LUCIEN BIRGÉ

Vitesses optimales de convergence des estimateurs

Astérisque, tome 68 (1979), p. 171-185
<http://www.numdam.org/item?id=AST_1979__68__171_0>

VITESSES OPTIMALES DE CONVERGENCE DES ESTIMATEURS

Lucien BIRGÉ

I - Introduction

Etant donné un problème statistique défini par un espace de paramètres Θ muni d'une distance d et par des probabilités P_θ sur un espace $(\Omega^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$, on peut s'intéresser aux performances des estimateurs $\hat{\theta}_n$ de P_θ^n où P_θ^n est la projection de P_θ sur $(\Omega^n, \mathcal{A}^n)$. En particulier si les P_θ^n se comportent bien en fonction de n , (et si la distance choisie sur Θ correspond à une distance entre les P_θ) on peut espérer un résultat du genre $d(\theta, \hat{\theta}_n) \xrightarrow[n\to+\infty]{} 0$ P_θ p.s. ou $\mathbb{E}_\theta(d(\hat{\theta}_n,\theta)) \xrightarrow[n\to+\infty]{} 0$.

On peut alors se demander s'il existe une puissance α de n telle que $\mathbb{E}_\theta[n^\alpha\, d(\hat{\theta}_n,\theta)]$ reste bornée uniformément en n et θ et quel est le meilleur α possible.

Le Cam dans [1] et [2] s'est intéressé à ce problème en supposant que les P_θ^n étaient des mesures produits $(P_\theta^1)^{\otimes n}$ et que la distance d était la distance de Hellinger h entre les P_θ^1 . Il a montré que,pourvu que l'espace Θ ne soit pas trop gros (de dimension métrique finie et compact),on pouvait construire des estimateurs tels que $\mathbb{E}_\theta[n\, h^2(\hat{\theta}_n,\theta)]$ soit borné et que ceci constituait la meilleure vitesse possible. On va ici démontrer un résultat analogue, mais qui englobe certains cas de variables dépendantes et permet de retrouver ces résultats de Le Cam en affaiblissant les hypothèses de compacité.

II - Les hypothèses

On suppose que l'on observe un processus $X = (X_1, \ldots, X_n, \ldots)$ de loi P_θ sur $(\Omega^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$. On notera P_θ^n la loi jointe des n premières coordonnées de X et on supposera que Θ est muni d'une pseudo-distance d , c'est-à-dire d'une application de $\Theta \times \Theta$ dans $\overline{\mathbb{R}^+}$ qui vérifie les propriétés suivantes :

D1) $d(s,t) = d(t,s)$

D2) il existe $M > 0$ et $B \geqslant 2$ tels que pour $d(s,t)$ et $d(t,u) \leqslant M$, on ait l'inégalité triangulaire $d(s,u) \leqslant \frac{B}{2}[d(s,t) + d(t,u)]$.

De plus la topologie induite par d sur Θ doit être liée à la topologie forte sur les P_θ^n de la manière suivante :

D3) il existe η et A_1 positifs tels que si $\sqrt{n}\, d(s,t) \leqslant \eta$, on ait :

$$\frac{1}{2}\,||P_s^n - P_t^n|| \leqslant A_1 \sqrt{n}\, d(s,t)$$

D4) il existe une constante A_2 telle que pour tous les couples s et t on ait :

$$\Pi(s^n,t^n) = 1 - \frac{1}{2}||P_s^n - P_t^n|| \leqslant A_2\, e^{-nd^2(s,t)} .$$

La notation Π est reprise de Le Cam et on peut remarquer (cf. Kraft [3]) que $\Pi(P,Q)$ représente la somme des erreurs du meilleur test entre P et Q . (D4) précise donc la vitesse de séparation des hypothèses P_s^n et P_t^n en fonction de n .

<u>Définition</u> : On appellera λ-réseau $\frac{\lambda}{2}$-discernable R_λ un ensemble de points de Θ vérifiant les propriétés suivantes :

a) Pour tout x de Θ , il existe y dans R_λ avec $d(x,y) \leqslant \lambda$.

b) Pour tous x et y dans R_λ on a $d(x,y) \geqslant \lambda/2$.

Le théorème de Zorn permet de démontrer sur tout espace muni d'une pseudo-distance l'existence de λ-réseaux λ-discernables et pour une suite λ_n décroissante l'existence de réseaux R_{λ_n} croissants .

On fait sur Θ les hypothèses topologiques suivantes :

Il existe des constantes positives α, δ et A_4 , $p \geqslant 1$ telles que l'on ait :

R1) pour tous θ dans Θ r et λ tels que $\lambda \leqslant \delta$ et $r\lambda \leqslant \alpha$, on aura :

$$\text{Cardinal}\left[R_\lambda \cap \mathcal{B}(\theta,r\lambda)\right] \leqslant A_4\, r^p$$

R2) pour tout entier $i \geq 2$ et tout θ dans Θ on a

$$\text{Cardinal}\,[R_\alpha \cap \mathcal{B}(\theta, i\alpha)] \leq e^{Q_i^2}$$

où l'on note $\mathcal{B}(x,r)$ la boule de centre x et rayon r. Quitte à changer la constante Q qui intervient dans (R2) on peut toujours remplacer α par une constante $\alpha' < \alpha$ ce qui nous permettra de supposer :

$$(1) \qquad \alpha < \frac{M}{B^2} .$$

L'hypothèse (R1) dit que Θ est localement de dimension métrique finie, (R2) qu'il ne croît pas trop vite à l'infini . Lorsque l'hypothèse (R2) n'est pas vérifiée, on pourra toujours se limiter à une boule dans Θ .

Généralement, on pourra vérifier l'hypothèse (R1) en comparant la distance d à une autre distance d' sur Θ (par exemple euclidienne lorsque Θ est une partie de $\mathbb{R}^n$) pour laquelle Θ est de dimension métrique finie. En particulier (R1) sera vérifiée si l'on peut trouver des constantes $a \leqslant b$ et $r \leqslant q$ telles que l'on ait :

$$a[d'(s,t)]^q \leqslant d(s,t) \leqslant b[d'(s,t)]^r$$

lorsque $d(s,t)$ est assez petit . cf [2] .

Une première conséquence de ces hypothèses est l'existence du maximum de vraisemblance sur un réseau.

<u>Lemme 1</u>

Soit un réseau R_λ de pas λ inférieur à α et $n > \frac{Q}{\alpha}$, sous les hypothèses précédentes l'estimateur du maximum de vraisemblance existe sur le réseau R_λ P_s^n p.s. lorsque s est un point de R_λ .

<u>Démonstration</u>

Si l'estimateur $\hat{\theta}$ du maximum de vraisemblance n'existe pas, pour toute boule $\mathfrak{B}(s,k\alpha)$ on peut trouver un point t de R_λ extérieur à la boule tel que le test de rapport de vraisemblance au seuil 0 de s contre t rejette s . On va majorer cette probabilité par la somme des probabilités des erreurs des tests de s contre t pour tous les points t hors de la boule $\mathfrak{B}(s,k\alpha)$. Pour cela, on découpe l'espace en couronnes $\mathfrak{B}(s,(k+1)\alpha) \setminus \mathfrak{B}(s,k\alpha)$. Si t est dans la couronne, d'après (D4) l'erreur du test de s contre t est au plus $A_2\, e^{-nk^2\alpha^2}$ et le nombre de points t de la couronne est majoré par celui de la boule $\mathcal{B}(s,(k+1)\alpha)$ c'est à dire $A_4(\frac{\alpha}{\lambda})^p \, e^{Q(k+1)^2}$. On obtient finalement :

$$P_s^n[\{\omega | \hat{\theta}(\omega) \text{ n'existe pas}\}] \leqslant \lim_k \searrow \sum_{i=k}^{+\infty} A_4(\frac{\alpha}{\lambda})^p \, e^{Q(i+1)^2} A_2 \, e^{-ni^2\alpha^2} .$$

D'après l'hypothèse, pour i assez grand on aura $Q(i+1)^2 < ni^2\alpha^2$ et la série du second membre converge, d'où le résultat.

III - Vitesses maximales d'estimation

On obtient ici un résultat tout à fait analogue à ceux de Le Cam [1] de majoration de la vitesse uniforme de convergence d'un estimateur, pourvu que l'espace Θ soit suffisamment riche, en particulier sans point isolé.

Proposition 2

Supposons que pour un point s de Θ, il existe une suite s_n de points de Θ différents de s et tels que $d(s,s_n)$ tende vers 0 . Soit K un réel positif inférieur à $(\frac{\eta \wedge A_1^{-1}}{2B})^2$ alors, on a pour tout estimateur $\hat{\theta}$ et tout réel $r > 0$:

$$\overline{\lim_n} \sup_{t \in \mathcal{B}(s,r)} P_t^n[n \, d^2(t,\hat{\theta}) > K] > 0$$

Démonstration

Soit K' tel que $\eta \wedge A_1^{-1} > K' > 2B\sqrt{K}$. Par hypothèse, on pourra trouver des entiers n arbitrairement grands tels qu'il existe un point t de Θ avec

$$r \geqslant \frac{K'}{\sqrt{n}} > d(s,t) > \frac{2B\sqrt{K}}{\sqrt{n}} .$$

Supposons en outre que $\sqrt{\frac{K}{n}}$ est inférieur à M et considérons le test de s contre t qui accepte s si $d(s,\hat{\theta}_n) \leqslant d(t,\hat{\theta}_n)$ et t dans le cas contraire.

D'après (D2) si $d^2(s,\hat{\theta}_n) < \frac{K}{n}$ alors $d^2(t,\hat{\theta}_n) > \frac{K}{n} > d^2(s,\hat{\theta}_n)$ puisque $d(s,t) > 2B\sqrt{\frac{K}{n}}$.

Donc la somme des erreurs de ce test sera inférieure à

$P_s^n[nd^2(s,\hat\theta_n) > K] + P_t^n[nd^2(t,\hat\theta_n) > K]$ et on aura alors :

$$\Pi(P_s^n, P_t^n) \leq \sup_{t \in \mathcal{B}(s,r)} P_t^n[nd^2(t,\hat\theta_n) > K]$$

Mais d'après (D3), on a $\Pi(P_s^n, P_t^n) \geq 1 - A_1 \sqrt{n}\, d(s,t) \geq 1 - A_1 K'$ et donc

$$\overline{\lim_n} \sup_{t \in \mathcal{B}(s,r)} P_t^n[nd^2(t,\hat\theta) \leq K] > 1 - A_1 K' > 0 .$$

Remarque : Si on suppose que le point s est tel que pour tout $r > 0$, on peut trouver t avec $d(s,t) = r$, le résultat précédent devient :

$$\underline{\lim_n} \sup_{t \in \mathcal{B}(s,r)} P_t^n[nd^2(t,\hat\theta) > K] > 0 .$$

Corollaire 3

Sous les hypothèses de la proposition 2, on a :

$$\overline{\lim_n} \sup_{t \in \mathcal{B}(s,r)} \mathbb{E}_t[nd^2(t,\hat\theta)] > 0 .$$

IV - Estimateurs atteignant la vitesse optimale

1) Construction des estimateurs

Soit n le nombre de variables observées. Alors, il existe b dans $[1,B[$ et k, q entiers tels que :

(2) $$\sqrt{n}\,\alpha = b\,B^k$$ et (3) $$B^q \leq \sqrt{n}\,\frac{M}{Bb} < B^{q+1} .$$

D'après (1) $k < q$.

On supposera en outre n assez grand pour que l'on ait $k \geq 2$ et $n\,\alpha^2 \geq 2Q\,B^4$ ce qui entraîne :

(4) $$b^2\,B^{2k} \geq 2Q\,B^4 .$$

Etant donné a tel que

(5) $$a \leqslant \inf(2^6\eta,\ 2^6\delta,\ 1)$$

on posera :

$$a_i = \frac{a}{\sqrt{n}} B^{-3(i+1)} \qquad b_i = \frac{b}{\sqrt{n}} B^i \qquad C = \frac{a}{b} B^{-7}$$

Alors $\alpha = b_k$, (3') $\frac{M}{B^2} < b_q \leqslant \frac{M}{B}$ et $a_i \leqslant C\, b_i \quad i$.

Etant donnée une suite croissante de réseaux R_i de pas a_i $(i = 1, \ldots, q)$ on construit une suite $\hat\theta_i$, des régions B_i, un estimateur $\hat\theta$ et une variable τ comme suit :

$\hat\theta_q$ est le maximum de vraisemblance sur R_q s'il existe, sinon on prend $\hat\theta$ arbitraire et $\tau = +\infty$.

On définit quand cela est possible $B_q = \mathcal{B}(\hat\theta_q, b_q)$. Le processus se poursuit en définissant par récurrence $\hat\theta_i$ le maximum de vraisemblance sur $B_{i+1} \cap R_i$ si cet ensemble est non vide et $B_i = B_{i+1} \cap \mathcal{B}(\hat\theta_i, b_i)$ (toujours non vide car $\hat\theta_i$ est dans B_{i+1}). Si $B_{i+1} \cap R_i$ est vide, on posera $\hat\theta = \hat\theta_{i+1}$ et $\tau = i+1$. Si on peut continuer le processus jusqu'au bout on posera $\hat\theta = \hat\theta_1$ et $\tau = 1$.

On a alors toujours si $\tau \leqslant i$ $\hat\theta_i \in B_i$ $B_i \subset B_{i+1}$ et et diamètre $B_i \leq Bb_i \leq M$.

2) Préliminaires

Lemme 4

Pour tout s de R_q, on a :

(6) $$P_s^n[s \notin B_q] \leqslant 2A_2\, A_4 \left(\frac{b}{a}\right)^p B^{(3q+k+3)p}\, e^{-\frac{7}{8} b^2 B^{2q}}.$$

Démonstration

Cette expression est majorée par la probabilité qu'il existe un point t avec $d(s,t) > b_q$ tel que le test de s contre t rejette s. Pour évaluer cette quantité on fait le même découpage en couronnes que pour le lemme 1 et on majore le nombre de points de la couronne par le nombre de points de la boule, ce qui donne le résultat :

$$P^n_s[s \notin B_q] \leq A_2A_4(\frac{\alpha}{a_q})^p \sum_{\ell=1+q-k}^{+\infty} e^{QB^{2\ell}} e^{-n(b_{k+\ell-1})^2}$$

$$= A_2A_4(\frac{b}{a})^p B^{(3q+k+3)p} \sum_{\ell=1+q-k}^{+\infty} e^{QB^{2\ell}-b^2B^{2(k+\ell-1)}}$$

or $$e^{QB^{2\ell}-b^2B^{2(k+\ell-1)}} = e^{(Q-b^2B^{2k-2})B^{2\ell}} \leq e^{-\frac{7}{8}b^2 B^{2k-2\ell-2}}$$

d'après (4) .

Cette série converge très rapidement et on peut la majorer par le double de son premier terme d'où le résultat.

Lemme 5

Pour tout s_i de R_i et $q-1 \geq j \geq i$, on a :

$$(7) \quad P^n_{s_i}[s_i \in (B_{j+1}-B_j)\cap \tau \leq j] \leq \begin{cases} A_2A_4(\frac{b}{a})^p B^{(4j+5)p} e^{-b^2B^{2j}} & \text{si } j+2 \leq k \\ A_2A_4(\frac{b}{a})^p B^{(3j+k+3)p} e^{-\frac{b^2}{2}B^{2j}} & \text{si } j+2>k \end{cases}$$

Démonstration

Pour tout point de cet ensemble, il existe un point $\hat{\theta}_j$ de R_j , inclus dans B_{j+1} (donc $d(s_i,\hat{\theta}_j) \leq B\, b_{j+1}$) tel que le test de s_i contre $\hat{\theta}_j$ accepte $\hat{\theta}_j$ et $d(s_i,\hat{\theta}_j) \geq b_j$. Donc la probabilité considérée est majorée par $e^{-n\, b_j^2}$ fois le nombre de points de la boule de centre s_i et rayon Bb_{j+1} qui appartiennent au réseau R_j . On a alors deux cas à considérer :

- si $j+2 \leq k$ on obtient la majoration :

$$A_2A_4(\frac{b_{j+2}}{a_j})^p e^{-nb_j^2} = A_2A_4(\frac{b}{a})^p B^{(4j+5)p} e^{-b^2B^{2j}}$$

- si $j+2>k$ il vient :

$$A_2A_4(\frac{b_k}{a_j})^p e^{QB^{2(j+2-k)}} e^{-nb_j^2} = A_2A_4(\frac{b}{a})^p B^{(3j+k+3)p} e^{B^{2j}[QB^{4-2k}-b^2]}$$

Comme la majoration obtenue ne dépend pas de i mais seulement de j , on notera H_j le terme du second membre de (7) et H_q celui de (6) .

Lemme 6

$$P^n_{s_i}\left[d(s_i,\hat\theta) > Bb_i\right] \leqslant \sum_{j=i}^{q-1} H_j + H_q \qquad s_i \in R_i \qquad i \leqslant q$$

Si $d(s_i,\hat\theta) > Bb_i$, alors s_i n'appartient pas à B_i et on peut donc majorer $P^n_{s_i}\left[d(s_i,\hat\theta) > Bb_i\right]$ par :

$$\sum_{j=i}^{q-1} P^n_{s_i}\left[s_i \in B_{j+1}-B_j) \cap \tau \leqslant j\right] + P^n_{s_i}\left[s_i \in B_q\right]$$

d'où le résultat .

IV - Vitesse atteinte par l'estimateur

Théorème 7

Soit un nombre $N \geqslant \frac{B}{4}\left[B+C\right] b_q$, et $\hat\theta$ l'estimateur précédemment construit, il existe une fonction $K(M)$ ne dépendant pas de n , et bornée sur tout intervalle $[\varepsilon, +\infty[$ telle que l'on ait :

$$\mathbb{E}_s\left[n(d(s,\hat\theta) \wedge N)^2\right] \leqslant K(M) .$$

Démonstration

On peut écrire :

$$\mathbb{E}_s\left[n(d \wedge N)^2\right] = n \int_0^{+\infty} P^n_s\left[(d \wedge N)^2 > t\right] dt$$

$$= n \int_0^{N} P^n_s\left[d > u\right] 2u\, du \qquad \text{où} \quad d = d(s,\hat\theta)$$

$$= n(I_1 + I_2 + I_3) \qquad \text{avec}$$

$$I_1 = \int_0^{B/2(B+C)\ell_1} P^n_s\left[d > u\right] 2u\, du \leqslant \frac{B^2}{4}(B+C)^2 b_1^2 = \frac{b^2 B^4 (B+C)^2}{4n}$$

$$I_2 = \int_{\frac{B}{2}[B+C]b_q}^{N} P_s^n[d(s,\hat{\theta}) > a]\ 2u\ du$$

$$I_3 = \sum_{i=1}^{q-1} \int_{\frac{B}{2}(B+C)b_i}^{B/2(B+C)b_{i+1}} P_s^n[d(s,\hat{\theta}) > u]\ 2u\ du$$

Au point s associons une suite s_i , $i = 1, \ldots, q$ telle que $d(s,s_i) \leqslant a_i \leqslant Cb_i$ pour tout i .

Par (D3) et (5) , on aura :

$$P_s^n[d(s,\hat{\theta}) > u] \leqslant P_{s_i}^n[d(s,\hat{\theta}) > u] + A_i \sqrt{n}\, a_i$$

et d'après (D2) :

$$P_{s_i}^n[d(s,\hat{\theta}) > \tfrac{B}{2}(B+C)b_i] \leqslant P_{s_i}^n[d(s_i,\hat{\theta}) > Bb_i] .$$

Ceci permet d'obtenir pour I_2 et I_3 les majorations :

$$I_2 \leqslant (N^2 - \tfrac{B^2}{4}(B+C)^2\, b_q^2)(A_1 \sqrt{n}\, a_q + P_{s_q}^n[d(s_q,\hat{\theta}) > Bb_q])$$

$$I_3 \leqslant \tfrac{B^2}{4}(B+C)^2 \sum_{i=1}^{q-1} (b_{i+1}^2 - b_i^2)(A_1 \sqrt{n}\, a_i + P_{s_i}^n[d(s_i,\hat{\theta}) > Bb_i])$$

On obtient donc finalement, en utilisant les lemmes 4 et 6 ainsi que (3) :

$$I_2 \leqslant [N^2 - B^2 (\tfrac{B+C}{4n})^2\, b^2 B^2][A_1 a\, B^{-3(q+1)} + H_q]$$

$$I_3 \leqslant \frac{B^2(B+C)^2}{4\ n} b^2 \sum_{i=1}^{q-1} (B^{2i+2} - B^{2i})[A_1 a\, B^{-3(i+1)} + \sum_{j=i}^{q-1} H_j + H_q]$$

Donc :

$$I_2+I_3 \leqslant [N^2 - \frac{B^4(B+C)^2 b^2}{4n}]H_q + [N^2 - \frac{B^2(B+C)^2 b^2}{4n} B^2 q]A_1 a\, B^{-3(q+1)}$$

$$+ \frac{B^2(B^2-1)(B+C)^2 b^2}{4n}\left[\sum_{i=1}^{q-1} A_1 a\, B^{-i-3} + \sum_{i=1}^{q-1} B^{2i} \sum_{j=i}^{q-1} H_j\right]$$

$$\leqslant N^2 H_q + \left[N^2 - \frac{M^2(B+C)^2}{4B^2}\right] A_1 \; a \frac{B^3 \; b^3}{M^3 \; n^{3/2}}$$

$$+ \frac{B^2(B^2-1)(B+C)^2 \; b^2}{4n} \left[A_1 a \; \frac{B^{-3} - B^{-q-2}}{B-1} + B^2 \sum_{j=1}^{q-1} H_j \; \frac{B^{2j}-1}{B^2-1}\right]$$

$$\leqslant N^2 H_q + \left[N^2 - \frac{M^2(B+C)^2}{4B^2}\right] A_1 a \left[\frac{Bb}{M\sqrt{n}}\right]^3 + A_1 a \; \frac{(B+1)(B+C)^2 b^2}{4nB}$$

$$+ \frac{B^4(B+C)^2 \; b^2}{4n} \sum_{j=1}^{q-1} B^{2j} \; H_j$$

(2), (3) et (3') donnent une majoration de H_q par

$$2A_4A_2\left(\frac{b}{a}\right)^p \; B^3\left(\frac{\sqrt{n}\;M}{Bb}\right)^{3p} \left(\frac{\sqrt{n}\;\alpha}{b}\right)^p \; e^{-\frac{7}{8}\frac{n\;M^2}{B^4}} \leqslant 2A_4A_2B^3\left[\frac{M^3 \;\alpha}{a\;B^3}\right]^p \; n^{2p} \; e^{-\frac{7}{8}\frac{n\;M^2}{B^4}}$$

Le lemme 3 va nous donner :

$$\sum_{j=1}^{k-2} B^{2j} \; H_j = A_2A_4\left(\frac{b}{a}\right)^p \; B^{5p} \sum_{j=1}^{k-2} B^{2j(2p+1)} \; e^{-b^2B^{2j}}$$

$$\sum_{j=k-1}^{q-1} B^{2j} \; H_j = A_2A_4\left(\frac{b}{a}\right)^p \; B^{4p} \sum_{j=k-1}^{q-1} B^{2j(2p+1)} \; B^{p(k-1-j)} \; e^{-\frac{b^2}{2}B^{2j}}$$

et donc $$\sum_{j=1}^{q-1} B^{2j} \; H_j \leqslant A_2A_4\left(\frac{b}{a} B^4\right)^p \sum_{j=1}^{q-1} B^{2j(2p+1)} \; e^{-\frac{B^{2j}}{2}}$$

Si on pose $$S = \sum_{j=1}^{+\infty} B^{2j(2p+1)} \; e^{-\frac{1}{2}B^{2j}} < +\infty \; ,$$

$K_1 = A_2A_4\left(\frac{B^5}{a}\right)^p$ $\quad$ $K_2 = 2A_4A_2B^3\left[\frac{\alpha}{\alpha B^3}\right]^p$, on pourra écrire :

$$\sum_{j=1}^{q-1} B^{2j} \; H_j \leqslant K_1 S \qquad M_n^2 \; H_q \leqslant K_2 \; M^{3p+2} \; n^{2p+1} \; e^{-\frac{7}{8}\frac{n\;M^2}{B^4}} .$$

Si K_3 est le maximum de la fonction $x^{2p+1} \; e^{-\frac{7x}{8B^4}}$ on aura

$M_n^2 \; H_q \leqslant K_2 \; K_3 \; M^{-p} \; H(n,M)$ où $H(n,M) \longrightarrow 0$ si n ou $M \longrightarrow +\infty$ et $H(n,M) \leqslant 1$

pour tout n et tout M .

On obtient finalement avec $K_4 = \frac{B^6(B+C)^2}{4}$

$$\mathbb{E}_s\left[n(d\wedge N)^2\right] < K_4 + K_2K_3 \frac{N^2}{M^{2+p}} H(n,M) + \left(N^2 - \frac{M^2 K_4}{B^8}\right) A_1 a \frac{B^6}{M^3\sqrt{n}}$$

$$+ A_1 a\, K_4 (B+1)\, B^{-S} + K_1 K_4 S$$

C.Q.F.D.

<u>Remarques</u> : Si (D2) est vérifiée pour tout M positif, en posant $N = \frac{M}{B^4}\sqrt{K_4}$, on voit pour la formule précédente que $\mathbb{E}_s\left[n(d(s,\hat{\theta})\wedge N)^2\right]$ admet une borne qui ne dépend pas de N et donc que pour tout entier N, on pourra construire un estimateur $\hat{\theta}(N)$ qui vérifiera :

$$\mathbb{E}_s\left[n(d(s,\hat{\theta}(N))\wedge N)^2\right] \leq K \qquad \text{uniformément en } N \text{ et } n .$$

- Tout ce qui précède peut se transposer aisément en remplaçant $\sqrt{n}$ par n^{γ} si on suppose que dans (D3) et (D4) les équations deviennent :

$$\frac{1}{2}\|P_s^n - P_t^n\| < A_1\, n^{\gamma}\, d(s,t) \qquad \Pi(s^n,t^n) < A_2\, e^{-n^{2\gamma} d^2(s,t)}$$

le résultat devient alors $\mathbb{E}_s\left[\left[n^{\gamma}(d(s,\hat{\theta})\wedge N)\right]^2\right] < K$. Cependant ce résultat reste vide faute d'exemples de processus vérifiant (D3) et (D4) dans le cas $\gamma \neq \frac{1}{2}$.

V - Applications

1) Cas des variables indépendantes

Dans ce cas P_θ^n s'écrit $(P_\theta^1)^{\otimes n}$. Conformément à Le Cam [1], on posera :

$$\rho(P,Q) = \int \sqrt{dP\, dQ} \qquad h^2(P,Q) = \frac{1}{2}\int (\sqrt{dP} - \sqrt{dQ})^2 = 1-\rho .$$

On a alors (8) $h^2(P,Q) \leq \frac{1}{2}\|P-Q\| < \sqrt{2}\, h(P,Q)$

et (9) $\rho(P^{\otimes n}, Q^{\otimes n}) = \rho^n(P,Q)$.

On posera $d^2 = -\log(1-h^2) = -\log\rho$.

Dans ce cas pour $h^2 < M' < \frac{1}{2}$ on pourra écrire $h < d < Bh$ ce qui permet avec (8) d'établir (D2) et (D3). (D4) se déduit de (9) et du fait que Π est inférieur à ρ .

2) Cas des chaîne de Markov

On suppose donnée sur $(\mathcal{X}, \mathcal{A})$ une famille $K_\theta(x,dy)$ de noyaux de transition. On notera $K_\theta^n(x,dy)$ le noyau itéré n fois de $K_\theta(x,dy)$. P_θ est alors la loi de la chaîne de Markov associée de loi initiale ν .

Proposition 8

Supposons qu'il existe une mesure μ et des réels positifs $0 < a < A$ tels que toutes les probabilités ν et $K_s(x,dy)$ soient dominées par μ pour tous les s et x , de densités $g\,(y)$ et $p_s(x,y)$ respectivement et que l'on ait pour tout s et tous x et x' dans $\mathcal{X}$

$$a \leqslant \frac{p_s(x,y)}{p_s(x',y)} \leqslant A \qquad \text{et} \qquad \frac{g\,(y)}{p_s(x,y)} \leqslant A$$

alors il existe sur Θ une pseudo-distance vérifiant les hypothèses (D1) à (D4) .

Démonstration

On notera $q_{s,t}(x,y) = \sqrt{p_s(x,y)\, p_t(x,y)}$ et on définira les noyaux itérés

$$p_s^n(x,y) = \int p_s(x,x_1)\, p_s(x_1,x_2) \ldots p_s(x_{n-1},y)\, d\mu(x_1) \ldots d\mu(x_n)$$

et de même $q_s^n(x,y)$.

On aura alors $\rho(P_s^n, P_t^n) = \int q_{s,t}^n(x,y)\ \nu(dx)\ \mu\,(dy)$.

On définit alors d par la relation :

$$(10) \qquad d^2(s,t) = -\frac{1}{3}\log\Big[\sup_x \int q_{s,t}^2(x,y)\ \mu(dy)\Big] .$$

On va tout d'abord vérifier (D3) :

$q_{s,t}(x,y)$ définit sur $\mathfrak{X}$ un noyau sous-Markovien R défini par

$$R(x,A) = \int_A q_{s,t}(x,y)\ \mu(dy)$$

qu'on prolonge en un noyau Markovien sur $\mathfrak{X} \cup \{\Delta\}$ en posant :

$$R(x,\{\Delta\}) = 1 - R(x,\ ) \quad \text{si } x \text{ est dans } \Omega \quad \text{et} \quad R(\Delta,\{\Delta\}) = 1 .$$

On aura alors $R^n(x,\mathfrak{X}) = \int q^n_{s,t}(x,y)\ \mu(dy)$, et d'après (10)

$$\inf_x R^2(x,\{\Delta\}) = 1 - e^{-3d^2(s,t)} .$$

Si on appelle β cette quantité, pour tout ε positif on pourra trouver un x tel que :

$$R^2(x,\{\Delta\}) = \int_{\mathfrak{X}} R(x,dy)\ R(y,\{\Delta\}) + R(x,\{\Delta\}) \leqslant \beta + \varepsilon .$$

Soit alors un point x' quelconque de $\mathfrak{X}$ on aura :

$$\int_{\mathfrak{X}} R(x',dy)\ R(y,\{\Delta\}) = \int_{\mathfrak{X}} R(x,dy)\ d\,\frac{R(x',dy)}{R(x\ ,dy)}\ R(y,\{\Delta\})$$

$$\leqslant \frac{A}{a} \int_{\mathfrak{X}} R(x,dy)\ R(y,\{\Delta\}) \leqslant \frac{A}{a}\ (\beta + \varepsilon) .$$

Comme ε est arbitraire , il vient :

$$\sup_x \int_{\mathfrak{X}} R(x,dy)\ R(y,\{\Delta\}) \leqslant \frac{A\beta}{a} \tag{11}$$

et par le même raisonnement :

$$\int_{\mathfrak{X}} R(y,\{\Delta\})\ \nu(dy) \leqslant \frac{A\beta}{a} . \tag{12}$$

On peut alors calculer $R^n(x,\mathfrak{X})$ par récurrence, pour tout x

$$R^2(x, \mathfrak{X}) = 1 - R^2(x,\{\Delta\}) \geqslant 1 - \frac{A\beta}{a} - R(x,\{\Delta\}) .$$

Si $R^i(x, \mathfrak{X}) \geqslant 1 - (i-1) \frac{A\beta}{a} - R(x,\{\Delta\})$ on aura :

$$R^{i+1}(x,\mathfrak{X}) = \int_{\mathfrak{X}} R(x,dy)\ R^i(y, \mathfrak{X})$$

$$\geqslant R(x, \mathfrak{X}) - (i-1) \frac{A\beta}{a} - \int_{\mathfrak{X}} R(x,dy)\ R(y,\{\Delta\})$$

$$\geqslant 1 - R(x,\{\Delta\}) - i \frac{A\beta}{a} \qquad \text{d'après (11)} .$$

Il s'en suivra que l'on a :

$$\rho(P_s^n,P_t^n) = \int R^n(x,\mathfrak{X})\nu(dx) \geqslant 1 - \frac{(n-1)A\beta}{a} - \int R(x,\{\Delta\})\nu(dx)$$

$$\geqslant 1 - \frac{nA\beta}{a} \qquad \text{par} \quad (12) .$$

En utilisant la relation (8) et en remplaçant β par sa valeur , on obtiendra ainsi :

$$\frac{1}{2}||P_s^n - P_t^n|| \leqslant \sqrt{\frac{6An}{a}}\ d(s,t) .$$

(D4) est immédait puisque l'on a :

$$\rho(P_s^{2n},P_t^{2n}) = \int q_{s,t}^{2n}(x,y)\ \nu(dx)\ \mu(dy) \leqslant \left[\sup_x \int q_{s,t}^2(x,y)\ \mu(dy)\right]^n \leqslant e^{-3nd^2(s,t)}$$

comme ρ décroît avec n on a la même inégalité pour $\rho(P_s^{2n+1},P_t^{2n+1})$ et le résultat s'ensuit puisque Π est inférieur à ρ .

Pour (D2) on utilise le fait que d est comparable à la distance de Hellinger. En effet, on a :

$$\rho(P_s^2,P_t^2) \geqslant 1 - \frac{2A}{a}\left[1 - e^{-3d^2(s,t)}\right]$$

et

$$\rho(P_s^2,P_t^2) \leqslant e^{-3d^2(s,t)} .$$

On aura donc :

$$1 - e^{-3d^2(s,t)} \leqslant H^2(P_s^2, P_t^2) \leqslant \frac{2A}{a}\left[1 - e^{-3d^2(s,t)}\right] .$$

Ce qui montre que d et H sont équivalentes tant que H n'est pas trop grand et donc que d vérifie bien (D2) pour M et B convenablement choisis.

Bibliographie :

[1] LE CAM L. : Convergence of estimates under dimensionnality restrictions
Annals of Statistics, 1973, vol. 1, n° 1, p. 38-53

[2] LE CAM L. : On local and global properties in the theory of asymptotic normality of experiments
Stochastic process and related Topics, vol. 1, Acad. Press, 1975

[3] KRAFT C. : Some conditions for consistency and uniform consistency of statistical procedures
University of California, Publications in Statistics, vol. 2, n° 6, p.125-142

[4] DACUNHA-CASTELLE D. : Vitesse de convergence pour certains problèmes statistiques
Ecole d'Eté de Probabilités de Saint-Flour, 1977 - Lecture Notes in Maths. n° 678, Springer-Verlag.

Lucien BIRGÉ
Université Paris VII
Mathématiques
ERA CNRS 532 "Statistique Appliquée"
2, Place Jussieu
75221 PARIS CEDEX 05