PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 12, Fasc. 1 (1991), pp. 67–83

MAXIMUM D'ENTROPIE ET PROBLÈME DES MOMENTS: CAS MULTIDIMENSIONNEL

PAR

F. GAMBOA ET E. GASSIAT (ORSAY)

Abstract. We study the maximum entropy method on the mean developed in [4] to reconstruct a density under linear and non-linear convex constraints. We extend the procedure to the case of a locally convex topological linear space under large deviation assumptions. We give various applications: the most important gives a solution to the generalized moment problem on a polytope.

1. INTRODUCTION

Le problème de reconstruction d'une densité à partir d'informations sur les moments est classique, la méthode dite du *maximum d'entropie* (ME) permet d'en donner une solution explicite simple. Rappelons les notions et le critère employés: Soit $\mathcal{M}$ l'ensemble des mesures positives sur R^n, et $\mathcal{P}$ l'ensemble des probabilités sur R^n. Pour P et Q éléments de $\mathcal{P}$, l'information de Kullbak de P par rapport à Q est donnée par

$$K(P, Q) = \begin{cases} \int \log \frac{dP}{dQ} dP & \text{si } P \ll Q \text{ et } \log \frac{dP}{dQ} \in L^1(P), \\ +\infty & \text{sinon.} \end{cases}$$

Pour μ élément de $\mathcal{M}$ et P élément de $\mathcal{P}$, la μ-entropie de P est définie par

$$S_\mu(P) = \begin{cases} -\int \log \frac{dP}{d\mu} dP & \text{si } P \ll \mu \text{ et } \log \frac{dP}{d\mu} \in L^1(P), \\ -\infty & \text{sinon.} \end{cases}$$

Si $\varphi(X)$ est la variable aléatoire, dont l'espérance est fixée, et si $\mathcal{A}$ désigne l'ensemble des lois de X possibles sous cette contrainte, le critère de choix ME(μ) est de choisir comme estimateur de cette loi la loi P qui maximise $S_\mu(P)$ lorsque cela est possible (ou, de manière équivalente, si μ est une probabilité Q, qui minimise $K(P, Q)$). Une condition suffisante à l'existence et l'unicité d'un tel estimateur est l'existence, dans $\mathscr{A}$, d'une probabilité P_0 équivalente à μ (i.e. qui charge les mêmes événements), voir le travail de Csiszár [3]. On peut alors noter

$$P^{\mathrm{ME}} = \operatorname{Arg}\operatorname*{Max}_{P\in\mathscr{A}} S_\mu(P)$$

ou encore, si μ est une probabilité Q,

$$P^{\mathrm{ME}} = \operatorname{Arg}\operatorname*{Min}_{P\in\mathscr{A}} K(P, Q).$$

Dans un précédent travail (voir [4]), Dacunha-Castelle et Gamboa ont montré que l'on peut introduire des contraintes convexes, à condition de discrétiser le problème et de le randomiser. Ils ont traité complètement le cas où l'on cherche une solution continue, la contrainte φ étant continue. Ces solutions sont interprétées comme des dérivées d'espérances de champs aléatoires, indexés par l'espace (discrétisé) de définition de la fonction, conduisant (par la randomisation induite) à une méthode dite du *maximum d'entropie en moyenne* (MEM).

Notre but est de généraliser la construction à des *espaces vectoriels topologiques localement convexes* (E.V.T.L.C.) (cf. [7]), mettant en évidence une condition faible de grandes déviations (condition sur la randomisation et la contrainte de moments). Nous appliquons alors les résultats aux espaces de fonctions bornées multidimensionnelles sur un espace probabilisé (on retrouve le cas des fonctions continues sur un compact à bornes fixes), en insistant sur les critères d'existence de solutions pour les problèmes de moments avec contraintes convexes, dont nous indiquerons une application à un problème de commande.

2. CAS GÉNÉRAL

Soit E un E.V.T.L.C. dont l'élément générique sera noté g, et E' son dual topologique. Soit $\mathscr{C}$ un convexe ouvert de E, et $\overline{\mathscr{C}}$ *sa fermeture faible.* Soit φ un élément fixé de $(E')^k$ (k est un entier fixé). Le problème que l'on se propose de résoudre est le suivant: trouver g dans E vérifiant les deux contraintes:

(C1) $g\in\mathscr{C}$ (ou C'1) $g\in\overline{\mathscr{C}}$;

(C2) $\langle\varphi, g\rangle = c^*$ (ou (C'2) $\langle\varphi, g\rangle\in Q$, Q étant un compact convexe de R^k donné; on appellera (C'2) *contrainte relaxée*).

2.1. Notations. On se donne une suite de projecteurs (T_n) linéaires continus et définis sur E. On pose $E_n = T_n(E)$, et l'on suppose que (E_n) est une suite (non nécessairement croissante) de sous-espaces de dimension finie et croissante. L'espace E_n est muni de la topologie trace. $\mathscr{C}_n$ est l'image par T_n du convexe $\mathscr{C}$, que l'on munit d'une mesure positive μ_n. On fait les hypothèses suivantes:

Pour tout n, $\mathscr{C}_n$ est inclus dans $\mathscr{C}$ (et donc $\mathscr{C}_n = E_n \cap \mathscr{C}$).
φ est de rang k au sens suivant:

$$\alpha \in R^k, \ \forall g \in E, \ \langle(\alpha, \varphi), g\rangle = 0 \Rightarrow \alpha = 0.$$

$(\, , \,)$ désigne le produit scalaire usuel sur R^k, et $\langle \, , \, \rangle$ la dualité E', E. Plus généralement on notera aussi $\langle \, , \, \rangle$ le crochet de dualité $(E')^k$, E.

On pose, pour θ de E',

$$\psi_n^\theta(t, u) = (a_n)^{-1} \log \int_{\mathscr{C}_n} \exp\big(a_n[t\langle\theta, g\rangle + (u, \langle\varphi, g\rangle)]\big) d\mu_n(g),$$

où $t \in R$, $u \in R^k$, et (a_n) est une suite donnée de réels positifs de limite ∞ (liée, nous le verrons dans les applications, à la discrétisation de l'espace).

On appelle $D_n(\theta)$ le *domaine de* ψ_n^θ, c'est-à-dire

$$D_n(\theta) = \{(t, u) \in R^{k+1} \colon \psi_n^\theta(t, u) < \infty\}.$$

D_n quant à lui est le domaine de la fonction $\psi_n^0(0, \cdot)$ que l'on notera simplement $\psi_n(u)$,

$$D_n = \{u \in R^k \colon \psi_n(u) < \infty\}.$$

On pose ensuite $\psi_\infty^\theta(t, u) = \lim_{n\to\infty} \psi_n^\theta(t, u)$ lorsque cette limite existe, et l'on appelle $D_\infty(\theta)$ (resp. D_∞) le *domaine de* ψ_∞^θ (resp. de $\psi_\infty(\cdot)$).

On note pour $u \in D_n$ et $c \in R^k$:

$$H_n(u, c) = \psi_n(u) - (u, c), \qquad H_n^1(u, Q) = \psi_n(u) - \inf_{c \in Q} (u, c)$$

et pour $u \in D_\infty$, $c \in R^k$:

$$H(u, c) = \psi_\infty(u) - (u, c), \qquad H^1(u, Q) = \psi_\infty(u) - \inf_{c \in Q} (u, c).$$

On convient que ces fonctions valent $+\infty$ à l'extérieur de leur domaine de définition.

On note enfin $Z_n(u) = \exp\big(a_n \cdot \psi_n(u)\big)$.

2.2. Résolution du problème projeté. La méthode adoptée est la suivante:

On remplace le problème général par le problème de dimension finie associé:

Trouver g_n de $\mathscr{C}_n$ qui vérifie la contrainte (C2) (ou (C′2)).

Pour résoudre ce problème de dimension finie, on adopte un critère MEM, c'est-à-dire que l'on cherche la probabilité μ_n^{ME} de μ_n-entropie maximale, dont l'espérance $g_n^{\mathrm{ME}} = \int_{\mathscr{C}_n} g \, d\mu_n^{\mathrm{ME}}(g)$ vérifie la contrainte de moments (C2) (ou (C′2)). Cette espérance est alors solution du problème de dimension finie énoncé ci-dessus. On cherche ensuite à montrer que la suite des solutions g_n ainsi obtenues converge (au sens fort ou faible) vers une solution du problème initial.

Donnons maintenant une définition.

Définition 1. Soit Q un sous-ensemble de R^k; on dit que Q est *φ réalisable* (resp. *strictement réalisable*) dans $\mathscr{C}$ s'il existe un élément f de $\overline{\mathscr{C}}$ (resp. $\mathscr{C}$) qui vérifie $\langle\varphi, f\rangle \in Q$. Dans le cas d'un singleton $\{c\}$ on dira que c est *φ réalisable.*

Nous résolvons d'abord le problème de dimension finie, et nous intéressons donc à la réalisation des contraintes linéaires sur la suite des convexes $(\mathscr{C}_n)$; c'est l'objet du lemme 1.

Lemme 1. *On suppose que les hypothèses suivantes sont vérifiées*:

(H1) $\forall g \in E$, $\lim_{n\to\infty} \langle\varphi, T_n g\rangle = \langle\varphi, g\rangle$.

(H2) *c^* est φ strictement réalisable dans $\mathscr{C}$.*

Alors, pour n assez grand, c^* est φ strictement réalisable dans $\mathscr{C}_n$.

Démonstration. Soit g un élément de $\mathscr{C}$ qui vérifie $\langle\varphi, g\rangle = c^*$, et soit $\mathscr{V}$ un voisinage convexe équilibré de g (cf. [7]) inclus dans $\mathscr{C}$. Notons $\mathscr{V}_n$ l'image par T_n de $\mathscr{V}$; alors $\mathscr{V}_n$ est un convexe inclus dans $\mathscr{C}_n$. Supposons que l'on ait

$$\forall n, \forall f \in \mathscr{C}_n, \langle\varphi, f\rangle \neq c^*.$$

L'ensemble $\{c \in R^k \colon c = \langle\varphi, f\rangle, f \in \mathscr{V}_n\}$ est convexe et ne contient pas le point c^*; il existe donc un hyperplan $\alpha_n = (\alpha_n^1, \ldots, \alpha_n^k)$, avec $\sum_{i=1}^k (\alpha_n^i)^2 = 1$, qui le sépare de c^*. On a donc, par exemple,

$$\forall \delta \in \mathscr{V}, (\alpha_n, c^*) \geqslant (\alpha_n, \langle\varphi, T_n\delta\rangle).$$

La suite (α_n) est relativement compacte sur la sphère unité de R^k. Soit α_∞ l'un de ses points d'adhérence; l'inégalité précédente et l'hypothèse (H1) entrainent que

$$\forall \delta \in \mathscr{V} - g, (\alpha_\infty, \langle\varphi, \delta\rangle) \leqslant 0.$$

Comme $\mathscr{V} - g$ est équilibré, on en déduit que

$$\forall \delta \in \mathscr{V} - g, (\alpha_\infty, \langle\varphi, \delta\rangle) = 0.$$

Pour tout f de E, il existe λ de R^* tel que $\lambda f \in \mathscr{V} - g$, et donc

$$\forall f \in E, (\alpha_\infty, \langle\varphi, f\rangle) = 0,$$

ce qui contredit l'hypothèse faite sur le rang de φ, et achève la démonstration.

Remarque. L'hypothèse faite sur le rang de φ n'est pas restrictive: si φ n'est pas de rang plein, on extrait de φ un système libre maximal, et on exprime les autres composantes de φ sur ce système.

Nous résolvons maintenant le problème projeté de dimension finie: c'est l'objet du lemme 2.

LEMME 2. *Dans le cas d'une contrainte d'égalité, sous les hypothèses* (H1), (H2), *et*

(H3) *l'intérieur de l'enveloppe convexe du support de* μ_n *est* $\mathscr{C}_n$,

(H4) D_n *est un ouvert non vide,*

le problème de maximisation de l'entropie projeté sur E_n *admet pour n assez grand l'unique solution* $d\mu_{u_n}$ *définie par*

$$d\mu_{u_n}(g) = Z_n^{-1}(u_n)\exp\{a_n(u_n, \langle\varphi, g\rangle)\}d\mu_n(g),$$

où u_n *est l'unique minimum de* $H_n(\cdot, c^*)$.

Posant $g_n^{\mathrm{ME}} = \int_{\mathscr{C}_n} g d\mu_{u_n}(g)$, *on a* $g_n^{\mathrm{ME}} \in \mathscr{C}_n$ *et* $\langle\varphi, g_n^{\mathrm{ME}}\rangle = c^*$.

COROLLAIRE 1. *Dans le cas d'une contrainte relaxée, sous les hypothèses* (H1), (H3), (H4), *et*

(H'2) Q *est* φ *strictement réalisable dans* $\mathscr{C}$,

le problème de maximisation de l'entropie projeté sur E_n *admet pour n assez grand l'unique solution* $d\mu_{u_n}$ *définie par*

$$d\mu_{u_n}(g) = Z_n^{-1}(u_n)\exp\{a_n(u_n, \langle\varphi, g\rangle)\}d\mu_n(g),$$

où u_n *est l'unique minimum de* $H_n^1(\cdot, Q)$.

Posant $g_n^{\mathrm{ME}} = \int_{\mathscr{C}_n} g d\mu_{u_n}(g)$, *on a* $g_n^{\mathrm{ME}} \in \mathscr{C}_n$ *et* $\langle\varphi, g_n^{\mathrm{ME}}\rangle \in Q$.

D é m o n s t r a t i o n s. Par le lemme 1, et pour n assez grand, si c est φ strictement réalisable dans $\mathscr{C}$, il l'est aussi dans $\mathscr{C}_n$. $\mathscr{C}_n$ est ouvert pour la topologie associée à une norme quelconque, ce qui nous permet de raisonner sur R^n muni de la norme euclidienne habituelle.

On conclut à l'aide de lemmes concernant la résolution de problèmes MEM sur R^n: Soit μ une mesure positive sur R^n, et soit C l'intérieur de l'enveloppe convexe de son support.

On note $D(\mu) = \{u \in R^n\colon Z_\mu(u) = \int \exp(u \cdot x) d\mu(x) < \infty\}$, où $\cdot$ désigne le produit scalaire sur R^n.

$A = (a_{ij})_{i=1,\dots,k; j=1,\dots,n}$ est une matrice donnée, et Q un convexe compact de R^k.

On s'intéresse à la résolution du problème suivant (MEM):

Maximiser $S_\mu(P)$ sur $E(Q)$, où

$$E(Q) = \{P \in \mathscr{P}\colon A \cdot E_P(X) = A \int x dP(x) \in Q\}.$$

Notons maintenant $\Sigma(\mu) = \{E_P(X)\colon P \in \mathscr{P}, P$ équivalente à μ et $S_\mu(P) < \infty\}$.

On a alors le résultat suivant (voir [1]):

LEMME 3. *Si* $D(\mu)$ *est ouvert non vide, alors* $\Sigma(\mu) = C$.

On résout tout d'abord le problème (MEM) lorsque Q est réduit à un point c.

LEMME 4. *On suppose que* $({}^tA)^{-1}D(\mu)$ *est ouvert non vide. Alors pour* y *élément de* C, *le problème* (MEM) *sur* $E(Ay)$ *possède l'unique solution* P_u, *avec*

$$\frac{dP_u}{d\mu}(x) = \frac{\exp^t Au \cdot x}{Z_\mu({}^tAu)},$$

où u *est l'unique minimum de la fonction* $H(u, c)$:

$$H(u, c) = \log Z_\mu({}^tAu) - \langle u, c\rangle \quad \text{pour } c = Ay.$$

Démonstration. Posons $c = Ay$. Soit $u_0 \in ({}^tA)^{-1}D(\mu)$. On voit facilement que pour tout élément P de $E(c)$ on a $K(P, P_{u_0}) = -S_\mu(P) + H(u_0, c)$ et la maximisation de l'entropie sur $E(c)$ est donc équivalente à la minimisation de $K(P, P_{u_0})$ sur $E(c)$. Par le lemme 3 il existe une probabilité équivalente à μ et élément de $E(c)$. On peut alors appliquer le résultat de Csiszár [3], et il existe u tel que

$$K(P, P_u) = \min_{P \in E(c)} K(P, P_u) = \min_v H(v, c) = H(u, c).$$

Comme $({}^tA)^{-1}D(\mu)$ est ouvert non vide, $\operatorname{Grad} H(u, c) = 0$, et $P_u \in E(c)$.

Passons maintenant au cas où Q n'est pas réduit à un point.

LEMME 5. *On suppose que* $({}^tA)^{-1}D(\mu)$ *est ouvert non vide. Alors si* $A^{-1}Q$ *a une intersection non vide avec* C, *le problème* (MEM) *sur* $E(Q)$ *possède l'unique solution* P_u *avec*

$$\frac{dP_u}{d\mu}(x) = \frac{\exp^t Au \cdot x}{Z_\mu({}^tAu)},$$

où u *est l'unique minimum de la fonction* $H^1(u, c)$:

$$H^1(u, c) = \log Z_\mu({}^tAu) - \sup_{c \in Q} \langle u, c\rangle.$$

Démonstration. La fonction $c \to \inf_u H(u, c)$ est strictement concave et pour domaine $D = \{c\colon A^{-1}c \cap C \neq \emptyset\}$ elle est s.c.s. On en déduit qu'il existe c^* élément de $D \cap Q$ et u^* tels que

$$S_\mu(P_{u^*}) = \sup_{P \in E(K)} S_\mu(P) = \sup_{c \in K} \inf H(u, c) = H(u^*, c^*).$$

Le théorème de permutation de minimax [2] donne le résultat.

2.3. Convergence des solutions du problème projeté. Nous nous intéressons maintenant à la convergence de la suite (u_n).

THÉORÈME 1. *Dans le cas d'une contrainte d'égalité, sous les hypothèses* (H1), (H2), (H3), *et*

(H4) *pour tout* n, $D_n = D_\infty$, *et est un ouvert non vide,*

(H5) H *est strictement convexe en sa variable* u, *et il existe un voisinage*

V de c^* et une fonction L définie et finie sur V qui vérifient

$$\forall u \in D_\infty,\ \forall c \in V,\ H(u, c) \geqslant L(c),$$

la suite (u_n) converge vers une limite u_∞ dans R^k.

COROLLAIRE 2. *Dans le cas d'une contrainte relaxée, sous les hypothèses* (H1), (H′2), (H3), (H′4), *et*

(H′5) *H^1 est strictement convexe en sa variable u, et il existe un élément c^* de Q, φ strictement réalisable dans $\mathscr{C}$, un voisinage V de c^* et une fonction L définie et finie sur $\bar{V}$ qui vérifient*

$$\forall u \in D_\infty,\ \forall c \in V,\ H^1(u, c) \geqslant L(c),$$

la suite (u_n) converge vers une limite u_∞ dans R^k.

Démonstration du théorème. Posons $K = \{\langle \varphi, g \rangle, g \in \mathscr{C}\}$, K est un convexe de R^k comme l'image linéaire d'un convexe. On va montrer que K est un ouvert de R^k. Pour cela montrons d'abord que la restriction de l'application $\langle \varphi, \cdot \rangle$ à E_n est pour n assez grand de rang plein. Supposons le contraire; alors il existe une suite de vecteurs (α_n) de la sphère unité de R^k telle que

$$\forall g \in E,\ (\alpha_n, \langle \varphi, T_n g \rangle) = 0.$$

Soit α_∞ un point d'adhérence de la suite (α_n); on a alors

$$\forall g \in E,\ (\alpha_\infty, \langle \varphi, g \rangle) = 0,$$

ce qui contredit l'hypothèse sur le rang de φ.

Posons maintenant $K_n = \{\langle \varphi, g \rangle, g \in \mathscr{C}_n\}$, K_n est un ouvert de R^k (théorème de l'image ouverte appliqué à E_n). Comme on a, d'après le lemme 1, $\bigcup_n K_n = K$, K est ouvert.

Montrons maintenant le théorème. u_n minimise la fonction strictement convexe $H_n(\cdot, c^*)$. La suite $H_n(\cdot, c^*)$ converge vers $H(\cdot, c^*)$ pour tous les points de R^k. Soit u' un point de R^k. Posons $c' = \operatorname{grad} \psi_n(u')$. Alors $H(\cdot, c')$ est minimum en u'. En utilisant l'hypothèse (H5) on conclut que $\lim_{\|u\| \to \infty} H(u, c^*) = \infty$ (voir [4]). Ceci entraine que H possède un unique minimum atteint en u_∞. La convergence de H_n vers H est uniforme sur tout compact inclus dans R^k (voir [5]); on en déduit que $\lim_{n \to \infty} u_n = u_\infty$.

Démonstration du corollaire. Soit $c^* \in Q \cap K$. On a, d'après ce qui précède, $\lim_{\|u\| \to \infty} H(u, c^*) = \infty$, et donc puisque $H^1(u, Q) \geqslant H(u, c^*)$, on a aussi

$$\lim_{\|u\| \to \infty} H^1(u, Q) = \infty.$$

On conclut ensuite de la même manière que dans le cas d'une contrainte d'égalité.

En général, dans les applications, ces deux résultats sont suffisants pour obtenir la convergence forte de la suite (g_n) vers une solution du problème initial, car on dispose, de par la structure de μ_n, d'une forme explicite pour g_n. C'est ce que nous verrons dans le paragraphe 3. (Voir aussi [4].) Nous allons cependant donner un résultat concernant la convergence de la suite (g_n) dans le cas général.

DÉFINITION 2. On dira que g est une *solution faible du problème initial* si $g \in \mathscr{C}$ et $\langle \varphi, g \rangle = c^*$ (ou $\in Q$).

DÉFINITION 3. Soit E un E.V.T.L.C. On dit que E est *faiblement séquentiellement complet* si la propriété suivante est vérifiée:
Si

$$\forall \theta \in E', \ \lim_{n\to\infty} \langle \theta, f_n \rangle = a(\theta), \qquad \text{avec } |a(\theta)| < \infty.$$

Alors

$$\exists f \in E, \ \forall \theta \in E', \ a(\theta) = \langle \theta, f \rangle.$$

THÉORÈME 2. *On suppose que E est faiblement séquentiellement complet. Sous les hypothèses* (H1), (H2), (H3), (H5), *et*

(H″4) $D_n(\theta) = D_\infty(\theta) = R^k$, *et* grad ψ_∞^θ *existe en tous points de l'hyperplan* $t = 0$,

la suite g_n^{ME} *converge faiblement vers une solution faible du problème initial.*

COROLLAIRE 3. *On suppose que E est faiblement séquentiellement complet. Sous les hypothèses* (H1), (H′2), (H3), (H″4) *et* (H′5), *la suite* g_n^{ME} *converge faiblement vers une solution faible du problème initial.*

Remarques. Lorsque E est un espace de Banach réflexif, le théorème de Banach Steinhaus donne la propriété de la définition précédente. Pour un espace de Banach, fermeture faible et forte d'un convexe coincident (voir [2]).

Démonstrations. Soit θ un élément de E'. Posons

$$f_n(t) = (a_n)^{-1} \log \int_{\mathscr{C}_n} \exp\{a_n t \langle \theta, g \rangle\} d\mu_{u_n},$$

où $d\mu_{u_n}$ a été défini dans le lemme 2. On a $f_n(t) = \psi_n^\theta(t, u_n) - \psi_n^\theta(0, u_n)$, et

$$\forall t \in R, \ \lim_{n\to\infty} = \psi_\infty^\theta(t, u_\infty) - \psi_\infty^\theta(0, u_\infty).$$

f_n est convexe sur R, différentiable en 0, et par hypothèse f l'est aussi. On a donc (voir [5]) $\lim_{n\to\infty} f_n'(0) = f'(0)$. Mais $f_n'(0) = \langle \theta, g_n \rangle$, et E est faiblement séquentiellement complet. Il existe donc un élément g_∞ de E tel que la suite (g_n) converge faiblement vers g_∞, et g_∞ est un élément de $\mathscr{C}$.

3. BANDE OUVERTE DE $(L^\infty(\Omega, \mathscr{A}, P))^r$

3.1. Enoncé du problème et résolution. Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé possédant la propriété de discrétisation suivante:

Il existe une famille d'événements $(A_n^m)_{n\geqslant 1, m=1,\ldots,b_n}$ de $\mathscr{A}$ qui vérifie les propriétés:

$A_n^m \cap A_n^{m'} = \emptyset$ pour $m \neq m'$.

Notant $\mathscr{A}_n$ la σ-algèbre engendrée par les événements A_n^m, $m = 1, \ldots, b_n$, on a

$$\forall g \in L^\infty(\Omega, \mathscr{A}, P), \ \lim_{n\to\infty} E_P(g/\mathscr{A}_n) = g \ \ P \text{ p.s.},$$

où $E_P(g/\mathscr{A}_n)$ désigne l'espérance conditionnelle de la variable aléatoire g sur la tribu $\mathscr{A}_n$ sous la probabilité P.

$a_n^{-1} = P(A_n^m)$, $m = 1, \ldots, b_n$.

L'espace considéré est $E = (L^\infty(\Omega, \mathscr{A}, P))^r$, que l'on munit de la norme

$$\|g\|_E = \sup_{j=1,\ldots,r} \|g_j\|_\infty .$$

Soit A un convexe ouvert non vide de R^r donné. Le convexe $\mathscr{C}$ est

$$\mathscr{C} = \{g \in E\colon \exists B, \ B \text{ fermé}, \ B \subset A, g \in B \ \ P \text{ p.s.}\}.$$

Soit φ un élément donné de E^k qui vérifie l'hypothèse de rang:

$$\alpha \in R^k, \ \forall g \in E, \ \alpha^t E_P(\varphi g) = 0 \Rightarrow \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0,$$

où α^t est le vecteur transposé de α. Remarquons que l'hypothèse précédente équivaut à

$$\alpha \in R^k, \ \alpha^t \varphi = 0 \ \ P \text{ p.s.} \Rightarrow \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0.$$

La contrainte linéaire est $\langle \varphi, g\rangle = E_P(\varphi g) = c^*$ ($E_P(\cdot)$ est l'espérance sous la probabilité P). On supposera implicitement par la suite que c^* est φ strictement réalisable dans $\mathscr{C}$.

L'opérateur T_n est le projecteur sur les variables $\mathscr{A}_n$ mesurables. Pour f appartenant à E on a donc

$$T_n g = E_P(g/\mathscr{A}_n) = a_n \sum_{m=1}^{b_n} 1_{A_n^m} E_P(g 1_{A_n^m}).$$

Cette dernière variable aléatoire converge P p.s. par hypothèse vers g. Par le théorème de Lebesgue, φ étant P p.s. bornée, $\langle \varphi, T_n g\rangle$ converge vers $\langle \varphi, g\rangle$. (On a l'hypothèse (H1).) De plus

$$E_P(\varphi T_n g) = E_P\{E_P(\varphi T_n g/\mathscr{A}_n)\} = E_P(T_n \varphi T_n g).$$

Dans la suite on assimilera $\mathscr{C}_n = T_n \mathscr{C}$ à A^n.

Mesure de référence sur $\mathscr{C}_n$. Soit F une mesure positive sur A. On pose pour t appartenant à R^r

$$\psi(t) = \log \int_A \exp(t \cdot y) dF(y),$$

où $\cdot$ désigne le produit scalaire usuel sur R^r. On note

$$D(F) = \{t \in R^k \colon \psi(t) < \infty\}.$$

Soit (L) les hypothèses:

(L1) L'intérieur de l'enveloppe convexe du support de F est A.

(L2) $D(F)$ est ouvert non vide.

(L3) Posant

$$D = \{u \in R^k \colon \exists B,\ B \text{ fermé},\ B \subset D(F),\ \varphi^t u \in B\ \ P \text{ p.s.}\},$$

et

$$D' = \{u \in R^k \colon E_P(|\psi(\varphi^t u)|) < \infty\},$$

D est non vide, et coincide avec D'.

Remarquons que sous l'hypothèse (L2) D est ouvert, et que sous les hypothèses (L2) et (L3) D' est ouvert non vide. La mesure de référence sur $\mathscr{C}_n$ est alors $\mu_n = F^{\otimes n}$

Posons

$$D_n = \{u \in R^k \colon (T_n \varphi)^t u \in D(F)\ \ P \text{ p.s.}\}.$$

Sous les hypothèses (L2) et (L3) D_n est ouvert non vide. Pour u de D_n on pose

$$\begin{aligned} Z_n(u) &= \log \frac{1}{a_n} \int_{\mathscr{C}_n} \exp\{a_n u^t \langle \varphi, g \rangle\} d\mu_n(g) \\ &= \frac{1}{a_n} \sum_{m=1}^{b_n} \psi\{a_n E_P(\varphi 1_{A_n^m})^t u\} = E_P(\psi\{(T_n \varphi)^t u\}). \end{aligned}$$

Les hypothèses (H1) et (H3) sont vérifiées. Pour appliquer le théorème 1, il reste à montrer que (H'4) et (H5) le sont aussi.

Pour étudier la convergence de Z_n nous avons besoin du lemme suivant:

Lemme 6. *On suppose que les hypothèses* (L2) *et* (L3) *sont vérifiées. Alors si u est un élément de D, il existe un compact B' inclus dans $D(F)$ tel que*

$$\forall n,\ \{T_n(\varphi(x))\}^t u \in B'\ \ P\ p.s.$$

Démonstration. Soit u un élément de D. Par définition il existe un ensemble fermé B inclus dans $D(F)$ tel que $\varphi^t u' \in B$, P p.s. Remarquons que puisque φ est borné P p.s. B peut-être choisi compact. Soit B' l'enveloppe convexe fermée de B, $D(F)$ étant convexe ouvert il contient B'.

Comme conséquence du lemme 6, pour u de D on a

$$\|\psi\{(T_n\varphi)^t u\}\|_\infty \leqslant \sup_{y\in B'} |\psi(y)|,$$

et donc

$$\lim_{n\to\infty} Z_n(u) = \lim_{n\to\infty} E_P\psi\{(T_n\varphi)^t u\} = E_P\psi(\varphi^t u).$$

Posons $H(u, c^*) = E_P\psi(\varphi^t u) - (u, c^*)$ si $u\in D$, et $H(u, c^*) = \infty$ sinon. On a alors par le théorème de Lebesgue

$$\operatorname{grad} H = E_P\{\varphi\psi'(\varphi^t u)\} - c^*, \qquad u\in D,$$

où l'on a posé $\psi'(t) = \operatorname{grad}\psi(t)$, $t\in D(F)$. On a alors par le lemme 2: $g_n^{\mathrm{ME}} = \psi'(\varphi^t u_n)$.

Reste à montrer que $H(u, c)$ est uniformément minorée (en sa variable u), pour c élément de $K = \{c\in R^k\colon c = E_P\{\varphi f\}, f\in\mathscr{C}\}$. Remarquons tout d'abord que K est ouvert par le théorème de l'image ouverte.

Posons pour y appartenant à A

$$(\psi')^{-1}(y) = t \text{ si } \psi'(t) = y, \quad \text{et} \quad \gamma(y) = -y\cdot(\psi')^{-1}(y) + \psi\{(\psi')^{-1}(y)\}.$$

On a pour n assez grand (voir [4])

$$\forall u,\ H_n(u, c^{(n)}) = E_P\{(T_n\varphi)^t u\} - (u, c^{(n)}) \geqslant E_P\gamma(T_n g)$$

avec $E_P(\varphi g) = c$, $g\in\mathscr{C}$, et $c^{(n)} = E_P(\varphi T_n g)$. Mais puisque g appartient à $\mathscr{C}$, il existe un compact B, inclus dans A dans lequel g prend P p.s. ses valeurs. Et donc

$$|\gamma(T_n g)| \leqslant \sup_{y\in \mathrm{env\,conv}(B)} |\gamma(y)|.$$

On a alors

$$H(u, c) \geqslant E_P\gamma(g) = \Gamma(g).$$

En appliquant le théorème 1, on obtient

THÉORÈME 3. *Sous les hypothèses* (L) *et* (H3), *la suite* (g_n^{ME}) *converge dans* E *vers* $g_\infty^{\mathrm{ME}} = \psi'(\varphi^t u_\infty)$, *où* u_∞ *est le minimum de* $H(\cdot, c^*)$. *De plus* g_∞^{ME} *maximise* $\Gamma(g)$ *dans* $\mathscr{C}$ *sous la contrainte linéaire* $E_P(\varphi g) = c^*$.

Remarque. Lorsque φ est continue, la solution g_∞ obtenue est continue, et l'on retrouve les résultats de [4].

3.2. Problème des moments sur un polytope. Soit A un polytope borné, convexe, ouvert, non vide de R^r, ce qui signifie que A est l'intérieur de l'enveloppe convexe d'un nombre fini de points qui ne se trouvent pas tous dans un même hyperplan. Pour c de R^k on se propose d'étudier la solvabilité du problème (P) suivant:

Trouver g de $(L^\infty(\Omega, \mathscr{A}, P))^r$ tel que

$$g \in A \; P \text{ p.s. (ou } g \in \bar{A} \; P \text{ p.s.)} \quad \text{et} \quad E_P\{\varphi g\} = c.$$

Notons $y_1, \ldots, y_m$ les sommets de A.

Le théorème qui suit donne une condition nécessaire et suffisante sur c, numériquement calculable, pour que le problème (P) ait une solution.

Posons

$$h^*(c) = \sup_{u \in R^k} \{(u, c) - h(u)\}$$

avec

$$h(u) = E_P\{\log(\sum_{j=1}^{m} \exp\{u^t \varphi y_j\})\}, \quad u \in R^k.$$

On a alors

THÉORÈME 4. *c est φ réalisable dans $\mathscr{C}$ si et seulement si $h^*(c) \leqslant 0$. Si c est φ strictement réalisable dans $\mathscr{C}$, alors $h^*(c) < 0$. c n'est pas φ réalisable dans $\mathscr{C}$ si et seulement si $h^*(c) = \infty$.*

COROLLAIRE 4. *On suppose que φ vérifie*

$$\forall \alpha \in R^k, \; \alpha \neq 0, \; P(\alpha^t \varphi(y_i - y_j) \neq 0) = 1, \; \forall i, j, \, i \neq j.$$

Alors c est φ strictement réalisable dans $\mathscr{C}$ si et seulement si $h^*(c) < 0$. c est φ réalisable dans $\mathscr{C}$ non φ strictement réalisable dans $\mathscr{C}$ si et seulement si $h^*(c) = 0$. Dans ce cas le problème (P) *possède une unique solution qui est de la forme*

$$g = \sum_{j=1}^{m} y_j 1_{\{\alpha^t \varphi y_j > \alpha^t \varphi y_i, i \neq j\}}.$$

Démonstration du théorème. Prenons comme mesure de référence $dF(y) = \sum_{i=1}^m \delta_{y_i}(y)$. Si c est φ strictement réalisable dans $\mathscr{C}$, le problème (MEM) admet une solution unique et $\sup_{u \in R^k}\{(u, c) - h(u)\}$ est atteint en u_∞. De plus

$$(u_\infty, c) - h(u_\infty) = - \sup_{g \in \mathscr{C}, E_P(\varphi g) = c} E_P \gamma(g)$$

avec, pour $y \in A$,

$$\gamma(y) = -\sup_{t \in R^l} \left\{(t \cdot y) - \log\{\sum_{j=1}^{m} \exp(t \cdot y_j)\}\right\} = -y \cdot \psi'^{-1}(y) + \psi\{\psi'^{-1}(y)\},$$

$$\psi(t) = \log\{\sum_{j=1}^{m} \exp(t \cdot y_j)\}, \quad \psi'(t) = \operatorname{grad} \psi(t),$$

et

$$\psi'^{-1}(y) = t \quad \text{si } \psi'(t) = y.$$

Pour y de A faisons le changement de variable $t = \psi'^{-1}(y)$. On obtient alors

$$\begin{aligned}\gamma(y) &= -t\cdot\psi'(t)+\psi(t)\\ &= -\{t\cdot E_{P_t}(Y)\}+\sup_{i=1,\dots,m}(t\cdot y_i)+\log\Big[1+\sum_{j=1}^{m}\exp\{(t\cdot y_j)-\sup_{i=1,\dots,m}(t\cdot y_i)\}\Big],\end{aligned}$$

où P_t désigne la probabilité sur A définie par

$$dP_t(Y) = \frac{\sum_{j=1}^{m}\exp(t\cdot y_j)\delta_{y_j}(Y)}{\sum_{j=1}^{m}\exp(t\cdot y_j)},$$

et $Y = (Y_1, \dots, Y_l)$ désigne le vecteur coordonnées. Mais

$$t\cdot E_{P_t}(Y) - \sup_{i=1,\dots,m}(t, y_i) < 0,$$

car $E_{P_t}(Y)$ appartient à l'intérieur du polytope de sommets $y_1, \dots, y_m$.

On en déduit que, pour y de A, $\gamma(t) > 0$. D'où le résultat pour c φ strictement réalisable dans $\mathscr{C}$. Pour c φ réalisable non φ strictement réalisable dans $\mathscr{C}$ on a ([5], pp. 214 et 221):

$$h^*(c) = \lim_{n\to\infty} h^*(c_n),$$

(c_n) étant une suite de limite c, telle que, pour tout n, c_n soit φ strictement réalisable dans $\mathscr{C}$, et donc $h^*(c) \leqslant 0$. On obtient que $h^*(c) = \infty$ lorsque c n'est pas φ réalisable dans $\mathscr{C}$ à l'aide de résultats de grandes déviations suivants.

Soit $(W_n)_{n\geqslant 1}$ une suite de vecteurs aléatoires à valeur dans R^k. On pose pour t de R^k:

$$c_n(t) = \frac{1}{a_n}\log E_{P_n}[\exp(t, W_n)],$$

où (a_n) est une suite de réels positifs tendant vers l'infini. Soit Q_n la distribution de W_n/a_n. On a alors le théorème de grandes déviations [5]:

THÉORÈME 5. *On suppose que pour tout t de R^k, $c_n(t)$ est finie et $c_W(t) = \lim_{n\to\infty} c_n(t)$ existe, est finie, et est différentiable. On a alors*:

(a) $$\overline{\lim_{n\to\infty}}\,\frac{1}{a_n}\log Q_n(K) \leqslant -\inf_{z\in K} I_W(z)$$

pour tout fermé K de R^k;

(b) $$\underline{\lim_{n\to\infty}}\,\frac{1}{a_n}\log Q_n(G) \geqslant -\inf_{z\in G} I_W(z)$$

pour tout ouvert G de R^k avec

$$I_W(z) = \sup_{t \in R_k} [(t, z) - c_W(t)].$$

Lemme 7. *Soit c^* non φ réalisable dans $\mathscr{C}$. Alors il existe ε strictement positif tel que*

$$\forall c \in \bar{B}(c^*, \varepsilon),\ \forall n,\ \forall g \in \mathscr{C}_n,\ \langle \varphi, g \rangle \neq c.$$

Démonstration. Notons B_ε la boule fermée centrée en c^* et de rayon ε. Supposons que

$$\forall \varepsilon > 0,\ \varphi^{-1}(B_\varepsilon) \cap \mathscr{C} \neq \varnothing.$$

$\varphi^{-1}(B_\varepsilon)$ étant une suite décroissante de fermés lorsque ε tend vers 0, on a

$$\exists g \in \bigcap_\varepsilon \varphi^{-1}(B_\varepsilon) \cap \bar{\mathscr{C}}$$

et donc

$$\exists g \in \varphi^{-1}(\{c^*\}) \cap \bar{\mathscr{C}},$$

ce qui contredit l'hypothèse. On a donc: $\exists \varepsilon > 0$ tel que $\varphi^{-1}(B_\varepsilon) \cap \bar{\mathscr{C}} = \varnothing$ et l'on a le résultat puisque $\mathscr{C} \subset \bar{\mathscr{C}}$ pour tout n.

Corollaire 5. *On suppose que, pour tout n, $D_n(0) = D_\infty(0) = R^k$, que μ_n est une probabilité, et que $\psi_\infty(u)$ est différentiable sur R^k. Alors pour c^* non φ réalisable on a*

$$\inf_{u \in R^k} H(u, c^*) = -\infty.$$

Démonstration. Notons Q_n la loi de $\langle \varphi, g \rangle$ lorsque g a la loi μ_n. On applique le lemme 8 et le point (b) du théorème 5.

Démonstration du corollaire 4. Soit c φ réalisable non φ strictement réalisable dans $\mathscr{C}$, alors $\sup\{(u, c) - h(u)\}$ n'est pas atteint. Soit donc (u_n) une suite de R^k qui vérifie

$$\lim_{n \to \infty} \|u_n\| = \infty, \qquad \lim_{n \to \infty} \frac{u_n}{\|u_n\|} = \alpha,$$

$$h^*(c) = \lim_{n \to \infty} \{(u_n, c) - h(u_n)\}.$$

On a

$$(u_n, c) - h(u_n)$$

$$= (u_n, c) - E_P\{\sup_{i=1,\dots,m} (u_n^T \varphi y_i)\} - E_P\{\log \sum_{i=1}^m \exp[u_n^T \varphi y_j - \sup(u_n^T \varphi y_i)]\}.$$

Par le théorème de Lebesgue on a

$$\lim_{n\to\infty}\left\{\frac{1}{\|u_n\|}E_P[\sup_{i=1,\ldots,m}(u_n^T\varphi y_i)]\right\} = E_P\{\sup_{i=1,\ldots,m}(\alpha^t\varphi y_i)\},$$

$$\lim_{n\to\infty} E_P\{\log\sum_{j=1}^{m}\exp[u_n^T\varphi y_j - \sup_{i=1,\ldots,m}(u_n^T\varphi y_i)]\} = 0.$$

Comme $h^*(c)$ est fini, on en déduit que

$$(\alpha, c) - E_P\{\sup_{i=1,\ldots,m}(\alpha^T\varphi y_i)\} = 0.$$

Prenant ensuite la suite $u_n = n\alpha$ on obtient $h^*(c) = 0$.

Soit (c_n) une suite de R^k convergeant vers c, telle que pour tout n c_n soit φ strictement réalisable dans $\mathscr{C}$. Il existe alors une suite (u_n) de R^k telle que

$$c_n = E_P\left\{\varphi\frac{\sum_{i=1}^{m} y_i\exp[u_n^T\varphi y_i]}{\sum_{i=1}^{m}\exp[u_n^T\varphi y_i]}\right\},\qquad (u_n, c_n) - h(u_n) = h^*(c_n).$$

Soit α un point d'adhérence de la suite $u_n/\|u_n\|$. On obtient par le théorème de Lebesgue

$$c = E_P\{\varphi\sum_{i=1}^{m} y_i 1_{\{\alpha^T\varphi y_i > \alpha^T\varphi y_j, i\neq j\}}\}.$$

3.3. Contrôlabilité d'un système sous contraintes. Soit le système dynamique

$$X'(\sigma) = MX(\sigma) + Nu(\sigma),\ \sigma \geqslant 0,\qquad X(0) = X_0 \neq 0,$$

où $X(\sigma)$ est un vecteur de R^k, $u(\sigma)$ un vecteur de R^r, et N est une matrice de rang plein. Pour T réel positif donné, et A ouvert de R^r fixé, on se propose de trouver un critère répondant à l'existence d'une commande $u(\sigma)$, $\sigma\in[0, T]$, à valeurs dans A, et ramenant le système en 0 à la date T ($X(T) = 0$), le critère donnant en plus lorsqu'elle existe une commande qui réalise les deux contraintes précédentes. Pour cela réécrivons ce problème comme un problème de moment.

On a

$$X(t) = \exp(tM)X_0 + \exp(tM)\int_0^t\exp(-\sigma M)Nu(\sigma)d\sigma,$$

et l'on veut que $X(T) = 0$.

Le problème se met donc sous la forme:

Trouver $u(\sigma)$ qui vérifie la contrainte linéaire

$$\int_0^T\exp(-\sigma M)Nu(\sigma)d\sigma = -X_0$$

et la contrainte non linéaire

$$\forall \sigma \in [0, T], \; u(\sigma) \in A.$$

EXEMPLE 1. A est ici l'ensemble $]0, 1[^r$. Plus généralement par translation et changement d'échelle on peut considérer n'importe quel pavé de R^r. Choisissons comme mesure de référence $F = (\delta_0 + \delta_1)^{\otimes r}$. Il résulte alors de l'étude du problème des moments sur un polytope que le problème de commande possède une solution si et seulement si $h^*(-X_0) \leqslant 0$, où l'on a posé

$$h^*(c) = - \inf_{u \in R^k} H(u, c),$$

$$H(u, c) = \int_0^T \sum_{j=1}^k \log\{1 + \exp([N^t \exp(-M^t \sigma) u]_j)\} d\sigma - (u, c),$$

$[v]_j$ désignant la $j^{\text{ème}}$ composante du vecteur v, pour v vecteur de R^k.

Dans le cas où $h^*(-X_0) < 0$, il existe une commande du type

$$[u]_j(\sigma) = \frac{\exp\{[N^t \exp(-M^t \sigma) u_\infty]_j\}}{1 + \exp\{[N^t \exp(-M^t \sigma) u_\infty]_j\}},$$

qui répond au problème.

EXEMPLE 2. On considère dans cet exemple le cas où A est la boule unité de R^r. On choisit comme mesure F de référence la probabilité uniforme sur A, on a alors (voir les calculs dans [4])

$$\psi(t) = \log \int_{-1}^{1} \exp\{\|t\| y\} (1 - y^2)^{(n-1)/2} \frac{v_{n-1}}{v_n} dy,$$

où v_n désigne le volume de la boule unité de R^n.

Le problème de commande possède une solution si et seulement s'il en admet une de la forme $u(\sigma) = \psi'(N^t \exp(-M^t \sigma) u_\infty)$.

TRAVAUX CITÉS

[1] R. Azencott et G. Ruget, *Mélange d'équations différentielles et grands écarts à la loi des grands nombres*, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 38 (1977), pp. 1–54.
[2] H. Brezis, *Analyse fonctionnelle, théorie et applications*, Masson 1983.
[3] I. Csiszár, *Sanov property, generalized I-projection and a conditional limit theorem*, Ann. Probab. 12 (1984), pp. 768–793.
[4] D. Dacunha-Castelle et F. Gamboa, *Maximisation de l'entropie sous contraintes non linéaires*, Annales de l'I.H.P. 4 (1990), pp. 567–596.

[5] R. S. Ellis, *Entropy, Large Deviations and Statistical Mechanics*, Springer-Verlag, 1985.
[6] M. G. Krein and A. A. Nudelmam, *The Markov moment problem and extremal problems*, Transl. Math. Monographs 50 (1977).
[7] K. Yoshida, *Functional Analysis*, Springer-Verlag, 1978.

Laboratoire de Statistique Appliquée URA 743
Université Paris-Sud
Mathématique Bât. 425
91405 Orsay Cédex

Received on 6.6.1989