

Astérisque

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

© Société mathématique de France, 1979, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

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.

$X = (\Omega, \mathcal{A}, (P_x)_{x \in \mathcal{H}}, (X_n)_{n \in \mathbb{N}})$ une chaîne de Markov sur un espace d'état mesurable $(\mathcal{H}, \underline{\mathcal{H}})$ où la tribu $\underline{\mathcal{H}}$ est supposée à base dénombrable. On note π le noyau de transition, et π_n la $n^{\text{ième}}$ itérée de π_0 . \mathcal{B}_n désigne la tribu des événements antérieurs à n : $\mathcal{B}_n = \sigma(X_0, \dots, X_n)$. On note \mathcal{P} l'ensemble des probabilités sur $(\mathcal{H}, \underline{\mathcal{H}})$. Lorsque $(\mathcal{H}, \underline{\mathcal{H}})$ est métrique, \mathcal{P} est muni de la topologie de la convergence étroite. On note \mathbb{P}_λ 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 $\| \cdot \|$, la norme uniforme sur \mathcal{P} et sur l'ensemble des variables aléatoires (v.a.) sur $\mathcal{H} \times \underline{\mathcal{H}}$.

EXPOSÉ 7

La chaîne de Markov X est supposée récurrente Doeblin, aperiodique, de probabilité invariante μ , c'est-à-dire : il existe $c > 0$, il existe ρ , $0 < \rho < 1$, tels que pour tout n , $\sup_{x \in \mathcal{X}} \|\pi_n(x, \cdot) - \mu(\cdot)\| \leq c \rho^n$ (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 \mathbb{N} , on notera $\mu \otimes \pi_n$ la mesure sur \mathcal{X}^{n+1} définie par

$$(\mu \otimes \pi_n)(A) = \int 1_A(x_0, \dots, x_n) d\mu(x_0) \pi(x_0, dx_1) \dots \pi(x_{n-1}, dx_n)$$

Théorème 1 :

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

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

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

Démonstration :

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

Alors $u_0 = h'(a)$ est strictement positif et vérifie

$$e^{-h(a)} = E_{\mu} \left(e^{(g(X_0, X_1) - a)u_0} \right)$$

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

STATISTIQUE SÉQUENTIELLE

$$\begin{aligned} \left| E_x (e^{u_0 (g(X_k, X_{k+1}) - a)}) - e^{-h(a)} \right| &= \left| \int e^{u_0 (g(y, z) - a)} \Pi(y, dz) (\Pi_k(x, dy) - \mu(dy)) \right| \\ &\leq \| e^{u_0 (g-a)} \| \| \Pi_k(x, \cdot) - \mu(\cdot) \| \\ &\leq \| e^{u_0 (g-a)} \| c \rho^k = b \rho^k \end{aligned}$$

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

$$\sup_x E_x (e^{u_0 (g(X_k, X_{k+1}) - a)}) \leq e^{-h(a)} + b \rho^k$$

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

$$\begin{aligned} E_x (e^{u_0 \sum_{p=1}^n (g(X_{kp}, X_{kp+1}) - a)}) \\ &= E_x e^{u_0 \sum_{p=1}^{n-1} (g(X_{kp}, X_{kp+1}) - a)} E_x (e^{u_0 (g(X_{kn}, X_{kn+1}) - a)}) / \mathcal{B}_{k(n-1)+1} \\ &= E_x (e^{u_0 \sum_{p=1}^{n-1} (g(X_{kp}, X_{kp+1}) - a)} E_{X_{k(n-1)}} (e^{u_0 (g(X_k, X_{k+1}) - a)}) \end{aligned}$$

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

$$\begin{aligned} E_x (e^{u_0 \sum_{p=1}^{n-1} (g(X_{kp}, X_{kp+1}) - a)}) (e^{-h(a)} + b \rho^k) \text{ et en itérant on trouve} \\ \sup_{x \in \mathcal{X}} E_x (e^{u_0 \sum_{p=1}^n (g(X_{kp}, X_{kp+1}) - a)}) \leq (e^{-h(a)} + b \rho^k)^n \end{aligned}$$

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

$$E_x (e^{u_0 \sum_{p=1}^n (g(X_{pk+r}, X_{pk+r+1}) - a)}) = E_x (E_{X_r} (e^{u_0 \sum_{p=1}^n (g(X_{kp}, X_{kp+1}) - a)}))$$

La dernière relation démontrée entraîne :

$$\sup_x E_x (e^{u_0 \{ \sum_{p=1}^n (g(X_{pk+r}, X_{pk+r+1}) - a) \}}) \leq (e^{-h(a)} + b \rho^k)^n$$

Puisque a est strictement supérieur à m , u_0 est strictement positif et on peut écrire :

EXPOSÉ 7

$$P_x \left(\sum_{p=k}^{nk-1} \{g(X_p, X_{p+1}) - a\} \geq 0 \leq \sum_{r=0}^{k-1} P_x \left(\sum_{p=1}^{n-1} (X_{pk+r}, X_{pk+r+1}) - a \geq 0 \right) \right. \\ \left. \leq \sum_{r=0}^{k-1} E_x \left(e^{u \sum_{p=1}^{n-1} (g(X_{pk+r}, X_{pk+r+1}) - a)} \right) \right)$$

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

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

$$P_x \left(\sum_{p=0}^n (g(X_p, X_{p+1}) - a) \geq n\epsilon \right) \leq P_x \left(\sum_{p=0}^{k-1} (g(X_p, X_{p+1}) - a) \geq \frac{n\epsilon}{2} \right) \\ + P_x \left(\sum_{p=k}^{\lfloor \frac{n}{k} \rfloor k - 1} (g(X_p, X_{p+1}) - a) \geq 0 \right) + P_x \left(\sum_{p=\lfloor \frac{n}{k} \rfloor k}^n (g(X_p, X_{p+1}) - a) \geq \frac{n\epsilon}{2} \right)$$

où pour tout réel s , $\lfloor s \rfloor$ désigne sa partie entière.

g étant une v.a. bornée sur $(\mathcal{H} \times \mathcal{H})$, il existe $n_0(k, \epsilon)$, tel que

$\forall n \geq n_0$,

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

$$\text{et } \sup_x P_x \left(\sum_{p=\lfloor \frac{n}{k} \rfloor k}^n (g(X_p, X_{p+1}) - a) \geq \frac{n\epsilon}{2} \right) = 0$$

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

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

Par conséquent, il existe $n_0(k, \epsilon)$, tel que pour tout $n \geq n_0$,

$$\sup_{x \in \mathcal{H}} P_x \left(\sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq n\epsilon \right) \leq k (e^{-h(a)} + b \rho^k)^{\lfloor \frac{n}{k} \rfloor - 1}$$

$$\frac{1}{n} \text{Log} \sup_{x \in \mathcal{H}} P_x \left(\sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq n\epsilon \right) \leq \frac{\text{Log } k}{n} + \frac{1}{n} (\lfloor \frac{n}{k} \rfloor - 1) \text{Log} (e^{-h(a)} + b \rho^k)$$

ce qui entraîne

STATISTIQUE SÉQUENTIELLE

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

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

$$\overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{X}} P_x \left(\sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq n\epsilon \right) \leq \inf_{k \in \mathbb{N}} \left(\frac{1}{k} \text{Log} (e^{-h(a)} + b\rho^k) \right).$$

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

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

$$\overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{X}} P_x \left(\sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq 0 \right) \leq \inf_{k \in \mathbb{N}} \left(\frac{1}{k} \text{Log} (e^{-h(a-\epsilon)} + b_{a-\epsilon} \rho^k) \right)$$

Faisons tendre ϵ vers 0.

h est une fonction convexe donc continue : $\lim_{\epsilon \rightarrow 0} h(a - \epsilon) = h(a)$.

On a : $b_{a-\epsilon} = c \left\| e^{h'(a-\epsilon) \cdot (g-a+\epsilon)} \right\|$

h' est continue donc : $\lim_{\epsilon \rightarrow 0} b_{a-\epsilon} = b_a$.

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

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

Cet inf est strictement négatif puisque $h(a)$ étant strictement positif ($a \neq m$) et ρ strictement inférieur à 1, il existe k_0 de \mathbb{N} tel que $(e^{-h(a)} + b_a \rho^{k_0})$ 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

$a \rightarrow \overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{X}} P_x \left(\sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq 0 \right)$ est croissante. ■

Remarque : Soit $a \in]m, \bar{g}[$. Si g et c vérifient : $\|g\| \geq a$ et $c \geq 1$, la constante $b_a = c \left\| e^{u_0(g-a)} \right\|$ est supérieure à 1 et $\inf_{k \in \mathbb{N}} \frac{1}{k} \text{Log} (e^{-h(a)} + b_a \rho^k)$ est minorée par $\text{Log} \rho$. Donc en général cet inf ne sera pas $-\infty$ même si $\|g\| \geq a > \bar{g}$. L'exemple suivant met en évidence cette minoration par $\text{Log} \rho$.

EXPOSÉ 7

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 $\bar{g} = 0$ et on a :

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

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

DEUXIÈME PARTIE : STATISTIQUE.

Maintenant, le noyau π dépend d'un paramètre. L'espace des paramètres $(\Theta \times \underline{\Theta})$ 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 $X^{\hat{\theta}} = (\Omega, \mathcal{A}, (P_{\theta, x})_{x \in \mathcal{H}}, (X_n)_{n \in \mathbb{N}}$ est une chaîne markovienne récurrente Doeblin, apériodique, de probabilité invariante $\mu_{\hat{\theta}}$. 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 $X^{\hat{\theta}}$ 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

$\int \mathbb{1}_{\hat{\theta}}(x) \pi(\hat{\theta}, x, dy) f(\hat{\theta}, x, y)$ est fini et $\int \mathbb{1}_{\hat{\theta}}(x) \pi(\hat{\theta}, x, dy) (f(\varphi, x, y) - f(\hat{\theta}, x, y)) \geq 0$ l'égalité n'ayant lieu que lorsque $\varphi = \hat{\theta}$.

STATISTIQUE SÉQUENTIELLE

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

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

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 $\Gamma \in \underline{\mathcal{H}}$,

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

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

Sous ces hypothèses, la v.a. $f : f(\varphi, x, y) = -\text{Log } p(\varphi, x, y)$ 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 $\{\pi(\varphi, x; \cdot)\}_{\varphi \in \Theta}$ est un ensemble de probabilités équivalentes. Aussi, si p est bornée,

$$I(\theta, \varphi, x) = \int \text{Log } \frac{p(\theta, x, y)}{p(\varphi, x, y)} \pi(\theta, x; dy) = I_{\pi(\varphi, x; \cdot)}(\pi(\theta, x; \cdot)),$$

où $I_\mu(\nu)$ est l'information de Kullback de ν par rapport à μ . Pour simplifier, on note $I_{\pi(\varphi, x; \cdot)}(\pi(\theta, x; \cdot))$, $I(\theta, \varphi, x)$, et on définit $I(\theta, \varphi)$ par $\int I(\theta, \varphi, x) \mu_\theta(x)$. 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 (Ω, \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 $\tilde{\theta}_n$ de \mathcal{H} dans Θ . On confondra $\hat{\theta}_n$ 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 $\alpha \rightarrow f(\alpha, x, y)$ 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_0 est la loi de probabilité initiale de la chaîne, une suite $\{\hat{\theta}_n\}_{n \in \mathbb{N}}$ 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_0 , tel que pour tout $n \geq n_0$, $\hat{\theta}_n$ 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 $\{\hat{\theta}_n\}_{n \in \mathbb{N}}$ 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 $\{\hat{\theta}_n\}_{n \in \mathbb{N}}$ d'estimateurs du minimum de contraste est exponentiellement consistante :

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

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 .

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

STATISTIQUE SÉQUENTIELLE

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

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

Pour $\epsilon = K(\theta, \varphi) > 0$, 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 : $\lim_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{H}} P_{\theta, x} (T_{\theta} > n) < 0$. ■

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 son cardinal). Le processus des observations X^{θ} est une chaîne markovienne, récurrente Doeblin, apériodique, de probabilité invariante μ_{θ} , de loi initiale λ_{θ} . 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 $\underline{\mathcal{H}}$,

$$\pi(\theta, x; \Gamma) = \int p(\theta, x, y) \lambda(x, dy) 1_{\Gamma}(y) .$$

On suppose aussi que dès que φ est différent de θ , l'événement $\{x, \pi(\theta, x, \cdot) \neq \pi(\varphi, x, \cdot)\}$ 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 :

EXPOSÉ 7

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

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

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

Sous les hypothèses ci-dessus, $f(\theta, \dots) = -\text{Log} p(\theta, \dots)$ est une fonction de contraste bornée. On désigne dans la suite par $(\hat{\theta}_n)_{n \in \mathbb{N}}$ une suite d'estimateurs du minimum de contraste ; d'après le théorème 2.1 précédent, elle est exponentiellement consistante : $\overline{\lim}_n \frac{1}{n} \text{Log} \sup_x P_{\theta, x}(T_\theta > n) < 0$. 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 $r(\cdot)$ 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 $a(\theta) = \Theta - h(\theta)$. $(\hat{\theta}_n)_{n \in \mathbb{N}}$ désigne alors une suite d'estimateurs, telle que pour tout n , $\hat{\theta}_n$ soit \mathcal{B}_n mesurable et soit défini par :

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

Les résultats obtenus ne dépendent pas de la loi initiale λ_0 : on notera P_{θ, λ_0} (respectivement E_{θ, λ_0}) 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 :

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

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

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

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

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

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

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

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

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

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

2. Définition d'un test ϕ_c .

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

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

<p>sa règle d'arrêt est ν_c</p> <p>si $n = \nu_c$, on arrête et on décide que $\hat{\theta}$ est élément de $h(\hat{\theta}_n)$</p> <p>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.</p>

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

$\beta_c(\hat{\theta})$ et $R_c(\hat{\theta})$ 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 $\Theta = q$), 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 :

$$E_{\theta}(g(X_0, \dots, X_{n-1})) = E_{\varphi}(g(X_0, \dots, X_{n-1})) \prod_{i=0}^{n-1} \frac{p(\theta, X_i, X_{i+1})}{p(\varphi, X_i, X_{i+1})}$$

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

D'où :
$$P_{\theta}(\Omega_{n,\varphi}) = E_{\varphi} \left\{ \prod_{k=0}^{n-1} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})} \leq c \right\} \quad \text{!} \quad (\nu=n) \prod_{k=0}^{n-1} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})}$$

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

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

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

3.b. La durée moyenne du test.

Théorème 3.b. : Pour tout θ de Θ , $E_{\theta}(v_c) \leq \frac{-(1+\frac{\epsilon}{2}) \text{Log } c}{I(\theta)}$

Démonstration : Soit $\epsilon > 0$ 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 :

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

De la relation : $P_{\theta}(v_{c, \varphi} = n) \leq P_{\theta}(S_{n-1}(\theta, \varphi) \leq -\text{Log } c)$, on déduit qu'il existe $k_1 > 0$ et $b_1 > 0$, tels que pour tout φ de $a(\theta)$, pour tout

$n \geq \frac{-(1 + \frac{\epsilon}{2}) \text{Log } c}{I(\theta)}$, on ait :

$$P_{\theta}(v_{c, \varphi} = n) \leq k_1 e^{-b_1 n}.$$

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

$$\{v_c = n\} \subset \{T_{\theta} > n\} \cup \left\{ \bigcup_{\varphi \in a(\theta)} v_{c, \varphi} = n \right\}$$

entraîne qu'il existe $k > 0$ et $b > 0$ tels que pour tout $n \geq \frac{-(1 + \frac{\epsilon}{2}) \text{Log } c}{I(\theta)}$

on ait :

$$P_{\theta}(v_c = n) \leq k e^{-bn}.$$

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

$$E_{\theta}(v_c) = E_{\theta}(v_c \cdot I_{(v_c \leq N_c)}) + E_{\theta}(v_c \cdot I_{(v_c > N_c)})$$

$$E_{\theta}(v_c) \leq N_c + \sum_{n > N_c} n P_{\theta}(v_c = n)$$

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

Comme $\sum_{n > N_c} n e^{-bn} = \xi_c$ tend vers 0 quand c tend vers 0, il existe

$c_0(\epsilon)$ tel que pour $c < c_0(\epsilon)$, on ait : $\xi_c \leq \frac{-\epsilon \text{Log } c}{2k I(\theta)}$.

EXPOSÉ 7

Aussi, il existe $c_0(\epsilon)$, tel que pour tout $c < c_0(\epsilon)$,
 $E_{\theta}(\nu_c) \leq \frac{-(1+\epsilon) \text{Log } c}{I(\theta)}$. ■

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

Démonstration :

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

Les 2 théorèmes précédents permettent de conclure facilement. ■

3.c. Optimalité asymptotique de la procédure ϕ_c

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) .

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

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

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

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

Alors :

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

où "h(θ)" signifie : la règle de décision décide h(θ).

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{D}_n , on a :

STATISTIQUE SÉQUENTIELLE

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

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

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

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

Comme :

$$\begin{aligned} P_{\theta}(S_T(\theta, \varphi) < -(1-\epsilon)\text{Log } b) &= P_{\theta} \left(\bigcup_{n=1}^{\infty} (T=n, S_n(\theta, \varphi) < -(1-\epsilon)\text{Log } b) \right) \\ &\leq \sum_{n=1}^{\infty} P_{\theta}(\Omega_n) + e(\theta) \end{aligned}$$

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

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

Démonstration : Soient ϵ et η positifs non nuls.

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

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

Il existe donc n_0 tel que pour tout $n \geq n_0$:

$$P_{\theta} \left(\max_{n_0 \leq m \leq n} \min_{\varphi \in a(\theta)} S_m(\theta, \varphi) \geq n(I(\theta) + \epsilon) \right) \leq \eta$$

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

$$P_{\theta} \left(\max_{0 \leq m \leq n_0} \min_{\varphi \in a(\theta)} S_m(\theta, \varphi) \geq n(I(\theta) + \epsilon) \right) = 0$$

EXPOSÉ 7

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

$$P_{\theta} \left(\max_{0 \leq m \leq n} \min_{\varphi \in a(\theta)} S_m(\theta, \varphi) \geq n(I(\theta) + \epsilon) \right) \leq \eta \quad \blacksquare$$

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 $\mathcal{O}(-b \text{Log } b)$, elle vérifie :

$$E_{\theta}(T) \geq \frac{-(1+\sigma(1)) \text{Log } b}{I(\theta)} .$$

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

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

On a :

$$P_{\theta}(T \leq n_b) \leq P_{\theta}(T \leq n_b \text{ et pour tout } \varphi \text{ de } a(\theta), S_T(\theta, \varphi) \geq -(1-\epsilon'') \text{Log } b) \\ + P_{\theta}(\text{il existe } \varphi \text{ de } a(\theta), S_T(\theta, \varphi) < -(1-\epsilon'') \text{Log } b)$$

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

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

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 $b_0(\epsilon)$, tel que pour tout $b \leq b_0(\epsilon)$, on ait :

$$P_{\theta}(T \leq n_b) \leq \epsilon'' .$$

Comme : $E_{\theta}(T) = \sum_{n=0}^{\infty} P_{\theta}(T_{\theta} > n) \geq n_b P_{\theta}(T > n_b)$, pour tout $b \leq b_0(\epsilon)$,

on a :

$$\begin{aligned}
 E_{\theta}(T) &\geq n_b (1-\epsilon'') \geq \frac{-(1-\epsilon'')^2 \text{Log } b}{I(\theta) + \epsilon'} \\
 &\geq \frac{-(1-\epsilon') \text{Log } b}{I(\theta) + \epsilon'} \\
 &\geq \frac{-(1+\epsilon) \text{Log } b}{I(\theta)} \quad \blacksquare
 \end{aligned}$$

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 :

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

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. ■

EXPOSÉ 7

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.