ACTA MATHEMATICA VIETNAMICA

TOM 4, N° 2 (1979)

SUR L'ESTIMATION QUASI-OPTIMALE

BÙI TRỌNG LIỄU

U. E. R. de Mathématiques, Université René Descartes de Paris

0 — INTRODUCTION ET RESUME. Le but de cet exposé est de présenter sous sa forme générale, la notion d'estimateurs quasi-optimaux, et à cette occasion, de passer en revue des travaux qui ont été effectués autour de cette notion. Après un examen du problème d'existence dans son contexte général, nous citons les divers cas qui ont été étudiés, les présentant aussi comme exemples d'illustration. Nous évoquons enfin, rapidement, le problème de la convergence au sens presque-sûr des estimateurs.

I — Soit $(\Omega, \mathcal{A}, Pr)$ un espace probabilisé. soit $(X_t)_{t\in T}$ une fonction aléatoire définie sur cet espace, à valeur dans un espace mesurable $(\mathcal{X}, \mathcal{B})$, à temps discret ou à temps continu (c'est à dire que T est un intervalle de $\mathbf{N}$ ou de $\mathbf{R}$), et dont la loi d'évolution dépend d'un paramètre θ appartenant à un ensemble Θ muni d'une tribu $\mathcal{C}$ (la vraie valeur θ_0 de θ est inconnue).

II — DEFINITION. Soient une famille $G = (G_t)_{t\in T}$ d'applications de $\Theta \times \Omega$ dans $\mathbf{R}$, $\mathcal{R} - \mathcal{A} \otimes \mathcal{C}$-mesurables, ($\mathcal{R}$ désignant la tribu borélienne de $\mathbf{R}$, et $\mathcal{A} \otimes \mathcal{C}$ désignant la tribu produit des tribus $\mathcal{A}$ et $\mathcal{C}$) et une famille monotone $c = (c_t)_{t\in T}$ de nombres strictement positifs, telles que quels que soient $(\theta, \omega) \in \Theta \times \Omega$ et $t \in T$,

$$(1) \qquad (1 - c_t)\, G_t(\theta, \omega) \geqslant 0.$$

Nous disons qu'un estimateur $\theta^* = (\theta^*_t)_{t\in T}$ est G-quasi-optimal relativement à la famille c des coefficients c_t, si quel que soit $t \in T$, et quel que soit $\omega \in \Omega$,

$$(2) \qquad G_t(\theta^*_t(\omega), \omega) - c_t \sup_{\theta \in \Theta} G_t(\theta, \omega) \geqslant 0.$$

Plus précisément, l'estimateur θ^* est dit G-quasi-maximal (resp. G-quasi-minimal) si quel que soit $t \in T$, G_t est à valeur positive (resp. G_t est à valeur négative).

Le cas optimal correspond à celui où quel que soit $t \in T$, $c_t = 1$.

C'est la nature de la famille G (donc son choix) qui détermine celle de l'estimateur θ^*. Nous verrons plus loin quelques cas particuliers d'estimateurs quasi-optimaux qui ont été étudiés.

III — PROBLEME D'EXISTENCE: Dans des travaux précédents, et dans des cas particuliers, (cf. par exemple [1], [2], [3]), nous avons alors proposé de régler le problème de l'existence des estimateurs par des théorèmes de sélection, en particulier par les versions dites de Himmelberg-Van Vleck [5], lesquelles sont des généralisations fines du théorème de Kuratowski-Ryll Nardzewski [4]. Ces versions peuvent être utilisées dans le cas général. Il existe aussi d'autres théorèmes de sélection utilisables, comme la version donnée par [6] qui est une généralisation de la version dite de Von Neumann-Aumann. Il ne parait cependant pas possible de prétendre que telle version est meilleure que d'autre: l'allègement d'une condition étant compensée par l'alourdissement d'une autre. Nous donnerons ci-dessous trois énoncés concernant l'existence d'estimateurs quasi-optimaux.

III. 1 — Avec la famille $G=(G_t)_{t\in T}$ et la famille $c=(c_t)_{t\in T}$ données dans la définition II, posons

$$W_t(\omega)=\{\theta\in\Theta:\ G_t(\theta,\omega)-c_t\sup_{\theta\in\Theta}G_t(\theta',\omega)\geqslant 0\}. \tag{3}$$

W_t est une application multivoque si pour tout $\omega\in\Omega$, $W_t(\omega)\neq\phi$. Une condition suffisante pour qu'il en soit ainsi (c'est à dire pour que W_t soit une application multivoque) est que, pour tout $t\in T$, $c_t\neq 1$. Une autre condition suffisante est que Θ soit un espace compact et que pour tout $\omega\in\Omega$, $G_t(.\,,\omega)$ soit une fonction semi-continue supérieurement de la variable θ, (ce qui laisse à c_t la possibilité de prendre la valeur 1, incluant ainsi dans ce cas les méthodes d'estimation optimale et non pas seulement quasi-optimale; dans [1] par exemple, nous avons évoqué ce problème ainsi qu'une condition générale pour la projection d'une tribu produit).

III.2 — PROPOSITION: Si Θ est un espace de Lusin, si $\mathscr{C}$ est sa tribu borélienne; et si pour tout $t\in T$, $W_t:\Omega\to\Theta$ (définie en III.1. (3)) est une application multivoque $\mathscr{A}$-$\mathscr{C}$-mesurable telle que, pour tout $\omega\in\Omega$, $W_t(\omega)$ soit un ensemble fermé dans Θ, alors il existe au moins un estimateur θ^*, G-quasi-optimal relativement à la famille c.

III.3 — PROPOSITION: Si Θ est un espace métrique séparable, si $\mathscr{C}$ est sa tribu borélienne, et si pour tout $t\in T$, $W_t:\Omega\to\Theta$ (définie en III.1. (3)) est une application multivoque $\mathscr{A}$-$\mathscr{C}$-mesurable telle que, pour tout $\omega\in\Omega$, $W_t(\omega)$ soit un ensemble complet dans Θ, alors il existe au moins un estimateur θ^*, G-quasi-optimal relativement à la famille c.

Le premier énoncé est une adaptation à la théorie de l'estimation du premier théorème de sélection de Himmelberg-Van Vleck (cf.[5]), tandis que le deuxième énoncé est une adaptation du corollaire du deuxième théorème de sélection de Himmelberg-Van Vleck (voir la même référence [5]). Ces deux énoncés n'exigent pas que l'espace probabilisé $(\Omega, \mathcal{A}, Pr)$ soit complet. Par contre, si l'espace $(\Omega, \mathcal{A}, Pr)$ est complet, on a alors l'énoncé suivant:

III.4 — PROPOSITION. Si $(\Omega, \mathcal{A}, Pr)$ est complet, si Θ est un espace de Souslin, si $\mathcal{C}$ est la tribu borélienne de Θ, et si pour tout $t \in T$, $W_t : \Omega \to \Theta$ (définie en III.1.(3)) est une application multivoque de graphe mesurable, alors il existe au moins un estimateur θ^*. G-quasi-optimal relativement à la famille c

Rappelons ici que dire que le graphe de W_t est mesurable veut dire que $\{(\omega, \theta) \in \Omega \times \Theta : \theta \in W_t(\omega)\} \in \mathcal{A} \otimes \mathcal{C}$. Le présent énoncé est une adaptation à la présente situation d'un théorème de [6] qui est une généralisation d'un théorème de sélection dit de Von Neumann-Aumann.

IV — CAS PARTICULIERS ET EXEMPLES D'ILLUSTRATION.

IV. 1. Cas des variables aléatoires indépendantes: C'est le cas où $T = \mathbf{N}$ et où $(X_t)_{t \in T}$ est une suite $(X_t)_{t \in \mathbf{N}}$ de variables aléatoires indépendantes et (disons, pour plus de commodité) de même loi $P_X(., \theta)$ qu'une variable aléatoire X.

a) En prenant pour G_t l'opposée de la fonction du χ^2 classique du cas des variables aléatoires indépendantes, c'est à dire

$$G_t(\theta,\omega) = - \chi^2_t(\theta, \omega)$$
$$= - \sum_{i \in I} \frac{[N_t(A_i, \omega) - tP_X(A_i, \theta)]^2}{tP_X(A_i, \theta)}$$

où $(A_i)_{i \in I}$ est une partition finie $\mathcal{B}$-mesurable de $\mathcal{X}$ telle que quel que soit $i \in I$, quel que soit $\theta \in \Theta$, $P_X(A_i, \theta) > 0$ (auquel cas, II.(1) entraîne que $c_t \geqslant 1$) et où $N_t(A_i, \omega) = \sum_{s=1}^{t} 1_{A_i} \circ X_s(\omega)$, relativement à un échantillon $(X_s(\omega))_{0 \leqslant s \leqslant t}$, ($1_{A_i}$ désignant ici l'indicatrice de l'ensemble A_i), on obtient les estimateurs dits du χ^2-quasi-minimum de coefficients c_t. Les θ^*_t vérifient alors

$$= \chi^2_t(\theta^*_t(\omega), \omega) \geqslant c_t \sup_{\theta \in \Theta} [- \chi^2_t(\theta, \omega)]$$

c'est à dire

$$\chi^2_t(\theta^*_t(\omega), \omega) \leqslant c_t \inf_{\theta \in \Theta} \chi^2_t(\theta, \omega) .$$

b) En prenant pour G_t la fonction de vraisemblance classique du cas des variables aléatoires indépendantes dont la loi commune $P_X(.,\theta)$ possède une densité $p(x, \theta)$,

$$G_t(\theta, \omega) = exp\ L_t(\theta, \omega)$$

où (sous des conditions adéquates) $L_t(\theta,\omega)$ peut s'écrire $\int_{\mathcal{X}} Log\ p(x, \theta)\ N_t(dx, \omega)$, (auquel cas, II. (1) entraîne que $c_t \leqslant 1$), nous avons les estimateurs dits du quasi-maximum de vraisemblance de coefficients (additifs) $\varepsilon_t = - Log\, c_t$, présentés souvent sous la forme équivalente

$$L_t(\theta_t^*(\omega), \omega) \geqslant \sup_{\theta \in \Theta} L_t(\theta, \omega) - \varepsilon_t .$$

IV.2. Cas markovien à temps discret: C'est le cas où $T = \mathbf{N}$ et où $(X_t)_{t \in \mathbf{N}}$ est une fonction aléatoire markovienne stationnaire, avec une probabilité absolue stationnaire unique $\pi(.,\theta)$ prise comme probabilité initiale et dont la probabilité de transition $P(., B, \theta)$, ces probabilités étant dépendantes du paramètre θ.

a) Prenons pour G_t la fonction négative

$$G_t(\theta, \omega) = - \psi_t^2(\theta, \omega)$$

$$= - \sum_{(i,j) \in I^2} \frac{\left[N_t(A_i \times A_j, \omega) - t \int_{A_i} \pi(dx, \theta)\, P(x, A_j, \theta)\right]^2}{t \int_{A_i} \pi(dx, \theta)\; P(x, A_j, \theta)}$$

(resp. $G_t(\theta, \omega) = - \chi_t^2(\theta, \omega)$

$$= - \sum_{(i,j) \in I^2} \frac{\left[N_t(A_i \times A_j, \omega) - \int_{A_i \times \mathcal{X}} P(x, A_j, \theta) N_t(dx dy, \omega)\right]^2}{\int_{A_i \times \mathcal{X}} P(x, A_j, \theta) N_t(dx dy, \omega)} \Bigg)$$

où $(A_i)_{i \in I}$ est une partition $\mathcal{B}$-mesurable de $\mathcal{X}$ telle que les dénominateurs de G_t ne soient pas nuls, et où $N_t(A_i \times A_j, \omega) = \sum_{s=0}^{t-1} 1_{A_i \times A_j}(X_s(\omega), X_{s+1}(\omega))$ désigne le nombre de passages directs de A_i à A_j relativement à $(X_s(\omega))_{0 \leqslant s \leqslant t}$, (auquel cas, II.1 entraîne que $c_t \geqslant 1$). On obtient alors les estimateurs dits du ψ^2-quasi-minimum (resp. les estimateurs dits du χ^2-quasi-minimum) de coefficients multiplicatifs) c_t. Ces estimateurs ont été étudiés sous la forme équivalente

$$\psi_t^2(\theta_t^*(\omega), \omega) \leqslant c_t \inf_{\theta \in \Theta} \psi_t^2(\theta, \omega)$$

(resp. $\chi_t^2(\theta_t^*(\omega), \omega) \leqslant c_t \inf_{\theta \in \Theta} \chi_t^2(\theta, \omega)$), dans notre article [1].

b) Prenons pour G_t la fonction de vraisemblance

$$G_t(\theta, \omega) = exp\, L_t(\theta, \omega)$$

où, sous des conditions adéquates et en négligeant un premier terme, $L_t(\theta, \omega) = \int_{\mathcal{X}^2} Log\ p(x, y, \theta) N_t(dxdy, \omega)$, p désignant la densité de la probabilité de transition P, (c_t étant $\leqslant 1$ d'après II. 1). Nous avons alors les estimateurs du quasi-maximum de vraisemblance de coefficients (additifs) $\varepsilon_t = - Log\ c_t$, présentés sous la forme équivalente

$$L_t(\theta_t^*(\omega), \omega) \geqslant \sup_{\theta \in \Theta} L_t(\theta, \omega) - \varepsilon_t.$$

Une étude de ce cas a été faite dans [7] sans que le problème d'existence soit abordé. Nous avons fait une étude plus complète dans [1].

IV.3—Cas markovien à temps continu : Considérons le cas où $T = \mathbf{R}_+$ et où $(X_t)_{t \in \mathbf{R}_+}$ est une fonction aléatoire markovienne stationnaire de type de sauts, dont l'espace des états $\mathcal{X}$ est un sous-ensemble de $\mathbf{R}^h$, muni de sa tribu borélienne $\mathcal{B}$, avec une probabilité absolue stationnaire unique $\pi(., \theta)$ prise comme probabilité initiale et dont les probabilités de transition $P_t(.,., \theta)$ vérifient la condition $\lim_{t \to 0} P_t(x, \{x\}, \theta) = 1$ pour tout θ et uniformément en x. Nous noterons $\tau_t(B, \omega)$ la durée de séjour dans $B \in \mathcal{B}$, et $N_t(B \times B', \omega)$ le nombre de passages directs de B dans $B' \in \mathcal{B}$, relativement à un échantillon $(X_s(\omega))_{0 \leqslant s \leqslant t}$; $Q(x, \theta) = \lim_{t \to 0} \frac{1}{t} [1 - P_t(x, \{x\}, \theta)]$ et $Q(x, B, \theta) = \lim_{t \to 0} \frac{1}{t} P_t(x, B, \theta)$, $x \notin B$.

a) Prenons pour G_t la fonction négative

$$G_t(\theta, \omega) = - \psi_t^2(\theta, \omega)$$

$$= - \sum_{(i, j) \in I^2} \frac{\left[N_t(A_i \times A_j, \omega) - t \int_{A_i} \pi(dx, \theta) Q(x, A_j, \theta)\right]^2}{t \int_{A_i} \pi(dx, \theta) Q(x, A_j, \theta)}$$

(resp. $G_t(\theta, \omega) = - \chi_t^2(\theta, \omega)$

$$= - \sum_{(i, j) \in I^2} \frac{\left[N_t(A_i \times A_j, \omega) - \int_{A_i \times \mathcal{X}} Q(x, A_j, \theta) \tau_t(dx, \omega)\right]^2}{\int_{A_i \times \mathcal{X}} Q(x, A_j, \theta) \tau_t(dx, \omega)},$$

où $(A_i)_{i \in I}$ est une partition finie $\mathcal{B}$-mesurable de $\mathcal{X}$ telle que les dénominateurs de G_t ne soient pas nuls, c_t étant $\geqslant 1$ d'après II.1. Nous avons alors les estimateurs dits du ψ^2-quasi-minimum (resp. χ^2-quasi-minimum) de coefficients (multiplicatifs) c_t, étudiés dans [2] et [8] sous la forme équivalente

$$\psi_t^2(\theta_t^*(\omega), \omega) \leqslant c_t \inf_{\theta \in \Theta} \psi_t^2(\theta, \omega)$$

(resp. $\chi^2_t(\theta^*_t(\omega), \omega) \leqslant c_t \inf_{\theta \in \Theta} \chi^2_t(\theta, \omega))$.

b) En prenant pour G_t la fonction de vraisemblance définie par [9], c'est à dire en prenant

$$G_t(\theta, \omega) = exp\ L_t(\theta, \omega),$$

où (sous des conditions adéquates et en négligeant un premier terme)

$$L_t(\theta, \omega) = -\int_{\mathcal{X}} Q(x, \theta)\, \tau_t(dx, \omega) + \int_{\mathcal{X}^2} \mathrm{Log}\ q(x, y, \theta)\, N_t(dxdy, \omega).$$

$q(x, y, \theta)$ désignant la densité de $Q(x, ., \theta)$ par rapport à une mesure, (c_t étant alors $\leqslant 1$ d'après II.I), nous avons alors les estimateurs du quasi-maximum de vraisemblance de coefficients (additifs) $\varepsilon_t = -Log\ c_t$, étudiés dans [9] et [10].

IV. 4 — D'autres cas de variables aléatoires dépendantes:

a) Le cas des systèmes aléatoires généralisés à liaisons complètes est étudié dans [3] et [12], avec le temps discret, sans que l'homogénéité soit requise. C'est le cas où $T = \mathbf{N}$ et où $(X_t)_{t \in \mathbf{N}} = ((U_t, Y_{t+1}))_{t \in \mathbf{N}}$ est une fonction aléatoire attachée à un s. a. g. l. c. non homogène $(\hat{W}_t, \mathcal{W}_t), (\mathcal{Y}_{t+1}, \mathcal{C}_{t+1}), {}^t\pi(., ., \theta), {}^tP(., ., \theta))_{t \in \mathbf{N}}$ dont les probabilités de transition ${}^t\pi$ et ${}^tP$ dépendent du paramètre θ, et dont les espaces $\hat{W}_t$ et $\mathcal{Y}_{t+1}$ sont des ensembles finis ou dénombrables. On utilise dans ce cas un échantillon $\left(((U^{(k)}_t(\omega), Y^{(k)}_{t+1}(\omega)))_{0 \leqslant t \leqslant n}\right)_{1 \leqslant k \leqslant h}$ provenant de h fonctions aléatoires, toutes attachées au s. a. g. l. c. telles que les sous-tribus de $\mathcal{A}$ qu'elles engendrent sont Pr-indépendantes (cf. [11]). Deux variantes de la méthode du χ^2-quasi-minimum sont données dans [12], et la méthode du quasi-maximum de vraisemblance est étudiée de [3].

b) L'article [13] étudie la méthode du quasi-maximum de vraisemblance dans le cas d'une suite de variables aléatoires dépendantes $(X_t)_{t \in \mathbf{N}}$ sur laquelle certaine condition sur le coefficient de dépendance d'Ibragimov est imposée et telle que quel que soit $n \in \mathbf{N}$, $(X_0, X_1, \dots, X_n)$ possède une densité de probabilité f_n de la forme

$$f_n(x_0, \dots, x_n, \theta) = f_0(x, \theta)\, g_n(x_0, \dots, x_n) \,.\, exp \sum_{i=0}^{n-1} Z(x_i, x_{i+1}, \theta).$$

V — REMARQUE SUR LA CONVERGENCE — Dans tous les cas cités en IV, outre l'existence, la convergence au sens presque-sûr des estimateurs est démontrée. La démonstration varie selon les cas et des conditions qui leur sont imposées. Néanmoins, on peut dire, grosso modo, que pour les estimateurs du ψ^2-quasi-minimum et du χ^2-quasi-minimum, le principe est de prouver que quel soit le voisinage V de θ_0 et pour presque tout $\omega \in \Omega$, t grand et $\theta^*_t(\omega) \in CV$ entraînent que $G_t(\theta^*_t(\omega), \omega) < c_t\, G(\theta_0, \omega)$. Il y a alors contradiction avec la définition de l'estimateur quasi-optimal θ^*_t, et par conséquent, on doit avoir $\theta\ (\omega) \in V$. La démonstration pour le cas des estimateurs du quasi-maximum de vraisemblance est plus délicate, (surtout pour des Markov à temps continu où la définition de la vraisemblance est elle-même un problème délicat), et utilise la forme quasi-optimale avec des coefficients additifs, c'est à dire avec $L_t(\theta, \omega)$ et non pas avec $G_t(\theta, \omega)$. Les détails se trouvent dans les travaux cités en références.

Received March 15, 1978

REFERENCES

[1] BUI-TRONG-LIEU et PHAN-THANH-LONG. — Sur l'existence et sur la convergence d'estimateurs concernant des processus de Markov, Acta Scientiarum Vietnamicarum VI, 1970, pp. 67 — 84.

[2] BUI-TRONG-LIEU et LANGRAND, C.— Deux méthodes d'estimation pour des processus de Markov à temps continu, Comptes Rendus Acad. Sci. Paris 270, 1970, A, pp. 722—724.

[3] BUI-TRONG-LIEU et THEODORESCU R. — Estimateurs du quasi-maximum de vraisemblance pour des s. a. g. l. c., Comptes Rendus Acad. Sci. Paris 278, 1974, A, pp. 901—903

[4] KURATOWSKI, K. et RYLL-NARDZEWSKI, C.— A general theorem on selectors, Bull. Acad. Pol, Sciences, série Math., XIII, 1965, pp. 397 — 403.

[5] HIMMELBERG, C. J. et VAN VLECK, F. S.— Some selection theorems for measurable functions, Canadian J, of Math. XXI, 1969, pp. 394 — 399.

[6] SAINTE-BEUVE, M. F. — On the extension of Von Neumann—Aumann's theorem, J Functional Analysis, 17, N°1, 1974, pp. 112 — 129.

[7] ROUSSAS, G.G. — Extension to Markov proce ses of a result of A. Wald about consistency of the maximum likelihood estimate, Z. Wahrscheinlichkeitstheorie, 4, 1965, pp. 69 — 73.

[8] ANDRIEU, C., BUI-TRONG-LIEU et LANGRAND, C.—Sélecteurs, et estimateurs du ψ^2-quasi-minimum et du χ^2-quasi-minimum pour des processus de Markov à temps continu, Revue R. de Math. pures et appli. XVII. N°10. 1972, pp. 1497 — 1511.

[9] ANDRIEU C.— La méthode du quasi-maximum de vraisemblance concernant des proccessus de Markov à temps continu, Comptes Rendus Acad. Sci Paris 272, 1971, A, pp. 334 —336.

[10] ANDRIEU C, et LANGRAND, C,—Estimateurs du quasi-maximum de vraisemblance pour des processus de Markov à temps continu, Revue R. de Math. pures et appli., XVIII, N°7, 1973, pp. 997 — 1012.

[11] BUI-TRONG-LIEU et FLAVIGNY, R.—Fonctions aléatoires attachées à un s. a. g. l. c. et application à un problème d'estimation, Comptes Rendus Acad. Sci. Paris 278, 1974, A, pp. 33 — 35.

[12] ANDRIEU C, et FLAVIGNY R.— Une méthode de moindres carrés pour les s. a. g. l. c,, Comptes Rendus Acad. Sci. Paris 278, 1974, A, pp. 965 — 967.

[13] FLAVIGNY R, — Un problème de Statistique relatif au paramètre dont dépend la loi d'evolution d'une suite de variables aléatoires dépendantes, Comptes Rendus Acad. Sci. Paris 281, 1975, A, pp. 921 — 924.