(\hat\theta)}}$$

C'est le meilleur risque en pire cas sur $\mathcal{P}$.

$\hat\theta$ est dit minimax optimal si $R(\hat\theta) = R(\mathcal{P})$. Les cas où $R(\mathcal{P})$ est calculable sont rare, et sa valeur exacte ne présente pas un grand intérêt, mais son comportement est important. On se contente de l'encadrer :

- Pour majorer $R(\mathcal{P})$: on étudie un estimateur $\hat\theta$ bien choisi et on borne $r(\hat\theta, \theta(P))$ uniformément en $P \in \mathcal{P}$
- Pour minorer $R(\mathcal{P})$: les méthodes classiques reposent sur des approches bayésiennes.

Remarque : Dans la suite, on pourra noter indifféremment P, $\theta(P)$ ou θ, et désigner le modèle par Θ au lieu de $\mathcal{P}$.

<u>Observation initiale</u> : Pour toute loi de probabilité π sur Θ et tout estimateur $\hat{\theta}$,

$$\sup_{\theta \in \Theta} r(\hat{\theta}, \theta) \geqslant \int_{\Theta} r(\hat{\theta}, \theta)\, \pi(d\theta) =: B(\hat{\theta}, \pi)$$

donc

$$\boxed{R(\Theta) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} r(\hat{\theta}, \theta) \geqslant \inf_{\hat{\theta}} B(\hat{\theta}, \pi) = B(\pi)}$$

- Toutes les bornes présentées dans la suite sont basées là dessus.
- Souvent, π est composée uniquement de Dirac.
- Le choix de π est primordial et se fait au cas par cas. On peut aussi optimiser sur une famille de priors Π pour obtenir

$$R(\Theta) \geqslant \sup_{\pi \in \Pi} B(\pi) =: B(\Pi) \qquad (*)$$

Bien que la définition du risque minimax soit purement fréquentiste, son étude est fortement dépendante de considérations bayésiennes. Dans certains cas, on peut donner un lien rigoureux entre les deux.

2/ Risques minimax fréquentiste et bayésien (Johnstone, Chapitre 4)

On considère le problème de l'estimation de $\theta = (\theta_i)_{i\in\mathbb{N}}$ dans le modèle gaussien hétéroscédastique

$$Y_i = \theta_i + \varepsilon \rho_i Z_i \quad , \; i \in \mathbb{N}$$

avec $Z_i \overset{iid}{\sim} N(0,1)$ et $\varepsilon, \rho_i > 0$ connus. L'espace des paramètres est $\Theta = l_{2,\rho} = \left\{ \theta : \sum_{i=0}^{\infty} \left(\frac{\theta_i}{\rho_i}\right)^2 < \infty \right\}$.

Th. (Théorème minimax)

Supposons que pour tout $\theta \in \overline{\mathbb{R}}^{\mathbb{N}}$, $a \mapsto l(a,\theta)$ est convexe et semi-continue inférieurement ($l(a,\theta) \leq \underline{\lim}_{b \to a} l(b,\theta) \; \forall \theta$).

Soit Π un ensemble convexe de mesures de probabilité sur $l_{2,\rho}$.

Alors
$$\operatorname{Inf}_{\hat\theta} \operatorname{Sup}_{\pi \in \Pi} B(\hat\theta, \pi) = \operatorname{Sup}_{\pi \in \Pi} \operatorname{Inf}_{\hat\theta} B(\hat\theta, \pi) = B(\Pi)$$

Dem : Cf Johnstone. C'est une application du théorème minimax de Von Neumann, en remarquant notament que $\pi \mapsto B(\hat\theta, \pi)$ est linéaire et donc concave.

Application : On note que $r(\hat{\theta},\theta) = B(\hat{\theta}, \delta_\theta) = \int_{\Theta} r(\hat{\theta},\theta')\, \delta_\theta(d\theta')$

Ainsi, si $\Pi \supset \{\delta_\theta, \theta \in \Theta\}$ on obtient

$$\sup_{\theta \in \Theta} r(\hat{\theta},\theta) \leq \sup_{\pi\in\Pi} B(\hat{\theta},\pi)$$

Et donc

$$\begin{aligned} R(\Theta) &= \inf_{\hat{\theta}} \sup_{\theta\in\Theta} r(\hat{\theta},\theta) \\ &\leq \inf_{\hat{\theta}} \sup_{\pi\in\Pi} B(\hat{\theta},\pi) \\ &= \sup_{\pi\in\Pi} \inf_{\hat{\theta}} B(\hat{\theta},\pi) \quad \text{par théorème minimax si } \Pi \text{ est convexe} \\ &= B(\Pi) \end{aligned}$$

D'autre part on a vu que $R(\Theta) \geq B(\Pi)$, donc

$$\boxed{R(\Theta) = B(\Pi)}$$

En prenant par exemple $\Pi = \mathrm{Conv}(\{\delta_\theta, \theta \in \Theta\})$ on a montré que l'on peut trouver une famille de lois a priori Π sur Θ réalisant l'égalité dans $(*)$. En particulier, dans ce contexte on ne perd rien à considérer le risque bayesien dans la minoration $(*)$.

Remarque : On peut faire plus précis : le Lemme 4.18 ("Least favorable priors are discrete") suggère que les lois π maximisant $B(\pi) = \inf_{\hat{\theta}} B(\hat{\theta},\pi)$ sont discrètes et sans points d'accumulation. Les méthodes où π est une somme finie de Dirac (Assouad, Fano, Lecam) sont donc potentiellement de bons choix

3/ Théorèmes usuels de bornes inférieures

a) Lemme de Lecam

Dans ce qui suit, la perte $l(\cdot,\cdot)$ est une (semi)-distance. On peut généraliser à des pertes composites ou à des fonctionnelles satisfaisant l'inégalité triangulaire faible $l(x,y)+l(z,y) \geq A\,l(x,y)$, $A<1$ facilement: Cf Tsybakov et Yu.

Th: (Lemme de Lecam)

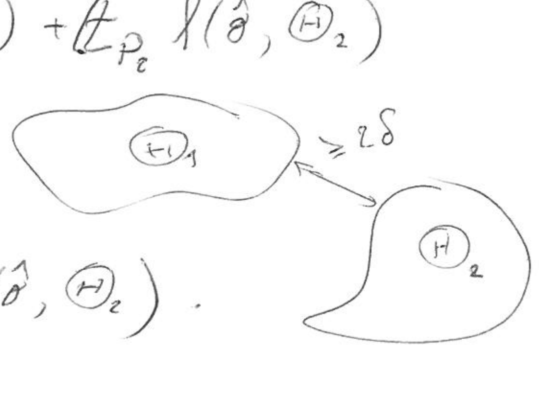
Soient $\Theta_1, \Theta_2 \subset \Theta$ 2δ-séparés, au sens où $l(\theta_1,\theta_2) \geq 2\delta$ pour tout $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$. On suppose que $\mathcal{P}_1, \mathcal{P}_2 \subset \mathcal{P}$ sont tels que $\theta(P_i) \in \Theta_i$ pour $P_i \in \mathcal{P}_i$, $i=1,2$. Alors

$$R(\mathcal{P}) = \inf_{\hat{\theta}} \sup_{P\in\mathcal{P}} \mathbb{E}_P\, l(\hat{\theta},\theta(P)) \geq \delta \cdot \sup_{P_i \in \mathrm{Conv}(\mathcal{P}_i)} \|P_1 \wedge P_2\|,$$

où $\|P \wedge Q\| = \int dP \wedge dQ$ et $\mathrm{Conv}(\mathcal{P}_i)$ est l'enveloppe convexe de $\mathcal{P}_i$.

Dem: Soit $\hat{\theta}$ un estimateur. Pour $P_i \in \mathcal{P}_i$,

$$M = 2 \sup_{P\in\mathcal{P}} \mathbb{E}_P\, l(\hat{\theta},\theta(P)) \geq \mathbb{E}_{P_1} l(\hat{\theta},\theta(P_1)) + \mathbb{E}_{P_2} l(\hat{\theta},\theta(P_2))$$
$$\geq \mathbb{E}_{P_1} l(\hat{\theta},\Theta_1) + \mathbb{E}_{P_2} l(\hat{\theta},\Theta_2)$$



Par combinaisons convexes, $\forall P_i \in \mathrm{Conv}(\mathcal{P}_i)$,

$$M \geq \mathbb{E}_{P_1} l(\hat{\theta},\Theta_1) + \mathbb{E}_{P_2} l(\hat{\theta},\Theta_2).$$

Comme $l(\hat{\theta}, \Theta_1) + l(\hat{\theta}, \Theta_2) \geq l(\Theta_1, \Theta_2) \geq 2\delta$, pour tout $P_i \in \mathrm{Conv}(\mathcal{P}_i)$

$$M \geq 2\delta \cdot \inf_{\substack{f_i \geq 0 \\ f_1 + f_2 = 1}} \left\{ \mathbb{E}_{P_1} f_1 + \mathbb{E}_{P_2} f_2 \right\}$$

le terme de droite est clairement minoré par

$$2\delta \inf_{\substack{f_i \geq 0 \\ f_1 + f_2 = 1}} \int f_1 dP_1 + f_2 dP_2 \geq 2\delta \inf_{\substack{f_i \geq 0 \\ f_1 + f_2 = 1}} \int \underbrace{(f_1 + f_2)}_{=1} dP_1 \wedge dP_2$$

$$= 2\delta \, \| P_1 \wedge P_2 \| \quad \square$$

a') Proximités stochastiques

Th : (Théorème de Scheffé)

$$\boxed{TV(P,Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)| = \frac{1}{2} \int |dP - dQ| = 1 - \|P \wedge Q\|}$$

Dem : (Seulement l'égalité de droite) $|p - q| = p + q - 2\, p \wedge q$ + intégration

* distance de Hellinger $H(P,Q)^2 = \int (\sqrt{dP} - \sqrt{dQ})^2$
* divergence de Kullback $K(P,Q) = \begin{cases} \int \log \frac{dP}{dQ} dP & \text{si } P \ll Q \\ +\infty & \text{sinon} \end{cases}$

Prop : (Inégalités de Lecam)

$$\boxed{\frac{1}{2} H^2(P,Q) \leq TV(P,Q) \leq H(P,Q)\sqrt{1 - \frac{H^2(P,Q)}{4}}}$$

Prop : (Pinsker)

$$\boxed{TV(P,Q) \leq \sqrt{\frac{K(P,Q)}{2}}}$$

Prop :
- $H^2(P,Q) \leq K(P,Q)$
- $TV(P,Q) \leq 1 - \frac{1}{2} \exp(-K(P,Q))$

Corollaire (Version Variation Totale de Lecam)

Si $\begin{cases} \ell(\Theta_1, \Theta_2) \geq 2\delta \\ \inf\limits_{P_i \in Conv(\mathcal{P}_i)} TV(P_1, P_2) \leq \alpha \end{cases}$, alors $R(\Theta) \geq \delta \cdot \dfrac{1-\alpha}{2}$

Corollaire (Version Hellinger)

Si $\begin{cases} \ell(\Theta_1, \Theta_2) \geq 2\delta \\ \inf\limits_{P_i \in Conv(\mathcal{P}_i)} H^2(P_1, P_2) \leq \alpha < 2 \end{cases}$, alors $R(\Theta) \geq \dfrac{\delta}{2}\left(1 - \sqrt{\alpha\left(1 - \frac{\alpha}{2}\right)}\right)$

Corollaire (Version Kullback)

Si $\begin{cases} \ell(\Theta_1, \Theta_2) \geq 2\delta \\ \inf\limits_{P_i \in Conv(\mathcal{P}_i)} K(P_1, P_2) \leq \alpha < \infty \end{cases}$, alors $R(\Theta) \geq \delta \cdot \max\left\{\dfrac{1}{4}\exp(-\alpha), \dfrac{1-\sqrt{\frac{\alpha}{2}}}{2}\right\}$

Rq : Pour des exemples montrant la nécessité de considérer les enveloppes convexes dans certains cas ; • Jisu Kim "Minimax dimension estimation", 2016

• Kim & Zhou "Tight minimax rates for Manifold Estimation ..." 2013

On a ici comparé deux "hypothèses" entre elles. Ça n'est parfois pas suffisant, et introduire plus d'alternatives peut être nécessaire pour obtenir des bornes optimales

b) Comparaison de plusieurs hypothèses

Prop : (Comparaisons de plusieurs hypothèses, Version Kullback)

Soit $N \geq 2$ et $\underset{\theta(P_0)}{\theta_0}, \underset{\theta(P_1)}{\theta_1}, \dots, \underset{\theta(P_N)}{\theta_N} \in \Theta$ tels que :

- $l(\theta_j, \theta_k) \geq 2\delta \quad \forall j \neq k$
- $P_j \ll P_0$ et $\displaystyle \frac{1}{M} \sum_{j=1}^{M} K(P_j, P_0) \leq \alpha \log M$

avec $\alpha < \frac{1}{8}$.

Alors $$\inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P\, l(\hat{\theta}, \theta(P)) \geq \delta \cdot \frac{\sqrt{M}}{1+\sqrt{M}} \left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log M}}\right)$$

Rq : Ici, seule importe la proximité stochastique entre P_0 et les autres P_j, mais pas les $(P_j)_{j\neq 0}$ entre elles.

c) lemme d'Assouad

Th : (lemme d'Assouad)

Soit $m \geq 1$. On suppose que $\mathcal{P} \supset \{P_\tau, \tau \in \{-1,1\}^m\}$ contient 2^m lois. Notons $\tau \sim \tau'$ si τ et τ' ne diffèrent que d'une coordonnée (exactement), et $\tau \sim_j \tau'$ si celle-ci est la coordonnée j. Supposons que la perte se décompose sous la forme
$$l(x,y) = \sum_{j=1}^{m} l_j(x,y).$$
On suppose que si $\tau \sim_j \tau'$, alors $l_j(\theta(P_\tau), \theta(P_{\tau'})) \geq 2\delta$.

Alors
$$\inf_{\hat\theta} \max_{\tau \in \{-1,1\}^m} \mathbb{E}_\tau \, l(\hat\theta, \theta(P_\tau)) \geq m \cdot \delta \cdot \min_{\tau \sim \tau'} \| P_\tau \wedge P_{\tau'} \|$$

Dem :
$$\max_\tau \mathbb{E}_\tau \, l(\hat\theta, \theta(P_\tau)) = \max_\tau \sum_{j=1}^{m} \mathbb{E}_\tau \, l_j(\hat\theta, \theta(P_\tau))$$
$$\geq \frac{1}{2^m} \sum_\tau \sum_{j=1}^{m} \mathbb{E}_\tau \, l_j(\hat\theta, \theta(P_\tau))$$
$$= \sum_{j=1}^{m} \frac{1}{2^m} \sum_\tau \mathbb{E}_\tau \, l_j(\hat\theta, \theta(P_\tau))$$
$$= \sum_{j=1}^{m} \frac{1}{2^{m+1}} \sum_\tau \mathbb{E}_\tau \, l_j(\hat\theta, \theta(P_\tau)) + \mathbb{E}_{\tau^j} \, l_j(\hat\theta, \theta(P_{\tau^j}))$$
où $\tau^j = (\tau_1, \dots, \tau_{j-1}, -\tau_j, \tau_{j+1}, \dots, \tau_m)$

Comme $l_j(\hat\theta, \theta(P_\tau)) + l_j(\hat\theta, \theta(P_{\tau^j})) \geq 2\delta$, le même argument que précédemment donne le résultat.

d) Autres méthodes

- Fano : Une "généralisation" d'Assouad (on peut déduire le second du premier) spécifique à Kullback

- Inégalité de Van Tree : prior paramétrique (segment $\subset \mathbb{R}$). Uniquement pour la perte quadratique dans $\mathbb{R}$

- Hypothèses floues : comparaison de deux posterior dont les priors sont stochastiquement éloignés. Écrit pour $\Theta \subset \mathbb{R}$ dans Tsybakov, mais facilement généralisable à une perte métrique générale.

- Théorème de convolution : résultat asymptotique en loi adapté à un cadre paramétrique. Cf Hajek ; van der Vaart.
  ($\simeq$ Borne de Cramer-Rao asymptotique locale)

4/ Exemples : regression sur les classes de Hölder

La <u>classe de Hölder</u> $\mathcal{C}^{\beta}(L)$ est l'ensemble des fonctions $f : [0,1] \to \mathbb{R}$ $\lfloor \beta \rfloor$ fois dérivables et telles que $\forall x, y \in [0,1]$,

$$|f^{(\lfloor\beta\rfloor)}(x) - f^{(\lfloor\beta\rfloor)}(y)| \leqslant L|x-y|^{\beta - \lfloor\beta\rfloor}$$

On considère le modèle $Y_i = f(X_i) + \xi_i$, $1 \leq i \leq n$

avec $\begin{cases} \xi_i \overset{iid}{\sim} \mathcal{N}(0,\sigma^2) \text{ et } f \in \mathcal{C}^{\beta}(L) \\ X_i \text{ déterministes} \end{cases}$

pour les pertes

$$\ell_{x_0}(f,g) = |f(x_0) - g(x_0)| \quad \text{ponctuelle}$$

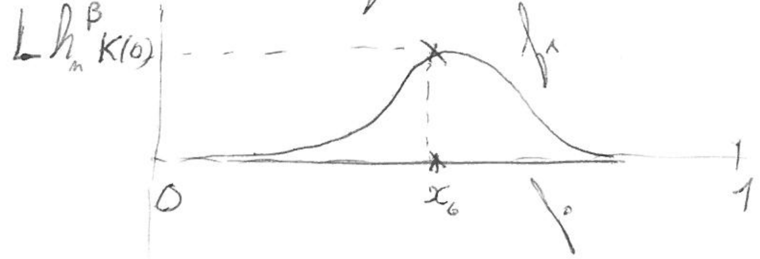
$$\ell_2(f,g) = \|f-g\|_{L^2} \quad L^2 \text{ (quadratique)}$$

$$\ell_\infty(f,g) = \|f-g\|_{L^\infty} \quad L^\infty \text{ (uniforme)}$$

Soit $K : \mathbb{R} \to \mathbb{R}_+$ tel que $\begin{cases} K \in \mathcal{C}^\infty(\mathbb{R}) \cap \mathcal{C}^{\beta}(\frac{1}{2}) \\ K(x) > 0 \iff -\frac{1}{2} < x < \frac{1}{2} \end{cases}$

a) <u>Perte ponctuelle</u> : On utilise <u>Lecam</u>, avec Kullback :

$$f_0(x) = 0 \quad ; \quad f_1(x) = L h_n^\beta K\left(\frac{x - x_0}{h_n}\right)$$



On a :

- $f_0, f_1 \in \mathcal{C}^\beta(L)$
- $\ell(f_0, f_1) = |f_1(x_0)| = L K(0) h_n^\beta$
- "$K(P_0, P_1) \leq \alpha$"

Sous P_j, $(Y_1, \ldots, Y_n)$ admet pour densité

$$p_j(u_1, \ldots, u_n) = \prod_{i=1}^n p_\xi(u_i - f_j(X_i)) \quad , j = 0, 1$$

$$K(P_0, P_1) = \int \log \frac{dP_0}{dP_1} \, dP_0$$

$$= \int \cdots \int \log \prod_i \frac{p_\xi(u_i)}{p_\xi(u_i - f_1(X_i))} \cdot \prod_i p_\xi(u_i) du_i$$

$$= \sum_i \int \log \frac{p_\xi(y)}{p_\xi(y - f_1(X_i))} p_\xi(y) dy$$

(Admis)

$$\leq p_* \sum_{i=1}^n f_1^2(X_i)$$

$$= p_* L^2 h_n^{2\beta} \sum_i K^2\left(\frac{X_i - x_0}{h_n}\right)$$

$$K(P_0,P_1) \leq p_\alpha L^2 h_n^{2\beta} \|K\|_\infty^2 \sum_i \mathbb{1}_{\left\{\left|\frac{X_i - x_0}{h_n}\right| \leq \frac{1}{2}\right\}}$$

$$\leq p_\alpha L^2 h_n^{2\beta} \|K\|_\infty \, (n h_n) \vee 1$$

$$= \underbrace{p_\alpha L^2 \|K\|_\infty \, n h_n^{2\beta+1}}_{= C n h_n^{2\beta+1} = \alpha} \quad \text{dès que } n h_n \geq 1$$

On obtient donc par Lecam :

$$\inf_{\hat{T}_n} \sup_{f \in \mathcal{C}^\alpha(L)} \mathbb{E}_f |f(x_0) - \hat{T}_n| \geq C' h_n^\beta \cdot \operatorname{Max}\left(\frac{1}{4}\exp(-\alpha) ; \frac{1-\sqrt{\alpha/2}}{2}\right)$$

En prenant $h_n = \left(\frac{1}{n}\right)^{\frac{1}{2\beta+1}}$ on obtient

ce terme devient constant.

$$\boxed{R_{l_{x_0}}\left(\mathcal{C}^\beta(L)\right) \geq C \left(\frac{1}{n}\right)^{\frac{\beta}{2\beta+1}}}$$

Rq : Vitesse atteignable avec un estimateur à noyau avec $h = \left(\frac{1}{n}\right)^{\frac{1}{2\beta+1}}$: vitesse optimale

b) Perte quadratique

Avec les mêmes deux hypothèses que précédement on a, pour $x_0 = \frac{1}{2}$

$$\|f_0 - f_1\|_{L^2}^2 = \left(L h_n^\beta \sqrt{\int_0^1 K^2\left(\frac{x-\frac{1}{2}}{h_n}\right)dx}\right)^2$$

$$= L^2 h_n^{2\beta+1} \|K\|_{L^2}^2$$

En appliquant Lecam on obtient

$$R_{\ell_2^2}(\mathcal{C}^\beta(L)) \geq c\, h_n^{2\beta+1} \operatorname{Max}\left(\frac{1}{4}\exp(-\alpha),\ \frac{1-\sqrt{\alpha/2}}{2}\right),$$

où $\alpha \asymp n h_n^{2\beta+1}$

et pour $h_n \asymp \left(\frac{1}{n}\right)^{\frac{1}{2\beta+1}}$ cela donne la vitesse $\asymp \left(\frac{1}{n}\right)^{\frac{\beta+\frac{1}{2}}{2\beta+1}} = \frac{1}{\sqrt{n}}$.

$\longrightarrow$ Vitesse paramétrique !

↳ Peut mieux faire.

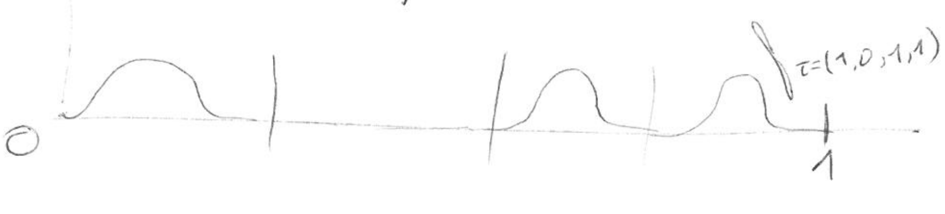
Utilisation d'Assouad : on note $[0,1] = \underset{x_1}{\underset{\Psi}{S_1}} \sqcup \underset{x_2}{\underset{\Psi}{S_2}} \sqcup \dots \sqcup \underset{x_m}{\underset{\Psi}{S_m}}$ la partition uniforme de $[0,1]$ de pas $\frac{1}{m}$.

$$\ell_2(f,g) = \|f-g\|_{L^2(0,1)} = \sum_j \|f-g\|_{L^2(S_j)}$$

On paramétrise le cube par $\{0,1\}^m$ (au lieu de $\{-1,1\}^m$)

$$\forall \tau \in \{0,1\}^m,\ f_\tau(x) = \sum_{j=1}^m \tau_j f_j(x)$$

$$\text{où } f_j(x) = L h_n^\beta K\left(\frac{x-x_j}{h_m}\right) \quad \text{avec } \boxed{h_m = \frac{1}{m}}$$



On a :

- $f_\tau \in \mathcal{C}^\beta(L)$
- $l_2^2(f_\tau, f_{\tau'}) = L\cancel{\|K\|} h_n^{2\beta+1}$ si $\tau \sim \tau'$
- "$K(P_\tau, P_{\tau'}) \leq \alpha$" : pareil que plus haut ;

$$\text{Si } \tau \sim \tau', \; K(P_\tau, P_{\tau'}) \leq c\, n h_n^{2\beta+1} =: \alpha$$

Par assouad :

$$R_{l_2}(\mathcal{C}^2(L)) \geq C \underbrace{h_n^{2\beta+1} \, . \, m}_{= h_n^{2\beta}} \, . \, \mathrm{Max}\left\{ \frac{1}{4}\exp(-\alpha) \, ; \, \frac{1 - \sqrt{\alpha/2}}{2} \right\}$$

$$h_n \simeq \left(\frac{1}{n}\right)^{\frac{1}{2\beta+1}} \longrightarrow \boxed{\text{Vitesse en } \left(\frac{1}{n}\right)^{\frac{\beta}{2\beta+1}}}$$

c'est la vitesse optimale, car atteignable par un estimateur à noyau.

$$\boxed{\inf_{\hat{T}_n} \sup_{f \in \mathcal{C}^\beta(L)} \sqrt{\mathbb{E}_f \|\hat{T}_n - f\|_{L^2}^2} \geq c \left(\frac{1}{n}\right)^{\frac{\beta}{2\beta+1}}}$$

c) Perte uniforme

Rq: les bornes supérieures pour les estimateurs à noyau donnent une vitesse $\left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}$. Lecam conclurait seulement $\left(\frac{1}{n}\right)^{\frac{\beta}{2\beta+1}}$

On utilise la version Kullback de la comparaison de M hypothèses pour aller chercher le log supplémentaire.

Comme précédemment, $\begin{cases} f_0(x) = 0 \\ f_j(x) = L h_n^\beta K\left(\frac{x - x_j}{h_n}\right) \\ \text{pour } j = 1, \dots, M \end{cases}$, $h_n = \frac{1}{M}$

- $f_j \in \mathcal{C}^\beta(L)$
- $\ell(f_j, f_k) = \|f_j - f_k\|_\infty \geq L h_n^\beta K(0) \quad (= 2\delta)$
- $\frac{1}{M} \sum_{j=1}^{M} K(P_j, P_0) \leq \frac{1}{M} \sum_{j=1}^{M} P_* \sum_{i=1}^{n} f_j^2(X_i)$

$$\leq P_* L^2 \|K\|_\infty^2 h_n^{2\beta} \cdot \frac{1}{M} \sum_{j=1}^{M} \left| \{ i \mid X_i \in \mathrm{Supp}(f_j) \} \right|$$

$$= P_* L^2 \|K\|_\infty^2 h_n^{2\beta} \cdot \frac{n}{M}$$

Pour $h_n = \frac{1}{M}$ et $M = \left\lceil c_0 \left(\frac{n}{\log n}\right)^{\frac{1}{2\beta+1}} \right\rceil$ on arrive à

$$\dots \leq C \cdot \log n .$$

Or, $\log M \geq \log\left(c_0 \left(\frac{n}{\log n}\right)^{\frac{1}{2\beta+1}}\right) \geq \frac{1}{2}\frac{\log n}{2\beta+1}$ pour n assez grand.

Ainsi, $\frac{1}{M}\sum_{j=1}^{M} K(P_j, P_0) \leq \alpha \log M$ où $\begin{cases} \alpha = p_* L^2 \|K\|_2^2 / c_0^{2\beta+1} \\ M \simeq \left(\frac{n}{\log n}\right)^{\frac{1}{2\beta+1}} \end{cases}$

En appliquant la version Kullback de la comparaison de M hypothèses,

$$\boxed{\inf_{\hat{T}_n} \sup_{f \in \mathcal{C}^\beta(L)} \mathbb{E}_f \|\hat{T}_n - f\|_\infty \geq c \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}}$$

pour n assez grand.