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

J.-P. GEORGIN

Contrôle des chaînes de Markov sur des espaces arbitraires

Annales de l'I. H. P., section B, tome 14, n° 3 (1978), p. 255-277

<http://www.numdam.org/item?id=AIHPB_1978__14_3_255_0>

Ann. Inst. Henri Poincaré, Section B :
Vol. XIV, n° 3, 1978, p. 255-277. *Calcul des Probabilités et Statistique.*

Contrôle des chaînes de Markov sur des espaces arbitraires

par

J.-P. GEORGIN

Université Paris-Nord, CSP Avenue J.-B. Clément
93430 Villetaneuse
(appartenant à l'équipe de recherche associée au C. N. R. S. n° 532,
« statistiques appliquées ».)

RÉSUMÉ. — Nous donnons un cadre d'hypothèses qui nous permet d'affirmer l'existence d'une stratégie stationnaire optimale pour une chaîne de Markov contrôlée dans le cas du critère dévalué. (L'espace des actions est polonais et l'espace des états est un espace mesurable quelconque, ou un espace polonais muni de sa tribu borélienne.)

L'étude du problème dévalué permet de trouver la solution au problème en moyenne, moyennant la résolution d'une équation de Poisson dont nous donnons des conditions d'existence.

Nous appliquons enfin ces résultats dans le cas où la chaîne contrôlée est une chaîne récurrente positive au sens de Harris.

INTRODUCTION

Nous avons repris dans [*4*] les principaux résultats de la théorie du contrôle, sur des espaces non dénombrables, basée sur les travaux de Striebel [*11*] et Hinderer [*5*], qui nous conduisent ici au rappel du théorème fondamental sur la caractérisation d'une stratégie optimale dans le cas du critère dévalué.

L'intérêt de l'étude du critère dévalué est qu'il nous permet de passer sous nos hypothèses à l'étude du critère en moyenne.

L'étude particulière que nous proposons est celle où l'espace d'état est un espace mesurable quelconque, alors que S. M. Ross dans [*10*] considérait un espace d'état polonais ; nous avons d'ailleurs dans [*4*] développé des conditions qui assurent la validité des hypothèses de S. M. Ross dans le cas d'espaces topologiques.

1. PRÉSENTATION DES OBJETS, NOTATIONS ET HYPOTHÈSES FONDAMENTALES

L'espace des temps est $\mathbb{N}$.

L'espace des épreuves est $(\Omega, \mathscr{A})$, un espace mesurable.

On étudie un système : processus défini sur $(\Omega, \mathscr{A})$ à valeurs dans un *espace mesurable des états* $(\mathscr{Y}, \mathscr{Y})$: on note $Y = (Y_t)_{t\in\mathbb{N}}$ ce processus.

A chaque instant t, on peut entreprendre une action A_t, fonction mesurable de $(\Omega, \mathscr{A})$ dans $(\mathscr{A}, \mathscr{A})$, $(\mathscr{A}, \mathscr{A})$ est l'espace mesurable des *actions*. Le processus $A = (A_t)_{t\in\mathbb{N}}$ est le processus des contrôles. On utilisera indifféremment les termes action ou contrôle.

On note pour toute suite $(a_t)_{t\in\mathbb{N}}$: $a^{(t)} = (a_0, \ldots, a_t)$.

La tribu des événements observables jusqu'à l'instant t est $\mathscr{B}_t$:

$$\mathscr{B}_t = \sigma(Y^{(t)}, A^{(t-1)})$$

On devra à chaque instant t choisir l'action A_t au vu des observations antérieures à t. Autrement dit, on choisira une *stratégie* $\delta = (\delta_t)_{t\in\mathbb{N}}$, c'est-à-dire un processus de $(\Omega, \mathscr{A})$ dans $(\mathscr{A}, \mathscr{A})$ adapté aux tribus $\mathscr{B}_t$. Seules seront acceptées certaines stratégies présentant des conditions de cohérence que l'on précisera. On notera $\mathscr{D}$ l'ensemble des stratégies cohérentes.

On se donne une probabilité de transition π de $\mathscr{Y} \times \mathscr{A}$ dans $\mathscr{Y}$ dont le sens intuitif est le suivant : si à l'instant t, $Y_t = y$, $A_t = a$; alors à l'instant $(t + 1)$, Y_{t+1} est dans Γ, $\Gamma \in \mathscr{Y}$, avec la probabilité $\pi(y, a, \Gamma)$. On donne enfin une probabilité λ_0 sur $(\mathscr{Y}, \mathscr{Y})$, la loi initiale du processus.

Les espaces produits seront munis des tribus produits associées. On ne mentionnera pas les tribus associées lorsqu'elles sont évidentes.

On se place dans le cadre suivant qui définit la notion de « chaîne de Markov contrôlée ».

Pour toute stratégie cohérente $d = (d_t)_{t\in\mathbb{N}}$, il existe une probabilité P^d sur $(\Omega, \mathscr{A})$ telle que pour tout t, et tout $B \in \mathscr{Y}$

$$P^d[A_t = d_t \text{ pour tout } t] = 1$$
$$P^d[Y_0 \in B] = \lambda_0(B)$$
$$P^d[Y_{t+1} \in B \mid \mathscr{B}_t] = \pi(Y_t, d_t, B)$$

Remarque. — La construction canonique d'un tel processus est possible (cf. [4]).

Définissons l'ensemble $\mathscr{D}$ des stratégies cohérentes :

A chaque instant t, le domaine où peuvent être choisies les actions ne dépend que de Y_t : à chaque $y \in \mathscr{Y}$, on associe $D(y)$ partie non vide de $\mathscr{A}$ tel que l'ensemble $\{(y, a) ; a \in D(y)\}$ soit élément de $\mathscr{Y} \otimes \mathscr{A}$.

Une stratégie $d = (d_t)_{t \in \mathbb{N}}$ est donc cohérente si et seulement si $d_t \in D(Y_t)$ pour tout t.

Le gain obtenu dans l'intervalle de temps $[t, t+1]$ est $\tilde{g}_t$, variable aléatoire réelle (v. a. r.) mesurable et bornée.

On suppose que pour toute $d \in \mathscr{D}$ et tout $t \in \mathbb{N}$

$$E^d[\tilde{g}_t \mid \mathscr{B}_t] = g(Y_t, d_t) = g(Y_t, A_t) \qquad [P^d \text{ p. s.}]$$

où g est une v. a. r. sur $\mathscr{Y} \times \mathscr{A}$.

Il existe une v. a. p sur $\mathscr{Y} \times \mathscr{A} \times \mathscr{Y}$ et une transition λ de $\mathscr{Y}$ dans $\mathscr{Y}$ telles que pour $y \in \mathscr{Y}$, $a \in \mathscr{A}$, $B \in \mathscr{Y}$

$$\pi(y, a ; B) = \int p(y, a, z)\lambda(y, dz)1_B(z)$$

Il existe alors une probabilité ρ sur $(\mathscr{Y}, \mathscr{Y})$ telle que :

$$\rho(B) = 0 \Leftrightarrow [P^d[Y_t \in B] = 0 \quad \text{pour toute } d \text{ et tout } t \geqslant 0]$$

car les mesures $Y_t(P^d)$ sont toutes absolument continues par rapport à :

$$\sum_{t \geqslant 0} \frac{1}{2^t} \lambda_0(\lambda_t) \qquad \text{où } \lambda_t \text{ est la } t\text{-ième itérée de } \lambda .$$

Notations : On notera :

$\mathscr{B}(\mathscr{Y})$ L'ensemble des v. a. bornées sur $\mathscr{Y}$ muni de la norme définie par :

$$\| \phi \| = \sup_x | \phi(x) |$$

$\mathscr{M}(\mathscr{Y})$ L'espace vectoriel des mesures signées sur $\mathscr{Y}$ muni de la norme

$$\| v \| = \sup_{\substack{\{\phi, \|\phi\| \leqslant 1\} \\ \phi \in \mathscr{B}(\mathscr{Y})}} \left| \int \phi dv \right|$$

2. SOLUTION DU PROBLÈME DÉVALUÉ ET INTRODUCTION AU PROBLÈME EN MOYENNE

Sous les conditions précédentes, introduisons les notions de « critère dévalué » et de « critère en moyenne ».

Définissons le gain maximin en moyenne

$$\underline{\mathscr{M}} = \varliminf_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} \tilde{g}_s$$

et le gain minimax en moyenne :

$$\bar{\mathscr{M}} = \varlimsup_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} \tilde{g}_s$$

Le problème que nous étudions est : à quelles conditions sur π et sur g existe-t-il une stratégie d_0 pour laquelle :

$$\lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} \tilde{g}_s = k \qquad \mathrm{P}^{d_0} \text{ p. s.}$$

et pour toute autre stratégie d :

$$\bar{\mathscr{M}} \leqslant k \qquad \mathrm{P}^{d} \text{ p. s.}$$

Un problème moins intéressant mais techniquement plus simple que celui posé précédemment est le cas du *problème dévalué :* on cherche à maximiser :

$$\mathrm{E}^d\left[\sum_{s=0}^{\infty} \beta^s \tilde{g}_s\right] \qquad \text{ou} \qquad \mathrm{E}^d\left[\sum_{s=0}^{\infty} \beta^s \tilde{g}_s \mid \mathrm{Y}_0\right]$$

où β, $0 < \beta < 1$, est le taux de dévaluation. Introduisons la définition :

Définition. — Une stratégie est dite *markovienne stationnaire* relativement à Y, si elle est de la forme :

$$\{ d(\mathrm{Y}_t) \}_{t\in\mathbb{N}}$$

où d est une fonction mesurable de $\mathscr{Y}$ dans $\mathscr{A}$ telle que $d(y) \in \mathrm{D}(y)$ pour tout $y \in \mathscr{Y}$.

On notera $\mathscr{S}$ *l'ensemble des stratégies markoviennes stationnaires* asso-

ciées à Y. Si $d \in \mathscr{S}$, $Y = (Y_s)_{s\in\mathbb{N}}$ est une *chaîne de Markov* sur $(\Omega, \mathscr{A}, P^d)$, dûment complété, de transition : $(y, B) \to \pi(y, d(y), B)$ notée Π^d.

On obtient alors le théorème que nous ne démontrons pas ici (cf. [4]).

THÉORÈME 1. — *Dans le cas « stationnaire dévalué ».*

1. *La classe $\mathscr{S}$ des stratégies markoviennes stationnaires associées à Y est complète. C'est-à-dire :*

$$\operatorname*{ess\,sup}_{d\in\mathscr{D}} E^d\left[\sum_{s=0}^{\infty} \beta^s \tilde{g}_s \mid Y_0\right] = \operatorname*{ess\,sup}_{d\in\mathscr{S}} E^d\left[\sum_{s=0}^{\infty} \beta^s \tilde{g}_s \mid Y_0\right] = l_\beta(Y_0)$$

2.

$$l_\beta(y) = Tl_\beta(y) = \rho - \operatorname*{ess\,sup}_{d\in\mathscr{S}} \left[g(y, d(y)) + \beta \int \pi(y, d(y), dz) l_\beta(z)\right]$$

et la valeur de l_β peut-être obtenue par approximation dans $L^\infty(\rho)$ à l'aide de la suite des $T^n h$ (n-ièmes itérées de T) pour $h \in L^\infty(\rho)$ quelconque.

3. *Si d_β est telle que :*

$$l_\beta(y) = g(y, d_\beta(y)) + \int \pi(y, d_\beta(y), dz) l_\beta(z) \qquad [\rho \text{ p. s.}]$$

alors d_β est une stratégie optimale :

$$l_\beta(Y_0) = E^{d_\beta}\left[\sum_{s\geqslant 0} \beta^s \tilde{g}_s \mid Y_0\right] \qquad [P^{d_\beta} \text{ p. s.}]$$

Quant au problème du critère en moyenne, nous recherchons la stratégie qui maximise l'espérance du gain moyen par unité de temps. Ce problème apparaît comme limite du précédent quand $\beta \to 1$.

Posons :

$$\underline{M}^d_{\lambda_0} = \varliminf_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} E^d_{\lambda_0}[\tilde{g}_s]$$

Rappelons que si (u_k) est une suite de nombres réels et $0 < \beta < 1$ on a :

$$\varliminf_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} u_k \leqslant \varliminf_{\beta\to 1_-} (1-\beta) \sum_{k=0}^{\infty} \beta^k u_k$$

et aussi si $\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} u_k$ existe, alors celle de $(1-\beta)\sum_{k=0}^{\infty} \beta^k u_k$ existe et ce sont les mêmes.

Donc :

$$\underline{M}^d_{\lambda_0} = \underline{\lim}\, \frac{1}{t} \sum_{s=0}^{\infty} E^d_{\lambda_0}[\tilde{g}_s] \leqslant \underline{\lim}_{\beta \to 1_-} (1-\beta) \sum_{s=0}^{\infty} E^d_{\lambda_0}[\beta^s \tilde{g}_s]$$

$$\leqslant \underline{\lim}_{\beta \to 1_-} (1-\beta) \sup_d E^d_{\lambda_0}\left[\sum_{s=0}^{\infty} \beta^s \tilde{g}_s\right]$$

ainsi :

$$\underline{M}^d_{\lambda_0} \leqslant \underline{\lim}_{\beta \to 1_-} (1-\beta) \int l_\beta d\lambda_0 = k(\lambda_0)$$

Ce résultat donne une idée du lien entre le critère en moyenne et le critère dévalué.

Une autre approche de ce critère est l'exemple suivant :

Exemple. — Chaîne de Markov non contrôlée où seul le gain dépend de l'action.

Cet exemple conduit à étudier le cas où Y *est une chaîne de Markov récurrente positive au sens de Harris non contrôlée.* Nous supposerons seulement que le gain *g dépend de l'action :* si à l'instant t, $Y_t = y$, $A_t = a$, on gagne $g(y, a)$.

On sait d'après un théorème ergodique que si μ est la probabilité invariante, alors pour toute $d \in \mathscr{S}$

$$\lim_{t \to \infty} \frac{1}{t} \sum_{s=0}^{t-1} g(Y_s, d(Y_s)) = \int \mu(dy) g(y, d(y)) = M^d \qquad \text{(p. s.)}$$

On notera :

$$g_d(\mu) = \int \mu(dy) g(y, d(y)) \qquad \text{et } g_d = g(., d(.))$$

Supposons qu'il existe une fonction $d_0 : \mathscr{Y} \to \mathscr{A}$, mesurable, telle que $d_0(x) \in D(x)$ et :

$$g(x, d_0(x)) = \sup_{a \in D(x)} g(x, a)$$

donc d_0 est une telle stratégie optimale. Alors :

$$M^{d_0} = k_{d_0} = \lim_{t \to \infty} \frac{1}{t} \sum_{s=0}^{t-1} g(Y_s, d_0(Y_s)) \qquad \text{(p. s.)}$$

et pour toute autre stratégie d

$$\overline{\lim}\, \frac{1}{t} \sum_{s=0}^{t-1} g(Y_s, d_s) \leqslant M^d \qquad \text{(p. s.)}$$

Si de plus, on a récurrence au sens de Doeblin [cf. Revuz, Orey [9] [8]] alors, comme $(g_{d_0} - k_{d_0})$ est d'intégrale nulle par rapport à μ, il existe une v. a. l, bornée, unique à une constante près, telle que l soit solution de l'équation de Poisson $l - \pi l = g_{d_0} - k_{d_0}$ ou

$$l(x) + k_{d_0} = g(x, d_0(x)) + \int \pi(x, dy) l(y) = \sup_a \left[g(x, a) + \int \pi(x, dy) l(y) \right]$$

on obtient donc, sous des conditions de récurrence très fortes, dans ce cas simple, des équations analogues à celles que nous avons obtenues dans le cas dévalué. Nous chercherons à exhiber des fonctions k et l analogues dans le paragraphe sur le critère en moyenne.

Le problème qui se pose est celui des sélections mesurables. A quelles conditions la borne essentielle supérieure pour $d \in \mathcal{S}$ peut-elle être remplacée par une borne supérieure pour $a \in \mathrm{D}(y)$ pour chaque y et existe-t-il une stratégie $d \in \mathcal{S}$ optimale ?

Envisageons maintenant des conditions qui répondent aux questions posées, que nous formulons dans les hypothèses suivantes :

HYPOTHÈSE 1. — *Conditions d'Evstigneev* (3)

$(\mathcal{Y}, \mathcal{Y})$ est un espace mesurable quelconque.

$\mathcal{A}$ est un espace polonais (métrique, complet, séparable) muni de la tribu borélienne $\mathcal{A}$.

Pour tout y $\mathrm{D}(y)$ est compact, et

$$\begin{cases} a \to g(y, a) & \text{est continue de } \mathrm{D}(y) \text{ dans } \mathrm{R} \\ a \to \int \pi(y, a, dz) h(z) & \text{est continue de } \mathrm{D}(y) \text{ dans } \mathrm{R} \end{cases}$$

quelle que soit la v. a. bornée h sur $\mathcal{Y}$.

On complète cette dernière hypothèse en prenant $\mathcal{Y}$ un espace polonais.

HYPOTHÈSE 2. — *Conditions d'Evstigneev et de Brown-Purves* (1)

Hypothèse 1
et $\mathcal{Y}$ est un espace polonais muni de la tribu borélienne $\mathcal{Y}$.

On a alors le théorème :

THÉORÈME 2. — *Dans le cadre de l'hypothèse 2* [resp. 1] *on peut choisir une version de l_β telle que :*

$$l_\beta(y) = \sup_{a \in \mathrm{D}(y)} \left[g(y, a) + \beta \int \pi(y, a, dz) l_\beta(z) \right]$$

et il existe une stratégie stationnaire markovienne optimale $d_\beta(.)$ mesurable sur $\mathscr{Y}$ telle que :

$$l_\beta(y) = g(y, d_\beta(y)) + \beta \int \pi(y, d_\beta(y); dz) l_\beta(z) \qquad [\text{resp. } \rho \text{ p. s.}]$$

Démonstration :

1) *Sous l'hypothèse 1*

Rappelons un théorème de sélection dû à Evstigneev [3] que nous transposons ici avec nos notations :

THÉORÈME DE EVSTIGNEEV. — *Soient $(\mathscr{Y}, \mathscr{Y})$ un espace mesurable quelconque muni d'une probabilité ρ, $\mathscr{A}$ un espace polonais muni de sa tribu borélienne $\mathscr{A}$.*

ϕ une v. a. r. sur $\mathscr{Y} \times \mathscr{A}$ telle que

$$\begin{cases} . \text{ pour tout } c, \text{ pour tout } y \in y, \{ a, \phi(y, a) \geqslant c \} \text{ est compact} \\ . \ \phi(y, a) \leqslant q(y) \qquad \text{où} \qquad \int | q | \, d\rho < \infty \end{cases}$$

Alors il existe $d : \mathscr{Y} \to \mathscr{A}$ mesurable telle que

$$\phi(y, d(y)) = \sup_{a \in \mathscr{A}} \phi(y, a) \qquad [\rho \text{ p. s.}]$$

Appliquons ce théorème à la fonction $\Phi(y, a)$

$$\Phi(y, a) = g(y, a) + \beta \int \pi(y, a, dz) l_\beta(z) \qquad \text{si} \quad a \in \mathrm{D}(y)$$
$$= -\infty \qquad \text{sinon}$$

Les conditions du théorème de Evstigneev sont vérifiées sous l'hypothèse 1, en effet :

— $a \to \Phi(y, a)$ est continue sur $\mathrm{D}(y)$, $a \to \Phi(y, a)$ est s. c. s. sur $\mathscr{A}$,

— $\Phi(y, a) \leqslant \| g \| \dfrac{1}{1 - \beta}$, pour tout c et pour tout y, l'ensemble $\{ a, \Phi(y, a) \geqslant c \}$ est fermé inclus dans le compact $\mathrm{D}(y)$, il est compact.

Ainsi il existe une $d_\beta \in \mathscr{S}$ telle que, d_β mesurable de $\mathscr{Y}$ dans $\mathscr{A}$ et :

$$\Phi(y, d_\beta(y)) = \sup_{a \in \mathrm{D}(y)} \Phi(y, a) \qquad [\rho \text{ p. s.}]$$

On avait :

$$l_\beta(y) = \rho - \operatorname*{ess\,sup}_{d\in\mathscr{S}} \Phi(y, d(y))$$

Pour toute $d \in \mathscr{S}$

$$\Phi(y, d(y)) \leqslant \sup_{a\in D(y)} \Phi(y, a)$$

mais ici :

$$\Phi(y, d_\beta(y)) = \sup_{a\in D(y)} \Phi(y, a) \qquad [\rho \text{ p. s.}]$$

Il en résulte que $\Phi(., d_\beta(.)) = \operatorname*{ess\,sup}_{d\in\mathscr{S}} \Phi(., d(.))$, en effet

$$\Phi(., d_\beta(.)) \leqslant \operatorname*{ess\,sup}_{d\in\mathscr{S}} \Phi(., d(.)) \qquad \text{car } d_\beta \in \mathscr{S}$$

et

$$\Phi(., d_\beta(.)) \geqslant \operatorname*{ess\,sup}_{d\in\mathscr{S}} \Phi(., d(.)) \qquad \text{car } \Phi(., d_\beta(.)) \geqslant \Phi(., d(.)) \qquad (\rho \text{ p. s.})$$

pour toute fonction de la famille.

On choisit désormais cette version de l_β et ainsi :

$$\begin{cases} l_\beta(y) = \sup_{a\in D(y)} \left[g(y, a) + \beta \int \pi(y, a, dz) l_\beta(z) \right] \\ \qquad = g(y, d_\beta(y)) + \beta \int \pi(y, d_\beta(y), dz) l_\beta(z) \qquad [\rho \text{ p. s.}] \end{cases}$$

2) *Sous l'hypothèse 2*

Le théorème de sélection est dû ici à Brown-Purves [*1*], que nous rappelons.

THÉORÈME DE BROWN-PURVES. — *Sous l'hypothèse 2, si ϕ est une v. a. r. sur $\mathscr{Y} \times \mathscr{A}$ telle que pour tout $y \in \mathscr{Y}$, l'application $a \to \phi(y, a)$ est continue de* D(y) *dans* R, *alors il existe une application $d : \mathscr{Y} \to \mathscr{A}$ mesurable telle que, pour tout y,*

$$\phi(y, d(y)) = \sup_{a\in\mathscr{A}} \phi(y, a)$$

Le raisonnement est identique au précédent et on obtient une égalité sûre de l_β et de $\Phi(., d_\beta(.))$.

Indépendance par rapport à la loi initiale

Il est intéressant de s'assurer de *l'indépendance des résultats par rapport à la loi initiale.*

Considérons l'opérateur T_β défini sur l'ensemble $\mathscr{B}(\mathscr{Y})$ des v. a. bornées sur $\mathscr{Y}$ en posant pour $f \in \mathscr{B}(\mathscr{Y})$

$$T_\beta f(y) = \sup_{a\in D(y)} \left[g(y, a) + \beta \int \pi(y, a, dz) f(z) \right]$$

T_β est une contraction dont l_β est l'unique point fixe, donc l_β est indépendante de la loi initiale sous les hypothèses 1 et 2.

Dans le cadre de l'hypothèse 2, d_β est aussi indépendante de la loi initiale.

Dans le cadre de l'hypothèse 1, formulons l'hypothèse la suivante.

HYPOTHÈSE 1 *a*. — *La transition* λ *est remplacée par une probabilité* λ : *pour* $(y, a, B) \in \mathcal{Y} \times \mathcal{A} \times \mathcal{Y}$

$$\pi(y, a, B) = \int p(y, a, z) 1_B(z) \lambda(dz)$$

et

. $p(y, a, .) \leqslant q(y, .)$ λ *intégrable*,

. *pour* $(y, z) \in \mathcal{Y}^2$, $a \to p(y, a, z)$ *est continue de* $D(y)$ *dans* R.

Conséquences

a) Si pour tout (y, z), la fonction $a \to p(y, a, z)$ est continue de $D(y)$ dans R, pour tout a_0 dans $D(y)$ on a :

$$\lim_{a \to a_0} \int | p(y, a, z) - p(y, a_0, z) | \lambda(dz) = 0$$

L'application $a \to \pi(y, a, .)$ de $D(y)$ dans $\mathcal{M}(\mathcal{Y})$ est continue, $D(y)$ étant compact la continuité est uniforme. La dernière condition de l'hypothèse 1 est automatiquement réalisée, et en outre :

b) Soit $\mathcal{H} \subset \mathcal{B}(\mathcal{Y})$ une famille de fonctions de $\mathcal{B}(\mathcal{Y})$ uniformément bornées, alors à y donné, la famille

$$\left\{ a \to \int p(y, a, z) h(z) \lambda(dz) \right\}_{h \in \mathcal{H}} \quad \text{est équicontinue.}$$

Pour toute mesure initiale λ_0, $\lambda_0 \ll \lambda$, on a $\rho \ll \lambda$. On peut appliquer le théorème d'Estigneev en remplaçant ρ par λ : on a λ p. s.

$$l_\beta(y) = g(y, d_\beta(y)) + \beta \int \pi(y, d_\beta(y) ; dz) l_\beta(z)$$

Donc d_β ne dépend pas de la loi initiale absolument continue par rapport à λ.

On notera N_β l'ensemble λ négligeable où la dernière relation n'est pas vraie.

Si $d \in \mathcal{D}$, on désignera par $P^d_{\lambda_0}$ la probabilité qui est désignée ci-dessus par P^d, si la mesure initiale est λ_0. Lorsque λ_0 est la mesure de Dirac au point $x \in \mathcal{Y}$, on notera P^d_x. Les espérances sont notées $E^d_{\lambda_0}$ ou E^d_x. *Lorsque*

les résultats énoncés seront indépendants de la loi initiale, on utilisera les notations P^d *et* E^d *sans préciser cette loi.*

Remarquons que sous l'hypothèse 2 l'existence de la fonction d_0, rencontrée lors de l'exemple, est assurée. D'après la formule obtenue dans le théorème 2, on a, pour un état 0 fixé, sous les hypothèses 1 et 2

$$(1 - \beta)l_\beta(0) + l_\beta(x) - l_\beta(0) = \sup_a \left[g(x, a) + \beta \int \pi(x, a, dy)(l_\beta(y) - l_\beta(0)) \right]$$

et on aimerait pouvoir *passer à la limite quand* $\beta \to 1$ afin de *retrouver les fonctions k et l de l'équation de Poisson* que nous avons rencontrées lors de l'exemple. Nous étudierons séparément les conditions de Evstigneev et les conditions de Brown-Evstigneev.

3. PROBLÈME EN MOYENNE

3.1. Conditions d'existence d'une équation de Poisson

Afin de simplifier l'exposé de ce qui suit, nous nous intéresserons aux deux hypothèses suivantes :

HYPOTHÈSE 1'. — *L'hypothèse 1 et l'hypothèse 1 a.*

. $\lambda_0 \ll \lambda$

. $L^1(\mathscr{Y})$ *est un espace de Banach séparable*

HYPOTHÈSE 2'. — *L'hypothèse 2 et l'hypothèse 1 a.*

3.1.1. Dans le cadre de l'hypothèse 1' :

On a

$$l_\beta(x) = \sup_{a \in D(x)} \left[g(x, a) + \beta \int \pi(x, a, dz) l_\beta(z) \right]$$

Considérons $(l_\beta(x) - l_\beta(0))$

$$l_\beta(x) - l_\beta(0) = \sup_{a \in D(x)} \left[g(x, a) + \beta \int \pi(x, a, dz)(l_\beta(z) - l_\beta(0)) + l_\beta(0)(\beta - 1) \right.$$

en posant :

$$f_\beta(x) = l_\beta(x) - l_\beta(0),$$

on a :

$$(1 - \beta)l_\beta(0) + f_\beta(x) = \sup_a \left\{ g(x, a) + \beta \int \pi(x, a, dz) f_\beta(z) \right\}$$

Supposons que la famille $\{ f_\beta \}$ est uniformément bornée par un nombre M.

$L^\infty(\lambda)$ est le dual de l'espace de Banach séparable $L^1(\lambda)$ donc la boule unité de $L^\infty(\lambda)$ est séquentiellement relativement compacte pour la topologie faible $\sigma(L^\infty(\lambda), L^1(\lambda))$. Par suite, il existe une suite (β_n) qui croît vers 1 telle que :

$$\begin{cases} \lim_{n\to\infty} (1 - \beta_n) l_{\beta_n}(0) = k \\ \lim_{n\to\infty} f_{\beta_n}(x) = \hat{l}(x) \qquad \text{dans } \sigma(L^\infty(\lambda), L^1(\lambda)) \end{cases}$$

C'est-à-dire que pour tout (y, a) en utilisant l'hypothèse 1 a,

$$\lim_{n\to\infty} \int p(y, a, z) f_{\beta_n}(z) \lambda(dz) = \int p(y, a, z) \hat{l}(z) \lambda(dz)$$

D'après la remarque suivant l'hypothèse 1 a, on sait alors que, pour tout y, la famille $\left\{ a \to \int p(y, a, z) f_{\beta_n}(z) \lambda(dz) \right\}_n$ est équicontinue et donc d'après Ascoli

$$a \to \int p(y, a, z) \hat{l}(z) \lambda(dz)$$

est continue et la convergence est uniforme sur le compact $D(y)$.

Ainsi pour tout $\varepsilon > 0$, il existe n_0 tel que $|1 - \beta_{n_0}| \leqslant \varepsilon/M$ et tel que pour $n \geqslant n_0$ et pour tout $a \in D(y)$ on ait :

$$\begin{cases} \left| \int p(y, a, z) f_{\beta_n}(z) \lambda(dz) - \int p(y, a, z) \hat{l}(z) \lambda(dz) \right| \leqslant \varepsilon \\ |k - (1 - \beta_n) l_{\beta_n}(0)| \leqslant \varepsilon \end{cases}$$

Montrons que $\{ f_{\beta_n}(y) \}$ est une suite de Cauchy.

Pour tout n, il existe d'après le théorème 2 et la remarque qui suit H_{1a}, un ensemble N_{β_n}, λ négligeable et une stratégie $d_{\beta_n} \in \mathscr{S}$ tels que pour tout $y \notin N_{\beta_n}$

$$f_{\beta_n}(y) + (1 - \beta_n) l_{\beta_n}(0) = g(y, d_{\beta_n}(y)) + \beta_n \int p(y, d_{\beta_n}(y), z) f_{\beta_n}(z) \lambda(dz)$$

Prenons $y \notin \dot{N} = \bigcup_n N_{\beta_n}$. Si n et m sont supérieurs à n_0

$$f_{\beta_m}(y) + (1 - \beta_m) l_{\beta_m}(0) \geqslant g(y, d_{\beta_n}(y)) + \beta_m \int p(y, d_{\beta_n}(y), z) f_{\beta_m}(z) \lambda(dz)$$

donc :

$$f_{\beta_n}(y) - f_{\beta_m}(y) + (1 - \beta_n)l_{\beta_n}(0) - (1 - \beta_m)l_{\beta_m}(0)$$
$$\leqslant \beta_n \int p(y, d_{\beta_n}(y), z) f_{\beta_n}(z)\lambda(dz) - \beta_m \int p(y, d_{\beta_n}(y), z) f_{\beta m}(z)\lambda(dz)$$

Ainsi :

$$f_{\beta_n}(y) - f_{\beta_m}(y) \leqslant (1-\beta_m)l_{\beta_m}(0) - (1-\beta_n)l_{\beta_n}(0)$$
$$+\beta_n \int p(y, d_{\beta_n}(y), z)(f_{\beta_n}(z) - \hat{l}(z))\lambda(dz) - \beta_m \int p(y, d_{\beta_n}(y), z)(f_{\beta_m}(z) - \hat{l}(z))\lambda(dz)$$
$$+(\beta_m - \beta_n)\int p(y, d_{\beta_n}(y), z)\hat{l}(z)\lambda(dz) \leqslant 4\varepsilon$$

Donc $\{ f_{\beta_n}(y) \}$ est une suite de Cauchy pour tout $y \notin \dot{\mathrm{N}}$.

Soit

$$l(y) = \lim_{n \to \infty} f_{\beta_n}(y)$$

On a $l = \hat{l}$, λ p. s.

On peut alors prendre la limite quand $n \to \infty$ de

$$(1 - \beta_n)l_{\beta_n}(0) + f_{\beta_n}(y) = \sup_a \left\{ g(y, a) + \beta_n \int \pi(y, a, dz) f_{\beta_n}(z) \right\}$$

On obtient :

$$k + l(y) = \lim_{n \to \infty} \sup_a \left[g(y, a) + \beta_n \int \pi(y, a, dz) f_{\beta_n}(z) \right]$$
$$\geqslant \sup_a \left[g(y, a) + \int \pi(y, a, dz) l(z) \right]$$

et on a aussi pour $y \notin \dot{\mathrm{N}} = \bigcup_n \mathrm{N}_{\beta_n}$

$$k + l(y) = \lim_{n \to \infty} \left[g(y, d_{\beta_n}(y)) + \beta_n \int \pi(y, d_{\beta_n}(y), dz) f_{\beta_n}(z) \right]$$

Pour $n \geqslant n_0$

$$g(y, d_{\beta_n}(y)) + \beta_n \int \pi(y, d_{\beta_n}(y), dz) f_{\beta_n}(z)$$
$$\leqslant g(y, d_{\beta_n}(y)) + \beta_n \int \pi(y, d_{\beta_n}(y), dz) l(z) + \varepsilon$$

donc :

$$k + l(y) \leqslant \varepsilon + \underline{\lim}_n \left\{ g(y, d_{\beta_n}(y)) + \beta_n \int \pi(y, d_{\beta_n}(y), dz) l(z) \right\}$$
$$\leqslant \varepsilon + \sup_{a \in \mathrm{D}(y)} \left\{ g(y, a) + \int \pi(y, a, dz) l(z) \right\}$$

et comme ceci est vrai pour tout ε

$$k + l(y) \leqslant \sup_{a \in D(y)} \left\{ g(y, a) + \int \pi(y, a, dz) l(z) \right\}$$

En conclusion :

$$k + l(y) \leqslant \sup_{a \in D(y)} \left\{ g(y, a) + \int \pi(y, a, dz) l(z) \right\}$$

3.1.2. Remarquons avant de nous résumer, qu'une des conditions imposées dans l'hypothèse 1' est de prendre $L^1(\mathscr{Y})$ un espace de Banach séparable, condition donnée dans l'hypothèse 2 puisque $\mathscr{Y}$ est polonais ($\mathscr{Y}$ espace métrique séparable, muni de sa tribu borélienne $\mathscr{Y}$, suffit).

Dans le cadre de l'hypothèse 2', l'ensemble N de la démonstration précédente est vide.

Résumons nos résultats dans la proposition suivante.

PROPOSITION 3. — *Dans le cadre de l'hypothèse 2' (respectivement dans le cadre de l'hypothèse 1'), si la condition suivante est réalisée.*

CONDITION 1. — *La famille* $\{ x \to l_\beta(x) - l_\beta(0) \}_\beta$ *est bornée uniformément. Alors pour une suite* $\{ \beta_n \}$ *croissant vers 1*

$$\begin{cases} \lim\limits_{\beta_n \uparrow 1} (1 - \beta_n) l_{\beta_n}(0) = k \\ \lim\limits_{\beta_n \uparrow 1} (l_{\beta_n}(.) - l_{\beta_n}(0)) = l(.) \qquad [\text{resp. } \lambda \text{ p. s.}] \end{cases}$$

et les fonctions k et l vérifient l'équation

$$k + l(x) = \sup_{a \in D(x)} \left[g(x, a) + \int \pi(x, a, dz) l(z) \right]$$

3.2. **Stratégies optimales en moyenne**

Montrons maintenant un théorème qui caractérise les stratégies optimales. Ce théorème est adapté de théorèmes dus à Ross [*10*] et à Mandl [*6*]. Nous réunissons dans ce théorème les cas où l'hypothèse 2 ou l'hypothèse 1 sont vérifiées, on notera [respectivement] pour l'hypothèse 1.

THÉORÈME 4. — *Dans le cadre de l'hypothèse 2* [resp. 1].

S'il existe une fonction aléatoire bornée l sur $\mathscr{Y}$ *et une constante k telles que pour tout* $x \in \mathscr{Y}$

$$k + l(x) = \sup_{a \in D(x)} \left[g(x, a) + \int \pi(x, a, dy) l(y) \right] \qquad (1)$$

alors :

a) *Il existe une stratégie stationnaire markovienne d telle que*

$$k = \lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} \tilde{g}_s \qquad [\mathrm{P}^d \text{ p. s.}]$$

et pour toute $d \in \mathscr{D}$

$$\overline{\lim_{t\to\infty}} \frac{1}{t} \sum_{s=0}^{t-1} \tilde{g}_s \leqslant k \qquad [\mathrm{P}^d \text{ p. s.}]$$

b) *Si*

$$\phi(x, a) = g(x, a) + \int \pi(x, a, dy) l(y) - k - l(x)$$

On a :

$$\lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} \tilde{g}_s = k \qquad [\mathrm{P}^d \text{ p. s.}] \Leftrightarrow \lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} \phi(\mathrm{Y}_s, \mathrm{A}_s) = 0 \qquad [\mathrm{P}^d \text{ p. s.}] \quad (3)$$

En particulier dans le cas des stratégies stationnaires markoviennes $d \in \mathscr{S}$, *si elles existent, telles que* $\phi(., d(.)) = 0$ [resp. λ p. s.], *c'est-à-dire :*

$$k + l(.) = g(., d(.)) + \int \pi(., d(.), dy) l(y) \qquad [\text{resp. } \lambda \text{ p. s.}] \qquad (4)$$

Démonstration. — La fonction ϕ introduite mesure la différence entre l'emploi d'un contrôle optimal et d'un contrôle quelconque.

On a posé :

$$\phi(x, a) = g(x, a) + \int \pi(x, a, dy) l(y) - k - l(x)$$

alors : $\phi \leqslant 0$.

Introduisons pour tout $t \in \mathbb{N}$

$$\mathrm{U}_t = \tilde{g}_t - k + l(\mathrm{Y}_{t+1}) - l(\mathrm{Y}_t) - \phi(\mathrm{Y}_t, \mathrm{A}_t) \qquad \text{et } \mathrm{M}_t = \sum_{s=0}^{t-1} \mathrm{U}_s$$

Montrons que $(\mathrm{M}_t)_{t\in\mathbb{N}}$ est une martingale sur $(\Omega, \mathscr{A}, \mathrm{P}^d)$ adaptée aux tribus $(\mathscr{B}_t)_{t\in\mathbb{N}}$ quelle que soit la stratégie d.

. U_t est $\mathscr{B}_{t+1}$ mesurable

. $\mathrm{E}^d[\mathrm{M}_{t+1} - \mathrm{M}_t \mid \mathscr{B}_t] = \mathrm{E}^d[\tilde{g}_t \mid \mathscr{B}_t] + \mathrm{E}^d[l(\mathrm{Y}_{t+1}) \mid \mathscr{B}_t] - k - l(\mathrm{Y}_t) - \phi(\mathrm{Y}_t, \mathrm{A}_t) = 0.$

On a donc d'après la loi des grands nombres pour les martingales si :

$$\sum_{t=1}^{\infty} \frac{1}{t^2} \mathrm{E}^d[\mathrm{U}_t^2] < \infty$$

(condition vérifiée car (U_t) est bornée uniformément en t)

$$\lim_{t\to\infty} \frac{\mathrm{M}_t}{t} = 0 \qquad [\mathrm{P}^d \text{ p. s.}]$$

Or

$$\frac{\mathrm{M}_t}{t} = \frac{1}{t}\sum_{s=0}^{t-1} \tilde{g}_s - k + \frac{l(\mathrm{Y}_t) - l(\mathrm{Y}_0)}{t} - \frac{1}{t}\sum_{s=0}^{t-1} \phi(\mathrm{Y}_s, \mathrm{A}_s)$$

donc

$$\lim_{t\to\infty}\left[\frac{1}{t}\sum_{s=0}^{t-1} \tilde{g}_s - k - \frac{1}{t}\sum_{s=0}^{t-1} \phi(\mathrm{Y}_s, \mathrm{A}_s)\right] = 0 \qquad [\mathrm{P}^d \text{ p. s.}]$$

1° Il existe sous les hypothèses du théorème une stratégie d telle que (4) soit vérifiée, en effet en posant :

$$\psi(y, a) = g(y, a) + \int \pi(y, a, dz) l(z),$$

il existe $d \in \mathscr{S}$, mesurable de $\mathscr{Y}$ dans $\mathscr{A}$ telle que pour tout y

$$\psi(y, d(y)) = \sup_{a\in \mathrm{D}(y)} \psi(y, a) \qquad [\text{resp. } \lambda \text{ p. s.}]$$

et pour cette d :

$$\phi(y, d(y)) = 0 \qquad [\text{resp. } \lambda \text{ p. s.})$$

Dans le cadre de l'hypothèse 2 le résultat a) du théorème est alors immédiat. Pour l'hypothèse 1, il suffit de vérifier que si :

$$\phi(., d(.)) = 0 \qquad [\lambda \text{ p. s.})$$

alors

$$\phi(\mathrm{Y}_s, d(\mathrm{Y}_s)) = 0 \qquad [\mathrm{P}^d \text{ p. s.}]$$

pour $s \geqslant 1$.

Soit $\Gamma = \{ ., \phi(., d(.)) \neq 0 \}$; on a $\lambda(\Gamma) = 0$.

Or $\mathrm{P}_x^d[\mathrm{Y}_s \in \Gamma] = \pi_s^d[x, \Gamma]$ et comme $\pi(x, d(x), .) \ll \lambda(.)$, $\mathrm{P}_x^d[\mathrm{Y}_s \in \Gamma]$ est nulle.

Alors d est telle que :

$$\lim_{t\to\infty} \frac{1}{t}\sum_{s=0}^{t-1} \tilde{g}_s = k \qquad [\mathrm{P}^d \text{ p. s.}]$$

Et comme $\phi \leqslant 0$, pour toute autre stratégie d

$$\overline{\lim_{t\to\infty}} \frac{1}{t} \sum_{s=0}^{t-1} \tilde{g}_s \leqslant k \qquad [\mathrm{P}^d \text{ p. s.}]$$

d est optimale.

2° Si

$$\lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} \phi(\mathrm{Y}_s, \mathrm{A}_s) = 0 \qquad [\mathrm{P}^d \text{ p. s.}]$$

alors d est optimale au sens du 1°.

3.3. Étude du cas où pour toute $d \in \mathscr{S}$, la chaîne de transition π^d est récurrente positive au sens de Harris

Étude de $k(.)$

Revenons à l'étude du comportement si β tend vers 1, de $(1-\beta)l_\beta(.)$.

Plaçons-nous dans le cadre de l'hypothèse 1′.

La famille $\{(1-\beta)l_\beta(.)\}_{0<\beta<1}$ est bornée par $\|g\|$ dans $\mathrm{L}^\infty(\lambda)$, reprenons le raisonnement vu plus haut. $\mathrm{L}^\infty(\lambda)$ est le dual de l'espace de Banach séparable $\mathrm{L}^1(\lambda)$, donc la boule unité de $\mathrm{L}^\infty(\lambda)$ est séquentiellement relativement compacte pour la topologie faible $\sigma(\mathrm{L}^\infty(\lambda), \mathrm{L}^1(\lambda))$, donc il existe une sous-suite $\{\beta_n\}$ qui tend vers 1 telle que :

$$\lim_{n\to\infty} (1-\beta_n)l_{\beta_n} = \hat{k} \qquad \text{dans } \sigma(\mathrm{L}^\infty(\lambda), \mathrm{L}^1(\lambda))$$

c'est-à-dire que pour tout (y, a)

$$\lim_{n\to\infty} \quad p(y, a, z)(1-\beta_n)l_{\beta_n}(z)\lambda(dz) = \int p(y, a, z)\hat{k}(z)\lambda(dz)$$

Or, d'après la conséquence b) de l'hypothèse 1 a, pour tout $y \in \mathscr{Y}$

$$\left\{ a \to \int p(y, a, z)(1-\beta_n)l_{\beta_n}(z)\lambda(dz) \right\}_{\beta_n}$$

est une famille équicontinue, et de plus, d'après Ascoli la convergence sur le compact $\mathrm{D}(y)$ de $\int p(y, a, z)l_{\beta_n}(z)(1-\beta_n)\lambda(dz)$ vers $\int p(y, a, z)\hat{k}(z)\lambda(dz)$ est uniforme en a.

Ainsi pour tout y dans le complémentaire de :

$$N = \bigcup_n N_{\beta_n}$$

[N_β étant l'ensemble de λ-mesure nulle où il n'y avait pas l'égalité dans le théorème 2, N est vide sous l'hypothèse 2] on a :

$$l_{\beta_n}(y) = \sup_{a \in D(y)} \left\{ g(y, a) + \beta_n \int p(y, a, z) l_{\beta_n}(z) \lambda(dz) \right\}$$

donc :

$$\lim_{n \to \infty} (1 - \beta_n) l_{\beta_n}(y) = \sup_{a \in D(y)} \int p(y, a, z) \hat{k}(z) \lambda(dz) = \sup_{a \in D(y)} \int \pi(y, a, dz) \hat{k}(z)$$

Donc sur N^c :

$$\hat{k}(.) = \sup_{a \in D(.)} \int \pi(., a, dz) \hat{k}(z) .$$

Remarquons que la borne supérieure étant prise sur $D(y)$, compact, et

$$a \to \int \pi(y, a, dz) \hat{k}(z)$$

étant continue, il existe une stratégie $d \in \mathcal{S}$ telle que

$$\pi^d \hat{k} = \sup_{a \in D(.)} \int \pi(., a, dz) \hat{k}(z) \qquad [\text{resp. } \lambda \text{ p. s.}]$$

Posons :

$$k = \pi^d \hat{k}$$

$$k = \begin{cases} \pi^d k \\ \sup\limits_{a \in D(.)} \int \pi(., a, dz) k(z) \qquad [\text{resp. } \lambda \text{ p. s.}] \end{cases}$$

Ainsi dans le cadre de l'hypothèse 1′ (en particulier sous l'hypothèse 2′) :

$$k(x) = \int \pi(x, d(x), dz) k(z)$$

Supposons, pour la suite de ce paragraphe, que pour toute $d \in \mathcal{S}$, la chaîne Y est récurrente Harris positive.

Ainsi comme k est harmonique pour une certaine stratégie stationnaire, k est constante. Pour tout y

$$\lim_{n \to \infty} (1 - \beta_n) \int \pi(y, a, dz) l_{\beta_n}(z) = k$$

et la convergence est uniforme en $a \in D(y)$.

Donc :

$$\lim_{n\to\infty}(1-\beta_n)\int \pi(y, d_{\beta_n}(y), dz)l_{\beta_n}(z) = k$$

Soit $\mu_{d_{\beta_n}}$ la probabilité invariante par $\pi^{d_{\beta_n}}$: $\mu_{d_{\beta_n}}\pi^{d_{\beta_n}} = \mu_{d_{\beta_n}}$, $\mu_{d_{\beta_n}} \ll \lambda$. Comme

$$\lim_{n\to\infty}(1-\beta_n)\pi^{d_{\beta_n}}l_{\beta_n} = k \qquad [\lambda \text{ p. s.}]$$

$$\lim_{n\to\infty}(1-\beta_n)\mu_{d_{\beta_n}}\pi^{d_{\beta_n}}l_{\beta n} = \lim_{n\to\infty}\mu_{d_{\beta_n}}k = k \ ,$$

$$\lim_{n\to\infty}(1-\beta_n)l_{\beta_n}(\mu_{d_{\beta_n}}) = k$$

en notant

$$l(\mu) = \int l(x)\mu(dx)\,.$$

Or pour $d_{\beta_n} \in \mathscr{S}$, en considérant le critère de l'espérance du gain moyen

$$\mathrm{M}^{d_{\beta_n}} = \lim_{t\to\infty}\frac{1}{t}\sum_{s=0}^{t-1} g(\mathrm{Y}_s, \mathrm{A}_s) = \int g(x, d_{\beta_n}(x))d\mu_{d_{\beta_n}}(x) \qquad [\mathrm{P}^{d_{\beta_n}} \text{ p. s.}]$$

Montrons que :

$$\lim_{n\to\infty}\mathrm{M}^{d_{\beta_n}} = k$$

Par définition

$$l_{\beta_n}(\mathrm{Y}_0) = \mathrm{E}^{d_{\beta_n}}\left\{\sum_{s=0}^{\infty}\beta_n^s g(\mathrm{Y}_s, \mathrm{A}_s)\,|\,\mathrm{Y}_0\right\}$$

donc :

$$\int l_{\beta_n}(x)\mu_{d_{\beta_n}}(dx) = \int \mu_{d_{\beta_n}}(dx)\sum_{s=0}^{\infty}\beta_n^s\left(\int \pi(x, d_{\beta_n}(x), dy)g(y, d_{\beta_n}(y)\right)$$

Comme $\mu_{d_{\beta_n}}$ est $\pi^{d_{\beta_n}}$ invariante

$$\int l_{\beta_n}(x)\mu_{d_{\beta_n}}(dx) = \sum_{s=0}^{\infty}\beta_n^s\int g(x, d_{\beta_n}(x))\mu_{d_{\beta_n}}(dx) = \sum_{s=0}^{\infty}\beta_n^s\mathrm{M}^{d_{\beta_n}} = \frac{\mathrm{M}^{d_{\beta_n}}}{1-\beta_n}$$

ainsi $(1-\beta_n)\int l_{\beta_n}(x)\mu_{d_{\beta_n}}(dx) = \mathrm{M}^{d_{\beta_n}}$ et $\lim_{n\to\infty}\mathrm{M}^{d_{\beta_n}} = k$.

Donc pour toute suite (β_n) qui tend vers 1 telle que $(1-\beta_n)l_{\beta_n}$ converge, la limite est $k = \sup_d \mathrm{M}^d$. Par suite $k = \lim_{\beta\to 1}(1-\beta)l_\beta$. S'il existe une $d_0 \in \mathscr{S}$ telle que $k = \mathrm{M}^{d_0}$ et si π^{d_0} est récurrente au sens de Doeblin comme $(g-k)$ est d'intégrale nulle par rapport à μ_{d_0} ($\mu_{d_0}(g) = \mathrm{M}^{d_0} = k$), l'équa-

tion de Poisson est vérifiée, et on sait qu'il existe une v. a. l bornée unique à une constante près telle que :

$$l - \pi^{d_0}(l) = g - k$$

On vient donc de généraliser le résultat de l'exemple d'une chaîne de Markov récurrente au sens de Harris, non contrôlée.

Résumons les résultats dans le théorème suivant :

THÉORÈME 5. — *Dans le cadre de l'hypothèse 1'.*

Si pour toute $d \in \mathscr{S}$, la chaîne Y est récurrente Harris positive, *alors,*

1. *il existe une constante* k *telle que :*

$$k = \sup_{d \in \mathscr{S}} \mathrm{M}^d = \lim_{\beta \to 1} (1 - \beta) l_\beta(x)$$

Si de plus il existe une $d_0 \in \mathscr{S}$, *telle que* $\mathrm{M}^{d_0} = k$ *et si* π^{d_0} *est récurrente au sens de Doeblin, alors il existe une v. a.* l *bornée, unique à une constante près telle que :* $l - \pi^{d_0} l = g_{d_0} - k$.

2. *Si*

$$\phi(x, a) = g(x, a) - k - l(x) + \int \pi(x, a, dz) l(z)$$

et si

$$\lim_{t \to \infty} \frac{1}{t} \sum_{s=0}^{t-1} \phi(\mathrm{Y}_s, \mathrm{A}_s) = 0 \qquad [\mathrm{P}^d \text{ p. s.}]$$

alors

$$\lim_{t \to \infty} \frac{1}{t} \sum_{s=0}^{t-1} g(\mathrm{Y}_s, \mathrm{A}_s) = k \qquad [\mathrm{P}^d \text{ p. s.}]$$

En particulier si pour $d \in \mathscr{S}$ *on a* $\phi(., d(.)) = 0$, *alors* d *est optimale.*

Remarque. — Des résultats récents de Numelin [7] assurent l'existence d'une fonction l, μ_{d_0} intégrable, solution de l'équation de Poisson $l - \pi^{d_0} l = g_{d_0} - k$ en supposant seulement que π^{d_0} est récurrente positive. Mais l peut ne pas être μ_d intégrable pour $d \in \mathscr{S}$ quelconque et l'application du théorème 4 pose des problèmes avec un tel l.

Exemple d'application

Nous nous plaçons dans le cadre où l'hypothèse 1' est réalisée, et nous supposons que la condition 2 suivante est réalisée :

CONDITION 2. — Il existe une constante k', $k' < 1$ telle que si :

$$\mathrm{K} = \{ (x, a) ; a \in \mathrm{D}(x) \}$$

alors

$$\sup_{((x,a),(y,b))\in K\times K} || \pi(x, a, .) - \pi(y, b, .) || = 2k'$$

Ce qui est réalisé en particulier sous la condition 2' suivante :

CONDITION 2'. — 1. Il existe un point 0 tel que pour tout $(x, a)\in K$

$$\pi(x, a, \{ 0 \}) \geqslant \alpha > 0$$

ou

2. Il existe une mesure ν sur $(\mathscr{Y}, \mathscr{B})$ non nulle telle que pour tout $(x, a)\in K$

$$\pi(x, a, .) \geqslant \nu(.)$$

En effet, pour la condition 2'.2 (qui implique 2'.1), il existe deux ensembles disjoints $\mathscr{Y}^+$ et $\mathscr{Y}^-$ de $\mathscr{B}$ tels que $\mathscr{Y}^+ \cup \mathscr{Y}^- = \mathscr{Y}$

$$\begin{aligned} || \pi(x, a, .) - \pi(y, b, .) || &= \pi(x, a, \mathscr{Y}^+) - \pi(y, b, \mathscr{Y}^+) \\ &\quad + \pi(y, b, \mathscr{Y}^-) - \pi(x, a, \mathscr{Y}^-) \\ &\leqslant 1 - \nu(\mathscr{Y}^+) + 1 - \nu(\mathscr{Y}^-) \leqslant 2 - || \nu || \end{aligned}$$

Nous allons montrer que nous sommes dans les conditions du théorème 5. Nous montrons tout d'abord qu'il existe $d_0 \in \mathscr{S}$, telle que $M^{d_0} = k$, ce qui revient à montrer suivant la proposition 3 que $\{ l_\beta(.) - l_\beta(0) \}_\beta$ est uniformément bornée. Nous montrerons ensuite que π^{d_0} est récurrente au sens de Doeblin.

Rappelons alors un *théorème de Ueno* [*12*].

Si P est une probabilité de transition, on note P_s ses itérées et alors :

$$|| P_s(x, .) - P_s(y, .) || \leqslant \frac{1}{2^{s-1}} \sup_{x,y} || P(x, .) - P(y, .) ||^s$$

Appliquons ceci aux transitions π

$$\begin{aligned} || \pi_s^d(x, .) - \pi_s^d(y, .) || &\leqslant \frac{1}{2^{s-1}} (2k')^s, \qquad k' < 1 \\ &\leqslant 2k'^s \qquad\qquad k' < 1 \end{aligned}$$

et dans le cadre de nos hypothèses il existe une stratégie d_β qui vérifie :

$$l_\beta(x) = E_x^{d_\beta}\left[\sum_s \beta^s g(Y_s, d_\beta(Y_s)) \right]$$

Donc :

$$l_\beta(x) = \sum_s \beta^s \int g(z, d_\beta(z)) \pi_s[x, d_\beta(x), dz]$$

que nous notons :

$$l_\beta(x) = \sum_s \beta^s \pi_s^{d_\beta}(x, g_{d_\beta})$$

$$|\, l_\beta(x) - l_\beta(o)\,| = \left| \sum_s \beta^s \left(\pi_s^{d_\beta}(x, g_{d_\beta}) - \pi_s^{d_\beta}(0, g_{d_\beta})\right)\right|$$

$$\leqslant \sum_s \beta^s \,||\, g \,||\quad ||\, \pi_s^{d_\beta}(x, .) - \pi_s^d(0, .)\,||$$

$$\leqslant \sum_s \beta^s \,||\, g \,||\, 2k'^s \qquad \text{avec } k' < 1$$

$$\leqslant \frac{2\,||\, g \,||}{1 - \beta k'} \leqslant \frac{2\,||\, g \,||}{1 - k'}$$

en considérant évidemment g uniformément borné, on vient de montrer que $\{\, l_\beta(.) - l_\beta(0)\,\}_\beta$, sous les conditions précédentes, est uniformément bornée.

Montrons maintenant que π^d est une chaîne de Markov récurrente au sens de Doeblin, pour tout $d \in \mathscr{S}$.

On vient de voir que :

$$||\, \pi_s^d(x, .) - \pi_s^d(y, .)\,|| \leqslant 2k'^s$$

Soit $\Gamma \in \mathscr{Y}$,

$$\left| \int \pi_t^d(x, dy)\left[\pi_s^d(x, \Gamma) - \pi_s^d(y, \Gamma)\right]\right| \leqslant 2k'^s$$

donc :

$$||\, \pi_s^d(x, .) - \pi_{t+s}^d(x, .)\,|| \leqslant 2k'^s$$

Comme $\mathscr{M}(\mathscr{Y})$ est un espace de Banach, on en déduit que

$$\lim_{s \to \infty} \pi_s^d(x, .) = \mu_d(.) \qquad \text{dans } \mathscr{M}(\mathscr{Y})$$

et

$$||\, \pi_s^d(x, .) - \mu_d(.)\,|| \leqslant 2k'^s$$

et on est alors dans le cadre du théorème 5.

D'où le corollaire

COROLLAIRE 6. — *On suppose que :*

$$\sup_{[(x,a),\,(y,b)] \in K \times K} ||\, \pi(x, a, .) - \pi(y, b, .)\,|| = 2k', \qquad k' < 1$$

Sous l'hypothèse 1', il existe une constante k et une v. a. bornée l sur $\mathscr{Y}$, et une stratégie $d_0 \in \mathscr{S}$, telles que :

$$\begin{cases} k = \sup_{d \in \mathscr{S}} \mathrm{M}^d = \lim_{\beta \to 1} (1 - \beta) l_\beta(.) = \mathrm{M}^{d_0} \\ k = l(.) = \sup_a \left[g(., a) + \int \pi(., a, dy) l(y) \right] \\ k + l(.) = g(., d_0(.)) + \pi^{d_0} l(.) \qquad [\text{resp. } \lambda \text{ p. s.}] \end{cases}$$

Pour toute $d \in \mathscr{S}$, π^d est récurrente au sens de Doeblin.

BIBLIOGRAPHIE

[1] BROWN-PURVES, Measurable selection of extrema. *Annals of statistics*, vol. **1**, n° 5, 1973, p. 902-912.

[2] DERMANN, *Finite state Markovian decision process.* Academic Press, 1970.

[3] EŠTIGNEEV, Measurable selection and dynamic programming. *Mathematics of operations research*, vol. **1**, n° 3, 1976, p. 267-272.

[4] GEORGIN, Contrôle des chaînes de Markov sur des espaces arbitraires. Estimation et contrôle optimal dans le cadre adaptatif. *Thèse de 3e Cycle.* Université Paris-Sud, Centre d'Orsay, 1977.

[5] K. HINDERER, Foundations of non-stationary dynamic programming with discrete time parameters. *Lecture notes in operations research and Mathematical systems.* Springer-Verlag, Berlin.

[6] MANDL, Estimation and control in Markov chains. *Adv. Appl. Prob.*, t. **6**, 1974, p. 40-60.

[7] NUMELIN, On the Poisson equation for ϕ recurrent Markov chains. *A paraître 1977.*

[8] OREY, *Limit theorems for Markov chains transitions probabilities.* Van Nostrand, 1971.

[9] D. REVUZ, *Markov chains.* North Holland, 1975.

[10] ROSS, Arbitrary state markovian decision processes. *The annals of Mathematical statistics*, vol. **39**, n° 6, 1968.

[11] STRIEBEL, *a*) Optimal control of discrete time stochastic systems. *Lectures notes in operations research and mathematical systems*, Springer (19).
b) Martingale conditions for the optimal control of continuous time stochastic systems. International workshop of stochastic-filtering and control, Los Angeles, Californy.

[12] UENO, Some limit theorems for temporally discrete Markov processes. *J. Fac. Sciences*, Université Tokyo (I), 1957, p. 7.

(Manuscrit reçu le 1er février 1978)