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

R. SENOUSSI

Statistique asymptotique presque-sûre de modèles statistiques convexes

Annales de l'I. H. P., section B, tome 26, nº 1 (1990), p. 19-44

<http://www.numdam.org/item?id=AIHPB_1990__26_1_19_0>

Ann. Inst. Henri Poincaré,

Vol. 26, n° 1, 1990, p. 19-44. *Probabilités et Statistiques*

Statistique asymptotique presque-sûre de modèles statistiques convexes

par

R. SENOUSSI

Bât. 425, Mathématiques, Université de Paris-Sud, 91405 Orsay, Cedex, France

RÉSUMÉ. — On propose d'abord un cadre général de modèles statistiques assurant consistance forte et lois du logarithme itéré : des exemples avec des hypothèses de convexité sont donnés. Ce cadre s'adapte bien au problème de l'identification, et éclaire les conditions d'utilisation de la méthode d'Akaïké.

Mots clés : quasi-vraisemblance, critère d'Akaïké, loi du logarithme itéré.

ABSTRACT. — We suggest a general frame of statistical models leading to strong consistency and laws of iterated logarithm: several examples, with convexity hypothesis, are given. This frame is well adapted to identification's problem, and makes clear the conditions of application of Akaïké's method.

INTRODUCTION

Pour les modèles de régression ou autorégressifs, il est courant de s'interroger sur la dimension du modèle : Comment éliminer les covariables

Classification A.M.S. : 62 F 10, 62 M 05, 62 M 09.

explicatives superflues sans apauvrir excessivement le modèle? Comment éviter donc une sur ou sous-paramétrisation?

Des situations analogues peuvent se présenter dans bien d'autres domaines. Un premier exemple réside dans l'analyse des données de survie liées à des covariables explicatives par le modèle de Cox, modèle qui fut à l'origine de notre étude et qui sera détaillé ailleurs [13]. Un autre exemple est la détermination de l'ordre de la propriété de Markov d'une chaîne aléatoire à un nombre fini d'états [14].

Il nous a semblé auparavant utile de montrer comment, en ajoutant quelques hypothèses simples, les modèles statistiques asymptotiquement normaux dont les propriétés de convergence en loi et en probabilité sont bien connues possèdent des propriétés presque-sûres de type loi du logarithme itéré. Outre leur intérêt propre, ces résultats éclairent la méthode de la vraisemblance compensée d'Akaïké [1] destinée à identifier la dimension du modèle. Une littérature fort développée ([3], [4], [8], [9]), traite ce problème de l'identification dans le cas des séries chronologiques AR et ARMA. Ce travail dégage dans une certaine mesure les outils utilisés et montre comment ils s'adaptent à de nombreux problèmes liés à l'identification de modèles.

PREMIÈRE PARTIE

PROPRIÉTÉS PRESQUE-SÛRES EN STATISTIQUE ASYMPTOTIQUE

I.1. Modèle asymptotiquement régulier pour la convergence presque-sûre

Nous commençons par énoncer le cadre d'hypothèses dans lequel nous nous plaçons et ses conséquences: l'essentiel sera de montrer ensuite ses nombreuses applications.

L'espace des paramètres Θ est une partie de $\mathbb{R}^q$, θ désignera toujours la vraie valeur, inconnue, du paramètre et θ est dans l'intérieur de Θ. Les observations sont effectuées à des instants $t \in \mathbb{T}$, $\mathbb{T}$ étant soit $\mathbb{N}$ soit $\mathbb{R}_+$. On suppose que l'espace mesurable des observations $(\Omega, \mathscr{A})$ est muni de la probabilité $\mathbb{P}_\theta$, et que, à chaque instant $t \in \mathbb{T}$ et pour chaque $\alpha \in \Theta$, on sait définir une « quasi-vraisemblance » $v_t(\alpha)$ observable à l'instant t. Dans de nombreux cas, si l'on sait calculer une vraisemblance $\mathscr{V}_t(\alpha)$, on aura $v_t(\alpha) = \operatorname{Log} \mathscr{V}_t(\alpha)$.

Pour x vecteur de $\mathbb{R}^q$, x désigne aussi la matrice colonne associée, et $x'=(x^1, x^2, \ldots, x^q)$ et pour une matrice symétrique définie positive Γ, on note par $\Gamma^{-1/2}$, $\Gamma^{1/2}$ les matrices $\Gamma^{-1/2}=\mathbb{O}'.\mathbb{D}^{-1/2}.\mathbb{O}$ et $\Gamma^{1/2}=\mathbb{O}'.\mathbb{D}^{1/2}.\mathbb{O}$ où $\mathbb{O}\Gamma\mathbb{O}'=\mathbb{D}$, $\mathbb{D}$ étant diagonale et $\mathbb{O}$ orthogonale. On note aussi par $\Gamma^{-}=\Gamma^{-1/2}.\Gamma^{-1/2}=\mathbb{O}'.\mathbb{D}^{-}.\mathbb{O}$ une inverse généralisée de Γ.

$\mathbb{H}1$. *Hypothèses relatives à la consistence*

(*a*) *Contraste entre lois de probabilités*

Il existe une fonction a de $\mathbb{T}$ dans $\mathbb{R}$, qui croît vers ∞ si $t\uparrow\infty$ et telle que pour tout α:

$$\mathscr{K}_t(\theta,\alpha)=\frac{1}{a(t)}(v_t(\theta)-v_t(\alpha))\xrightarrow[t\uparrow\infty]{\mathbb{P}_\theta\text{ p. s.}}\mathrm{K}(\theta,\alpha);$$
$$0\leqq\mathrm{K}(\theta,\alpha)\leqq\infty \qquad \text{et} \qquad (\mathrm{K}(\theta,\alpha)=0\Leftrightarrow\theta=\alpha).$$

La fonction limite $\mathrm{K}(.,.)$ joue le rôle de l'information de Kullback usuelle entre deux lois de probabilités.

(*b*) *Consistance forte des estimateurs du maximum de quasi-vraisemblance*

On suppose que, pour chaque $t\in\mathbb{T}$, on peut trouver

$$\hat\theta_t\in\Theta \quad \text{tel que} \quad v_t(\hat\theta_t)=\sup_{\alpha\in\Theta} v_t(\alpha) \quad \text{avec } \hat\theta_t\xrightarrow[t\uparrow\infty]{\mathbb{P}_\theta\text{ p. s.}}\theta.$$

Remarque 1. — Dans bien des cas, où l'on sait calculer $\hat\theta_t$, la consistance forte est vérifiée directement. Cependant on utilisera fréquemment la situation suivante: si les fonctions $\alpha\mapsto\mathscr{K}_t(\theta,\alpha)$ sont, pour tout t, $\mathbb{P}_\theta$ p. s. convexes sur Θ convexe et si la fonction limite $\alpha\mapsto\mathrm{K}(\theta,\alpha)$ est finie et strictement convexe alors $\mathbb{H}1.a$ implique $\mathbb{H}1.b$ (*voir* appendice 1).

$\mathbb{H}2$. *Modèles asymptotiquement réguliers*

(*a*) Pour tout t, $v_t(.)$ est de classe C^2 au voisinage de θ; et si $\mathscr{D}\mathbf{v}_t(.)$ et $\mathscr{D}^2 v_t(.)$ désignent le gradient et la matrice des dérivées secondes de $v_t(.)$.

(*b*) Il existe une matrice $\mathbb{I}(\theta)$ inversible telle que, si $(\theta_t^*)_{t\in\mathbb{T}}$ est une suite de v.a. à valeurs dans Θ vérifiant $\theta_t^*\xrightarrow[t\uparrow\infty]{\mathbb{P}_\theta\text{ p. s.}}\theta$, on ait

(*c*)
$$\frac{\mathscr{D}^2 v_t}{a(t)}(\theta_t^*)\xrightarrow[t\uparrow\infty]{\mathbb{P}_\theta\text{ p. s.}}-\mathbb{I}(\theta)$$
$$\frac{\mathscr{D}v_t(\theta)}{\sqrt{a(t)}}\xrightarrow[t\uparrow\infty]{\mathscr{L}(\mathbb{P}_\theta)}\mathrm{N}(0,\mathbb{I}(\theta))$$

où $N(m, \Gamma)$ désigne la loi gaussienne de moyenne m et de matrice de covariance Γ.

Remarque 2. — Les hypothèses $\mathbb{H}1$ et $\mathbb{H}2$ impliquent, on le sait [6, T 2]

(i) $\sqrt{a(t)}\,(\hat{\theta}_t - \theta) \xrightarrow[t \uparrow \infty]{\mathscr{L}(\mathbb{P}_\theta)} N(0, \mathbb{I}^{-1}(\theta))$.

(ii) $2\,(v_t(\hat{\theta}_t) - v_t(\theta)) \xrightarrow[t \uparrow \infty]{\mathscr{L}(\mathbb{P}_\theta)} \chi^2(q)$.

$\mathbb{H}3$. *Loi du logarithme itéré*

On suppose que la famille $\left(\dfrac{\mathbb{I}^{-1/2}(\theta)}{\sqrt{2\,a(t)\,\mathrm{Log\,Log}\,a(t)}}\,.\,\mathscr{D}\,v_t(\theta)\right)_{t \in \mathbb{T}}$ est $\mathbb{P}_\theta$ p.s. relativement compacte et que l'ensemble de ses valeurs d'adhérence est égal à la boule unité $B_q(0, 1)$ de $\mathbb{R}^q$.

Cette hypothèse souvent vérifiée permet de préciser la vitesse de convergence presque-sûre dans la remarque précédente.

THÉORÈME $\mathbb{I}$.1. — *Les hypothèses* $\mathbb{H}1$, $\mathbb{H}2$, $\mathbb{H}3$ *impliquent,*

(i) *La famille* $\left(\dfrac{\sqrt{a(t)}\,.\,\mathbb{I}^{1/2}(\theta)}{\sqrt{2\,\mathrm{Log\,Log}\,a(t)}}\,.\,(\hat{\theta}_t - \theta)\right)_{t \in \mathbb{T}}$ *est* $\mathbb{P}_\theta$ p.s. *relativement compacte dans* $\mathbb{R}^q$ *avec pour ensemble de valeurs d'adhérence la boule unité* $B_q(0, 1)$ *de* $\mathbb{R}^q$

(ii) $\limsup\limits_{t \uparrow \infty} \dfrac{v_b(\hat{\theta}_t) - v_t(\theta)}{\mathrm{Log\,Log}\,a(t)} = 1 \quad \mathbb{P}_\theta$ p.s.

Démonstration. — (i) pour $1 \leqq i \leqq q$

$$\mathscr{D}_i v_t(\theta) = \sum_{j=1}^{q} \mathscr{D}_i \mathscr{D}_j v_t(\theta^*_{i,t})\,(\theta^j - \hat{\theta}^j_t), \qquad \theta^*_{i,t} \in [\theta, \hat{\theta}_t];$$

$$\frac{\mathscr{D}_i v_t(\theta)}{\sqrt{2\,a(t)\,\mathrm{Log\,Log}\,a(t)}} = \sum_{j=1}^{q} \frac{\mathscr{D}_i \mathscr{D}_j v_t(\theta^*_{i,t})}{a(t)}\,.\,\sqrt{\frac{a(t)}{2\,\mathrm{Log\,Log}\,a(t)}}\,(\theta^j - \hat{\theta}^j_t)$$

Et l'on conclut facilement grâce à $\mathbb{H}2\,b$ et $\mathbb{H}3$.

(ii) $$v_t(\hat{\theta}_t) - v_t(\theta) = -\,a(t)\,(\hat{\theta}_t - \theta)'\left(\frac{\mathscr{D}^2 v_t(\theta^*_t)}{2\,a(t)}\right)(\hat{\theta}_t - \theta),$$
$$\theta^*_t \in [\theta, \hat{\theta}_t].$$

Les valeurs d'adhérence de $\dfrac{v_t(\hat{\theta}_t) - v_t(\theta)}{\mathrm{Log\,Log}\,a(t)}$ sont celles de

$$\left\| \frac{\sqrt{a(t)}}{\sqrt{2\,\mathrm{Log\,Log}\,a(t)}}\,.\,\mathbb{I}^{1/2}(\theta)\,(\hat{\theta}_t - \theta) \right\|^2$$

dont 1 est la plus grande valeur d'adhérence d'après (i).

$\mathbb{I}$.2. Échantillon de modèles réguliers

Un premier cadre d'application du $\mathbb{I}$.1 sera le suivant: $\mathbb{T}=\mathbb{N}$ et l'on observe des copies indépendantes d'un modèle régulier, autrement dit, à l'instant $n=1$:

1. La quasi-vraisemblance est $v_1(\alpha)$, v. a. observable et intégrable pour tout α; on suppose que:

$$\mathbb{E}_\theta(v_1(\theta)-v_1(\alpha))=\mathrm{K}(\theta,\alpha)\geqq 0,$$

où

$$\mathrm{K}(\theta,\alpha)=0 \text{ équivaut à } \theta=\alpha \text{ (identifiabilité).}$$

Si v_1 est une log-vraisemblance, $\mathrm{K}(.,.)$ est l'information de Kullback.

2. (i) $v_1(.)$ est de classe C^2 au voisinage de θ.

(ii) $\mathscr{D}v_1(\theta)$ est de carré intégrable, centrée et de covariance $\mathbb{I}(\theta)$ inversible avec $\mathbb{E}_\theta(\mathscr{D}^2 v_1(\theta))=-\mathbb{I}(\theta)$.

(iii) Il existe enfin une v. a. Z intégrable et $\eta>0$ tels que:

$$\sup_{\|\alpha-\theta\|\leqq\eta}\|\mathscr{D}^2 v_1(\alpha)-\mathscr{D}^2 v_1(\theta)\|\leqq \mathrm{Z}.$$

Alors l'observation de n copies indépendantes de ce modèle conduit à v_n somme de n v. a. indépendantes de même loi que v_1. Et, dès que $\hat\theta_n \xrightarrow[n\uparrow\infty]{\mathbb{P}_\theta\text{ p. s.}} \theta$, les hypothèses $\mathbb{H}1$, $\mathbb{H}2$ en résultent aisément pour $a(n)=n$.

3. Pour $\mathbb{H}3$, il suffit de remarquer que si Y_n est une somme de vecteurs aléatoires indépendants, centrés, de carré intégrable de matrice de covariance Γ inversible, la suite $\left(\dfrac{\Gamma^{-1/2}.\mathrm{Y}_n}{\sqrt{2n\,\mathrm{Log\,Log}\,n}}\right)$ est p. s. relativement compacte et l'ensemble de ses valeurs d'adhérence coïncide avec la boule unité (*voir* appendice 2).

Exemples. — Nous développons ici les exemples les plus simples pour montrer la généralité du modèle.

$\mathbb{E}$.1. *Échantillons de modèles réguliers classiques*

Dans le cas particulier des familles exponentielles bien paramétrées, lorsque $v_n(.)$ est le log-vraisemblance, on a

$$v_n(\alpha)=\langle\alpha,\mathrm{S}_n\rangle-n\varphi(\alpha)$$

où $S_n = \sum_{i=1}^{n} T_i$ est la somme de n vecteurs aléatoires i. i. d. de R^q dont la loi n'est pas concentrée sur un hyperplan et

$$\varphi(\alpha) = \text{Log} \int \exp\{\langle \alpha, T_i(\omega) \rangle\} d\mu(\omega).$$

Si $\Theta = \{\alpha; \varphi(\alpha) < \infty\}$, les hypothèses $\mathbb{H}1$ et $\mathbb{H}2$ sont vérifiées [6, T2]; dès que θ est un point intérieur de Θ, $\hat{\theta}_n$ existe et $\mathscr{D}^2 \varphi(\theta)$ est inversible; l'hypothèse $\mathbb{H}3$ est satisfaite, puisque $\mathbb{E}_\theta(T_i) = \mathscr{D}\varphi(\theta)$, $\text{cov}_\theta(T_i) = -\mathscr{D}^2 \varphi(\theta)$ et $\mathscr{D}v_n(\theta) = \sum_{i=n}^{n} (T_i - \mathscr{D}\varphi(\theta))$.

$\mathbb{E}2$. *Modèle linéaire unidimensionnel: estimateurs des moindres carrés*

Soient (Z_i) et (ε_i) deux suites indépendantes entre elles de v. a. indépendantes et équidistribuées; Z_i est une covariable observable à valeurs dans $\mathbb{R}^q$ et de carré intégrable: ε_i est un bruit non observable à valeurs réelles et centré. On observe alors X_i selon le modèle: $X_i = \langle \theta, Z_i \rangle + \varepsilon_i$, θ étant un paramètre inconnu.

La méthode des moindres carrés introduit la quasi-vraisemblance:

$$v_n(\alpha) = \frac{-1}{2} \sum_{i=1}^{n} (X_i - \langle \alpha, Z_i \rangle)^2 = - \sum_{i=1}^{n} (\varepsilon_i + \langle \theta - \alpha, Z_i \rangle)^2$$

et

$$\mathscr{K}_n(\theta, \alpha) = \frac{1}{2n}(v_n(\theta) - v_n(\alpha)) = \frac{-1}{2n} \sum_{i=1}^{n} (\langle \theta - \alpha, Z_i \rangle)^2 + 2\varepsilon_i \langle \theta - \alpha, Z_i \rangle$$

$$\xrightarrow[t \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} K(\theta, \alpha) = \frac{1}{2}(\theta - \alpha)' \Gamma (\theta - \alpha) \qquad \text{avec} \quad \Gamma = \mathbb{E}(Z . Z').$$

On voit que $\mathbb{H}1.a$ et $\mathbb{H}1.b$ sont vérifiées dès que Z_1 a une loi qui n'est pas concentrée sur un hyperplan de $\mathbb{R}^q$, puisque les fonctions $\alpha \mapsto \mathscr{K}_n(\theta, \alpha)$ et $\alpha \mapsto K(\theta, \alpha)$ sont convexes. Γ étant définie positive $\mathbb{H}2$ et $\mathbb{H}3$ sont aussi satisfaites car: $\mathscr{D} v_n(\alpha) = \sum_{i=1}^{n} Z_i (X_i - \langle \alpha, Z_i \rangle)$ et $\mathscr{D} v_n(\theta) = \sum_{i=1}^{n} Z_i \varepsilon_i$ est une somme de vecteurs aléatoires i. i. d. centrés et de matrice de covariance Γ. Mais $\mathscr{D}^2 v_n(\alpha) = -\sum_{i=1}^{n} Z_i Z_i'$ est indépendante de α avec $\frac{1}{n} \mathscr{D}^2 v_n(\alpha) \xrightarrow[n \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} -\Gamma$.

$\mathbb{E}3$. *Estimateurs du χ^2 minimum*

On note $(p_1, \ldots, p_s)$ une probabilité sur $E = \{1, \ldots, s\}$. On tire un n-échantillon $X_1, \ldots, X_n$ de $(E, \mathscr{P}(E), \mathbb{P}(\alpha))$ où $\mathbb{P}$ est une application de classe C^2, $\mathbb{P} : \Theta \subset \mathbb{R}^q \mapsto \Delta^s$ avec $\Delta^s = \{(p_1, \ldots, p_j)$ tel que $p_h > 0, j = 1, \ldots, s$ et $\sum_{j=1}^{s} p_j = 1\}$. L'estimateur du χ^2 minimum est celui qui minimise la quasi-vraisemblance :

$$v_n(\alpha) = \frac{1}{2} \sum_{j=1}^{s} \frac{(N_n^j - np_j(\alpha))^2}{N_n^j}$$

avec les contraintes $n = \sum_{j=1}^{s} N_n^j$, $\mathbb{P}(\alpha) = (p_1(\alpha), \ldots, p_s(\alpha)) \in \Delta^s$ et où $N_n^j = \sum_{i=1}^{n} \mathbf{1}_{\{X_i = j\}}$.

Et, si l'on tient compte de ces contraintes, on obtient :

$$\mathscr{D} v_n(\alpha) = \sum_{j=1}^{s} \frac{n}{N_n^j} (N_n^j - np_j(\alpha)) \mathscr{D} p_j(\alpha) = -n \sum_{j=1}^{s} \frac{p_j(\alpha)}{N_n^j/n} \mathscr{D} p_j(\alpha)$$

$$\mathscr{D}^2 v_n(\alpha) = -n \sum_{j=1}^{s} \frac{p_j(\alpha)}{N_n^j/n} \left(\left(\frac{\mathscr{D} p_j(\alpha)}{\sqrt{p_j(\alpha)}} \right) \left(\frac{\mathscr{D} p_j(\alpha)}{\sqrt{p_j(\alpha)}} \right)' - \mathscr{D}^2 p_j(\alpha) \right)$$

Vérifions $\mathbb{H}1\,a$:

$$\mathscr{K}_n(\theta, \alpha) = \sum_{j=1}^{s} \left(\frac{1}{2} \frac{(p_j(\theta) - p_j(\alpha))^2}{N_n^j/n} + \left(1 - \frac{p_j(\theta)}{N_n^j/n} \right) (p_j(\theta) - p_j(\alpha)) \right)$$

$$\xrightarrow[n \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} K(\theta, \alpha) = \frac{1}{2} \sum_{j=1}^{s} \frac{(p_j(\theta) - p_j(\alpha))^2}{p_j(\theta)}.$$

Dès lors que $\mathbb{P}$ est injective $\mathbb{H}1\,a$ est réalisée et $\mathbb{H}1\,b$ l'est aussi dans de nombreux cas intéressants, le plus évident étant l'identité de Δ^r dans Δ^r. $\mathbb{H}2\,a$ est évidente par hypothèse puisque $\mathbb{P}$ est de classe C^2. Étudions $\mathbb{H}2\,b$:

$$-\frac{1}{n} \mathscr{D}^2 v_n(\theta_n^*) \xrightarrow[n \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} \mathbb{I}(\theta) = \sum_{j=1}^{s} \left(\frac{\mathscr{D} p_j}{\sqrt{p_j}}(\theta) \right) \left(\frac{\mathscr{D} p_j}{\sqrt{p_j}}(\theta) \right)' = \mathbb{J}(\theta)\, \mathbb{J}'(\theta)$$

où $\mathbb{J}(\theta) = \left(\frac{\partial (\sqrt{p_l})}{\partial \alpha_k}(\theta) \right)_{k=1,\ldots,q;\, l=1,\ldots,s}$ est la matrice jacobienne de $\sqrt{\mathbb{P}}$ au point θ.

Alors, [7], $\mathbb{I}(\theta)$ est inversible si et seulement si l'application $\alpha \mapsto \sqrt{\mathbb{P}}(\alpha) = (\sqrt{p_1}(\alpha), \ldots, \sqrt{p_s}(\alpha))$ est de rang q au voisinage de θ. En suivant encore [6, T2], on peut écrire

$$\frac{\mathscr{D} v_n(\theta)}{\sqrt{n}} = \mathbb{J}(\theta) \cdot Z_n^*$$

où

$$Z_n^* = \mathbb{D}_n(\theta) \cdot Z_n(\theta) \qquad \text{avec} \quad Z_n(\theta) = \left(\frac{N_n^j - n p_j(\theta)}{\sqrt{n p_j}(\theta)} \right)_{j=1\ldots s}$$

et

$$\mathbb{D}_n(\theta) = \operatorname{diag}\left(\frac{p_j(\theta)}{N_n^j/n} ; j = 1 \ldots s \right).$$

Ainsi $\mathbb{D}_n(\theta) \xrightarrow[n \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} I_s$, I_s matrice identité $Z_n(\theta) \xrightarrow[n \uparrow \infty]{\mathscr{L}(\mathbb{P}_\theta)} Z$, Z de loi $N(0, \Gamma(\theta))$ avec $\Gamma(\theta) = I_s - (\sqrt{\mathbb{P}(\theta)})(\sqrt{\mathbb{P}(\theta)})'$ projecteur sur $(\sqrt{\mathbb{P}(\theta)})^\perp$. Autrement dit: $\frac{\mathscr{D} v_n(\theta)}{\sqrt{n}} \xrightarrow[n \uparrow \infty]{\mathscr{L}(\mathbb{P}_\theta)} \mathbb{J}(\theta) \cdot Z$, de loi $N(0, \mathbb{I}(\theta))$. $\mathbb{H}3$ s'ensuit du moment que $\mathbb{I}(\theta)$ est de rang plein, car $\frac{Z_n}{\sqrt{2 n \operatorname{Log} \operatorname{Log} n}}$ est une somme de vecteurs aléatoires indépendants, centrés et de covariance $\Gamma(\theta)$ de rang $(s-1)$. On en déduit (appendice 2) que l'ensemble des valeurs d'adhérence est l'ellipsoïde $\mathscr{E}(0, \Gamma^-) \cap \mathscr{I}m(\Gamma)$ ici $\mathscr{I}m(\Gamma) = (\sqrt{\mathbb{P}}(\theta))^\perp$. Donc

$$\left(\frac{\mathbb{I}(\theta)^{-1/2} \cdot \mathscr{D} v_n(\theta)}{\sqrt{2 n \operatorname{Log} \operatorname{Log} n}} \right)_{n \geqq 3} = \left(\mathbb{I}(\theta)^{-1/2} \cdot \mathbb{J}(\theta) \cdot \mathscr{D} v_n(\theta) \cdot \frac{Z_n}{\sqrt{2 n \operatorname{Log} \operatorname{Log} n}} \right)_{n \geqq 3}$$

est p. s. relativement compact, ayant pour ensemble de valeurs d'adhérence la boule $B_q(0, 1)$ puisque

$$\mathbb{I}(\theta)^{-1/2} \cdot \mathbb{J}(\theta) \cdot \Gamma(\theta) \cdot \mathbb{J}(\theta)' \cdot \mathbb{I}(\theta)^{-1/2} = \mathbb{I}(\theta)^{-1/2} \cdot \mathbb{J}(\theta) \cdot \mathbb{J}(\theta)' \cdot \mathbb{I}(\theta)^{-1/2} = I_q.$$

I3. Modèle de Cox

Suite à la longueur excessive dans la vérification des différentes hypothèses, ce problème est traité à part dans [13].

I4. Exemples en statistique des processus

I.4.1. *Statistique des chaînes de Markov récurrentes*

On peut se placer sans peine dans le cadre d'étude des modèles statistiques réguliers pour les chaînes de Markov [6, T2], autrement dit dès que l'on a pour la chaîne des lois des grands nombres, des théorèmes de limite centrale et de loi du logarithme itéré, on débouche aisément sur le cadre d'étude du I1. La procédure se déroule bien aussitôt que la chaîne est récurrente positive : voir à ce propos [16] et [17]. Traitons à titre d'exemples les cas simples suivants.

I.4.1.*a*. *Chaîne de Markov à espace d'état fini* $E=\{1, \ldots, s\}$

On considère une chaîne de Markov canonique $(\Omega, \mathscr{A}, (\mathbb{P}_x)_{x\in E} : (X_n)_{n\geqq 0})$ associée à une transition $\pi(.,.)$. On suppose la loi initiale connue, mais la transition inconnue et, pour tout $(l, k)\in E^2$, considérons les statistiques :

$N_n^{l,k}=\sum_{i=1}^{n} \mathbf{1}_{\{X_{i-1}=l\,:\,X_i=k\}}$ le nombre de transitions de l à k.

$N_n^{l\cdot}=\sum_{i=1}^{n-1} \mathbf{1}_{\{X_i=l\}}$ le nombre de passages en l.

On suppose que E constitue une seule classe récurrente, et soit μ la probabilité invariante. La log-vraisemblance est :

$$v_n(\pi)=\sum_{(l,k)\in D} N_n^{l,k}\operatorname{Log}\pi_{lk}$$

où

$$\pi_{lk}=\pi(l,k) \qquad \text{et} \qquad D=\{(l,k) \text{ tels que } \pi_{lk}>0\}$$

Sous les contraintes : $\pi_{ls}=1-\sum_{k=1}^{s-1}\pi_{lk}$, $l=1, \ldots, s$, on a :

$$v_n(\pi)=\sum_{l=1}^{s}\left(\sum_{k=1}^{s-1} N_n^{lk}\operatorname{Log}(\pi_{lk})+N_n^{ls}\operatorname{Log}\left(1-\sum_{k=1}^{s-1}\pi_{lk}\right)\right),$$

$$(\mathscr{D}v_n(\pi))_{lk}=(N_n^{lk}/\pi_{lk})-\left(N_n^{ls}/\left(1-\sum_{n=1}^{s-1}\pi_{ln}\right)\right),$$

$$1\leqq l\leqq s \qquad \text{et} \qquad 1\leqq k\leqq s-1.$$

Par suite $\mathscr{D}v_n(\pi)$ est une $(\mathbb{P}_\pi, \mathbb{F})$ martingale à accroissements bornés où $\mathbb{F}$ est la filtration naturelle. Son processus croissant est :

$$\langle \mathscr{D}v_n(\pi)\rangle_{lk;\,l'k'}=\begin{matrix} 0 & \text{si } l\neq l' \\ N_n^{l\cdot}/\pi_{ls} & \text{si } l=l';\ k\neq k' \\ N_n^{l\cdot}(1/\pi_{lk}+1/\pi_{ls}) & \text{si } l=l';\ k=k' \end{matrix}$$

Dans ce cas on sait [6, T2] que :

$$\left(\frac{\mathscr{D}v_n(\pi)}{\sqrt{n}}\right)_{lk}=\sqrt{\mathrm{N}_n^{l.}/n}\left\{\left(\frac{\mathrm{N}_n^{lk}-\mathrm{N}_n^{l.}\pi_{lk}}{\pi_{lk}\sqrt{\mathrm{N}_n^{l.}}}\right)-\left(\frac{\mathrm{N}_n^{ls}-\mathrm{N}_n^{l.}\pi_{ls}}{\pi_{ls}\sqrt{\mathrm{N}_n^{l.}}}\right)\right\}\underset{n\uparrow\infty}{\xrightarrow{\mathscr{L}(\mathbb{P}_\theta)}}\mathrm{N}(0,\Gamma).$$

Les estimateurs de maximum de vraisemblance sont fortement consistants

$$\hat{\pi}_{lk}^n=\frac{\mathrm{N}_n^{lk}}{\mathrm{N}_n^{l.}}\underset{n\uparrow\infty}{\xrightarrow{\mathbb{P}_\theta\text{ p. s.}}}\pi_{lk}$$

et

$$\left(\frac{\mathrm{N}_n^{lk}-\mathrm{N}_n^{l.}\pi_{lk}}{\sqrt{\mathrm{N}_n^{l.}\pi_{lk}}}\right)_{(lk)\in \mathrm{D}}\underset{n\uparrow\infty}{\xrightarrow{\mathscr{L}(\mathbb{P}_\theta)}}\mathrm{N}(0,\Delta),$$

$$\Delta=(\delta_{lk;\,l'k'})=\begin{array}{ll} 0 & \text{si } l\neq l' \\ -\sqrt{\pi_{lk}\pi_{lk'}} & \text{si } l=l';\ k\neq k' \\ 1-\pi_{lk} & \text{si } l=l';\ k=k'. \end{array}$$

D'autre part

$$-\left(\frac{\mathscr{D}^2 v_n(\pi)}{n}\right)_{lk;\,l'k'}=\begin{array}{ll} 0 & \text{si } l\neq l' \\ \mathrm{N}_n^{ls}/\pi_{ls}^2 & \text{si } l=l';\ k\neq k' \\ (\mathrm{N}_n^{lk}/\pi_{lk}^2)+(\mathrm{N}_n^{ls}/\pi_{ls}^2) & \text{si } l=l';\ k=k' \end{array}$$

$$\underset{n\uparrow\infty}{\xrightarrow{\mathbb{P}_\pi\text{ p. s.}}}\Gamma=\begin{cases} 0 & \text{si } l\neq l' \\ \mu(l)/\pi_{ls} & \text{si } l=l';\ k\neq k' \\ \mu(l)(1/\pi_{lk}+1/\pi_{ls}) & \text{sinon} \end{cases}$$

et $\langle \mathscr{D} v_n(\pi)\rangle \underset{n\uparrow\infty}{\xrightarrow{\mathbb{P}_\pi\text{ p. s.}}} \Gamma$ aussi.

$\mathscr{D} v_n(\pi)$ étant une martingale vectorielle de processus croissant, $\langle \mathscr{D} v_n(\pi)\rangle$, on tire de [17] une loi du logarithme itéré pour $\mathscr{D} v_n(\pi)$.

$$\left(\frac{\mathscr{D} v_n(\pi)}{\sqrt{2n\operatorname{Log}\operatorname{Log} n}}\right)_{lk}=\frac{\sqrt{2\mathrm{N}_n^{l.}\operatorname{Log}\operatorname{Log}\mathrm{N}^{l.}}}{\sqrt{2n\operatorname{Log}\operatorname{Log} n}}\times\frac{\sqrt{1/\pi_{ls}+1/\pi_{lk}}\,.\,\mathscr{D} v_n(\pi)_{lk}}{\sqrt{2\langle\mathscr{D} v_n(\pi)\rangle_{lk}\,.\operatorname{Log}\operatorname{Log}\langle\mathscr{D} v_n(\pi)\rangle_{lk}}}.$$

La suite $(\mathscr{D} v_n(\pi))_{n\geq 3}$ est p. s. relativement compacte et a pour ensemble de valeurs d'adhérence « l'ellipsoïde » $\mathscr{C}(0,\Gamma^-)\cap\mathscr{I}m(\Gamma)$.

I.4.1. *b. Chaîne* AR (1)

Si $(\varepsilon_n)_{n\geq 1}$ est une suite de v. a. i. i. d. centrées de variance 1 et de densité l strictement positive par rapport à la mesure de Lebesgue et si X_0 est de loi ν, indépendante de la suite (ε_n). On définit la chaîne de Markov sur $\mathbb{R}$

par:

$$X_n = \theta X_{n-1} + \varepsilon_n$$

de probabilité de transition $\pi_\theta(x, \Gamma) = \int_\Gamma l(y - \theta x)\, dy$ pour tout $\Gamma \in \mathscr{B}(\mathbb{R})$. Lorsque $\theta \in]-1, +1[$, la chaîne $(X_n)_{n \geq 1}$ est récurrente positive pour la probabilité $\mu(\theta)$ loi de la v.a. $Z = \sum_1^\infty \theta^{n-1} \varepsilon_n$ de loi invariante pour π_θ. Si on estime θ par la méthode des moindres carrés, au vu de l'observation $X_1, \ldots, X_n$ on a

$$v_n(\alpha) = -\frac{1}{2} \sum_1^n (X_i - \alpha . X_{i-1})^2;$$

$$\mathscr{D} v_n(\alpha) = \sum_1^n X_{i-1} (X_i - \alpha . X_{i-1})$$

$$\mathscr{D}^2 v_n(\alpha) = -\sum_1^n X_{i-1}^2.$$

L'estimateur des moindres carrés est fortement consistant:

$$\hat{\theta}_n = \left(\sum_1^n X_i X_{i-1} \Big/ \sum_1^n X_{i-1}^2 \right) \xrightarrow[n \uparrow \infty]{\mathbb{P}_\theta \text{ p.s.}} \theta$$

et $\mathscr{K}_n(\theta, \alpha) \xrightarrow[n \uparrow \infty]{\mathbb{P}_\theta \text{ p.s.}} \dfrac{(\theta - \alpha)^2}{2(1 - \theta^2)}$ strictement convexe.

$\mathscr{D} v_n(\theta)/\sqrt{n}$ est une $\mathbb{P}_\theta$ martingale dont le processus croissant

$$\langle \mathscr{D} v_n(\theta) \rangle / n \to 1/(1 - \theta^2)$$

$$\mathscr{D} v_n(\theta)/\sqrt{n} \xrightarrow[n \uparrow \infty]{\mathscr{L}(\mathbb{P}_\theta)} N(0, 1/(1 - \theta^2))$$

$(\mathscr{D} v_n(\theta)/\sqrt{2 n \operatorname{Log} \operatorname{Log} n})_{n \geq 3}$ est p.s. relativement compacte avec pour ensemble de valeurs d'adhérence l'intervalle $[-\sqrt{1 - \theta^2}; +\sqrt{1 - \theta^2}]$.

I.4.2. *Processus de Ornstein Uhlenbeck*

Soit $(\Omega, \mathbb{F}, \mathbb{P}_{\theta, x}, (X_t)_{t \geq 0})$ une diffusion canonique solution forte de l'équation différentielle stochastique:

$$dX_t = -\theta X_t\, dt + dB_t$$

avec la condition initiale $X_0 = x$, où B est le mouvement brownien réel issu de 0 et θ un paramètre réel strictement positif. X est alors récurrent positif au sens de Harris [16]; par la formule de Girsanov, la log-vraisemblance sur $\mathcal{F}_t$ s'écrit :

$$v_t(\alpha) = -\alpha \int_0^t X_s \, dX_s - (\alpha^2/2) \int_0^t X_s^2 \, dX_s;$$

par la formule d'Ito,

$$v_t(\alpha) = -(\alpha/2)(X_t^2 - t) - (\alpha^2/2) \int_0^t X_s^2 \, ds.$$

Le comportement asymptotique de l'estimateur du maximum de vraisemblance $\hat{\theta}_t = (t - X_t^2)/2 \left(\int_0^t X_s^2 \, d_s \right)$ peut être étudié directement, mais si l'on veut suivre le cadre d'étude présenté ici, on obtient

$$\mathcal{K}_t(\theta, \alpha) = \frac{v_t(\alpha) - v_t(\theta)}{t} = \frac{1}{2t}(\theta - \alpha)^2 \int_0^t X_s^2 \, ds + \frac{1}{t}(\theta - \alpha) \int_0^t X_s \, dB_s.$$

Les fonctions $\alpha \mapsto \mathcal{K}_t(\theta, \alpha)$ sont convexes pour tout t, et par la loi forte on a :

$$\frac{\int_0^t X_s^2 \, ds}{t} \xrightarrow[t \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} \frac{1}{2\theta}; \qquad \frac{\int_0^t X_s \, dB_s}{\int_0^t X_s^2 \, ds} \xrightarrow[t \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} 0$$

et

$$\mathcal{K}_t(\theta, \alpha) \xrightarrow[t \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} K(\theta, \alpha) \qquad \text{où} \quad \alpha \mapsto K(\theta, \alpha) = (\theta - \alpha)^2 / 4\theta$$

est convexe ayant un point minimum unique.

$$\mathcal{D} v_t(\alpha)/\sqrt{t} = \left(-\int_0^t X_s \, dX_s - \alpha \int_0^t X_s^2 \, ds \right) / \sqrt{t}$$

$$-\mathcal{D}^2 v_t(\alpha)/t = \left(\int_0^t X_s^2 \, ds \right) \Big/ t \xrightarrow[t \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} -1/2\theta \quad \text{pour tout } \alpha$$

Ainsi $\mathcal{D} v_t(\theta)/\sqrt{t} = \left(\int_0^t X_s \, dB_s \right) \Big/ \sqrt{t}$ est une $(\mathbb{F}, \mathbb{P}_\theta)$ martingale de processus croissant $-\mathcal{D} v_t^2(\theta)/t$. D'une part on a, $\mathcal{D} v_t(\theta)/\sqrt{t} \xrightarrow[t \uparrow \infty]{\mathcal{L}(\mathbb{P}_\theta)} N(0, 1/2\theta)$, et d'autre part $(\mathcal{D} v_t(\theta)/\sqrt{2t \operatorname{Log} \operatorname{Log} t})_{t \geqq 3}$ est $\mathbb{P}_\theta$ p. s. relativement compacte ayant pour ensemble de valeurs d'adhérence l'intervalle $[-1/\sqrt{2\theta}, +1/\sqrt{2\theta}]$.

I.4.3. *Regression simple à temps continu*

Soit B_t un mouvement brownien et X_t la diffusion définie par $X_t = \theta t + B_t$; l'estimateur des moindres carrés qui minimise $v_t(\theta) = \frac{1}{2t}(X_t - \theta t)^2$ est $\hat{\theta}_t = \frac{X_t}{t} = \theta + \frac{B_t}{t}$ lequel est fortement consistant, et de l'autre côté

$$\mathcal{K}_t(\theta, \alpha) = \frac{v_t(\theta) - v_t(\alpha)}{t} = \frac{1}{2}(\theta - \alpha)^2 + (\theta - \alpha)\frac{B_t}{t} \xrightarrow[t \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} \frac{(\theta - \alpha)^2}{2}$$

ici $K(\theta, \alpha) = (\theta - \alpha)^2/2$ est convexe avec un minimum strict. $\mathcal{D} v_t(\alpha)/\sqrt{t} = -(X_t - \alpha t)$ vaut $-B_t/\sqrt{t} \sim N(0, 1)$ pour tout $t > 0$ au point $\alpha = \theta$ et $\mathcal{D}^2 v_t(\alpha)/t = \alpha$.

Ainsi $\sqrt{t}(\hat{\theta}_t - \theta)/\sqrt{2 \operatorname{Log} \operatorname{Log} t} = \mathcal{D} v_t(\theta)/\sqrt{2 t \operatorname{Log} \operatorname{Log} t}$ et l'hypothèse $\mathbb{H}3$ se trouve satisfaite, puisque $\mathcal{D} v_t(\theta)/\sqrt{2 t \operatorname{Log} \operatorname{Log} t}$ vérifie la loi du logarithme itéré.

I.4.4. *Statistiques pour les séries chronologiques*

Une abondante littérature traite du problème de l'identification de modèles, ARMA essentiellement ([1], [3], [4], [8], [9], [12]). Les nombreux résultats sont établis principalement à partir d'une loi du logarithme itéré pour le périodogramme. Le schéma d'étude établi ici constitue en quelque sorte l'extension naturelle de ces techniques.

SECONDE PARTIE

APPLICATIONS AUX PROBLÈMES DE L'IDENTIFICATION ET AUX TESTS

On a vu que la loi du logarithme itéré précise l'étude asymptotique de la quasi-vraisemblance ainsi que celle de l'estimateur associé. Elle va nous permettre de résoudre certains problèmes de statistique, essentiellement pour le moment celui de l'identification du modèle, lorsque le statisticien se trouve en présence d'une famille de modèles compétitifs et sans information *a priori*.

II 1. Tests des rapports de quasi-vraisemblances

THÉORÈME II.1. — *Soit* $r<q$, $\Lambda \subset \mathbb{R}^r$ *et* $\Theta \subset \mathbb{R}^q$ *des voisinages respectifs de* λ *et* θ *et soit* $\varphi : \Lambda \to \Theta$ *une fonction connue de classe* C^2 *et de rang* r *au voisinage de* λ *avec* $\varphi(\lambda)=\theta$. *On suppose vérifiées les hypothèses* $\mathbb{H}1$, $\mathbb{H}2$ *et* $\mathbb{H}3$ *pour* v_t *et l'hypothèse suivante pour* $w_t = v_t \circ \varphi$:

$$\mathbb{H}^*1.b \left\{ \begin{array}{l} \text{Pour tout } t, \text{ on peut trouver } \hat{\lambda}_t \in \Lambda \text{ tel que} \\ w_t(\hat{\lambda}_t) = \sup\limits_{\lambda \in \Lambda} w_t(\lambda) \qquad \text{avec} \quad \hat{\lambda}_t \xrightarrow[t \uparrow \infty]{\mathbb{P}_\theta \text{ p.s.}} \lambda. \end{array} \right.$$

Alors

1. $2(v_t(\hat{\theta}_t) - w_t(\hat{\lambda}_t)) \xrightarrow[t \uparrow \infty]{\mathscr{L}(\mathbb{P}_\theta)} \chi^2(q-r)$.

2. $\limsup\limits_{t \uparrow \infty} \dfrac{v_t(\hat{\theta}_t) - w_t(\hat{\lambda}_t)}{\operatorname{Log}\operatorname{Log} a(t)} = 1 \quad \mathbb{P}_\theta \text{ p.s.}$

Démonstration. — 1. Le résultat est classique [6, T2].

2. On démontre sans peine que si le couple $(v_t, a(t))$ satisfait aux hypothèses $\mathbb{H}1$, $\mathbb{H}2$, $\mathbb{H}3$ alors il en est de même pour $(w_t, a(t))$ sous $\mathbb{H}^*1.b$.

3. La démonstration suit approximativement celle de la partie 1 [7].

(i) On traite le cas où φ est affine de rang r, ce qui, à un changement d'origine et de base près, revient à considérer le cas,

$$\lambda = (\lambda^1, \ldots, \lambda^r)' \mapsto \varphi(\lambda) = \theta = (\lambda^1, \ldots, \lambda^r, 0, \ldots, 0)'.$$

Soient $\alpha = (\alpha^1, \alpha^2)$ et $\theta = (\lambda, 0)$ des éléments de Θ, avec les conditions $\mathbb{H}1$, $\mathbb{H}2$, $\mathbb{H}3$ vérifiées par les deux modèles. Considérons les notations suivantes dans la décomposition $\mathbb{R}^q = \mathscr{I}m(\varphi) \oplus \mathscr{I}m(\varphi)^\perp$:

$$\mathscr{D} v_t(\alpha) = M_t(\alpha) = (M_t^1, M_t^2)(\alpha)$$

et

$$-\mathscr{D}^2 v_t(\alpha)/a(t) = \mathbb{I}_t(\alpha) = \begin{pmatrix} \mathbb{I}_t^{11} & \mathbb{I}_t^{12} \\ \mathbb{I}_t^{21} & \mathbb{I}_t^{22} \end{pmatrix}(\alpha).$$

Soient $\hat{\theta}_t = (\hat{\theta}_t^1, \hat{\theta}_t^2)$ vérifiant $M_t(\hat{\theta}_t) = 0$ et $\tilde{\theta}_t = (\hat{\lambda}_t, 0)$ vérifiant $M_t^1(\tilde{\theta}_t) = 0$ les estimateurs respectifs de θ dans les deux modèles. Un premier développement donne

$$(\bigstar) \qquad \begin{array}{l} 2(v_t(\theta) - v_t(\hat{\theta}_t)) = -a(t)(\theta - \hat{\theta}_t)' \, \mathbb{I}_t(\theta_t^{**})(\theta - \hat{\theta}_t) \\ \text{où } \theta_t^{**} \in [\theta, \hat{\theta}_t] \\ 2(v_t(\theta) - v_t(\tilde{\theta}_t)) = -a(t)(\lambda - \hat{\lambda}_t)' \, \mathbb{I}_t^{11}(\theta_t^*)(\lambda - \hat{\lambda}_t) \\ \text{où } \theta_t^* \in [\theta, \tilde{\theta}_t]. \end{array}$$

D'un autre côté, l'égalité suivante

$$M_t^1(\theta) = a(t)(\hat{\lambda}_t - \lambda)' \, \mathbb{I}_t^{11}(\bar{\theta}_t) = a(t)\{(\hat{\theta}_t^1 - \lambda)' \, \mathbb{I}_t^{11}(\ddot{\theta}_t) + (\hat{\theta}_t^2)' \, \mathbb{I}_t^{21}(\ddot{\theta}_t)\}$$

avec $\bar{\theta}_t \in [\theta, \tilde{\theta}_t]$ et $\ddot{\theta}_t \in [\theta, \hat{\theta}_t]$ conduit à l'égalité

$$(\bigstar\bigstar) \quad (\hat{\theta}_r^2)' \mathbb{I}_t^{21} (\ddot{\theta}_t) + (\hat{\theta}_t^1 - \hat{\lambda}_t)' \mathbb{I}_t^{11} (\bar{\theta}_t) = (\hat{\theta}_t^1 - \lambda)' \{ \mathbb{I}_t^{11} (\bar{\theta}_t) - \mathbb{I}_t^{11} (\ddot{\theta}_t) \}$$

et en arrangeant les termes de ($\bigstar$), on obtient :

$$\frac{v_t(\hat{\theta}_t) - v_t(\tilde{\theta}_t)}{\operatorname{Log}\operatorname{Log} a(t)} = \frac{a(t)}{2 \operatorname{Log}\operatorname{Log} a(t)} \{ (\hat{\theta}_t^1 - \lambda)' \{ \mathbb{I}_t^{11} (\theta_t^{**}) - \mathbb{I}_t^{11} (\theta_t^*) \} (\hat{\theta}_t^1 - \lambda)$$
$$+ 2 \{ (\hat{\theta}_t^2)' \mathbb{I}_t^{21} (\theta_t^{**}) + (\hat{\theta}_t^1 - \hat{\lambda}_t)' \mathbb{I}_t^{11} (\theta_t^*) \} (\hat{\theta}_t^1 - \lambda)$$
$$+ (\hat{\theta}_t^2)' \mathbb{I}_t^{22} (\theta_t^{**}) (\hat{\theta}_t^2) - (\hat{\theta}_t^1 - \hat{\lambda}_t)' \mathbb{I}_t^{11} (\theta_t^*) (\hat{\theta}_t^1 - \hat{\lambda}_t) \}.$$

(*a*) Sous $\mathbb{H}1$, θ_t^{**}, θ_t^*, $\bar{\theta}_t$ et $\ddot{\theta}_t$ sont fortement consistants, ce qui nous assure que sous $\mathbb{H}2b$ $\mathbb{I}_t(\theta^{**})$, $\mathbb{I}_t(\theta^*)$, $\mathbb{I}_t(\bar{\theta}_t)$, $\mathbb{I}_t(\ddot{\theta}_t)$ convergent $\mathbb{P}_\theta$ p. s. vers $\mathbb{I}(\theta)$.

(*b*) La loi du logarithme itéré montre que le premier terme converge $\mathbb{P}_\theta$ p. s. vers 0 puisque $\sqrt{a(t)/2 \operatorname{Log}\operatorname{Log} a(t)}\,(\hat{\theta}_t^1 - \lambda)$ est $\mathbb{P}_\theta$ p. s. borné.

(*c*) De l'égalité ($\bigstar\bigstar$), on tire la même conclusion pour le second terme :

$$\frac{a(t)}{2 \operatorname{Log}\operatorname{Log} a(t)} \{ (\hat{\theta}_t^2)' \{ \mathbb{I}_t^{21} (\theta_t^{**}) - \mathbb{I}_t^{21} (\ddot{\theta}_t) \}$$
$$+ (\hat{\theta}_t^1 - \lambda_t)' \{ \mathbb{I}_t^{11} (\theta_t^*) - \mathbb{I}_t^{11} (\ddot{\theta}_t) \}$$
$$+ (\lambda - \hat{\lambda}_t)' \{ \mathbb{I}_t^{11} (\theta_t^*) - \mathbb{I}_t^{11} (\bar{\theta}_t) \} \} (\hat{\theta}_t^1 - \lambda) \xrightarrow[t \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} 0.$$

(*d*) Enfin pour le dernier terme, $\mathbb{P}_\theta$ p. s. à partir d'un certain rang la matrice $\mathbb{I}_t^{11} (\theta_t^*)$ est inversible, soit alors

$$Z_t = (\hat{\theta}_t^1 - \hat{\lambda}_t)' \mathbb{I}_t^{11} (\theta_t^*),$$
$$(\hat{\theta}_t^1 - \hat{\lambda}_t)' \mathbb{I}_t^{11} (\theta_t^*) (\hat{\theta}_t^1 - \hat{\lambda}_t) = (Z_t)' (\mathbb{I}_t^{11} (\theta_t^*))^{-1} . Z_t$$

et on tire de ($\bigstar\bigstar$)

$$Z_t = (\hat{\theta}_t^1 - \lambda)' \{ \mathbb{I}_t^{11} (\theta_t^*) - \mathbb{I}_t^{11} (\ddot{\theta}_t) \}$$
$$+ (\lambda - \hat{\lambda}_t)' \{ \mathbb{I}_t^{11} (\theta_t^*) - \mathbb{I}_t^{11} (\bar{\theta}_t) \} - \hat{\theta}_t^2 \mathbb{I}_t^{21} (\ddot{\theta}_t) = - \hat{\theta}_t^2 \mathbb{I}_t^{21} (\ddot{\theta}_t) + R_t$$

où $\| R_t \| = o(\sqrt{\operatorname{Log}\operatorname{Log} a(t)/a(t)})$ p. s. pour $t \uparrow \infty$ c'est-à-dire :

$$(\hat{\theta}_t^1 - \hat{\lambda}_t)' \mathbb{I}_t^{11} (\theta_t^*) (\hat{\theta}_t^1 - \hat{\lambda}_t) = (\hat{\theta}_t^2)' \mathbb{I}^{21} (\ddot{\theta}_t) (\mathbb{I}^{11} (\theta_t^*))^{-1} \mathbb{I}^{21} (\ddot{\theta}) \hat{\theta}_t^2 + S_t$$

avec $S_t = o(\sqrt{\operatorname{Log}\operatorname{Log} a(t)/a(t)})$ p. s. pour $t \uparrow \infty$. Autrement dit :

$$v_t(\hat{\theta}_t) - v_t(\tilde{\theta}_t)$$
$$= a(t) (\hat{\theta}_t^2)' \{ \mathbb{I}_t^{22} (\theta_t^{**}) - \mathbb{I}_t^{21} (\ddot{\theta}_t) (\mathbb{I}_t^{11} (\theta_t^*))^{-1} \mathbb{I}_t^{21} (\ddot{\theta}_t) - \mathbb{J}_{22}^{-1} (\theta) \} \hat{\theta}_t^2$$
$$+ a(t) (\hat{\theta}_t^2)' \mathbb{J}_{22}^{-1} (\theta) \hat{\theta}_t^2 + L_t$$

où $L_t = o\,(\operatorname{Log}\operatorname{Log} a(t))$ quand $t \uparrow \infty$ et

$$\mathbb{I}(\theta)^{-1} = \begin{pmatrix} \mathbb{J}_{11} & \mathbb{J}_{12} \\ \mathbb{J}_{21} & \mathbb{J}_{22} \end{pmatrix} (\theta)$$

avec

$$\mathbb{J}_{22}^{-1}(\theta)=(\mathbb{I}^{22}-\mathbb{I}^{21}\,.\,(\mathbb{I}^{11})^{-1}\,.\,\mathbb{I}^{12})(\theta)$$

et ceci achève la démonstration d'après le théorème $\mathbb{I}$. 1, avec

$$\limsup_{t\uparrow\infty}\frac{v_t(\hat{\theta}_t)-v_t(\tilde{\theta}_t)}{\operatorname{Log}\operatorname{Log}a(t)}=1.$$

(ii) Dans le cas général où φ est quelconque de rang r, si $\mathscr{D}\varphi(u)$ est sa différentielle au point u, à un changement de base et d'origine près, on peut supposer que pour $\varphi=(\varphi_1,\ldots,\varphi_q)$, $\lambda=0$, $\varphi(\lambda)=\theta=0$ et que $\mathrm{L}(u)=\mathscr{D}\varphi(0)\,(u)=(u,0)$ pour $u\in\Lambda$. Ainsi

$$\mathscr{D}_l\varphi_k(0)=\delta_{lk}=\begin{cases}0 & \text{si } l\neq k\\ 1 & \text{sinon}\end{cases}\qquad l=1,\ldots,r;\quad k=1,\ldots,q$$

$$\mathscr{D}_l w_t(u)=\mathscr{D}_l v_t(\varphi(u))=\sum_{k=1}^{q}\mathscr{D}_k v_t(\varphi(u))\,.\,\mathscr{D}_l\varphi_k(u)$$

$$\mathscr{D}_{ll'}^2 w_t(u)=\sum_{k=1}^{q}\sum_{k'=1}^{q}\mathscr{D}_l\varphi_k(u)\,.\,\mathscr{D}_{kk'}^2 v_t(\varphi(u))\,.\,\mathscr{D}_{l'}\varphi_{k'}(u)+\sum_{k=1}^{q}\mathscr{D}_k v_t(\varphi(u))\,.\,\mathscr{D}_{ll'}^2\varphi_k(u)$$

c.à.d.

$$(\bigstar\bigstar\bigstar)\qquad\begin{cases}\mathscr{D}_l w_t(\lambda)=\mathscr{D}_l v_t(\theta) & \text{pour } l=1,\ldots,r\\ \mathscr{D}_{ll'}^2 w_t(\lambda)=\mathscr{D}_{ll'}^2 v_t(\theta) & \text{pour } l,l'=1,\ldots,r\end{cases}$$

$\varphi(\hat{\lambda}_t)\in\mathscr{I}m\,(\mathscr{D}\varphi(0))$ à partir d'un certain rang t et soit $\tilde{\theta}_t=\varphi(\hat{\lambda}_t)=(\hat{\lambda}_t,0)$, alors de ($\bigstar\bigstar\bigstar$), on se retrouve dans le cadre précédent avec les résultats suivants:

$$\tilde{\theta}_t\xrightarrow[t\uparrow\infty]{\mathbb{P}_\theta\text{ p. s.}}\theta,\qquad\sqrt{a(t)}\,(\tilde{\theta}_t-\theta)\xrightarrow[t\uparrow\infty]{\mathscr{L}(\mathbb{P}_\theta)}\mathrm{N}(0,\mathbb{I}_{11}^{-1}(\theta))$$

et

$$\mathscr{D}w_t(\lambda)/\sqrt{a(t)}\xrightarrow[t\uparrow\infty]{\mathscr{L}(\mathbb{P}_\theta)}\mathrm{N}(0,\mathbb{I}_{11}(\theta)),$$

$$\mathscr{D}_{ll'}^2 w_t(\lambda)/a(t)\xrightarrow[t\uparrow\infty]{\mathbb{P}_\theta\text{ p. s.}}-\mathbb{I}_{11}(\theta)$$

ainsi qu'une loi du logarithme itéré pour $\tilde{\theta}_t=\varphi(\hat{\lambda}_t)$.

C.Q.F.D.

Remarque 5. — (*a*) On applique souvent ce théorème lorsque $\Lambda=\{\alpha\in\theta,\ \alpha_{r+1}=,\ldots,\alpha_q=0\}$ et $\theta\in\Lambda$ avec $\varphi(\alpha_1,\ldots,\alpha_r)=(\alpha_1,\ldots,\alpha_r,0,\ldots,0)$.

(*b*) Si $\theta \notin \varphi(\Lambda)$, il arrive souvent que :

$$(v_t(\theta) - w_t(\hat{\lambda}_t))/a(t) \underset{t \uparrow \infty}{\overset{\mathbb{P}_\theta \text{ p. s.}}{\longrightarrow}} K(\theta, \lambda_\theta),$$

par exemple avec des hypothèses de convexité convenables, $\lambda_\theta \neq \theta$ et $K(\theta, \lambda_\theta) = \inf_{\lambda \in \Lambda} K(\theta, \varphi(\lambda))$ alors

$$\frac{v_t(\hat{\theta}_t) - w_t(\hat{\lambda}_t)}{a(t)} \underset{t \uparrow \infty}{\overset{\mathbb{P}_\theta \text{ p. s.}}{\longrightarrow}} K(\theta, \lambda_\theta) > 0,$$

$$\frac{v_t(\hat{\theta}_t) - w_t(\hat{\lambda}_t)}{\operatorname{Log}\operatorname{Log} a(t)} \underset{t \uparrow \infty}{\overset{\mathbb{P}_\theta \text{ p. s.}}{\longrightarrow}} +\infty$$

et ceci entraîne le résultat suivant :

Tout test de la forme $\left(\dfrac{v_t(\hat{\theta}_t) - w_t(\hat{\lambda}_t)}{\operatorname{Log}\operatorname{Log} a(t)} \geqq c_t\right)$ avec $1 < c \leqq c_t \leqq d < \infty$ est asymptotiquement de niveau 0 et de puissance 1 pour tester « $\theta \in \varphi(\Lambda)$ » contre « $\theta \notin \varphi(\Lambda)$ ».

Applications

$\mathbb{E}$1. *Test de Neyman et Pearson*

Soit un n-échantillon $X_1, \ldots, X_n$ de loi soit F_0 soit $F_1 \neq F_0$ telles que $F_0 = f_0 . \nu$ et $F_1 = f_1 . \nu$. Si $\operatorname{Log}\left(\dfrac{f_0}{f_1}\right)$ est de carré intégrable pour F_0 et F_1 simultanément, alors on a

$$K(F_0, F_1) < \infty, \qquad K(F_1, F_0) < \infty$$

et

$$I(F_0, F_1) = \int (\operatorname{Log}(f_0/f_1))^2 \, dF_0 < \infty.$$

Si

$$T_n = \frac{\sum_{i=1}^{n} (\operatorname{Log}(f_0/f_1)(X_i) - K(F_0, F_1))}{\sqrt{2n \operatorname{Log}\operatorname{Log} n}}$$

alors tout test de région de rejet $(T_n \leqq c_n)$ avec $c_n \leqq c < -I(F_0, F_1)$ est asymptotiquement de niveau 0 et de puissance 1 puisque, sous F_0, l'ensemble des valeurs d'adhérence de T_n est $[-I(F_0, F_1), +I(F_0, F_1)]$ alors que, sous F_1, $T_n \xrightarrow[n \uparrow \infty]{\text{p. s.}} -\infty$.

$\mathbb{E}2$. *Tests de comparaison de moyennes de q échantillons*

Soient $(X_{il})\, i=1, \ldots, q,\ l=1, \ldots, n_i$ q-échantillons indépendants entre eux, de tailles respectives n_i, de moyenne θ_i et de même variance 1 pour simplifier. Si l'on prend les estimateurs des moindres carrés ($\mathbb{I}2$, $\mathbb{E}2$)

$$v_n(\theta) = -\frac{1}{2} \sum_{i=1}^{q} \sum_{l=1}^{n_i} (X_{il} - \theta_i)^2$$

$$\hat{\theta}_n = \left(\frac{X_{1.}}{n_1}, \ldots, \frac{X_{q.}}{n_q} \right)$$

si

$$\bar{X}_i = X_{i.}/n_i = \sum_{l=1}^{n_i} X_{il}/n_i \quad \text{et} \quad \bar{X} = \left(\sum_{i=1}^{q} n_i X_{i.} \right) \Big/ n$$

$n_i/n \underset{n \uparrow \infty}{\rightarrow} p_i$ avec $n = \sum_{i=1}^{q} n_i$, $i=1, \ldots, q$, alors toutes les hypothèses $\mathbb{H}1$, $\mathbb{H}2$ et $\mathbb{H}3$ sont vérifiées pour les deux modèles: $\theta \in \mathbb{R}^q$ où $\theta = \varphi(\lambda)$, $\varphi : \mathbb{R} \rightarrow \mathbb{R}^q$ et $\varphi(\lambda) = (\lambda, \ldots, \lambda)$. Ainsi

$$\left(\frac{v_n(\hat{\theta}_n) - v_n(\varphi(\hat{\lambda}_n))}{\operatorname{Log}\operatorname{Log} n} \right)_{n \geq 3} = \left(\frac{\sum_{i=1}^{q} n_i (\bar{X}_i - \bar{X})^2}{\operatorname{Log}\operatorname{Log} n} \right)_{n \geq 3}$$

a pour ensemble de valeurs d'adhérence l'intervalle $[0, 1]$. Or $\mathcal{K}_n(\theta, \alpha) \xrightarrow{\mathbb{P}_\theta \text{ p. s.}} K(\theta, \alpha) = \frac{1}{2} \sum_{i=1}^{q} p_i (\theta_i - \alpha_i)^2$ et $\alpha \mapsto K(\theta, \alpha)$ est strictement convexe et admet un minimum strict sur le sous-espace engendré par le vecteur $(1, \ldots, 1)$, $\alpha_\theta = (\bar{\theta}, \ldots, \bar{\theta})$ où $\bar{\theta} = \sum_{i=1}^{q} p_i \theta_i$ et la remarque 2 précédente s'applique ici.

$\mathbb{II}$ 2. Identification de modèles

Dans la pratique courante, le nombre de covariables explicatives dans un modèle de régression est souvent trop grand pour être utile. Si on pense que seule une partie de ces variables est réellement « pronostique », il s'agit de déterminer le vrai espace des paramètres ainsi que sa dimension r_0. On peut alors partir de l'hypothèse que cette dimension est majorée par une valeur q, qui serait le nombre total de facteurs éventuellement explicatifs envisagés.

Convenons en plus des notations suivantes:

Soit $\delta:\mathbb{R}^q \to \Delta=\{0,1\}^q$ telle que, pour $\alpha=(\alpha^1,\ldots,\alpha^q)\in\mathbb{R}^q$ $\delta(\alpha)=(\delta_1,\ldots,\delta_q)$ où $\delta_l=\mathbf{1}_{\{\alpha_l\neq 0\}}$; on lui associe la relation d'ordre partielle, $\delta\subset\delta'$ si et ssi $\delta_l>0\Rightarrow\delta'_l>0$. On désigne par $|\delta|$ la « longueur » du paramètre $|\delta|=\sum_{l=1}^{q}\delta_l$. Considérons par ailleurs, une famille de mesures de cette longueur, $(c_t)_{t\geqq 1}$, $c_t:\mathbb{N}\to\mathbb{R}_+$ que l'on choisit de la forme $c_t(r)=h(r).c^*(t)$ où c^* vérifie

$$(\bigstar)\qquad c^*(t)/a(t)\underset{t\uparrow\alpha}{\to}0 \quad\text{et}\quad \underline{\lim}\, c^*(t)/\operatorname{Log}\operatorname{Log} a(t)>1$$
$$h(r)-h(r-1)\geqq 1.$$

Soit $\Theta_\delta=\{\alpha\in\Theta,\delta(\alpha)=\delta\}$. La vraie valeur θ du paramètre est dans un de ces sous-espaces: $\theta\in\Theta_{\delta_0}$ avec $|\delta_0|=r_0\leqq q$. Il s'agit d'estimer δ_0. On cherche à maximiser la quasi-vraisemblance, tout en minimisant la longueur $|\delta|$, d'où l'idée d'Akaïké [1] d'introduire la vraisemblance compensée suivante:

$$W_t(\alpha)=v_t(\alpha)-c_t(|\delta(\alpha)|);$$

un estimateur du maximum de quasi-vraisemblance compensée sera une paire $(\hat\delta_t,\hat\theta_t)$ vérifiant:

$$W_t(\hat\theta_t)=\sup_{\delta\in\Delta}\sup_{\alpha\in\Theta_\delta}W_t(\alpha).$$

Notons aussi, par $\bar v_{t,\delta}=\sup_{\alpha\in\Theta_\delta}v_t(\alpha)$, $\bar W_{t,\delta}=\bar v_{t,\delta}-c_t(|\delta|)$ et $\hat\alpha_{t,\delta}$ un élément de Θ_δ tel que $\bar v_{t,\delta}=v_t(\hat\alpha_{t,\delta})$.

Nous ferons aussi l'hypothèse suivante satisfaite dans la plupart des exemples classiques et qui permet de vérifier la condition $\mathbb{H}^*1.b$ liée aux différents sous-modèles.

$\mathbb{H}4$ — Θ est un convexe de $\mathbb{R}^q$, voisinage de θ. La fonction $\alpha\mapsto v_t(\alpha)$ est, $\mathbb{P}_\theta$ p. s., concave pour tout t et la fonction $\alpha\mapsto K(\theta,\alpha)$ (forcément convexe), possède pour tout $\delta\in\Delta$ un minimum unique $\alpha_\delta\in\Theta_\delta$ ($\alpha_\delta=\theta$ si et ssi $\delta_0\subset\delta$).

Le théorème de Rockafellar et son corollaire (appendice 1) impliquent alors que $\hat\alpha_{t,\delta}\xrightarrow[t\uparrow\infty]{\mathbb{P}_\theta\text{ p. s.}}\alpha_\delta$ et $(v_t(\theta)-\bar v_{t,\delta})/a(t)\xrightarrow[t\uparrow\infty]{\mathbb{P}_\theta\text{ p. s.}}K(\theta,\alpha_\delta)$, la convergence étant uniforme sur tout compact.

THÉORÈME II.2. — *Sous* $\mathbb{H}1$, $\mathbb{H}2$, $\mathbb{H}3$ *et* $\mathbb{H}4$ $(\hat{\delta}_t, \hat{\theta}_{t,\hat{\delta}_t}) \underset{t \uparrow \infty}{\xrightarrow{\mathbb{P}_\theta \text{ p. s.}}} (\delta_0, \theta)$.

Démonstration. — Soit $N = \bigcup_{\delta \in \Delta} N_\delta$ l'ensemble $\mathbb{P}_\theta$ négligeable de Ω où

$$N_\delta = \{ (v_t(\theta) - \bar{v}_{t,\delta})/a(t) \not\longrightarrow K(\theta, \alpha_\delta) \}.$$

Alors, Δ étant fini, pour tout $\omega \notin N$, la suite $\hat{\delta}_t$ admet toujours au moins une valeur d'adhérence δ c.à.d. $\hat{\delta}_{t_k} \to \bar{\delta}$ et $\hat{\delta}_{t_k} = \delta$ à partir d'un certain rang t_{k_0}. De plus par hypothèse:

$$(\bigstar\bigstar) \qquad \bar{W}_{t,\hat{\delta}_t} - \bar{W}_{t,\hat{\delta}_0} \geqq 0$$

(*a*) Pour éviter la sous-paramétrisation, la condition $c^*(t)/a(t) \underset{t \uparrow \infty}{\longrightarrow} 0$ suffit, puisque pour $t > t_{k_0}$

$$\begin{aligned}\frac{1}{a(t)}(\bar{W}_{t,\hat{\delta}_t} - \bar{W}_{t,\delta_0}) &= \frac{1}{a(t)}(\bar{W}_{t,\bar{\delta}} - \bar{W}_{t,\delta_0}) \\ &= \frac{1}{a(t)}\{ (\bar{v}_{t,\bar{\delta}} - v_t(\theta)) + (v_t(\theta) - \bar{v}_{t,\delta_0}) + c_t(|\delta_0|) - c_t(|\bar{\delta}|) \} \\ &\underset{t \uparrow \infty}{\xrightarrow{\mathbb{P}_\theta \text{ p. s.}}} -K(\theta, \alpha_{\bar{\delta}}) \leqq 0.\end{aligned}$$

Ceci contredirait $(\bigstar\bigstar)$ sauf si $\delta_0 \subset \bar{\delta}$.

(*b*) pour éviter la sur-paramétrisation, la condition

$$\underline{\lim}\, c^*(t)/\text{Log}\,\text{Log}\,(a(t)) > 1$$

suffit aussi, car pour $t > t_{k_0}$ et $\delta_0 \subset \delta$

$$\begin{aligned}&\overline{\lim} \frac{1}{\text{Log}\,\text{Log}\,a(t)}(\bar{W}_{t,\hat{\delta}_t} - \bar{W}_{t,\delta_0}) \\ &\qquad = \overline{\lim} \frac{1}{\text{Log}\,\text{Log}\,a(t)}(\bar{W}_{t,\bar{\delta}} - \bar{W}_{t,\delta_0}) \\ &\qquad = \overline{\lim} \frac{1}{\text{Log}\,\text{Log}\,a(t)}\{ (v_t(\hat{\alpha}_{t,\delta}) - v_t(\hat{\theta}_t)) + c_t(|\delta_0|) - c_t(|\bar{\delta}|) \} \\ &\qquad \leqq 1 - (h(|\bar{\delta}|) - h(|\delta_0|)) \,.\, \underline{\lim} \frac{c^*(t)}{\text{Log}\,\text{Log}\,a(t)} \\ &\qquad \leqq (h(|\bar{\delta}|) - h(|\delta_0|)) \,.\, \left(1 - \underline{\lim} \frac{c^*(t)}{\text{Log}\,\text{Log}\,a(t)}\right) < 0\end{aligned}$$

et ceci contredirait aussi, sauf pour $h(|\bar{\delta}|) = h(|\delta_0|)$ c.à.d $\delta_0 = \bar{\delta}$.

On peut facilement généraliser le théorème précédent:

COROLLAIRE. — *Soit une famille* $(\varphi^l)_{l \in J}$, J *fini, de fonctions de classe* C^2, $\varphi^l : \Lambda^l \mapsto \mathbb{R}^q$ *de rang* r_l *et telle qu'il existe au moins un indice* l_0 *et* $\lambda \in \Lambda^{l_0}$

satisfaisant $\varphi^{l_0}(\lambda)=\theta$. *Si pour tout* $l \neq l_0$, $r_l \leqq r_{l_0}$ *entraîne que* $\theta \notin \varphi^l(\Lambda^l)$, *alors avec les mêmes hypothèses de convexité on a*:

$$(\hat{l}_t, \hat{\theta}_{t, \hat{l}_t}) \xrightarrow[t \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} (l_0, \theta).$$

Remarque 6. — 1. La fonction $h(t)$ de la condition (★) vérifie l'inégalité : $h(r)-h(r') \geqq r'-r$ si $r' \geqq r$. On peut donc choisir la « plus petite » de ces fonctions $h^*(r)=r$ et poser par exemple $c_t(r)=r \operatorname{Log} t$ ou $c_t(r)=(\mathrm{A}\, r+b) \operatorname{Log} t$, $\mathrm{A}>0$.

2. Si l'expérimentateur possède un classement donné *a priori* des différents modèles en compétition, il peut réduire notablement les calculs de maxima. Par exemple dans le cas d'une régression où les covariables sont déjà classées, on passe de 2^q à q calculs de maxima.

Exemples:

$\mathbb{E}1$. *Identification de modèles dans le cas d'une famille exponentielle*

Dans le cadre $\mathbb{I}2$, exemple 1, G. Schwartz [11] montre par des arguments bayésiens que la procédure du maximum de probabilité *a posteriori* donne la fonction suivante $c_n(r)=(r/2) \operatorname{Log} n$ qui permet d'éviter p. s. les problèmes de sous ou sur-paramétrisation de modèle dans le cas d'une famille exponentielle. Mais il introduit des hypothèses peu naturelles sur la loi *a priori* et l'étude asymptotique qui y est faite reste assez incomplète. Le théorème II.2 s'applique ici sans mal dès que θ est à l'intérieur du convexe naturel où la famille exponentielle est définie; ici $\mathrm{K}(\theta, \alpha)=\varphi(\alpha)-\varphi(\theta)-\langle \alpha-\theta, \mathbb{E}_0(\mathrm{T}) \rangle$ et φ est strictement convexe. L'existence de α_δ (dans $\mathbb{H}4$) est assurée dès que Θ_δ est fermé et $\alpha \rightarrow \mathrm{K}(\theta, \alpha)$ strictement convexe.

$\mathbb{E}2$. *Test du* χ^2 *paramétrique*

On reprend l'exemple $\mathbb{I}2.\mathbb{E}3$, dans le cas d'une famille finie $(\mathrm{P}^l)_{l \in \mathrm{J}}$ de fonctions vérifiant les conditions du corollaire précédent: pour $l \in \mathrm{J}$,

$$\mathrm{P}^l: \quad \Lambda^l \subset \mathbb{R}^l \mapsto \Delta=\left\{(p_1, \ldots, p_s): p_j>0, j=1, \ldots, s ; \sum_{j=1}^{s} p_j=1\right\}$$

est de rang l. Alors on obtient pour la quasi-vraisemblance compensée:

$$\sup_{l \in \mathrm{J}} \sup_{\lambda \in \Lambda^l}\left\{\left(\sum_{j=1}^{s}(\mathrm{N}_n^j-n \mathrm{P}_j^l(\lambda))^2 / 2 \mathrm{N}_n^j\right)-c_n(l)\right\}.$$

Et si, par exemple, les P_j^l sont convexes on a:

$$\frac{v_n(\theta)-\bar{v}_{n, l}}{n} \xrightarrow[n \uparrow \infty]{\mathbb{P}_\theta \text{ p. s.}} \mathrm{K}(\theta, \alpha_l)=\inf_{\lambda \in \Lambda^l} \mathrm{K}(\theta, \mathrm{P}^l(\lambda)).$$

Or $K(\theta, P^l(\lambda)) = \frac{1}{2} \sum_{j=1}^{s} \frac{(\theta_j - P_j^l(\lambda))^2}{\theta_j}$, ainsi α_l est la projection de θ sur $\mathscr{I}m(P^l)$ lorsque $\mathbb{R}^s$ est muni de la norme $\| x \|^2 = \sum_{j=1}^{s} \frac{(x_j)^2}{\theta_j}$.

$\mathbb{E}3$. *Regression, méthode des moindres carrés*

Dans le cas de la régression linéaire ou polynomiale, l'information de Kullback $K(\theta, \alpha) = \frac{1}{2}(\theta - \alpha)\Gamma(\theta - \alpha)'$ est une fonction quadratique. Si Γ est de rang plein. $\mathbb{H}4$ est automatiquement satisfaite, car la fonction $\alpha \mapsto K(\theta, \alpha)$ possède un minimum unique sur tout sous-espace Θ_δ, au point α_δ qui est la projection de θ sur Θ_δ selon le produit scalaire $u . v = u \Gamma v$.

$\mathbb{E}4$. *Modèle de régression de Cox*

$\mathbb{H}4$ est vérifiée pour le modèle de Cox, sous certaines conditions d'identifiabilité assez naturelles sur les covariables explicatives [13].

$\mathbb{E}5$. *Détermination de l'ordre d'une chaîne à un nombre d'états fini*

H. Tong montre assez formellement [15] que le critère initial d'Akaïké, à savoir le cas $c_n(r) = n . r$ qui conduit, on le sait maintenant [12], dans le cas des séries autorégressives à une sur-paramétrisation presque-sûre du modèle, pouvait servir à déterminer l'ordre (ou longueur de la mémoire) d'une chaîne $(X_n)_{n \geqq 1}$ d'espace d'état fini $E = (1, \ldots, s\}$. Ce problème s'inscrit aussi dans notre cadre général comme on le voit ci-dessous. L'ordre de la chaîne est r si r est le plus petit entier k tel que :

$$\begin{aligned} &\mathbb{P}(X_n = x_n / X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \\ &\qquad = \mathbb{P}(X_n = x_n / X_{n-1} = x_{n-1}, \ldots, X_{n-k} = x_{n-k}). \end{aligned}$$

On suppose que la chaîne est ergodique, avec une distribution stationnaire μ unique et que E est la seule classe récurrente, il est alors facile de se ramener à l'exemple $\mathbb{I}4 . \mathbb{E}1$. Le problème qui se pose ici est de déterminer l'ordre inconnu r_0 de la chaîne, et l'on peut supposer *a priori* que cet ordre est inférieur à un ordre p maximal raisonnable. Pour une chaîne d'ordre r, soit π_r^* sa transition :

$$\pi_r^*(x_n; x_{n-1}, \ldots, x_{n-r}) = P(X_n = x_n / X_{n-1} = x_{n-1}, \ldots, X_{n-r} = x_{n-r})$$

on peut construire alors une « vraie » chaîne de Markov $(Y_n)_{n \geqq 1}$ (d'ordre 1) équivalente à $(X_n)_{n \geqq 1}$, $Y_n = (X_n, \ldots, X_{n-r})$ d'espace d'état E^{r+1}

et de transition

$$\pi_r(x_{i_0}, \ldots, x_{i_r}; y_{i_0}, \ldots, y_{i_r}) = \begin{cases} \pi_r^*(y_{i_r}: x_{i_n}, \ldots, x_{i_1}) & \text{si} \quad x_{i_1}=y_{i_0}, \ldots, x_{i_r}=y_{i(r-1)} \\ 0 & \text{sinon} \end{cases}$$

L'espace des paramètres Θ_r est de dimension $q=s^r(s-1)$. Mais si la chaîne est seulement d'ordre $r' \leqq r$, cela signifie que:

$$\pi_r(x_{i_0}, \ldots, x_{i_r}; y_{i_0}, \ldots, y_{i_r}) = \pi_{r'}(x_{i(r-r')}, \ldots, x_{i_r}; y_{i(r-r')}, \ldots, y_{i_r})$$

ou, en d'autres termes, que $\Theta_{r'}$ est la projection canonique de Θ_r sur un sous-espace de dimension $s^{r'}(s-1)$. Ainsi pour l'observation $(X_1, \ldots, X_n)$ avec les conditions initiales $x_{-p}, \ldots, x_{-1}, x_0$, la log-vraisemblance est

$$v_{n,r}(\pi_r) = \sum_{x_{i_0} \ldots x_{i_n}} N_n^{i_0 \ldots i_r} \operatorname{Log} \pi_r^*(x_{i_r}; x_{i(r-1)}, \ldots, x_{i_0}); \qquad \pi_r \in \Theta_r$$

et en se ramenant aux notations de l'exemple 4.1.a, on a

$$\mathscr{K}_n(\pi, \pi') = \frac{1}{n} \sum_{(lk) \in D} N_n^{lk} \operatorname{Log}\left(\frac{\pi_{lk}}{\pi'_{lk}}\right) \xrightarrow[n \uparrow \infty]{\mathbb{P}_\pi \text{ p. s.}} K(\pi, \pi') = \sum_{(lk) \in D} \mu(l) \pi_{lk} \operatorname{Log}\left(\frac{\pi_{lk}}{\pi'_{lk}}\right)$$

ainsi $\pi' \mapsto K(\pi, \pi')$ est convexe strictement et nulle pour $\pi=\pi'$ seulement. Dire que la chaîne est d'ordre $r'<r$ revient donc à supposer qu'il existe une partition de D en $s^{r'}(s-1)$ parties composées chacunes de $s^{r-r'}$ éléments. Cherchons dans ce cas à déterminer le minimum sur un sous-espace $\Theta_{r'}$. Si $\pi' \in \Theta_r$ et $\pi^* = (\pi^*_{pm})$, $(p, m) \in D'$

$$K(\pi, \pi') = \sum_p \left\{ \sum_{m \in D'_p} \left(\sum_{(l,k) \in D_{pm}} \mu(l) \pi_{lk} (\operatorname{Log}(\pi_{lk}) - \operatorname{Log}(\pi^*_{pm}))\right)\right\}$$

où $D'_p = \{ m : (p, m) \in D' \}$ et $D_{pm} = \{ (l, k) : \pi'_{lk} = \pi^*_{pm} \}$ ce qui, compte tenu de la contrainte $\sum_{m \in D'_p} \pi^*_{pm} = 1$, donne lorsque l'on dérive, la solution suivante:

$\pi^*_{pm} = \sum_{(l,k) \in D_{pm}} \mu(l) \pi_{lk}$, donc H4 est bien vérifiée.

APPENDICE 1

Tout au long de l'exposé, on a utilisé à maintes reprises le théorème suivant dû à Rockafellar [10].

THÉORÈME. — *Si une suite $(l_n)_{n \geq 1}$ de fonctions convexes d'un ouvert convexe $\mathscr{U}$ de $\mathbb{R}^q$ dans $\mathbb{R}$ converge simplement vers l,* alors la convergence est uniforme sur tout compact de *$\mathscr{U}$ et la limite est convexe.*

Le corollaire qui suit est largement utilisé dans l'exposé comme d'ailleurs dans [2].

COROLLAIRE. — *De plus si l possède un unique point minimum $\bar{x}$ et si, pour tout n, x_n réalise le minimum de l_n alors $x_n \to \bar{x}$.*

Preuve. — Soit $r>0$ tel que la boule de centre $\bar{x}$ et de rayon r soit dans $\mathscr{U}$. Supposons que (x_n) ne tende pas vers $\bar{x}$, alors pour une sous-suite, encore notée (x_n), $\| x_n - \bar{x} \| \to a$, $0<a\leqq\infty$. Posons $y_n = \alpha_n x_n + (1-\alpha_n)\bar{x}$ avec $\alpha_n = \rho/\| x_n - \bar{x} \|$; on a $\| y_n - \bar{x} \| = \rho$, et prenons $\rho = r$ si $a=\infty$ et $0<\rho<\inf(r,a)$ si $a<\infty$. Pour n assez grand, on a $0<\alpha_n<1$. Soit $\bar{y}$ une valeur d'adhérence de la suite bornée (y_n): $\bar{y}\neq\bar{x}$ et, pour $0<\alpha_n<1$:

$$
\begin{aligned}
l_n(y_n) = l_n(\alpha_n x_n + (1-\alpha_n)\bar{x}) &\leqq \alpha_n l_n(x_n) + (1-\alpha_n) l_n(\bar{x}) \\
&\leqq \alpha_n l_n(\bar{x}) + (1-\alpha_n) l_n(\bar{x}) = l_n(\bar{x}).
\end{aligned}
$$

La convergence étant uniforme sur la boule de centre $\bar{x}$ et de rayon r

$$
l_n(y_n) \to l(\bar{y}) \qquad \text{et} \qquad l(\bar{y}) \leqq \lim l_n(\bar{x}) = l(\bar{x}),
$$

mais ceci contredit l'unicité du point minimum de l.

APPENDICE 2
LOI DU LOGARITHME ITÉRÉ FONCTIONNELLE DE STRASSEN

Soit

$$
\mathscr{K} = \left\{ \text{fonctions } l \text{ absolument continues de } \mathscr{C}([0,1]; \mathbb{R}^q) \right.
\left. \text{telles que } l(0)=0 \text{ et } \int_0^1 \| \mathscr{D} l \|^2 \, ds \leqq 1 \right\}
$$

où $\mathscr{D}l$ désigne la dérivée de l, définie presque-partout par rapport à la mesure de Lebesgue. $\mathscr{K}$ est un compact pour la norme uniforme. Soit une suite $(Y_i)_{i\geqq 1}$ de v. a. i. i. d. à valeurs dans $\mathbb{R}^q$, centrées et de covariance la matrice identité I_q. Notons la suite des sommes partielles

$$
S_n = (S_n^1, \ldots, S_n^q), \qquad \text{où} \quad S_n^j = \sum_{i=1}^{n} Y_i^j, \quad j=1,\ldots,q, \quad n\geqq 1
$$

et $s_n(t)$ l'interpolation linéaire de S_n sur $[0,1]$: pour $0\leqq t\leqq 1$ $s_n(t) = ([nt]+1-nt)\, S_{[nt]} + (nt-[nt])\, S_{[nt]+1}$ où $[nt]$ désigne la partie entière de nt. Il est alors facile de généraliser le résultat de Strassen [14] au cas multidimensionnel:

La suite de fonctions aléatoires $(s_n(t)/\sqrt{2n \operatorname{Log}\operatorname{Log} n})_{n\geqq 3}$ pour $0\leqq t\leqq 1$ est relativement compacte et l'ensemble de ses points limites coïncide p. s. avec $\mathscr{K}$.

COROLLAIRE. — *Si* $(Y_i)_{i \geqq 1}$ *est une suite de v.a. i.i.d. à valeurs dans* $\mathbb{R}^q$, *centrées et de matrice de covariance* Γ *de rang* r, *alors la suite* $(S_n/(\sqrt{2n \operatorname{Log} \operatorname{Log} n}))_{n \geqq 3}$ *est p.s. relativement compacte, ayant pour ensemble de valeurs d'adhérence « l'ellipsoïde »* $\mathscr{C}_q(0, \Gamma^-) \cap \mathscr{I}m(\Gamma)$, *indépendante du choix de l'inverse* Γ^-

$$\mathscr{C}_q(0, \Gamma^-) = \{ u \in \mathbb{R}^q : u' \Gamma^- u \leqq 1 \}$$

et

$$\mathscr{I}m(\Gamma) = \{ v \in \mathbb{R}^q : \exists\, u \in \mathbb{R}^q \text{ et } \Gamma u = v \}.$$

Démonstration. — 1. Lorsque la covariance est la matrice identité I_q, remarquons que $s_n(1) = S_n$. Or pour toute valeur d'adhérence $l \in \mathscr{K}$ de $s_n/\sqrt{2n \operatorname{Log} \operatorname{Log} n}$ on a $\| l(1) \| \leqq 1$ et pour tout élément de la boule unité: $a \in B_q(0, 1)$ la fonction $l : u \mapsto a.u$ est dans $\mathscr{K}$ avec $l(1) = a$. Ce qui signifie que p.s. l'ensemble des valeurs d'adhérence de $S_n/\sqrt{2n \operatorname{Log} \operatorname{Log} n}$ est $B_q(0, 1)$.

2. Pour $\Gamma = \operatorname{cov}(Y)$, on sait que Y prend ses valeurs p.s. sur le sous-espace $\mathscr{I}m(\Gamma)$ de dimension r, et qu'il existe une matrice orthogonale $\mathbb{O}$ telle que: $\mathbb{O} \Gamma \mathbb{O}' = \mathbb{D}$ où $\mathbb{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$ est la matrice diagonale de valeurs propres $\lambda_1 \geqq, \ldots, \geqq \lambda_r > 0$ de Γ. Désignons par

$$\mathbb{D}^{1/2} = \operatorname{diag}(\lambda_1^{1/2} \ldots \lambda_r^{1/2}, 0 \ldots 0)$$

et

$$\mathbb{D}^{-1/2} = \operatorname{diag}(\lambda_1^{-1/2} \ldots \lambda_r^{-1/2}, 0 \ldots 0)$$

$\Gamma^{-1/2} = \mathbb{O}' \, \mathbb{D}^{-1/2} \mathbb{O}$ et $\Gamma^{1/2} = \mathbb{O}' \, \mathbb{D}^{1/2} \mathbb{O}$ sont de rang r. Et soit une inverse généralisée $\Gamma^- = \Gamma^{-1/2} . \Gamma^{-1/2} = \mathbb{O}' \, \mathbb{D}^- \, \mathbb{O}$ où

$$\mathbb{D}^- = \operatorname{diag}(\lambda_1^{-1} \ldots \lambda_r^{-1}, 0 \ldots 0). \text{ La variable } Z_i = \Gamma^{-1/2} Y_i$$

est centrée et de covariance $\mathbb{P} = \mathbb{O}' \, \mathbb{I}_r^* \, \mathbb{O}$ où $\mathbb{I}_r^* = \begin{pmatrix} I_r, 0 \\ 0, 0 \end{pmatrix}$. On remarquera aussi que $\mathbb{P}$ est le projecteur orthogonal sur $\mathscr{I}m(\Gamma)$.

Si la loi du logarithme itéré tient pour la suite $(\mathbb{O} . Z_i)$, l'ensemble de ses points limite est la « boule » $B_r^*(0, 1)$ de $\mathbb{R}^r \times \{ 0 \}^{q-r}$. Soit $\mathbb{Q} = \mathbb{O}' \, \mathbb{D}^{1/2}$ on a $\mathbb{Q} \, \mathbb{O} \, Z_i = \mathbb{O}' \, \mathbb{I}_r^* \, \mathbb{O} \, Y_i = \mathbb{P} \, Y_i = Y_i$ p.s. Il suffit donc de démontrer que $\mathbb{Q}(B_r^*(0, 1)) = \mathscr{E}_q(0, \Gamma^-) \cap \mathscr{I}m(\Gamma)$. On a $\mathscr{I}m(\mathbb{Q}) = \mathscr{I}m(\Gamma)$ puisque $\mathbb{Q} \, \mathbb{Q}' = \Gamma$.

(i) Pour $w \in B_r^*(0, 1)$, $w = (w_1, 0)$ on a $u = \mathbb{Q} \, w \in \mathscr{I}m(\Gamma)$ et de plus

$$u' \Gamma^- u = w' \, \mathbb{Q}' \, \Gamma^- \, \mathbb{Q} \, w = w' \, \mathbb{D}^{1/2} \, \mathbb{O} \, \mathbb{O}' \, \mathbb{D}^- \, \mathbb{O} \, \mathbb{O}' \, \mathbb{D}^{1/2} \, w = w' \, \mathbb{I}_r^* \, w = w' \, w \leqq 1$$

(ii) Pour $u \in \mathscr{E}_q(0, \Gamma^-) \cap \mathscr{I}m(\Gamma)$ alors

$$1 \geqq u' \Gamma^- u = u' \, \mathbb{O}' \, \mathbb{D}^{-1/2} . \mathbb{D}^{-1/2} \, \mathbb{O} \, u.$$

Si on pose $w = \mathbb{D}^{-1/2} \, \mathbb{O} \, u$ on a

$$w' w \leqq 1 \qquad \text{et} \qquad \mathbb{Q} \, w = \mathbb{O}' \, \mathbb{I}_r^* \, \mathbb{O} \, u = \mathbb{P} \, u = u$$

puisque $u \in \mathcal{I}m(\Gamma)$.

RÉFÉRENCES

[1] H. AKAÏKÉ, A new look at the statistical model identification, *I.E.E.E. Trans. Automatic Control*, vol. **AC-19**, n° 6, 1974.
[2] P. K. ANDERSEN et R. D. GILL, Cox regression model for counting process: A large sample study, *Ann. Stat.*, vol. **10**, n° 4, 1982.
[3] R. AZENCOT et D. DACUNHA-CASTELLE, Series of irregular observations, forecasting and model building, Springer Verlag, 1986.
[4] M. BOUAZIZ, Vitesse de convergence de mesures spectrales empiriques, *C.R. Acad. Sci. Paris*, t. **301**, série I, n° 10, 1985.
[5] D. R. COX, Partial likelihood, *Biometrika*, 62, 1975.
[6] D. DACUNHA-CASTELLE et M. DUFLO, Probabilités et Statistiques, Tome 1 : Problèmes à temps fixe (cours, exercices); Tome 2 : Problèmes à temps mobile (cours, exercices), Masson, Paris, 1982, 1983, 1984.
[7] M. DUFLO, Cours de D.E.A., Orsay, 1986.
[8] E. J. HANNAN et B. G. QUINN, The determination of the order of an autoregression, *J.R. Stat. Soc. B*, vol. **41**, n° 2, 1979.
[9] E. J. HANNAN, The estimation of the order of an ARMA process, *Ann. Stat.*, vol. **8**, n° 5, 1980.
[10] R. T. ROCKAFELLAR, Convex analysis, Princeton University Press, 1970.
[11] G. SCHWARTZ, Estimating the dimension of a model, *Ann. Stat.*, vol. **6**, n° 2, 1978.
[12] R. SHIBATA, Selection of the order of an autoregressive model by Akaïké's information criterion, *Biometrika*, vol. **63**, n° 1, 1976.
[13] R. SENOUSSI, Problème d'identification dans le modèle de Cox, *Ann. Inst. Henri Poincaré*, Probabilités et Statistiques, vol. **26**, n° 1, 1989.
[14] V. STRASSEN, An invariance principle for the law of the iterated logarithm, *Z. W. Theor.*, vol. **3**, 1964.
[15] H. TONG, Determination of the order of a Markov chain by Akaïké's information criterion, *J. App. Prob.*, vol. **12**, n° 3, 1975.
[16] A. TOUATI, Théorèmes limites pour les processus de Markov récurrents, *Prob. Th and Rel. Fields*. 1990 (à paraître).
[17] A. TOUATI, Loi fonctionnelle du logarithme itéré pour des processus de Markov récurrents, *C.R. Acad. Sci. Paris*, t. **306**, série I, 1988.

(*Manuscrit reçu le* 14 *février* 1989,
version révisée le 19 *juin* 1989.)