ANNALES DE L'I. H. P., SECTION B

JEAN DESHAYES
DOMINIQUE PICARD

Principe d'invariance sur le processus de vraisemblance

Annales de l'I. H. P., section B, tome 20, nº 1 (1984), p. 1-20
<http://www.numdam.org/item?id=AIHPB_1984__20_1_1_0>

Ann. Inst. Henri Poincaré,
Vol. 20, n° 1, 1984, p. 1-20. *Probabilités et Statistiques*

Principe d'invariance sur le processus de vraisemblance

par

Jean DESHAYES et Dominique PICARD
E. R. A. 532, « Statistique Appliquée »,
Mathématiques, Bât. 425, 91405 Orsay Cedex

RÉSUMÉ. — Nous montrons un principe d'invariance sur le processus de vraisemblance d'un modèle régulier en adjoignant un paramètre temporel au paramètre statistique classique. La convergence étroite dans un espace de fonctions continues où sont liés ces deux types de paramètres fournit un outil fondamental à l'étude asymptotique des méthodes de vraisemblance.

ABSTRACT. — We show an invariance principle for the likelihood process of a regular model, adding a time-parameter to the classical statistical one. The weak convergence in a space of continuous functions where these two types of parameters are linked provides a fundamental tool for the asymptotic investigation of likelihood methods.

1. INTRODUCTION

Le but de cet article est de démontrer un Principe d'Invariance sur le processus de vraisemblance (Théorème 1) dans le cadre d'un modèle régulier classique. Ce type de résultat est indispensable pour certaines applications statistiques dans lesquelles la chronologie des observations est importante. Pour s'en convaincre, regardons rapidement sa signification dans le cadre simplifié des variables gaussiennes. (Les supréma utilisés se calculent exactement).

1.1. Quelques notations préliminaires.

Nous noterons $E_n(P_\theta, \theta \in \Theta)$ l'expérience statistique consistant à observer n variables aléatoires $X_1, X_2, \ldots, X_n$ à valeurs dans $\mathscr{X}$, espace Polonais, indépendantes, identiquement distribuées, de loi P_θ. C(E) désignera l'ensemble des fonctions réelles continues sur l'espace (métrique) E. $[y]$ désignera la partie entière de y. Sup (f, D) désignera le suprémum de la fonction réelle f sur la région D incluse dans son domaine de définition. Enfin Argsup (f, D) désignera le point d'atteinte de ce suprémum, quand celui-ci est unique.'

Considérons la suite d'expériences E_n pour laquelle l'espace $\mathscr{X}$ est la droite réelle et les lois P_θ sont les gaussiennes N_θ de moyenne θ et de variance 1, et la suite de processus Z_n :

$$\begin{aligned} \mathbb{R} \times [0,1] &\to \mathbb{R}^+ \\ (m,t) &\to Z_n(m,t) = \exp \left\{ \frac{m}{\sqrt{n}} \sum_{i=1}^{[nt]} X_i + (nt - [nt])X_{[nt]+1} - \frac{1}{2} m^2 t \right\} \end{aligned}$$

PROPOSITION 1. — Sous P_0, Z_n converge étroitement dans $C(\mathbb{R} \times [0,1])$ muni de la topologie de la convergence uniforme sur les compacts vers le processus $Z = \exp\left\{ mW(t) - m^2 \frac{t}{2} \right\}$, où $W(\cdot)$ est un mouvement Brownien réel sur $[0,1]$.

De plus, si φ est une fonction continue par morceaux et positive sur $[0,1]$ et $m_n(t) = \operatorname{Argsup}(Z_n(.,t), \mathbb{R})$, on a :

i) $\operatorname{Lim}_n \sup_t |\varphi(t)m_n(t)| = +\infty$ P_0 p. s. si

$$\lim_{t\to 0} \frac{\varphi(t)}{\sqrt{t}} \sqrt{\operatorname{Log}\log\frac{1}{t}} = +\infty \quad (1.1)$$

ii) $\{ \sup_t |\varphi(t)m_n(t)|, n \in \mathbb{N} \}$ est borné en P_0 probabilité si

$$\lim_{t\to 0} \frac{\varphi(t)}{\sqrt{t}} \sqrt{\operatorname{Log}\log\frac{1}{t}} = 0 \quad (1.2)$$

$Z_n\left(m, \frac{k}{n}\right)$ représente évidemment le rapport de vraisemblance de la loi $P_{\frac{m}{\sqrt{n}}}$ à la loi P_0 calculé sur les k premières observations. La convergence cylindrique traduit la convergence de la suite d'expériences séquentialisées $E_n^* = \{ X_1, \ldots, X_k \ iid\, N_{\frac{\theta}{\sqrt{n}}},\ k \in \{1, \ldots, n\},\ \theta \in \mathbb{R} \}$ vers $E^* = \{ W(t) - \theta t,\ \theta \in \mathbb{R},\ t \in [0,1] \}$

ce qui est une trivialité, puisque au sens de la déficience (cf. Le Cam [7]) E_n^* et $E_n(N_{\frac{\theta}{\sqrt{n}}}, \theta \in \mathbb{R})$ sont équivalentes ainsi que E^* et $E_1(N_\theta, \theta \in \mathbb{R})$.

Il est à noter que notre point de vue sera, par ailleurs relativement étranger à la notion de déficience entre expériences qui assimile expérience séquentialisée et expérience classique, alors que nous nous attacherons à étudier les comportements qui les différencient.

L'estimateur du maximum de vraisemblance calculé sur les k premières observations est $\hat{\theta}_k = \frac{1}{\sqrt{n}} m_n\left(\frac{k}{n}\right)$ et l'égalité (1.1) traduit, par exemple l'instabilité de la famille $\sqrt{n}\,\hat{\theta}_k$ lorsque $\frac{k}{n}$ approche 0, de même que celle de $\sqrt{k}\,\hat{\theta}_k$ ce qui est, peut-être, moins intuitif. Ces faits ne sont pas sans conséquences dans la pratique : Donnons-en un exemple :

1.2. **Application aux ruptures de modèles.**

Supposons que l'on désire tester l'hypothèse H_0 : $X_1, \ldots, X_n$ *iid* N_θ contre l'hypothèse H_1 : Il existe k dans $\{1, \ldots, n-1\}$ tel que $X_1, \ldots, X_k$ *iid* N_{θ_1}, $X_{k+1}, \ldots, X_n$ *iid* N_{θ_2} avec $\theta_1 \neq \theta_2$.

La statistique du test du rapport de vraisemblance de H_0 contre H_1 (généralisation aux hypothèses multiples du test de Neyman-Pearson) s'écrit si $g(X_1, \ldots, X_k, \theta)$ désigne la vraisemblance du k échantillon de P_θ :

$$L_n = \frac{\sup_k \sup_{\theta_1,\theta_2 \in \mathbb{R}} g(X_1, \ldots, X_k, \theta_1) g(X_{k+1}, \ldots, X_n, \theta_2)}{\sup_{\theta \in \mathbb{R}} g(X_1, \ldots, X_n, \theta)}$$
$$= \sup_{k \in \{1, \ldots n-1\}} \exp\left\{ \frac{1}{2} \frac{k}{n} \left(1 - \frac{k}{n}\right)^{-1} \left(m_n\left(\frac{k}{n}\right) - m_n(1)\right)^2 \right\}$$

L'égalité (1.1) se traduit par le fait que le niveau de ce test n'est pas approximable asymptotiquement puisque L_n tend vers l'infini p. s. sous H_0 ; ce qui est fondamentalement différent des problèmes classiques où le log du rapport de vraisemblance est généralement asymptotiquement un χ^2.

D'autre part, l'égalité (1.2) ainsi que la première partie de la proposition affirment, qu'en revanche, si l'on teste une rupture apparaissant loin des bords : pour tout ε de $]0, 1[$, la statistique

$$L_n(\varepsilon) = \sup_{k \in \{n\varepsilon, \ldots, n(1-\varepsilon)\}} \exp \frac{1}{2} \frac{k}{n} \left(1 - \frac{k}{n}\right)^{-1} \left(m_n\left(\frac{k}{n}\right) - m_n(1)\right)^2$$

est approximable en loi sous H_0.

Ces assertions conjointes permettent d'affirmer qu'il ne faut pas chercher une renormalisation de Log L_n par une suite de constantes c_n tendant vers 0, car c_n Log $L_n(\varepsilon)$ tendrait alors vers 0 en probabilité sous H_0, ce qui ferait perdre tout espoir de sensibilité du test face à une rupture intervenant sur le domaine $\{ n\varepsilon, \ldots, n(1-\varepsilon) \}$.

La seule renormalisation possible est alors une pénalisation aux bords de la fonction de vraisemblance du type :

$$\sup_{k \in \{1,\ldots,n-1\}} \exp \frac{1}{2} \left\{ \psi\left(\frac{k}{n}\right)\left(\psi(1) - \psi\left(\frac{k}{n}\right)\right)\left[m_n\left(\frac{k}{n}\right) - m_n(1)\right] \right\}^2$$

On trouvera une étude détaillée de ces types de problèmes dans [*12*]. Les égalités (1.1) et (1.2) traduisent le degré d'instabilité de la famille des estimateurs du maximum de vraisemblance au bord de l'intervalle d'observation. Mais ce phénomène n'est pas spécifique à la classe d'estimateurs choisie. C'est en particulier pour élargir cette classe (par exemple aux estimateurs bayésiens) que nous nous sommes penchés sur le comportement asymptotique de la suite des processus de vraisemblance dans des espaces fonctionnels adaptés. Le résultat le plus précis sur le comportement asymptotique dans le cadre général du processus de vraisemblance classique $(Z_n(.,1))$ — Ibragimov, Hasminskii [5] — démontre la convergence étroite de celui-ci lorsque $\Theta = \mathbb{R}^p$, sous des conditions de régularité relativement faibles, dans $C_0(\mathbb{R}^p)$ (fonctions de $C(\mathbb{R}^p)$ tendant vers 0 à l'infini). L'espace C_0 n'est, visiblement, plus adapté ici : (1.1) ; il est donc nécessaire d'introduire un espace C_φ pour arbitrer la dépendance entre les paramètres statistique et temporel : θ et t au voisinage du point 0 de l'intervalle [0, 1] :

1.3. **Espace C_φ.**

Soit φ une fonction continue par morceaux sur [0, 1], nulle en 0, valant 1 en 1 et telle que $\psi(t) = \dfrac{\varphi(t)}{\sqrt{t}}$ soit croissante sur [0, 1]. Nous appellerons C_φ l'ensemble des fonctions z continues sur $\mathbb{R}^p \times [0,1]$ valant 1 sur $\mathbb{R}^p \times \{0\}$ et telles que, si D_M désigne l'ensemble

$$\{ (\theta, t), \varphi(t) \mid \theta \mid > M \}, \lim_{M \to \infty} \sup (z, D_M) = 0 .$$

C_φ est muni de la topologie définie comme suit :

On dira que z_n converge vers z si et seulement si :

i) pour tout ε positif, pour tout compact K de $\mathbb{R}^p$, il existe un entier n_0 à partir duquel, $\sup(|z_n - z|, K \times [0,1]) \leq \varepsilon$.

ii) Pour tout ε positif, il existe un réel M_0 à partir duquel pour tout n, $\sup(|z_n|, D_M) \leq \varepsilon$.

Nous ne détaillerons pas plus cette topologie qui est d'ailleurs métrisable par la distance :

$$d(z_1, z_2) = \sum_{j=0}^{\infty} \frac{1}{2^j} \sup(|z_1 - z_2|, j \leq |\theta| < j+1) + \lim_{M \to \infty} \sup(|z_1 - z_2|, D_M)$$

Nous nous bornerons à en décrire les ensembles relativement compacts ; d'après le théorème d'Arzela-Ascoli, un ensemble $\mathscr{X}$ inclus dans C_φ est relativement compact si et seulement si :

i) pour tout ε positif, pour tout compact K, il existe δ tel que :

$$\sup_{z \in \mathscr{X}} \sup_{\substack{(\theta_1, t_1) \in K x[0,1] \\ (\theta_2, t_2) \in K x[0,1] \\ |(\theta_1, t_1) - (\theta_2, t_2)| \leq \delta}} |z(\theta_1, t_1) - z(\theta_2, t_2)| \leq \varepsilon$$

ii) pour tout ε positif, il existe M tel que :

$$\sup_{z \in \mathscr{X}} \sup(|z|, D_M) \leq \varepsilon .$$

1.4. **Hypothèses sur le modèle.**

Nous considérons la suite d'expériences $E_n(P_\theta, \theta \in \Theta)$ où les lois P_θ sont absolument continues par rapport à une mesure σ-finie μ, de densité $g(.,\theta)$, l'ensemble Θ étant identifié à un ouvert de $\mathbb{R}^p$. On considère le processus de vraisemblance défini comme suit :

Pour tout θ appartenant à $\mathbb{R}^p$, $Z_n(\theta, .)$ est la ligne polygonale joignant les points $(0;1)$ et $\left(\frac{k}{n}, \left[\prod_{i=1}^{k} \frac{g\left(X_1, \theta_0 + \frac{\theta}{\sqrt{n}}\right)}{g(X_i, \theta_0)} \right]^{\frac{1}{q}} \right)$ $k = 1, 2, \ldots, n$.

Le point θ_0 est la « vraie valeur » du paramètre et les autres points correspondent aux autres valeurs possibles du paramètre, renormalisées autour de θ_0 pour que la suite d'expériences E_n converge lorsque le modèle est régulier. Le nombre q n'a d'intérêt que technique ; sa valeur sera fixée par les hypothèses sur le modèle. Nous supposerons vérifiées les hypothèses suivantes :

A. *i*) la fonction $\theta \rightsquigarrow g(x, \theta)$ est absolument continue sur $\mathbb{R}^p$, μ-presque sûrement en x.

ii) $I(\theta) = \displaystyle\int \frac{\partial}{\partial\theta} g^{1/2}(x, \theta) \frac{{}^t\partial}{\partial\theta} g^{1/2}(x, \theta) d\mu(x)$ est continue sur $\mathbb{R}^p$ et définie positive en θ_0.

iii) $\forall\delta > 0, \quad \forall w, \quad |w| = 1,$

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^p} \int_{|v - \theta_0| \leq \varepsilon} d\lambda(v) \int_{\{x, |\log \frac{g(x,\theta_0 + \varepsilon w)}{g(x,\varepsilon_0)}| > \delta\}} \frac{{}^t\partial}{\partial v} g^{1/2}(x, v) \frac{\partial}{\partial v} g^{1/2}(x, v) d\mu(x) = 0$$

où λ est la mesure de Lebesgue sur $\mathbb{R}^p$.

Pour un sous-ensemble u non vide de $\{1, 2, \ldots, p\}$, on notera ∂_u ou $\dfrac{\partial^{l(u)}}{\partial\theta^u}$ l'opérateur de dérivation par rapport aux coordonnées d'indices appartenant à u : ($l(u)$ désigne le cardinal de u).

Si $u = \{j_1, j_2, \ldots, j_{l(u)}\}, \qquad \partial_u = \dfrac{\partial^{l(u)}}{\partial\theta^{(j_1)} \ldots \partial\theta^{(j_{l(u)})}}.$

B. Il existe $a > 0$, un entier $m \geq 1$ et un entier pair q tels que $m.q > p$ et $\displaystyle\sup_{\theta \in \mathbb{R}^p} \frac{1}{1 + |\theta|^a} \int |\partial_u g^{\frac{1}{q}}(x, \theta_0 + \theta)|^q d\mu(x) < \infty$ pour tout sous-ensemble u vérifiant $l(u) \leq m$.

C. Il existe $\eta > 0$ tel que $\displaystyle\overline{\lim_{|\theta| \to \infty}} |\theta|^\eta \cdot \rho(\theta_0, \theta_0 + \theta) < \infty$ où ρ désigne l'affinité de Hellinger $\rho(\theta_0, \theta_0 + \theta) = \displaystyle\int_{\mathscr{X}} [g(x, \theta_0) g(x, \theta_0 + \theta)]^{\frac{1}{2}} d\mu(x)$.

D. Pour tout ε positif,

$$P_{\theta_0} \left\{ \sup_{\substack{\theta \\ |\theta| > M}} \frac{g(X, \theta_0 + \theta)}{g(X, \theta_0)} > \varepsilon \right\}$$

tend vers 0 lorsque M tend vers l'infini.

E. La matrice Hessienne en θ de $L(X, \theta) = \operatorname{Log} g(X, \theta_0 + \theta)$ est p. s. continue en $\theta = 0$ et il existe $\delta > 0$ tel que $E \sup(|\ddot{L}(X, .)|, |\theta| \leq \delta) < +\infty$.

Remarques sur les hypothèses de régularité :

i) Dans le cas où Θ n'est pas l'espace $\mathbb{R}^p$ tout entier, le processus Z_n

n'est défini *a priori* que sur l'ensemble $\mathscr{D}_n = \left\{ \theta, \theta_0 + \frac{\theta}{\sqrt{n}} \in \Theta \right\}$ mais il est facile de poser :

$$Z_n(\theta\,; 0) \equiv 1, \qquad \forall \theta \in \mathbb{R}^p$$
$$Z_n\left(0, \frac{k}{n}\right) = 0$$

pour θ appartenant au complémentaire d'un voisinage de $\mathscr{D}_n$, $k=1, \ldots, n$ et de le prolonger en un processus continu sur $\mathbb{R}^p$ tout entier sans changer les propriétés de convergence ; désormais, nous supposons $\Theta = \mathbb{R}^p$.

ii) Ces hypothèses peuvent sembler restrictives mais en fait, elles coïncident à peu de choses près avec celles posées par Ibragimov-Hasminskii [5] pour montrer la convergence dans $\mathscr{C}_0$ de la suite de processus $Z_n(.\,, 1)$. Nous pourrions poser des hypothèses plus larges en particulier pour montrer la convergence uniforme sur les compacts, cela sera développé dans l'annexe (partie III).

iii) L'hypothèse B écrite dans le cas particulier : $m = 1$ ne nécessite plus la parité de l'entier q et c'est alors exactement l'hypothèse prise par Ibragimov-Kasminskii [5]. Il peut arriver que le modèle soit suffisamment dérivable en $\theta (m > 1)$ mais ne soit pas intégrable avec une puissance suffisante ($q < p$), ce qui justifie la généralisation.

Nous sommes en mesure, maintenant, d'énoncer le

THÉORÈME 1. — Sous les hypothèses A, B, C, D, E, la suite de processus Z_n converge étroitement sous P_{θ_0} dans C_φ vers le processus Z, dès que φ vérifie la condition (1.3) :

$$\int_0^1 \left[\frac{\varphi(t)}{t}\right]^2 dt < +\infty$$

$$Z(\theta, t) = \exp \frac{1}{q} \left\{ {}^t\theta I(\theta_0)^{\frac{1}{2}} W(t) - \frac{t}{2} \cdot {}^t\theta I(\theta_0)\theta \right\}$$

où $W(\cdot)$ est un mouvement Brownien réel à p dimensions sur $[0, 1]$.

2. DÉMONSTRATION DU THÉORÈME 1

La démonstration se fait en 2 parties : convergence sur $C(\mathbb{R}^p \times [0, 1])$ muni de la topologie de la convergence uniforme sur les compacts et contrôle du comportement à l'infini.

2.1. Convergence sur les compacts.

La première partie, que l'on peut résumer par la proposition suivante, est de loin, la plus simple : sa portée autonome ne dépasse pas fondamentalement celle du résultat d'Ibragimov-Hasminskii.

PROPOSITION 2. — Supposons que :

i) l'hypothèse D est vérifiée ;

ii) la matrice d'information de Fisher $I(\theta)$ existe et est définie au point θ_0 ;

iii) la suite de processus $Z_n(.,1)$ converge étroitement sous P_{θ_0} vers $Z(.,1)$, dans $C(\mathbb{R}^p)$ (muni de la topologie de la convergence uniforme sur les compacts).

Alors la suite Z_n converge, elle aussi vers Z dans $C(\mathbb{R}^p \times [0,1])$ muni de la même topologie.

Remarque. — Le fait que les hypothèses A et B impliquent la tension de la suite $Z_n(.,1)$ sera développé au paragraphe 3.

Démonstration de la Proposition 2. — Bien que très apparentée aux classiques théorèmes d'Invariance sur C(R) ou D(R) avec R pavé de $\mathbb{R}^p$ (cf., par exemple, Bickel-Wichura [*1*]), cette preuve garde toutefois un caractère particulier : la démonstration est, comme toujours une adaptation de celle de Donsker (cf. [*2*]), mais, d'une part, la sous-jacence d'un développement de Taylor au second ordre rend les choses plus délicates, alors que la liaison entre les paramètres θ et t (cf. inégalité (2.3)) a plutôt tendance à aplanir les difficultés.

* Dans ce paragraphe, nous nous intéresserons, en fait, à la suite $Y_n = \operatorname{Log} Z_n$: les problèmes sont, ici, équivalents mais Y_n plus maniable.

* La convergence des marginales finies est facile, nous concentrerons notre attention sur la preuve de la tension de la suite Y_n : nous montrerons donc que pour tout compact K de $\mathbb{R}^p$ et tout ε positif, il existe δ positif, tel que pour tout n, on ait :

$$P\Big(\sup_{\substack{t_1,t_2\in[0,1]\\ |t_1-t_2|\le\delta}} \sup_{\theta\in K} |Y_n(\theta,t_1) - Y_n(\theta,t_2)| \ge \varepsilon\Big) \le \varepsilon \tag{2.1}$$

et

$$P\Big(\sup_{\substack{\theta_1,\theta_2\in K\\ |\theta_1-\theta_2|\le\delta}} \sup_{t\in[0,1]} |Y_n(\theta_1,t) - Y_n(\theta_2,t)| \ge \varepsilon\Big) \le \varepsilon \tag{2.2}$$

Un point essentiel de cette démonstration est le lemme suivant :

LEMME 1. — Soient $\xi_1(.), \dots, \xi_n(.)$ des fonctions aléatoires indépendantes, définies et continues sur un compact K de $\mathbb{R}^p$, à valeurs dans $\mathbb{R}^r$.

Posons $S_k(\theta) = \sum_{j=1}^{k} \xi_j(\theta)$, et $\tau_A = \inf \{ j, \sup_{\theta \in K} | S_j(\theta) | > A \}$. Alors pour tout A positif :

$$P(\tau_A \leq n) \leq P \left\{ \operatorname{Sup}_{\theta \in K} | S_n(\theta) | > \frac{A}{2} \right\} \left[1 - \operatorname{Sup}_{k=1,\dots,n-1} P \left\{ \operatorname{Sup}_{\theta \in K} | S_n(\theta) - S_k(\theta) | > \frac{A}{2} \right\} \right]^{-1}$$

Démonstration du lemme 1. — Il est facile de vérifier que :

$$\begin{aligned} P(\tau_A \leq n) - P \left\{ \operatorname{Sup}_{\theta \in K} | S_n(\theta) | > \frac{A}{2} \right\} & \\ \leq \sum_{k=1}^{n-1} P\left(\tau_A = k \cap \operatorname{Sup}_{\theta \in K} | S_n(\theta)) | \leq \frac{A}{2} \right) & \\ \leq \sum_{k=1}^{n-1} P\left(\tau_A = k \cap \operatorname{Sup}_{\theta \in K} | S_n(\theta) - S_k(\theta) | > \frac{A}{2} \right) & \\ \leq \sum_{k=1}^{n-1} P(\tau_A = k) \cdot \operatorname{Sup}_{k=1,\dots,n-1} P \left\{ \operatorname{Sup}_{\theta \in K} | S_n(\theta) - S_k(\theta) | > \frac{A}{2} \right\} & \\ \leq P(\tau_A \leq n) \cdot \operatorname{Sup}_{k=1,\dots,n-1} P \left\{ \operatorname{Sup}_{\theta \in K} | S_n(\theta) - S_k(\theta) | > \frac{A}{2} \right\} & \end{aligned}$$

ce qui conduit au résultat annoncé.

Nous ne donnerons ici que les principales étapes qui permettent d'obtenir (2.1) (L'inégalité (2.2) s'obtient presque directement à partir du lemme 1 et du critère de tension établi au paragraphe 3).

En considérant une partition de l'intervalle $[0, 1]$ de pas δ :

$$\tau_1 = 0, \dots, \tau_{\left[\frac{1}{\delta}\right]+1} = 1,$$

et en utilisant la « stationnarité » en t de Y_n, on remplace le premier membre de (2.1) par l'évaluation de l'oscillation en 0 :

$$\frac{1}{\delta} P\left(\sup (Y_n, A(n, \delta)) > \frac{\varepsilon}{3} \right) \quad \text{si} \quad A(n, \delta) = \left\{ \left(\theta, \frac{j}{n} \right), \theta \in K, j \in \{ 1, \dots, [n\delta] \} \right\}$$

En utilisant le lemme 1 et en remarquant que $Y_n\left(\theta, \frac{j}{n}\right) = Y_j\left(\theta\sqrt{\frac{j}{n}}, 1\right)$

$$\frac{1}{\delta} P\left(\sup (Y_n, A(n, \delta)) > \frac{\varepsilon}{3}\right) \leq \frac{1}{\delta} P\left(\sup (Y_{[n\delta]}(\,.\,, 1), K\sqrt{\delta}) > \frac{\varepsilon}{6}\right) \times$$
$$\times \left[1 - \sup_{j=1,\ldots,[n\delta]} P(\sup (Y_j(\,.\,, 1), K\sqrt{\delta}) > \frac{\varepsilon}{6}\right]^{-1} \quad (2.3)$$

On a : $\sup_n P\left(\sup (Y_n(\,.\,, 1), K\sqrt{\delta}) > \frac{\varepsilon}{6}\right) \leq \frac{1}{2}$ pour δ suffisamment petit si Z_n $(\,.\,, 1)$ est une suite tendue, ce qui minore le dénominateur. Pour le numérateur, il convient de séparer les petites valeurs de n des grandes et de remarquer que : $(Y = \text{Log } Z)$,

— Il existe δ_0 tel que $P\left(\sup (Y(\,.\,, 1), K\sqrt{\delta_0}) > \frac{\varepsilon}{6}\right) \leq \frac{\varepsilon}{2} \delta_0$ (en utilisant le moment d'ordre 4 de $W(1)$).

— Il existe n_0 tel que, pour $n \geq n_0$,

$$\left| P\left(\sup (Y_n(\,.\,, 1), K\sqrt{\delta_0}) > \frac{\varepsilon}{6}\right) - P\left(\sup (Y(\,.\,, 1), K\sqrt{\delta_0}) > \frac{\varepsilon}{6}\right) \right| \leq \frac{\varepsilon}{2} \delta_0$$

(puisque $Z_n(\,.\,, 1)$ converge étroitement).

— La tension des n_0 premiers éléments de la suite Z_n provient directement de l'hypothèse D.

2.2. **Comportement à l'infini.**

Montrons maintenant la tension de la suite Z_n dans l'espace C_φ, c'est-à-dire vérifions que : pour tout ε positif, il existe M tel que uniformément en n,

$$P_{\theta_0} \{ \text{Sup} (Z_n ; D_M) \geq \varepsilon \} \leq \varepsilon \quad (2.4)$$

Le contrôle de ce suprémum repose sur le découpage du domaine D_M en deux régions donnant lieu à des techniques très différentes :

— la région $D_M(1)$ où le contrôle à l'infini sur θ est suffisamment rigoureux pour autoriser un traitement brutal sur k

$$D_M(1) = \left\{ (\theta, k), |\theta| > \sqrt{n\delta} \vee \frac{M}{\varphi\left(\frac{k}{n}\right)}, 1 \leq k \leq n \right\}$$

où $x_\vee y$ désigne le maximum de x et de y.

— La région $D_M(2)$ où les valeurs petites de $\dfrac{\theta}{\sqrt{n}}$ permettent un développement de Taylor mais nécessitent un traitement plus fin sur k :

$$D_M(2) = \left\{ (\theta, k), |\theta| \leq \sqrt{n}\delta, \quad \varphi\left(\frac{k}{n}\right)|\theta| > M, \quad 1 \leq k \leq n \right\}$$

Pour la première partie, nous utilisons fondamentalement le lemme suivant :

LEMME 2. — Pour tout A (supérieur à 2) et N positif, il existe des constantes B et γ positives et un entier $n(N)$ à partir duquel

$$P\{ \operatorname{Sup}(Z_n(.,1)\,;|\theta| > A) \geq \varepsilon \} \leq \frac{B}{\varepsilon^\gamma \cdot A^N}.$$

La démonstration nécessite l'introduction des processus de rectangles pour utiliser le lemme de Csensov : nous ferons donc la preuve détaillée de ce lemme dans la partie III.

Le contrôle uniforme du processus Z_n sur le domaine $D_M(1)$ est alors établi par application de l'hypothèse D pour les petites valeurs de k et par sommation pour les grandes valeurs de k (supérieures à n_0) :

LEMME 3. — Pour tout ε positif et tout N (supérieur à 2), il existe des constantes B et γ et un entier n_0 tels que pour tout n supérieur à n_0 et tout M supérieur à 2,

$$\sum_{k=n_0}^{n} P\left\{ \operatorname{Sup}\left(Z_n\left(.,\frac{k}{n}\right); |\theta| > \sqrt{n}\delta \vee \frac{M}{\varphi\left(\frac{k}{n}\right)} \right) \geq \varepsilon \right\} \leq \frac{B}{\varepsilon^\gamma \cdot \delta^2} \cdot M^{2-N}$$

Démonstration du lemme 3. — Lorsque M est supérieur à $\sqrt{n}\delta$, nous devons majorer :

$$\sum_{k=n_0}^{n} P\left\{ \operatorname{Sup}\left(Z_k(.,1)\,; |\theta| > \frac{M}{\psi\left(\frac{k}{n}\right)} \right) \geq \varepsilon \right\}$$

et le lemme 2 fournit directement la majoration $\dfrac{B}{\varepsilon^\gamma \delta^2} \cdot M^{2-N}$.

Lorsque M est inférieur à $\sqrt{n}\delta$, nous découpons la somme en deux parties avant d'appliquer le lemme 2 : avec ν désignant la partie entière de $n\cdot\varphi^{-1}\left(\frac{M}{\sqrt{n}\delta}\right)$:

$$\sum_{k=n_0}^{\nu} P\left\{\operatorname{Sup}\left(Z_k(.,1);|\theta|>\frac{M}{\psi\left(\frac{k}{n}\right)}\right)\geq\varepsilon\right\}$$

$$+\sum_{k=\nu+1}^{n} P\{\operatorname{Sup}(Z_k(.,1);|\theta|>\sqrt{k}\delta)\geq\varepsilon\}$$

$$\leq B\cdot\varepsilon^{-\gamma}\left[\delta^{-2}M^{2-N}+\frac{1}{\frac{N}{2}-1}\cdot\delta^{-N}\nu^{1-\frac{N}{2}}\right]$$

le second terme est équivalent au premier car le facteur multiplicatif

$$\frac{1}{\frac{N}{2}-1}\cdot\left(\frac{M}{\sqrt{n}\delta}\right)^{N-2}\left[\varphi^{-1}\left(\frac{M}{\sqrt{n}\delta}\right)\right]^{\frac{2-N}{2}}$$

est borné du fait que $\psi(t)$ est croissante et donc $\frac{t^2}{\varphi^{-1}(t)}$ aussi.

Pour achever la démonstration de (2.4) : convergence dans C_φ, il nous reste à contrôler la suite Z_n sur les domaines $D_M(2)$:

LEMME 4. — Pour tout ε positif $\left(\text{inférieur à } \frac{1}{2}\right)$, il existe une constante B, un entier n_0 et un réel M_0 à partir duquel

$$\operatorname*{Sup}_{n\geq n_0} P\{\operatorname{Sup}(Z_n;D_M\cap\{k\geq n_0\})\geq\varepsilon\}\leq\varepsilon+\frac{B}{M^2}.$$

Démonstration. — Nous établissons d'abord un encadrement de

$$Y_n=\operatorname{Log} Z_n$$

par un développement de Taylor à l'ordre 2 et l'utilisation de la loi des grands nombres :

$$Y_k(\theta,1)=\frac{t\theta}{\sqrt{k}}\sum_{i=1}^{k}\dot{L}(X_i,0)+\frac{1}{2k}t\theta\sum_{i=1}^{k}\ddot{L}\left(X_i,\frac{\theta_k}{\sqrt{k}}\right)\theta$$

avec $\dot{L}(X,\theta)$ désignant le gradient en θ de $\operatorname{Log} g(X,\theta_0+\theta)$ et $\theta_k\in[0,\theta]$.

Si $\lambda_{\min}$ et $\lambda_{\max}$ désignent respectivement les plus petites et les plus grandes valeurs propres de la matrice d'information $I(\theta_0)$, l'hypothèse E fournit la majoration :

$$E \operatorname{Sup} (|\ddot{L}(X, .)|, |\theta| \leq \delta) \leq -\left(1 - \frac{\varepsilon}{2}\right)\lambda_{\min}$$

et la loi forte des grands nombres appliquée aux variables indépendantes $\operatorname{Sup} (|\ddot{L}(X_i, .)|, |\theta| \leq \delta)$ permet d'affirmer

$$P\left\{ \exists k \geq n_0, \left| \frac{1}{k} \sum_{i=1}^{k} \ddot{L}\left(X_i, \frac{\theta_k}{\sqrt{k}}\right) \right| < (1 - \varepsilon)\lambda_{\min} \right\} \leq \varepsilon$$

et par conséquent d'obtenir la majoration en probabilité sur $D_M(2)$:

$$Y_k(\theta, 1) \leq \frac{t\theta}{\sqrt{k}} \sum_{i=1}^{k} \dot{L}(X_i, 0) - \frac{1 - \varepsilon}{2} \cdot \frac{\lambda_{\min}}{\lambda_{\max}} \cdot {}^t\theta I(\theta_0)\theta$$

Nous sommes ramenés au cas gaussien au facteur $\alpha = (1 - \varepsilon) \cdot \dfrac{\lambda_{\min}}{\lambda_{\max}}$ près ; il convient alors de distinguer deux cas : la valeur

$$\tilde{\theta}_k = \frac{1}{\alpha} I(\theta_0)^{-1} \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \dot{L}(X_i, 0)$$

où le suprémum non contraint en θ est atteint, appartient ou non à l'intérieur du domaine.

$$P\{ \operatorname{Sup} (Z_n ; D_M(2) \cap \{ k \geq n_0 \}) \geq \varepsilon \} = P\{ \operatorname*{Sup}_{v \leq k \leq n} \operatorname{Sup} (Z_k(., 1) ; \mathscr{D}_k) \geq \varepsilon \}$$

avec $\mathscr{D}_k = \left\{ \theta, \dfrac{M}{\psi\left(\frac{k}{n}\right)} < |\theta| < \delta\sqrt{k} \right\}$.

Uniformément en n supérieur à $\dfrac{M^2}{\delta^2}$, nous avons le majorant :

$$\varepsilon + \operatorname*{Sup}_{n \geq \frac{M^2}{\delta^2}} P\left\{ \bigcup_{k=v}^{n} |\tilde{\theta}_k| > \frac{M}{\psi\left(\frac{k}{n}\right)} \right\}$$

$$+ \operatorname*{Sup}_{n \geq \frac{M^2}{\delta^2}} P\left\{ \operatorname*{Sup}_{v \leq k \leq n} \operatorname*{Sup}_{|\theta| = \frac{M}{\psi\left(\frac{k}{n}\right)}} \frac{{}^t\theta}{\sqrt{k}} \sum_{i=1}^{k} \dot{L}(X_i, 0) - \frac{\alpha}{2} {}^t\theta I(\theta_0)\theta > q \log \varepsilon \right\}$$

en décomposant le 2^e terme suivant les coordonnées, le majorant devient :

$$\varepsilon + \operatorname*{Sup}_{n \geq \frac{M^2}{\delta^2}} \sum_{l=1}^{p} P \left\{ \operatorname*{Sup}_{k} \frac{\psi\left(\frac{k}{n}\right) | I(\theta_0)^{-1} |}{M \cdot \sqrt{k} \cdot \alpha} \left| \sum_{i=1}^{k} \dot{L}^{(l)}(X_i, 0) \right| > 1 \right\}$$
$$+ \operatorname*{Sup}_{n \geq \frac{M^2}{\delta^2}} P \left\{ \operatorname*{Sup}_{k} \frac{M}{\psi\left(\frac{k}{n}\right)} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \dot{L}(X_i, 0) \right| \geq \operatorname{Log} \varepsilon^q + \frac{\alpha}{2} \cdot \frac{M^2}{| I(\theta_0)^{-1} | \psi\left(\frac{k}{n}\right)^2} \right\} \tag{2.5}$$

L'inégalité de Birnbaum-Marshall [*3*] sur les martingales majore (2.5) par :

$$\varepsilon + \operatorname*{Sup}_{n \geq \frac{M^2}{\delta^2}} \sum_{l=1}^{p} \sum_{k=v}^{n} \frac{| I(\theta_0)^{-1} |^2 \cdot | I(\theta_0) |}{\alpha^2 \cdot M^2 \cdot n} \cdot \left[\frac{\varphi\left(\frac{k}{n}\right)}{\frac{k}{n}} \right]^2$$
$$+ \operatorname*{Sup}_{n \geq \frac{M^2}{\delta^2}} \frac{p \cdot | I(\theta_0) | \cdot M^2}{n} \cdot \sum_{k=v}^{n} \varphi\left(\frac{k}{n}\right)^{-2} \left[\frac{1-\varepsilon}{4} \cdot \frac{M^2}{| I(\theta_0)^{-1} | \cdot \left[\psi\left(\frac{k}{n}\right) \right]^2} \right]^2$$

pour M assez grand ; ce qui conduit au résultat annoncé :

$$\varepsilon + \frac{p}{\alpha^2} \cdot | I(\theta_0)^{-1} |^2 \cdot | I(\theta_0) | \int_0^1 \left[\frac{\varphi(t)}{t} \right]^2 dt \cdot \frac{1}{M^2} .$$

3. ANNEXE : TENSION DES PROCESSUS ET CRITÈRE DE CSENSOV

Les hypothèses A et B posées au théorème 1 sont suffisantes pour assurer la tension du processus $Z_n(.,1)$ sur les compacts et aussi au voisinage de l'infini. Elles peuvent être remplacées par des conditions plus faibles qui mettent mieux en valeur le critère de tension utilisé : généralisation multidimensionnelle du critère de Kolmogorov :

Dans un premier temps, nous redonnons une brève démonstration de ce critère de Csensov basé sur des « processus de pavés ». Nous serons alors en mesure de vérifier que les hypothèses A et B impliquement la condition *iii*) de la proposition 2 et enfin nous ferons une démonstration complète du lemme 2 énoncé au paragraphe précédent.

3.1. Critère de tension de Csensov.

L'extension multidimensionnelle du contrôle en probabilité des différences $Z(\eta_1)-Z(\eta_0)$ pour un processus Z à une dimension est effectuée en considérant des processus de pavés. Si η_0 et η_1 sont maintenant deux points bien ordonnés de $\mathbb{R}^r(\eta_0^{(j)} \leq \eta_1^{(j)}, j=1, \ldots, r)$ ils définissent sans ambiguïté un pavé R dans $\mathbb{R}^r$ et le processus Z est défini sur le pavé R par :

$$Z(R) = \Delta_1\Delta_2 \ldots \Delta_r Z(\eta)$$

où $\Delta_j Y(\eta^{(1)}, \ldots, \eta^{(r)})$ désigne la différence

$$Y(\eta^{(1)}, \ldots, \eta_1^{(j)}, \ldots, \eta^{(r)}) - Y(\eta^{(1)}, \ldots, \eta_0^{(j)}, \ldots, \eta^{(r)}).$$

Cela peut encore s'écrire :

$$Z(R) = \sum_{\varepsilon \in \{0,1\}^r} (-1)^{r-|\varepsilon|} Y(\eta_\varepsilon)$$

avec $\eta_\varepsilon = \eta_{\varepsilon_1}^{(1)}, \ldots, \eta_{\varepsilon_r}^{(r)}$ et $|\varepsilon| = \sum_{i=1}^{r} \varepsilon_i$; et on vérifie immédiatement que ces processus de pavés sont additifs sur les pavés ayant une face commune.

La condition de tension d'une famille de processus repose sur le lemme suivant :

LEMME 5. — Soit Z un processus de $\mathscr{C}([0,1]^r$ tel qu'il existe des constantes $B > 0$, $\gamma \geq 0$ et $\alpha > 1$ et une mesure ν finie sur $[0,1]^r$ satisfaisant pour tout A positif et tout pavé R inclus dans $[0,1]^r$:

$$P\{|Z(R)| > A\} \leq \frac{B}{A^\gamma}[\nu(R)]^\alpha$$

Alors il existe une constante C (ne dépendant que de la dimension r) telle que pour tout A positif et tout pavé R′ de $[0,1]^r$:

$$P\{\sup_{R, R \subset R'} |Z(R)| > A\} \leq \frac{C \cdot B}{A^\gamma}[\nu(R')]^\alpha$$

Démonstration. — Pour $r = 1$, ce résultat est classique : c'est par exemple le théorème (12.2) de Billingsley [*2*]. La démonstration se fait par récurrence sur la dimension r : nous la redonnons brièvement pour préciser celle de Bickel-Wichura [*1*] par l'usage indispensable en dimension $r > 2$

d'un lemme de section, omis dans [*1*] et généraliser la démonstration de Yor [*10*] en dimension 2.

Supposons le résultat vrai pour les dimensions inférieures ou égales à $r-1$ et fixons un pavé $R' = R'_{r-1} \times [s', t']$; avec des notations évidentes, nous avons :

$$\begin{aligned}\operatorname*{Sup}_{R \subset R'} |Z(R)| &= \operatorname*{Sup}_{\substack{s,t \\ s' \leq s \leq t \leq t'}} \ \operatorname*{Sup}_{R_{r-1} \subset R'_{r-1}} |Z(R_{r-1}, t) - Z(R_{r-1}, s)| \\ &= \operatorname*{Sup}_{R_{r-1} \subset R'_{r-1}} |Z(R_{r-1}, \hat{t}) - Z(R_{r-1}, \hat{s})|\end{aligned}$$

pour des variables aléatoires $\hat{s}$ et $\hat{t}$ (dépendant du rectangle R') définies grâce au théorème de section mesurable (voir par exemple Parthasarathy [*8*]).

Le processus $Z(R_{r-1}, \hat{t}) - Z(R_{r-1}, \hat{s})$ est un processus de pavé de dimension $r-1$ qui satisfait les hypothèses du lemme avec la mesure $\nu_{r-1} = \nu(. \times [s', t'])$; cette dernière assertion résultant de l'application, à R_{r-1} fixé, du théorème (12.2) de Billingsley [*2*].

La tension d'une famille $(Z_n)_{n \in \mathcal{N}}$ de processus de $\mathscr{C}(\mathbb{R}^p)$ va être réalisée grâce à celle des 2^p processus de pavés de dimension $r \in \{0, 1, \ldots, p\}$ définis sur les « faces » passant par un point θ_0 ($\theta_0 = 0$ sans restriction) car Z_n en est la somme (nous définissons une « face » de dimension r passant par 0 comme un sous-espace de $\mathbb{R}^p$ de dimension r où les p-r autres coordonnées sont nulles).

Critère de Csensov :

Une famille $(Z_n)_{n \in \mathcal{N}}$ de processus de $\mathscr{C}(\mathbb{R}^p)$ est tendue dès que :

i) la famille $Z_n(0)$ est tendue.

ii) Pour tout $r \in \{1, \ldots, p\}$, pour toute « face » de dimension r passant par 0, il existe des constantes $\gamma \geq 0$ et $\alpha > 1$ et une mesure de Radon ν_r sur $\mathbb{R}^r$ à marginales diffuses telles que pour tout pavé K de $\mathbb{R}^r$, il existe une constante B satisfaisant : $\forall R \subset K$, $\forall A > 0$, $\forall n \in \mathcal{N}$

$$P\{|Z_n(R)| > A\} \leq \frac{B}{A^\gamma} [\nu_r(R)]^\alpha$$

Démonstration. — Montrons la tension de l'un quelconque des processus de pavés sur une face, passant par 0, de dimension $r \in \{1, \ldots, p\}$. Par définition d'un processus de pavé de dimension r, il est nul sur les faces de dimension $r-1$ passant par 0 ; il suffit donc de montrer l'équicontinuité sur les (pavés) compacts K de $\mathbb{R}^r$.

L'additivité sur les pavés ayant une face commune nous réduit à prouver que pour tout ε positif il existe δ tel que pour tout pavé R de K dont l'un des côtés est inférieur à δ ($\| R \| < \delta$) satisfait :

$$P\{\operatorname{Sup}(|Z_n(R)|;\{R, \|R\| < \delta) > \varepsilon\} < \varepsilon \tag{3.1}$$

De façon classique, on considère dans K un treillis de pas δ dont les points seront les sommets inférieurs de pavés R_j à considérer ($j \in J$). Le premier membre de (3.1) est alors majoré par :

$$\sum_{j\in J} P\left\{\operatorname{Sup}(|Z_n(R)|;\{R \subset R_j\}) \geq \frac{\varepsilon}{3}\right\}$$

$$\leq B.C.\left(\frac{3}{\varepsilon}\right)^{\gamma} \sum_{j\in J} [\nu(R_j)]^{\alpha} \qquad \text{d'après le lemme 5.}$$

$$\leq B.C.\left(\frac{3}{\varepsilon}\right)^{\gamma} \nu(K)\cdot[\operatorname{Sup}(\nu(R);\{R, \|R\|<\delta\})]^{\alpha-1}$$

et le résultat suit du fait que les mesures ν sont à marginales diffuses. On constate par exemple que les conditions du critère de Csensov sont remplies avec $\nu = \lambda$: mesure de Lebesgue par un processus Z satisfaisant :

$$E \operatorname{Sup}_{\theta\in K}\left|\frac{\partial^p}{\partial\theta^{(1)}\ldots\partial\theta^{(p)}} Z(\theta)\right|^{\gamma} < \infty$$

avec $\gamma > 1$ car pour un pavé R de $\mathbb{R}^p$, on peut écrire :

$$Z(R) = \int\ldots\int_R \frac{\partial^p}{\partial\theta^{(1)}\ldots\partial\theta^{(p)}} Z(\theta)d\lambda(\theta)$$

3.2. **Tension de la suite** $Z_n(.\,;1)$.

Nous sommes maintenant en mesure de vérifier que les hypothèses A et B assurent la convergence de la suite de processus $Z_n(.,1)$ c'est-à-dire la condition *iii*) de la proposition 2. L'hypothèse A impliquant classiquement la convergence des marginales finies, montrons la tension de la suite $Z_n(.,1)$ sur les compacts à l'aide du

LEMME 7. — L'hypothèse B implique que pour toute dimension $r \in \{1,\ldots,p\}$, pour tous les processus de pavés définis sur des faces de dimension r

$$EZ_n(R\,;1)^q \leq B\cdot|R|^{m.a}[\lambda_r(R)]^{\alpha}$$

avec $\quad |R| = \sup\{|\theta|, \theta\in R\} \quad$ et $\quad \alpha = \dfrac{m\cdot q}{m \vee r}$.

Démonstration. — Pour tout sous-ensemble u d'indices inclus dans $\{1, \ldots, p\}$ de cardinal $l(u) = r \leq m$ et pour tout pavé de dimension r, on peut écrire :

$$Z_n(R\,;1) = \int \ldots \int_R \partial_u Z_n(\theta) d\lambda_r(\theta)$$

Le théorème de Fubini suivi de l'inégalité de Hölder donnent immédiatement :

$$EZ_n(R, 1)^q \leq [\lambda_r(R)]^q \cdot \operatorname*{Sup}_{\theta\in R} E[\partial_u Z_n(\theta\,;1)]^q \tag{3.2}$$

Pour les sous-ensembles d'indices de cardinal $r > m$, nous envisageons tous les sous-ensembles u de cardinal m exactement, $Z_n(R, 1)$ s'écrivant dans ce cas comme la somme de 2^{r-m} termes où nous appliquons la majoration (3.2) et nous obtenons :

$$EZ_n(R, 1)^q \leq 2^{q(r-m)} [\lambda_r(R)]^{\frac{m.q}{r}} \cdot \operatorname*{Sup}_{\theta\in R} E[\partial_u Z_n(\theta\,;1)]^q \tag{3.3}$$

Pour achever la démonstration de ce lemme, il reste à prouver l'existence d'une constante B satisfaisant pour tout $r \in \{1, \ldots, p\}$, pour tout u de cardinal $l(u)$ inférieur à $m \wedge r$,

$$\operatorname*{Sup}_{\theta\in R} E[\partial_u Z_n(\theta\,;1)]^q \leq B \cdot |R|^{m.a} \tag{3.4}$$

Sans restreindre la généralité, prenons $u = \{1, 2, \ldots, l\}$ et développons

$$\partial_u Z_n(\theta\,;1) = \sum_{j_1=1}^{n} \ldots \sum_{j_l=1}^{n} \prod_{k=1}^{n} \partial_1^{\delta(j_1,k)} \ldots \partial_l^{\delta(j_l,k)} \frac{g^{\frac{1}{q}}\left(X_k, \theta_0 + \frac{\theta}{\sqrt{n}}\right)}{g^{\frac{1}{q}}(X_k, \theta_0)}$$

où $\delta(.,.)$ désigne le symbole de Kronecker.

En élevant à la puissance q puis en intégrant, nous obtenons :

$$E[\partial_u Z_n(\theta, 1)]^q = n^{-\frac{l.q}{2}} \sum_{j_1^1=1}^{n} \ldots \sum_{j_l^q=1}^{n} \prod_{k=1}^{n} \int \prod_{i=1}^{q} \partial_1^{\delta(j_1^i,k)} \ldots \partial_l^{\delta(j_l^i,k)} g^{\frac{1}{q}}\left(x_k, \theta_0 + \frac{\theta}{\sqrt{n}}\right) d\mu(x_k)$$

Parmi les $n^{l.q}$ termes de la somme, beaucoup sont nuls car

$$\int \partial_j g^{\frac{1}{q}}\left(x, \theta_0 + \frac{\theta}{\sqrt{n}}\right) g^{\frac{q-1}{q}}\left(x, \theta_0 + \frac{\theta}{\sqrt{n}}\right) d\mu(x) = \frac{1}{q} \int \partial_j g\left(x, \theta_0 + \frac{\theta}{\sqrt{n}}\right) d\mu(x) = 0, \qquad j = 1, \ldots, p.$$

Ne seront donc éventuellement non nuls que les termes d'indices

$$(j_1^1, \ldots, j_l^1, \ldots, j_1^q, \ldots, j_l^q) \quad \text{vérifiant} \quad \sum_{i=1}^{q} \sum_{\nu=1}^{l} \delta(j_\nu^i, k) \neq 1, \quad \forall k = 1, \ldots, n\,;$$

il faut que les $l.q$ dérivations soient groupées par « paquets de taille au moins 2 » sur les indices k ; comme le nombre de paquets est inférieur ou égal à $\frac{l.q}{2}$, le nombre de termes non nuls est inférieur à $C(l, q) \,.\, n^{\frac{l.q}{2}}$ et chacun des termes restants est majoré facilement :

. Pour les indices k tels que $\sum_{i=1}^{q} \sum_{\nu=1}^{l} \delta(j_\nu^i, k) = 0$, l'intégrale vaut 1.

. Pour les indices k tels que $\sum_{i=1}^{q} \sum_{\nu=1}^{l} \delta(j_\nu^i, k) \geq 2$, l'inégalité de Hölder permet d'aboutir au résultat grâce à l'hypothèse B :

$$\left| \int \prod_{i=1}^{q} \partial_1^{\delta(j_1^i, k)} \ldots \partial_l^{\delta(j_l^i, k)} g^{\frac{1}{q}}\left(x_k, \theta_0 + \frac{\theta}{\sqrt{n}} \right) d\mu(x_k) \right|$$
$$\leq \prod_{i=1}^{q} \left[\int \left| \partial_1^{\delta(j_1^i, k)} \ldots \partial_l^{\delta(j_l^i, k)} g^{\frac{1}{q}}\left(x_k, \theta_0 + \frac{\theta}{\sqrt{n}} \right) \right|^q d\mu(x) \right]^{\frac{1}{q}}.$$

Nous terminons cet article par la démonstration laissée en suspens au paragraphe 2.

3.3. **Démonstration du lemme 2.**

Nous utilisons d'abord une majoration de l'affinité de Hellinger

$$EZ_n(\theta\,; 1)^{\frac{q}{2}} \leq B \,.\, |\,\theta\,|^{-N-p}$$

pour N positif arbitraire et n supérieur à $n(N)$. Cette majoration qui découle des hypothèses A et C en suivant Ibragimov-Hasminskii [5], nous la combinons avec le résultat du lemme 7 par l'inégalité de Hölder pour obtenir que pour tout N, il existe des réels α' compris entre 1 et α, γ inférieur à q et un entier $n(N)$ à partir duquel

$$EZ_n(R\,; 1)^\gamma \leq B \cdot [\,|\,R\,|^{-N-p} \vee 1\,] \cdot [\lambda_r(R)]^{\alpha'}$$

Le lemme 5 permet d'en déduire que pour tout cube R' de côté unité,

$$P\left\{\sup_{R\subset R'} |Z_n(R;1)| > \varepsilon\right\} \leq C.B.\varepsilon^{-\gamma}[|R'|^{-N-p} \vee 1]$$

et la sommation sur le domaine $\{\theta, |\theta| > A\}$ découpé en cubes de côté unité achève la démonstration du lemme 2.

REMERCIEMENTS

Nous remercions R. Azencott qui a bien voulu lire une première version de cet article et y a apporté de substantielles améliorations.

BIBLIOGRAPHIE

[1] P. J. BICKEL, M. J. WICHURA, Convergence criteria for multiparameter stochastic processes and some applications, *A. M. S.*, t. **42**, 1971, p. 1656-1670.
[2] P. BILLINGSLEY, *Convergence of probability measures*, Wiley (1968).
[3] Z. W. BIRNBAUM, A. W. MARSHALL, Some multivariate Chebyshev inequalities with extensions to continuous parameter processes, *A. M. S.*, t. **32**, 1961, p. 687-703.
[4] N. N. CHENSOV, Limit theorems for some classes of random functions. Proceedings All Union Conf. Theory Proba and Math. Statistics (Erevan, 1958). *Izdat. Akad. Nauk Armjan SSR. Erevan*, 1960, p. 280-285.
[5] I. A. IBRAGIMOV, R. Z. KHASMINSKII, Asymptotic behaviour of statistical estimators in the smooth case I. Study of the likelihood ratio. *Theory of probability and applications*, t. **17**, 1972, p. 445-462.
[6] I. A. IBRAGIMOV, R. Z. KHASMINSKII, Statistical Estimation. Asymptotic theory. *Applications of mathematics*, t. **16**, Springer-Verlag (1981).
[7] L. LECAM, On the assumption used to prove asymptotic normality of maximum likelihood estimates, *A. M. S.*, t. **41**, 1970, p. 802-828.
[8] T. PARTHASARATHY, Selection theorems and their applications, *Lecture Notes in Mathematics*, t. **263**, Springer-Verlag.
[9] R. PYKE, G. R. SHORACK, Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems, *A. M. S.*, t. **39**, 1968.
[10] M. YOR, *Le drap brownien comme limite en loi de temps locaux linéaires*, 1981 *(à paraître)*.
[11] J. DESHAYES, D. PICARD, Testing for a change point in statistical models, *Prépublication d'Orsay*, t. **52**, 1980.
[12] J. DESHAYES, D. PICARD, *Ruptures de modèles en statistique*, thèses d'État, Université Paris-Sud (mai 1983).

(Manuscrit reçu le 21 mars 1982).
(Modifié le 20 septembre 1983).