NELLY MAIGRET

Majorations de Chernoff et statistique séquentielle pour des chaînes de Markov récurrentes au sens de Doeblin

Astérisque, tome 68 (1979), p. 125-142
<http://www.numdam.org/item?id=AST_1979__68__125_0>

MAJORATIONS DE CHERNOFF ET STATISTIQUE SÉQUENTIELLE POUR DES CHAÎNES DE MARKOV RÉCURRENTES AU SENS DE DOEBLIN.

Nelly MAIGRET.

Résumé :

On établit un théorème de grandes déviations pour une chaîne de Markov récurrente Doeblin, semblable aux théorèmes de DONSKER et VARADHAN [5] . On construit (toujours pour une chaîne récurrente Doeblin) un test analogue à celui de Chernoff dans le cas indépendant [4] , [1] , [2], le théorème de grandes déviations obtenu permet de montrer que ce test est asymptotiquement le plus économique parmi les tests de force au moins égale.

PREMIÈRE PARTIE : MAJORATIONS DE CHERNOFF.

$X = (\Omega, \mathcal{A}, (P_x)_{x \in \mathcal{H}}, (X_n)_{n \in N})$ une chaîne de Markov sur un espace d'état mesurable $(\mathcal{H}, \underline{\underline{\mathcal{H}}})$ où la tribu $\underline{\underline{\mathcal{H}}}$ est supposée à base dénombrable. On note π le noyau de transition, et π_n la $n^{\text{ième}}$ itérée de π_o . $\mathcal{B}_n$ désigne la tribu des événements antérieurs à n : $\mathcal{B}_n = \sigma(X_o, \dots X_n)$. On note $\mathcal{P}$ l'ensemble des probabilités sur $(\mathcal{H}, \underline{\underline{\mathcal{H}}})$. Lorsque $(\mathcal{H}, \underline{\underline{\mathcal{H}}})$ est métrique, $\mathcal{P}$ est muni de la topologie de la convergence étroite. On note $\mathbb{P}_\lambda$ la distribution de la chaîne et E_λ l'espérance lorsque λ de $\mathcal{P}$ est la loi initiale. Pour simplifier, P_x et E_x désigneront respectivement P_{δ_x} et E_{δ_x} , δ_x étant la mesure de Dirac au point x . On notera de la même façon, par $\| \ \|$, la norme uniforme sur $\mathcal{P}$ et sur l'ensemble des variables aléatoires (v.a.) sur $\mathcal{H} \times \mathcal{H}$.

La chaîne de Markov X est supposée récurrente Doeblin, apériodique, de probabilité invariante μ ,c'est-à-dire : il existe $c > 0$, il existe ρ , $0 < \rho < 1$, tels que pour tout n, $\sup_{x \in \mathcal{H}} \|\pi_n (x,.) - \mu (.)\| \leq c \rho^n$ (cf. [7]).

Sauf précision contraire, les espaces topologiques seront munis de leur tribu borélienne, et les espaces produit de la tribu produit associée.

Enfin, pour tout n de N, on notera $\mu \otimes \pi_n$ la mesure sur $\mathcal{H}^{n+1}$ définie par

$$(\mu \otimes \pi_n) (A) = \int 1_A (x_o,\dots x_n)\, d\mu (x_o)\, \pi (x_o, dx_1) \dots \pi (x_{n-1}, dx_n)$$

Théorème 1 :

Soit g une v.a. bornée sur $\mathcal{H} \times \mathcal{H}$,m, $\overline{g}$ et $\underline{g}$ définis respectivement par $\int g\, d\mu \otimes \pi$, $\mu \otimes \pi$ - ess sup g, $\mu \otimes \pi$ ess inf g. On note h la transformée de Cramer de $g(\mu \otimes \pi)$ définie sur $[\underline{g},\overline{g}]$. Pour tout réel a strictement supérieur à m, on a :

$$\overline{\lim_n}\ \frac{1}{n} \log \sup_{x \in \mathcal{H}} P_x \left(\sum_{p=0}^{n-1} (g (X_p, X_{p+1}) - a) \geq 0\right) < 0$$

En particulier si a est élément de $]m,\overline{g}[$, le terme ci-dessus est majoré par le nombre négatif $\inf_k \{\frac{1}{k} \log (b\rho^k + e^{-h(a)})\}$ où b est une constante dépendant de c, de g et de a.

Démonstration :

. Supposons que a est élément de $]m,\overline{g}[$.

Alors $u_o = h'(a)$ est strictement positif et vérifie

$$e^{-h(a)} = E_\mu \left(e^{(g(X_o,X_1)-a)u_o}\right)$$

Pour tout x de $\mathcal{H}$, pour tout k entier naturel, on a :

$$\left|E_x(e^{u_o(g(X_k,X_{k+1})-a)}) - e^{-h(a)}\right| = \left|\int e^{u_o(g(y,z)-a)} \Pi(y,dz\ (\Pi_k(x,dy) - \mu(dy))\right|$$
$$\leq \|e^{u_o(g-a)}\|\ \|\Pi_k(x,.) - \mu(.)\|$$
$$\leq \|e^{u_o(g-a)}\|\ \|c\ \rho^k = b\rho^k$$

c'est-à-dire, pour tout k entier naturel

$$\sup_x E_x\ (e^{u_o(g(X_k,X_{k+1})-a)}) \leq e^{-h(a)} + b\rho^k$$

Soient k et n des entiers naturels fixes quelconques. Pour tout a, on a :

$$E_x\ (e^{u_o \sum_{p=1}^{n} (g(X_{kp},X_{kp+1})-a)}$$
$$= E_x\ e^{u_o \sum_{p=1}^{n-1} (g(X_{kp},X_{kp+1})-a)} \quad E_x\ (e^{u_o(g(X_{kn},X_{kn+1})-a)} / \mathcal{B}_{k(n-1)+1})$$
$$= E_x\ (e^{u_o \sum_{p=1}^{n-1} (g(X_{kp},X_{kp+1})-a)} \quad E_{X_{k(n-1}}\ (e^{u_o(g(X_k,X_{k+1})-a}))$$

En appliquant la relation démontrée ci-dessus, cette expression est majorée par

$$E_x\ (e^{u_o \sum_{p=1}^{n-1} (g(X_{kp},X_{kp+1})-a)})\ (e^{-h(a)} + b\ \rho^k)$$ et en itérant on trouve

$$\sup_{x\in\mathcal{H}} E_x\ (^{u_o \sum_{p=1}^{n} (g(X_{kp},X_{kp+1})-a)}) \leq (e^{-h(a)} + b\ \rho^k)^n \quad .$$

Soit r entier naturel, $1 \leq r \leq k-1$. Pour tout x,

$$E_x\ (e^{u_o \sum_{p=1}^{n} (g(X_{pk+r},X_{pk+r+1})-a)} = E_x\ (E_{X_r}(e^{u_o \sum_{p=1}^{n} (g(X_{kp},X_{kp+1})-a)}))$$

La dernière relation démontrée entraîne :

$$\sup_x E_x\ (e^{u_o \{\sum_{p=1}^{n} (g(X_{pk+r},X_{pk+r+1})-a)\}}) \leq e^{-h(a)} + b\ \rho^k)^n$$

Puisque a est strictement supérieur à m, u_o est strictement positif et on peut écrire :

$$P_x\Big(\sum_{p=k}^{nk-1}\{g(X_p,X_{p+1})-a\}\geq 0\Big)\leq \sum_{r=0}^{k-1} P_x\Big(\sum_{p=1}^{n-1}(X_{pk+r},X_{pk+r+1})-a)\geq 0\Big)$$

$$\leq \sum_{r=0}^{k-1} E_x\Big(e^{u_o\sum_{p=1}^{n-1}(g(X_{pk+r},X_{pk+r+1})-a)}\Big)$$

$$\sup_x P_x\Big(\sum_{p=k}^{nk=1}(g(X_p,X_{p+1})-a)\geq 0\Big)\leq k\,(e^{-h(a)}+b\,\rho^k)^{n-1}$$

D'autre part, pour tout ε strictement positif et pour tout a

$$P_x\Big(\sum_{p=0}^{n}(g(X_p,X_{p+1})-a)\geq n\varepsilon\Big)\leq P_x\Big(\sum_{p=0}^{k-1}(g(X_p,X_{p+1})-a)\geq \frac{n\varepsilon}{2}\Big)$$

$$+\,P_x\Big(\sum_{p=k}^{[\frac{n}{k}]k-1}(g(X_p,X_{p+1})-a)\geq 0\Big)+P_x\Big(\sum_{[\frac{n}{k}]k}^{n}(g(X_p,X_{p+1})-a)\geq \frac{n\varepsilon}{2}\Big)$$

où pour tout réel s, $[s]$ désigne sa partie entière.

g étant une v.a. bonrée sur $(\mathcal{H}\times\mathcal{H})$, il existe $n_o(k,\varepsilon)$, tel que $\forall n\geq n_o$,

$$\sup_x P_x\Big(\sum_{p=0}^{k-1}(g(X_p,X_{p+1})-a)\geq \frac{n\varepsilon}{2}\Big)=0$$

et $$\sup_x P_x\Big(\sum_{p=[\frac{n}{k}]k}^{n}(g(X_p,X_{p+1})-a)\geq \frac{n\varepsilon}{2}\Big)=0$$

(cette dernière égalité étant vraie puisque $\sum_{p=[\frac{n}{k}]k}^{n}$ comprend au plus k termes).

De plus, $$\sup_x P_x\Big(\sum_{p=k}^{[\frac{n}{k}]k-1}(g(X_p,X_{p+1})-a)\geq 0\Big)\leq k\,(e^{-h(a)}+b\,\rho^k)^{[\frac{n}{k}]-1}$$

Par conséquent, il existe $n_o(k,\varepsilon)$, tel que pour tout $n\geq n_o$,

$$\sup_{x\in\mathcal{H}} P_x\Big(\sum_{p=0}^{n-1}(g(X_p,X_{p+1})-a)\geq n\varepsilon\Big)\leq k\,(e^{-h(a)}+b\,\rho^k)^{[\frac{n}{k}]-1}$$

$$\frac{1}{n}\,\mathrm{Log}\sup_{x\in\mathcal{H}} P_x\Big(\sum_{p=0}^{n-1}(g(X_p,X_{p+1})-a)\geq n\varepsilon\Big)\leq \frac{\mathrm{Log}\,k}{n}+\frac{1}{n}\Big(\Big[\frac{n}{k}\Big]-1\Big)\,\mathrm{Log}\,(e^{-h(a)}+b\rho^k)$$

ce qui entraîne

$$\overline{\lim_n}\ \frac{1}{n}\ \text{Log}\ \sup_x\ P_x\ (\sum_{p=0}^{n-1}\ (g(X_p,X_{p+1})-a)\geq n\varepsilon)\leq\frac{1}{k}\ \text{Log}\ (e^{-h(a)}+b\rho^k)$$

Cette dernière relation étant vraie pour tout k, on en déduit

$$\overline{\lim_n}\ \frac{1}{n}\ \text{Log}\ \sup_{x\in\mathcal{H}}\ P_x\ (\sum_{p=0}^{n-1}\ (g(X_p,X_{p+1})-a)\geq n\varepsilon)\leq\inf_{k\in N}\ (\frac{1}{k}\ \text{Log}\ (e^{-h(a)}+b\rho^k))\ .$$

.. Soit maintenant $\varepsilon > 0$ tel que : $a-\varepsilon > m$. (On a bien sûr $a-\varepsilon < \overline{g}$).

Appliquons alors la dernière inégalité démontrée à $a-\varepsilon$, en notant la constante b correspondant à un réel a, b_a. On obtient :

$$\overline{\lim_n}\ \frac{1}{n}\ \text{Log}\ \sup_{x\in\mathcal{H}}\ P_x\ (\sum_{p=0}^{n-1}\ (g(X_p,X_{p+1})-a)\geq 0)\leq\inf_{k\in N}\ (\frac{1}{k}\ \text{Log}\ (e^{-h(a-\varepsilon)}+b_{a-\varepsilon}\ \rho^k))$$

Faisons tendre ε vers 0.

h est une fonction convexe donc continue : $\lim\limits_{\varepsilon\to 0} h(a-\varepsilon) = h(a)$.

On a : $b_{a-\varepsilon} = c\ \|e^{h'(a-\varepsilon).(g-a+\varepsilon)}\|$

h' est continue donc : $\lim\limits_{\varepsilon\to 0}\ b_{a-\varepsilon} = b_a$.

On en déduit que pour tout a, $m < a < \overline{g}$:

$$\overline{\lim_n}\ \frac{1}{n}\ \text{Log}\ \sup_{x\in\mathcal{H}}\ P_x\ (\sum_{p=0}^{n-1}\ (g(X_p,X_{p+1})-a)\geq 0)\leq\inf_{k\in N}\ (\frac{1}{k}\ \text{Log}\ (e^{-h(a)}+b_a\rho^k))$$

Cet inf est strictement négatif puisque $h(a)$ étant strictement positif ($a \neq m$) et ρ strictement inférieur à 1, il existe k_o de N tel que $(e^{-h(a)} + b_a\ \rho^{k_o})$ soit strictement inférieur à 1.

Ceci termine la démonstration de la 4ème partie du théorème. Le résultat de la 1ère se déduit immédiatement puisque la fonction

$$a \to \overline{\lim_n}\ \frac{1}{n}\ \text{Log}\ \sup_{x\in\mathcal{H}}\ P_x\ (\sum_{p=0}^{n-1}\ (g(X_p,X_{p+1})-a)\geq 0$$ est croissante. ∎

<u>Remarque</u> : Soit $a \in]m,\overline{g}[$. Si g et c vérifient : $\|g\| \geq a$ et $c \geq 1$, la constante $b_a = c\ \|e^{u_o(g-a)}\|$ est supérieure à 1 et $\inf\limits_{k\in N}\ \frac{1}{k}\ \text{Log}\ (e^{-h(a)} + b_a\rho^k)$ est minorée par $\text{Log}\ \rho$. Donc en général cet inf ne sera pas $-\infty$ même si $\|g\| \geq a > \overline{g}$. L'exemple suivant met en évidence cette minoration par $\text{Log}\ \rho$.

Soit la chaîne markovienne à 2 états 0 et 1 de probabilité de transition : $\Pi(0,0) = \rho$, $\Pi(1,1) = 1$. Sa mesure invariante μ est δ_1. Supposons $\rho > \frac{1}{2}$, alors les constantes c et ρ sont respectivement 1 et $\sup(\rho, 1-\rho) = \rho$.

La v.a. g définie par : $g(1) = 0$, $g(0) = 1$ est telle que $\overline{g} = 0$ et on a :

$$P_o\left(\sum_{p=0}^{n-1} g(X_p) \geq n\right) = \rho^n .$$

$$\frac{1}{n} \operatorname{Log} P_o\left(\sum_{p=0}^{n-1} g(X_p) \geq n\right) = \operatorname{Log} \rho = \frac{1}{n} \operatorname{Log} P_o\left(\sum_{p=0}^{n-1} (g(X_p)-1) \geq 0\right).$$

DEUXIÈME PARTIE : STATISTIQUE.

Maintenant, le noyau π dépend d'un paramètre. L'espace des paramètres $(\Theta \times \underline{\underline{\Theta}})$ est un espace topologique muni de sa tribu borélienne. La transition π est une transition de $\Theta \times \mathcal{H}$ dans $\mathcal{H}$. Le processus des observations $X^\theta = (\Omega, \mathcal{A}, (P_{\theta,x})_{x\in\mathcal{H}}, (X_n)_{n\in N}$ est une chaîne markovienne récurrente Doeblin, apériodique, de probabilité invariante μ_θ. On se pose des problèmes d'estimation et de test d'un point de vue séquentiel.

I. ESTIMATION

1. Définitions. Exemples.

Ce paragraphe reste valable si la chaîne X^θ est seulement récurrente Harris positive (cadre 1).

1.1. Fonction de contraste : la définition est celle de Gänsler [6].

Une v.af. sur $\Theta \times \mathcal{H} \times \mathcal{H}$ est une fonction de contraste si $\int d\mu_\theta(x)\, \Pi(\theta,x,dy)\, f(\theta,x,y)$ est fini et $\int d\mu_\theta(x)\, \Pi(\theta,x,dy)\, (f(\varphi,x,y)-f(\theta,x,y)) \geq 0$ l'égalité n'ayant lieu que lorsque $\varphi = \theta$.

On notera $h(\theta,\varphi,x,y) = f(\varphi,x,y) - f(\theta,x,y)$

et $K(\theta,\varphi) = \int d\mu_\theta(x)\, \Pi(\theta,x,dy)\, h(\theta,\varphi,x,y)$

Exemple : On suppose qu'il existe une v.a. p sur $\Theta \times \mathcal{H} \times \mathcal{H}$, strictement positive et une transition λ de $\mathcal{H}$ dans $\mathcal{H}$ telles que pour $x \in \mathcal{H}$ et $\Gamma \in \underline{\underline{\mathcal{H}}}$,

$$\Pi(\theta,x;\Gamma) = \int p(\theta,x,y)\, \lambda(x,dy)\, 1_\Gamma(y)$$

On suppose aussi que si φ est différent de θ, l'événement $\{x;\Pi(\varphi,x,.) \neq \Pi(\theta,x,.)\}$ n'est pas négligeable pour μ_θ.

Sous ces hypothèses, la v.a. f : $f(\varphi,x,y) = -\operatorname{Log} p(\varphi,x,y)$ est une fonction de contraste dès que les propriétés d'intégrabilité sont vraies. En outre, p étant strictement positive, pour tout x de $\mathcal{H}$, l'ensemble $\{\Pi(\varphi,x;.)\}_{\varphi\in\Theta}$ est un ensemble de probabilités équivalentes. Aussi, si p est bornée,

$$I(\theta,\varphi,x) = \int \operatorname{Log} \frac{p(\theta,x,y)}{p(\varphi,x,y)}\, \Pi(\theta,x;dy) = I_{\Pi(\varphi,x,;.)}(\Pi(\theta,x;.)) ,$$

où $I_\mu(\nu)$ est l'information de Kullback de ν par rapport à μ. Pour simplifier, on note $I_{\Pi(\varphi,x;.)}(\Pi(\theta,x;.))$, $I(\theta,\varphi,x)$, et on définit $I(\theta,\varphi)$ par $\int I(\theta,\varphi,x)\, d\mu_\theta(x)$. Dès que φ est différent de θ, $I(\theta,\varphi)$ est strictement positif.

1.2. Estimateur du minimum de contraste :

A l'instant n, un estimateur de θ est une fonction mesurable $\hat{\theta}_n$ de $(\Omega, \mathcal{B}_n)$ dans Θ, c'est-à-dire que $\hat{\theta}_n = \tilde{\theta}_n(X_1,\dots X_n)$ pour une fonction mesurable θ_t de $\mathcal{H}$ dans Θ. On confondra θ_n et $\tilde{\theta}_n$.

Etant donnée f une fonction de contraste on pose :

$$L_n(\alpha) = \sum_{k=0}^{n-1} f(\alpha,X_k,X_{k+1})$$

A l'instant n, un estimateur $\hat{\theta}_n$, $\mathcal{B}_n$ mesurable est un estimateur du minimum de contraste s'il vérifie :

$$L_n(\hat{\theta}) = \inf_{\alpha\in\Theta} L_n(\alpha)$$

Dans l'exemple précédent cet estimateur est l'estimateur du maximum de vraisemblance. Le théorème de sélection de BROWN [3] assure l'existence d'une suite d'estimateurs du minimum de contraste si $\mathcal{H}$ est polonais, si Θ est un espace métrique compact, et si l'application $\alpha \to f(\alpha,x,y)$ est semi continue inférieurement de Θ dans R pour tout x,y de $\mathcal{H}$.

On pose :

$$S_n(\alpha,\varphi) = \sum_{k=0}^{n-1} (f(\varphi,X_k,X_{k+1}) - f(\alpha,X_k,X_{k+1})) = \sum_{k=0}^{n-1} h(\alpha,\varphi,X_k,X_{k+1})$$

2. Consistance exponentielle.

Si λ_o est la loi de probabilité initiale de la chaîne, une suite $\{\hat{\theta}_n\}_{n\in N}$ d'estimateurs du minimum de contraste est dite consistante si pour tout voisinage $U(\theta)$ de θ, $P_{\theta,\lambda_\theta}$ ps, il existe n_o, tel que pour tout $n \geq n_o$, $\hat{\theta}_n$ appartient à $U(\theta)$. Dans [6] on trouve des hypothèses sous lesquelles une suite d'estimateurs du minimum de contraste est consistante dans le cadre 1.

Le but est d'établir ici, la consistance exponentielle de toute suite $(\hat{\theta}_n)_{n\in N}$ d'estimateurs du minimum de contraste lorsque Θ est fini, et la fonction de contraste bornée.

On définit : $T_\theta = \inf\{n;\ \forall p \geq n,\ \hat{\theta}_p = \theta\}$

Théorème 2 : Lorsque Θ est fini, et la fonction de contraste bornée, toute suite $(\hat{\theta}_n)_{n\in N}$ d'estimateurs du minimum de contraste est exponentiellement consistante :

$$\overline{\lim_n}\ \frac{1}{n} \operatorname{Log} \sup_{x\in\mathcal{H}} P_{\theta,x}(T_\theta > n) < 0.$$

Démonstration :

Pour tout x de $\mathcal{H}$, $P_{\theta,x}(T_\theta > n) \leq \sum_{m \geq n} \sum_{\varphi\neq\theta} P_{\theta,x}(S_m(\theta,\varphi) \leq 0)$

Soit φ différent de θ.

$$P_{\theta,x}(S_m(\theta,\varphi) \leq 0) = P_{\theta,x}\left(\sum_{k=0}^{m-1} h(\theta,\varphi,X_k,X_{k+1}) \leq 0\right)$$

Le théorème 1.1. du paragraphe I, appliqué à la fonction bornée $-h(\theta,\varphi,x,y)$ entraîne que pour tout $\varepsilon > 0$, il existe $k > 0$ et $b > 0$, tels que pour tout m de N,

$$\sup_x P_{\theta,x}\left(\sum_{k=0}^{m-1}\{h(\theta,\varphi,X_k,X_{k+1}) - (K(\theta,\varphi)-\varepsilon)\} \leq 0\right) \leq k\, e^{-bm}$$

Pour $\varepsilon = K(\theta,\varphi) > 0$, il existe $k > 0$ et $b > 0$, tels que pour tout m de N,

$$\sup_x P_{\theta,x}\left(\sum_{k=0}^{m-1} h(\theta,\varphi,X_k,X_{k+1}) \leq 0\right) \leq k\, e^{-bm}$$

ce qui entraîne qu'il existe $k' > 0$ et $b > 0$ tels que pour tout n de N,

$$\sup_{x\in\mathcal{H}} P_{\theta,x}(T_\theta > n) \leq k'\, e^{-bn}$$

c'est-à-dire : $\overline{\lim_n}\ \frac{1}{n}\ \mathrm{Log}\ \sup_{x\in\mathcal{H}} P_{\theta,x}(T_\theta > n) < 0$. ■

II. TEST SÉQUENTIEL DU RAPPORT DE VRAISEMBLANCE.

0. Introduction :

L'idée de ce paragraphe est inspirée du problème de test résolu par CHERNOFF dans le cas indépendant avec choix d'expériences. Ici, l'espace des paramètres Θ est fini (on notera q sont cardinal). Le processus des observations X^θ est une chaîne markovienne, récurrente Doeblin, apériodique, de probabilité invariante μ_θ, de loi initiale λ_o. On suppose qu'il existe une v.a. sur $\Theta \times \mathcal{H} \times \mathcal{H}$ strictement positive p, bornée, et une transition λ de $\mathcal{H}$ dans $\mathcal{H}$, telles que pour tout x de $\mathcal{H}$, pour tout Γ de $\underline{\underline{\mathcal{H}}}$,

$$\Pi(\theta,x;\Gamma) = \int p(\theta,x,y)\,\lambda(x,dy)\,1_\Gamma(y) \quad .$$

On suppose aussi que dès que φ est différent de θ, l'événement $\{x, \Pi(\theta,x,.) \neq \Pi(\varphi,x,.)\}$ n'est pas négligeable.

C'est le cadre de l'exemple du paragraphe I, dont on reprend les notations : Pour tout x de $\mathcal{H}$, pour tout θ et φ de Θ :

$$I(\theta,\varphi,x) = \int \mathrm{Log}\ \frac{p(\theta,x,y)}{p(\varphi,x,y)}\ \Pi(\theta,x,dy)$$

et $$I(\theta,\varphi) = \int I(\theta,\varphi,x)\ \mu_\theta(dx) \quad .$$

On définira de plus : $I(\theta) = \inf\limits_{\varphi\in a(\theta)} I(\theta,\varphi)$.

Sous les hypothèses ci-dessus, $f(\theta,.,.) = -\mathrm{Log}\ p\ (\theta,.,.)$ est une fonction de contraste bornée. On désigne dans la suite par $(\hat\theta_n)_{n\in N}$ une suite d'estimateurs du minimum de contraste ; d'après le théorème 2.1 précédent, elle est exponentiellement consistante : $\overline{\lim\limits_n}\ \frac{1}{n}\ \mathrm{Log}\ \sup\limits_x P_{\theta,x}\ (T_\theta > n) < 0$. On se pose le problème de test suivant : (Θ_1, Θ_2) étant une partition de Θ, on veut savoir à quel élément de la partition appartient la vraie valeur du paramètre. c est le coût commun à chaque expérience, $0 < c < 1$, et $r(.)$ la perte due à l'erreur de décision.

On note : $h(\theta) = \Theta_i$ si θ appartient à Θ_i ($i=1$ ou 2), et $a(\theta) = \Theta - h(\theta)$. $(\hat{\hat\theta}_n)_{n\in N}$ désigne alors une suite d'estimateurs, telle que pour tout n, $\hat{\hat\theta}_n$ soit $\mathcal{B}_n$ mesurable et soit défini par :

$$\sum_{k=0}^{n-1} \mathrm{Log}\ p(\hat{\hat\theta}_n, X_k, X_{k+1}) = \sup_{\alpha\in a(\hat\theta_n)} \sum_{k=0}^{n-1} \mathrm{Log}\ p\ (\alpha, X_k, X_{k+1}) \quad .$$

Les résultats obtenus ne dépendent pas de la loi initiale λ_o : on notera P_{θ,λ_o} (respectivement E_{θ,λ_o}) P_θ (E_θ).

La théorie développée ici est une théorie asymptotique, pour des valeurs de c petites.

1. Etude d'un temps d'arrêt.

Soit b positif, θ et φ de Θ. On définit :

$$\zeta_b = \inf\{n;\ S_n(\hat\theta_n, \hat{\hat\theta}_n) > b\}$$

et $$\zeta_{b,\varphi} = \inf\{n;\ \forall p \geq n, S_p(\theta,\varphi) > b\} \quad .$$

On a la relation : $\zeta_b \leq \max\ (T_\theta, \max\limits_{\varphi\in a(\theta)} \zeta_{b,\varphi})$

la suite $(\hat{\theta}_n)_{n\in N}$ étant exponentiellement consistante, il existe $k > 0$ et $b > 0$, tels que pour tout n de N, on ait : $P_\theta(T_\theta > n) \leq ke^{-bn}$, donc $P_\theta(T_\theta = \infty) = 0$.

La chaîne étant récurrente Harris positive et $\mathrm{Log}\,\frac{p(\theta,.,.)}{p(\varphi,.,.)}\,\mu_\theta \otimes \Pi_\theta$ intégrable, un théorème d'ergodicité implique :

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathrm{Log}\,\frac{p(\theta,X_k,X_{k+1})}{p(\varphi,X_k,X_{k+1})} \xrightarrow[n\uparrow\infty]{P_\theta\ \text{ps}} \int \mathrm{Log}\,\frac{p(\theta,x,y)}{p(\varphi,x,y)}\,\Pi(\theta,x,dy)\mu_\theta(dx) = I(\theta,\varphi)$$

donc pour tout φ différent de θ, puisque $I(\theta,\varphi)$ est positif non nul,

$$S_n(\theta,\varphi) \xrightarrow[n\uparrow\infty]{P_\theta\ \text{ps}} \infty$$

ce qui entraîne $P_\theta(\zeta_{b,\varphi} = \infty) = 0$.

Finalement : $P_\theta(\zeta_b = \infty) = 0$. Le temps d'arrêt ζ_b est fini P_θ presque sûrement. ■

2. Définition d'un test ϕ_c .

On note ν_c le temps d'arrêt $\zeta_{-\mathrm{Log}\,c}$ défini dans le paragraphe précédent. De la même façon on note $\nu_{c,\varphi}$ le temps d'arrêt $\zeta_{-\mathrm{Log}\,c,\varphi}$.

On définit le test de Chernoff, noté ϕ_c, ainsi :

> sa règle d'arrêt est ν_c
>
> si $n = \nu_c$, on arrêt et on décide que θ est élément de $h(\hat{\theta}_n)$
>
> si $n < \nu_c$, on continue, c'est-à-dire on refait une expérience et on regarde à nouveau dans quelle situation on se trouve.

Le test ϕ_c est clos pour toute valeur c du coût.

$\beta_c(\theta)$ et $R_c(\theta)$ désigneront respectivement la probabilité d'erreur et le risque associés au test de Chernoff ϕ_c .

3. Propriétés du test.

Le but est de montrer que le test ϕ_c possède une propriété d'optimalité analogue à celle obtenue par Chernoff dans le cas indépendant avec contrôle.

3.a. La probabilité d'erreur

Théorème 3.a : Pour tout θ de Θ, (card Θ = q), la probabilité d'erreur $\beta_c(\theta)$ associée au test ϕ_c vérifie : $\beta_c(\theta) \leq qc$.

Démonstration : Soit θ de Θ.

Pour φ élément de $a(\theta)$, on définit : $\Omega_{n,\varphi} = \{\hat{\theta}_n = \varphi,\ \nu_c = n\}$.

Alors : $\beta_c(\theta) = \sum_{n=1}^{\infty} \sum_{\varphi \in a(\theta)} P_\theta(\Omega_{n,\varphi})$, avec :

$$P_\theta(\Omega_{n,\varphi}) = P_\theta\left(\prod_{k=0}^{n-1} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})} \leq c \cap \nu = n\right)$$

Mais pour toute fonction g mesurable :

$$E_\theta(g(X_o \ldots, X_{n-1})) = E_\varphi(g(X_o, \ldots, X_{n-1}) \prod_{i=0}^{n-1} \frac{p(\theta, X_i, X_{i+1})}{p(\varphi, X_i, X_{i+1})}$$

c'est-à-dire $\quad P_\theta = \prod_{i=0}^{n-1} \frac{p(\theta, X_i, X_{i+1})}{P(\varphi, X_i, X_{i+1})} P_\varphi \quad$ sur $\mathcal{B}_n$

D'où :

$$P_\theta(\Omega_{n,\varphi}) = E_\varphi\left\{ 1_{\left\{ \prod_{k=0}^{n-1} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})} \leq c\right\}} 1_{(\nu=n)} \prod_{k=0}^{n-1} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})}\right\}$$

$$P_\theta(\Omega_{n,\varphi}) \leq c\, P_\varphi\left(\prod_{k=0}^{n-1} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})} \leq c, \cap \nu = n\right)$$

$$\sum_{n=1}^{\infty} P_\theta(\Omega_{n,\varphi}) \leq c$$

Par conséquent : $\beta_c(\theta) \leq cq$ ■

3.b. La durée moyenne du test.

Théorème 3.b. : Pour tout θ de Θ, $E_\theta\,(\nu_c) \le \frac{-(1+o(1))\ \mathrm{Log}\ c}{I(\theta)}$

Démonstration : Soit $\varepsilon > 0$ et θ un élément de Θ.

D'après le théorème 2.1, il existe $b_1 > 0$ et $k_1 > 0$ tels que pour tout n de N, pour tout φ de $a(\theta)$, on ait :

$$P_\theta\ (S_n\ (\theta,\varphi) \le \frac{n\ I(\theta)}{1+\frac{\varepsilon}{2}}) \le k_1\ e^{-b_1\ n}$$

De la relation : $P_\theta\ (\nu_{c,\varphi}=n) \le P_\theta\ (S_{n-1}\ (\theta,\varphi) \le -\ \mathrm{Log}\ c)$, on déduit qu'il existe $k_1 > 0$ et $b_1 > 0$, tels que pour tout φ de $a(\theta)$, pour tout $n \ge \frac{-(1+\frac{\varepsilon}{2})\ \mathrm{Log}\ c}{I(\theta)}$, on ait :

$$P_\theta\ (\nu_{c,\varphi} = n) \le k_1\ e^{-b_1\ n} \quad .$$

Ce résultat d'une part, la consistance exponentielle de la suite $(\hat{\theta}_n)_{n\in N}$ d'autre part, et la relation :

$$\{\nu_c = n\} \subset \{T_\theta > n\} \cup \{\bigcup_{\varphi\in a(\theta)} \nu_{c,\varphi} = n\}$$

entraîne qu'il existe $k > 0$ et $b > 0$ tels que pour tout $n \ge \frac{-(1+\frac{\varepsilon}{2})\ \mathrm{Log}\ c}{I(\theta)}$ on ait :

$$P_\theta\ (\nu_c = n) \le k\ e^{-bn} \quad .$$

Soit alors : $N_c = \frac{-(1+\frac{\varepsilon}{2})\ \mathrm{Log}\ c}{I(\theta)}$

$$E_\theta(\nu_c) = E_\theta\ (\nu_c\ 1_{(\nu_c \le N_c)}) + E_\theta\ (\nu_c\ 1_{(\nu_c > N_c)})$$

$$E_\theta(\nu_c) \le N_c + \sum_{n>N_c} n\ P_\theta\ (\nu_c = n)$$

$$E_\theta(\nu_c) \le N_c + k \sum_{n>N_c} n\ e^{-bn}$$

Comme $\sum_{n>N_c} n\ e^{-bn} = \xi_c$ tend vers 0 quand c tend vers 0, il existe $c_o(\varepsilon)$ tel que pour $c < c_o(\varepsilon)$, on ait : $\xi_c \le \frac{-\ \varepsilon\ \mathrm{Log}\ c}{2k\ I(\theta)}$.

Aussi, il existe $c_o(\varepsilon)$, tel que pour tout $c < c_o(\varepsilon)$,
$E_\theta(\nu_c) \leq \frac{-(1+\varepsilon) \operatorname{Log} c}{I(\theta)}$. ■

<u>Corollaire</u> : Pour tout θ de Θ, $R_c(\theta) \leq \frac{-(1+\sigma(1)) c \operatorname{Log} c}{I(\theta)}$

<u>Démonstration</u> :

$$R_c(\theta) = c\, E_\theta(\nu_c) + \beta_c(\theta)\, r(\theta)$$

Les 2 théorèmes précédents permettent de conclure facilement. ■

3.c. <u>Optimalité asymptotique de la procédure ϕ_c</u>

On note (T,δ) une procédure séquentielle de temps d'arrêt T et de règle de décision δ. $e(\theta)$ sera la probabilité d'erreur de la procédure (T,δ).

<u>Lemme 3.a</u> : Soit b, $0 < b < 1$ et θ de Θ. Soit (T,δ) une procédure telle que pour tout φ, $e(\varphi)$ soit un $\mathcal{O}(-b \operatorname{Log} b)$. Pour tout φ de $a(\theta)$, pour tout ε, $0 < \varepsilon < 1$, on a :

$$P_\theta\ (S_T(\theta,\varphi) < -(1-\varepsilon) \operatorname{Log} b) = \mathcal{O}(-b^\varepsilon \operatorname{Log} b) \quad .$$

<u>Démonstration</u> : Soit θ de Θ, φ appartenant à $a(\theta)$, et ε, $0<\varepsilon<1$.

On définit : $\Omega_n = \{T=n,\ S_n(\theta,\varphi) < -(1-\varepsilon) \operatorname{Log} b,\ \delta \text{ accepte } h(\theta)\}$.

Alors :
$$e(\varphi) \geq \sum_{n=1}^{\infty} P_\varphi\ (\Omega_n)$$
$$\geq \sum_{n=1}^{\infty} P_\varphi\ (\{\prod_{k=0}^{n-1} \frac{p(\varphi,X_k,X_{k+1})}{p(\theta,X_k,X_{k+1})} > b^{1-\varepsilon}\}\cap\{T=n\}\cap\{\text{"}h(\theta)\text{"}\})$$

où "$h(\theta)$" signifie : la règle de décision décide $h(\theta)$.

Comme $P_\varphi = \prod_{k=0}^{n-1} \frac{p(\varphi,X_k,X_{k+1})}{p(\theta,X_k,X_{k+1})} P_\theta$ sur $\mathcal{B}_n$, on a :

$$e(\varphi) \geq \sum_{n=1}^{\infty} E_\theta \left(1_{\left\{ \prod_{k=0}^{n-1} \frac{p(\varphi, X_k, X_{k+1})}{p(\theta, X_k, X_{k+1})} > b^{1-\varepsilon} \right\}} 1_{\{T=n\}} 1_{\{"h(\theta)"\}} \times \prod_{k=0}^{n-1} \frac{p(\varphi, X_k, X_{k+1})}{p(\theta, X_k, X_{k+1})} \right)$$

$$e(\varphi) \geq b^{1-\varepsilon} \sum_{n=1}^{\infty} P_\theta (\Omega_n)$$

$$b^{\varepsilon-1} e(\varphi) \geq \sum_{n=1}^{\infty} P_\theta (\Omega_n)$$

$$\sum_{n=1}^{\infty} P_\theta (\Omega_n) = \mathcal{O}(-b^{\varepsilon} \text{ Log } b)$$

Comme :

$$P_\theta (S_T(\theta,\varphi) < -(1-\varepsilon)\text{Log } b = P_\theta \left(\bigcup_{n=1}^{\infty} (T=n, S_n(\theta,\varphi) < -(1-\varepsilon)\text{Log } b) \right)$$

$$\leq \sum_{n=1}^{\infty} P_\theta (\Omega_n) + e(\theta)$$

On conclut : $P_\theta (S_T(\theta,\varphi) < -(1-\varepsilon)\text{Log } b) = \mathcal{O}(-b^{\varepsilon} \log b)$ ■

<u>Lemme 3.b.</u> : $\forall \varepsilon > 0, \quad P_\theta \left(\max_{0 \leq m \leq n} \min_{\varphi \in a(\theta)} S_m(\theta,\varphi) \geq n(I(\theta) + \varepsilon) \right) \xrightarrow[n \uparrow \infty]{} 0$

<u>Démonstration</u> : Soients ε et η positifs non nuls.

La chaîne étant récurrente Harris positive, d'après un théorème ergodique on a :

$$\forall \varphi, \quad \frac{1}{n} S_n(\theta,\varphi) \xrightarrow[n \uparrow \infty]{} I(\theta,\varphi) \quad \text{sur } (\Omega, \mathcal{A}, P_\theta) \quad \text{en probabilité.}$$

Il existe donc n_o tel que pour tout $n \geq n_o$:

$$P_\theta \left(\max_{n_o \leq m \leq n} \min_{\varphi \in a(\theta)} S_m(\theta,\varphi) \geq n(I(\theta) + \varepsilon) \right) \leq \eta$$

Pour tout m, $S_m(\theta,\varphi)$ est fini. Il existe donc n_1 tel que pour tout $n \geq n_1$:

$$P_\theta \left(\max_{0 \leq m \leq n_o} \min_{\varphi \in a(\theta)} S_m(\theta,\varphi) \geq n(I(\theta) + \varepsilon) \right) = 0$$

D'où , il existe n_2 tel que pour tout n supérieur à n_2 :

$$P_\theta\left(\max_{0\le m\le n}\ \min_{\varphi\in a(\theta)} S_m(\theta,\varphi) \ge n(I(\theta)+\varepsilon)\right) \le \eta \qquad \blacksquare$$

Lemme 3.c. : Soit b, $0 < b < 1$. Si (T,δ) est une procédure telle que pour tout φ , la probabilité d'erreur $e(\varphi)$ soit un $\mathcal{O}(-b \operatorname{Log} b)$, elle vérifie :

$$E_\theta(T) \ge \frac{-(1+o(1)) \operatorname{Log} b}{I(\theta)} .$$

Démonstration : Soit $\varepsilon > 0$, $0 < b < 1$ et soit (T,δ) une procédure vérifiant l'hypothèse du lemme. On définit les constantes ε', ε'' et n_b par :

$$\varepsilon' = \frac{I(\theta)}{1+\varepsilon+I(\theta)} , \quad (1-\varepsilon'')^2 = (1-\varepsilon') \quad \text{et} \quad n_b = \frac{-(1-\varepsilon'') \operatorname{Log} b}{I(\theta)+\varepsilon'}$$

On a :

$$P_\theta(T \le n_b) \le P_\theta(T \le n_b \text{ et pour tout } \varphi \text{ de } a(\theta),\ S_T(\theta,\varphi) \ge -(1-\varepsilon'')\operatorname{Log} b)$$

$$+ P_\theta(\text{il existe } \varphi \text{ de } a(\theta),\ S_T(\theta,\varphi) < -(1-\varepsilon'')\operatorname{Log} b)$$

Notons $K_1^b [K_2^b]$ le 1er $[2^{\text{ème}}]$ terme du membre de gauche.

$$K_1^b \le P_\theta\left(\max_{1\le m\le n_b}\ \min_{\varphi\in a(\theta)} S_m(\theta,\varphi) \ge n_b\,(I(\theta)+\varepsilon')\right)$$

D'après le lemme 3.b , K_1^b tend vers 0 quand n_b tend vers l'infini, c'est-à-dire quand b tend vers 0.

Quant au lemme 3.a, il permet de conclure que K_2^b tend vers 0 quand b tend vers 0.

Par conséquent, il existe $b_o(\varepsilon)$, tel que pour tout $b \le b_o(\varepsilon)$, on ait :

$$P_\theta\ (T \le n_b) \le \varepsilon'' .$$

Comme : $E_\theta(T) = \sum_{n=0}^{\infty} P_\theta\ (T_\theta > n) \ge n_b\ P_\theta\ (T > n_b)$, pour tout $b \le b_o(\varepsilon)$, on a :

$$E_\theta(T) \geq n_b\ (1-\varepsilon'') \geq \frac{-(1-\varepsilon'')^2\ \text{Log}\ b}{I(\theta)+\varepsilon'}$$

$$\geq \frac{-(1-\varepsilon')\ \text{Log}\ b}{I(\theta)+\varepsilon'}$$

$$\geq \frac{-(1+\varepsilon)\ \text{Log}\ b}{I(\theta)} \qquad \blacksquare$$

Tous ces résultats permettent de montrer la propriété d'optimalité asymptotique suivante.

Théorème 3.c. : Le test ϕ_c est asymptotiquement le plus économique parmi les tests de force au moins égale.

Démonstration : Soit (T,δ) une procédure telle que la probabilité d'erreur $e(\varphi)$ soit un $\mathcal{O}$ $(-c\ \text{Log}\ c)$ pour tout φ. D'après le lemme 3.3.3, elle vérifie :

$$E_\theta(T) \geq \frac{-(1+o(1)\ \text{Log}\ c}{I(\theta)} \quad .$$

Ce résultat, comparé aux résultats des théorèmes 3.a et 3.b précédents permet de conclure que ϕ_c est asymptotiquement le plus économique parmi les tests de force au moins égale. $\blacksquare$

Bibliographie :

[1] ALBERT A.E. : "The sequential design of experiments for infinitely many states of nature". Ann. Math. Stat., vol. 32 (1961),774-799

[2] BESSLER S. :"Theory and applications of the sequential design of experiments, k actions and infinitely many experiments". Part I, II. Mech. Report. N° 55, 56. 1960

[3] BROWN :"Measurable selections of extrema" Annals of Statistics, vol. 1, n° 5, 902-912 (1973)

[4] CHERNOFF H. :"Sequential design of experiments" Ann. Math. Stat., vol. 30 (1959). 755-770.

[5] DONSKER et VARADHAN :"Asymptotic evaluation of certain process expectations for large time". Communication on pure and applied mathematics I. vol. 28, 1-47 (1975)

[6] GÄNSLER :"Note on minimum contrast estimates for Markov processes" Metrika 19, 115-130 (1972)

[7] GREY :"Limit theorems for markov chains transitions probabilities" Van Nostrand (1971)

Nelly MAIGRET
Université Paris Nord
Mathématiques
ERA CNRS 532 "Statistique Appliquée"
Avenue J.B. Clément
93430 VILLETANEUSE.