TABLE DES MATIÈRES

PRÉFACE

Nous tenons à remercier la direction de l'École polytechnique, et tout particulièrement la Direction des Études, pour l'aide matérielle importante qu'elle a apportée à la préparation des journées X-UPS. Nous remercions les Éditions de l'École polytechnique qui ont bien voulu accueillir la série *Journées mathématiques X-UPS* au sein de leurs collections.

Nous remercions aussi les secrétaires du Centre de mathématiques, notamment Claudine Harmide et Michèle Lavallette, pour leur contribution à l'organisation de ces journées.

Nicole Berline, Alain Plagne et Claude Sabbah

INTRODUCTION

La théorie des jeux est une discipline qui étudie la *prise de décision interactive* : plusieurs personnes, ou joueurs, doivent prendre des décisions, choisir des actions, qui vont induire un résultat, les intérêts des joueurs étant potentiellement divergents.

Si les mathématiciens ont toujours montré un fort intérêt pour les jeux de hasard ou de stratégie, les premiers travaux théoriques sur les jeux de stratégie apparaissent au début du XX^e siècle avec Zermelo (1912), Borel (1921), Von Neumann (1928). La théorie des jeux naît réellement comme discipline sous l'impulsion du mathématicien John Von Neumann et de l'économiste Oskar Morgenstern, qui écrivent en 1944 un livre fondateur : *Games and Economic Behavior*. Les travaux de John Nash (1950), en donnant une notion de solution pour les jeux à somme non nulle, confortent cette fondation. Depuis, la théorie des jeux a connu un développement mathématique important et de nombreuses applications dans diverses disciplines : biologie, informatique, économie. Le succès est particulièrement remarquable en économie et plusieurs théoriciens des jeux ont reçu le prix Nobel d'économie : John C. Harsanyi, John F. Nash et Reinhardt Selten en 1994, Robert J. Aumann et Thomas C. Schelling en 2005. La théorie des jeux comporte aujourd'hui plusieurs branches : jeux coopératifs, jeux stratégiques, jeux à information incomplète, jeux dynamiques, jeux différentiels. Nous présentons ici les fondements mathématiques de la théorie des jeux stratégiques et développons un des thèmes principaux : les jeux répétés.

Il y a deux façons de décrire un jeu, la première étant de donner la règle du jeu, c'est-à-dire décrire précisément son déroulement, comment les joueurs interviennent, quelles sont les issues du jeu et les gains — ou paiements — des joueurs une fois le jeu terminé. C'est la façon la plus courante de décrire un jeu, et on parle alors de jeu *sous forme extensive*. On peut alors définir la notion de *stratégie* : une stratégie est un plan d'action qui prévoit ce que doit faire le joueur dans chaque éventualité qu'il va rencontrer. On peut voir une stratégie comme un programme informatique, une liste d'instructions données à un ordinateur, lui permettant de jouer le jeu. Lorsque chaque joueur choisit une telle stratégie, le déroulement du jeu est fixé et on peut, au moins théoriquement, calculer l'issue du jeu et les paiements. La seconde façon de décrire un jeu consiste à donner, pour chaque joueur, l'ensemble de ses stratégies, ainsi que les applications qui associent aux vecteurs — ou profils — de stratégies, les paiements des joueurs. On parle alors de jeu *sous forme stratégique*. Suivant la complexité du jeu, donner la forme stratégique peut être beaucoup plus difficile que donner la forme extensive : dans le jeu Pierre, Feuille, Ciseaux, où les choix sont simultanés, chaque joueur a trois stratégies et les fonctions de paiement sont faciles à écrire. Pour le jeu d'échecs, décrire la règle du jeu est assez simple, alors que donner l'ensemble des stratégies est humainement impossible.

L'avantage de la forme stratégique est de permettre une formalisation mathématique claire et compacte : un jeu à n joueurs est une application d'un produit cartésien de n facteurs dans $\mathbb{R}^n$. Grâce à des techniques d'analyse réelle et convexe, on obtient des théorèmes d'existence relativement généraux. On peut obtenir des résultats plus précis de structure voire de caractérisation des solutions dans les jeux répétés, qui sont des jeux dynamiques ayant des propriétés de stationnarité dans le temps.

Nous nous sommes inspirés de plusieurs sources : l'ouvrage « Repeated Games » de Jean-François Mertens, Sylvain Sorin et Schmuel Zamir (1994) , le Handbook of Game theory, « A First Course on Zero-Sum Repeated Games » de Sylvain Sorin (2002). Citons aussi quelques manuels classiques de théorie des jeux : « A course in game theory », Osborne et Rubinstein (1994), « Stability and Perfection of Nash Equilibria », Van Damme (1987), « Game Theory », Myerson (1991). Un article récent dans MATAPLI (Janvier 2006) présente brièvement certains travaux importants de R.J. Aumann.

Dans le premier texte, Tristan Tomala présente les jeux sous forme stratégique et leur analyse mathématique. Dans le second, il décrit le modèle de jeu répété et en étudie une classe assez simple : les jeux répétés à information complète et observation parfaite. Dans le troisième texte, Jérôme Renault présente les principaux résultats concernant les jeux répétés à information incomplète. Dans le quatrième, Rida Laraki traite des jeux stochastiques.

Nous remercions chaleureusement Sylvain Sorin pour son aide à la préparation des journées X-UPS 2006.

Rida Laraki, Jérôme Renault et Tristan Tomala

JEUX SOUS FORME NORMALE

par

Tristan Tomala

Table des matières

1. Préliminaires

On appelle *jeu sous forme normale* ou jeu *sous forme stratégique* la donnée d'un ensemble N de joueurs, d'une famille d'ensembles de stratégies (ou d'actions) $(A^i)_{i\in N}$ et d'une famille de fonctions de paiements $(g^i)_{i\in N}$ avec $g^i : \prod_{j\in N} A^j \to \mathbb{R}$. L'ensemble des joueurs sera toujours supposé fini et non-vide. Les ensembles d'actions seront toujours supposés non vides et on parlera de *jeu fini* lorsque A^i est fini pour tout i.

Un jeu sous forme normale représente une interaction entre joueurs rationnels : chaque joueur $i \in N$ choisit une action $a^i \in A^i$, les choix étant simultanés, et si $a = (a^i)_{i\in N}$ est le *profil* d'actions choisi, le joueur i reçoit le paiement $g^i(a)$. Tous les joueurs connaissent le jeu et le but du joueur i est d'obtenir un paiement le plus grand possible.

Un jeu à un joueur est donc simplement un problème de maximisation. Dès qu'il y a au moins deux joueurs, le joueur i ne contrôle que partiellement son paiement et la notion de *bonne* stratégie n'est pas claire. Les exemples usuels suivants permettent de s'en convaincre. Les matrices ci-dessous représentent des jeux à deux joueurs dans lesquels le joueur 1 choisit la ligne, le joueur 2 la colonne et l'entrée de la matrice est le couple de paiements (g^1, g^2).

Le Dilemme du Prisonnier. — Deux criminels sont arrêtés et interrogés dans des pièces séparées. Ils ont le choix entre dénoncer leur complice (D) où se taire et donc coopérer avec leur complice (C). Un criminel dénoncé par son complice se verra infliger une lourde peine et une peine légère dans le cas contraire. De plus le fait de dénoncer l'autre permet d'obtenir une remise de peine, que l'on soit soi-même dénoncé ou pas. Chaque joueur classe les issues du jeu par préférence décroissante selon l'ordre suivant : (ne pas être dénoncé et dénoncer), (ne pas être dénoncé et ne pas dénoncer), (être dénoncé et dénoncer), (être dénoncé et ne pas dénoncer). Attribuant des paiements numériques à ces alternatives, nous formalisons cette situation par le jeu suivant :

	C	D
C	$3,3$	$0,4$
D	$4,0$	$1,1$

Un jeu de coordination. — Deux amis veulent se rencontrer au lieu (A) ou au lieu (B). Leurs paiements sont égaux et valent 1 s'ils se rencontrent effectivement et 0 sinon. Ceci se représente par le jeu :

	A	B
A	$1,1$	$0,0$
B	$0,0$	$1,1$

Le jeu « Matching Pennies ». — Chaque joueur possède une pièce de monnaie et choisit secrètement de la mettre sur Pile (P) ou sur Face (F). Le joueur 1 gagne si son choix est le même que celui du joueur 2 et, dans ce cas, le joueur 2 perd. Dans le cas contraire c'est 2 qui gagne et 1 qui perd. Ceci se représente par le jeu :

	P	F
P	$1,-1$	$-1,1$
F	$-1,1$	$1,-1$

Commençons par donner quelques notions simples de *bonne* stratégie. Nous adopterons les notations suivantes. Pour tout joueur i, $-i$ désigne l'ensemble des autres joueurs $N \setminus \{i\}$. Si $(E^i)_{i\in N}$ est une famille d'ensembles indexée par N, nous notons $E = \prod_{i\in N} E^i$, $E^{-i} = \prod_{j\neq i} E^i$. Un élément e de E pourra se noter $e = (e^1, \dots, e^n) = (e^i)_{i\in N} = (e^i, e^{-i})$ cette dernière notation étant utilisée lorsque l'on veut séparer le joueur i des autres.

Définition 1.1

– Une stratégie $a^i \in A^i$ du joueur i est *dominée* si
$$\exists\, b^i \in A^i,\ \forall\, a^{-i} \in A^{-i},\quad g^i(a^i, a^{-i}) \leqslant g^i(b^i, a^{-i}).$$
– Une stratégie $a^i \in A^i$ du joueur i est *faiblement dominée* si
$$\exists\, b^i \in A^i,\ \begin{cases} \forall\, a^{-i} \in A^{-i}, & g^i(a^i, a^{-i}) \leqslant g^i(b^i, a^{-i}) \quad \text{et} \\ \exists\, a^{-i} \in A^{-i}, & g^i(a^i, a^{-i}) < g^i(b^i, a^{-i}). \end{cases}$$
– Une stratégie $a^i \in A^i$ du joueur i est *strictement dominée* si
$$\exists\, b^i \in A^i,\ \forall\, a^{-i} \in A^{-i},\quad g^i(a^i, a^{-i}) < g^i(b^i, a^{-i}).$$
– Une stratégie $a^i \in A^i$ du joueur i est *dominante* si
$$\forall\, b^i \in A^i,\ \forall\, a^{-i} \in A^{-i},\quad g^i(a^i, a^{-i}) \geqslant g^i(b^i, a^{-i}).$$
– Une stratégie $a^i \in A^i$ du joueur i est *faiblement dominante* si
$$\forall\, b^i \in A^i,\ \begin{cases} \forall\, a^{-i} \in A^{-i}, & g^i(a^i, a^{-i}) \geqslant g^i(b^i, a^{-i}) \quad \text{et} \\ \exists\, a^{-i} \in A^{-i}, & g^i(a^i, a^{-i}) > g^i(b^i, a^{-i}). \end{cases}$$
– Une stratégie $a^i \in A^i$ du joueur i est *strictement dominante* si
$$\forall\, b^i \in A^i,\ \forall\, a^{-i} \in A^{-i},\quad g^i(a^i, a^{-i}) > g^i(b^i, a^{-i}).$$

Un joueur rationnel ne jouera jamais de stratégie strictement dominée, jouera à coup sur une stratégie strictement dominante si elle existe (et alors elle est unique) et ne perd rien à jouer une stratégie dominante. On peut remarquer que dans le jeu du Dilemme du Prisonnier, la stratégie D est strictement dominante pour chaque joueur, l'issue rationnelle du jeu est donc (D, D).

Lorsque tous les joueurs sont rationnels et savent que leurs adversaires le sont, chacun peut supprimer ses propres stratégies strictement dominées et s'attendre à ce que les autres fassent de même.

De nouvelles stratégies strictement dominées peuvent alors apparaître dans le jeu réduit. On est donc conduit à itérer cette opération.

Procédure d'élimination itérée des stratégies strictement dominées (EISSD). — Pour tout jeu $G = (N, (A^i)_{i\in N}, (g^i)_{i\in N})$ et tout joueur i, on note $SD^i(G)$ l'ensemble des stratégies du joueur i strictement dominées dans G. Partons d'un jeu $G_0 = (N, (A^i_0)_{i\in N}, (g^i)_{i\in N})$.

– Pour tout $i \in N$, on pose

$$A^i_1 = A^i_0 \setminus SD^i(G_0) \quad \text{et} \quad G_1 = (N, (A^i_1)_{i\in N}, (g^i)_{i\in N}),$$

jeu dans lequel les fonctions de paiement sont définies par restriction.

– Pour tout entier $k > 1$ et tout $i \in N$, on pose

$$A^i_k = A^i_{k-1} \setminus SD^i(G_{k-1}) \quad \text{et} \quad G_k = (N, (A^i_k)_{i\in N}, (g^i)_{i\in N}).$$

– Pour tout $i \in N$ on pose enfin

$$A^i_\infty = \bigcap_k A^i_k \quad \text{et} \quad G_\infty = (N, (A^i_\infty)_{i\in N}, (g^i)_{i\in N}).$$

On dit que G_0 est *résoluble par EISSD* si pour tout joueur i, la restriction de g^i à $\prod_{i\in N} A^i_\infty$ est une application constante.

Le jeu de concurrence de Cournot. — Deux entreprises produisent le même bien et choisissent la quantité à produire. Le prix de vente est une fonction décroissante de la somme des quantités, le bénéfice de chaque entreprise s'écrit comme la différence entre ses recettes et le coût total de production.

Prenons des paramètres $\alpha, \beta, \gamma > 0$ avec $\alpha > \gamma$ et définissons le jeu suivant : $G = (A^1, A^2, g^1, g^2)$ dans lequel $A^1 = A^2 = \mathbb{R}_+$ et pour chaque joueur i et paire de stratégies (a_1, a_2) :

$$g^i(a^1, a^2) = a^i(\alpha - \beta(a^1 + a^2))^+ - \gamma a^i$$

Ce jeu est résoluble par EISSD et on montre que, pour $i = 1, 2$, $A^i_\infty = \left\{\frac{\alpha-\gamma}{3\beta}\right\}$.

Deviner le demi-moyenne. — Un autre exemple est le jeu à n joueurs dans lequel chacun choisit un réel entre 0 et 100, le but étant d'être le plus proche possible de la demi-moyenne :

$$g^i(a) = -\left| a^i - \frac{1}{2n} \sum_{j=1}^{n} a^j \right|$$

Ce jeu est résoluble par EISSD et on a, pour tout $i \in N$, $A^i_\infty = \{0\}$.

Comme le montrent les jeux de coordination et Matching Pennies ci- dessus, bon nombre de jeux ne sont pas résolubles par cette simple méthode. Les deux parties suivantes donnent des notions de solutions pour lesquels on dispose de théorèmes d'existence relativement généraux. Nous commençons par traiter les jeux *à somme nulle.*

2. Jeux à somme nulle

Un jeu à somme nulle est un jeu à deux joueurs $G = (A^1, A^2, g^1, g^2)$ tel que pour tous $(a^1, a^2) \in A$, $g^1(a^1, a^2) + g^2(a^1, a^2) = 0$. Pour cette partie posons, $A^1 = S$, $A^2 = T$, $g^1 = g$. Ainsi, un jeu à somme nulle est déterminé par une application $g : S \times T \to \mathbb{R}$. Dans la suite, par souci de simplicité, nous supposerons g bornée.

2.1. Notions de solutions

Définition 2.1

– Le joueur 1 *garantit* le paiement $d \in \mathbb{R}$ si :

$$\exists\, s \in S,\ \forall\, t \in T, \quad g(s,t) \geqslant d.$$

– Le joueur 1 *défend* le paiement $d \in \mathbb{R}$ si :

$$\forall\, t \in T,\ \exists\, s \in S, \quad g(s,t) \geqslant d.$$

– Le joueur 2 *garantit* le paiement $d \in \mathbb{R}$ si :

$$\exists\, t \in T,\ \forall\, s \in S, \quad g(s,t) \leqslant d.$$

– Le joueur 2 *défend* le paiement $d \in \mathbb{R}$ si :

$$\forall\, s \in S,\ \exists\, t \in T, \quad g(s,t) \leqslant d.$$

Les propriétés suivantes sont immédiates :

Proposition 2.2. — *Posons*

$$\underline{v}(g) = \sup_{s \in S} \inf_{t \in T} g(s,t) \quad \textit{et} \quad \overline{v}(g) = \inf_{t \in T} \sup_{s \in S} g(s,t).$$

On a $\underline{v}(g) \leqslant \overline{v}(g)$ *et*

$$\underline{v}(g) = \sup\{d \mid 1 \textit{ garantit } d\} = \inf\{d \mid 2 \textit{ défend } d\},$$
$$\overline{v}(g) = \sup\{d \mid 1 \textit{ défend } d\} = \inf\{d \mid 2 \textit{ garantit } d\}.$$

On voit donc que, à ε près, le joueur 1 peut faire en sorte que le paiement soit au moins $\underline{v}(g)$ alors que le joueur 2 peut assurer que le paiement ne soit pas plus que $\overline{v}(g)$. Lorsque ces deux quantités sont égales et que les deux joueurs sont rationnels, on peut penser que l'issue du jeu sera très proche de leur valeur commune. Ceci conduit aux définitions suivantes :

Définition 2.3

– On dit que le jeu (S, T, g) *a une valeur* lorsque $\underline{v}(g) = \overline{v}(g)$ et on note $v(g)$ cette valeur.

– Soit $\varepsilon \geqslant 0$, on dit que $s \in S$ est une *stratégie ε-optimale* (ou simplement optimale si $\varepsilon = 0$) du joueur 1 si la stratégie s garantit $\underline{v}(g) - \varepsilon$:

$$\forall\, t \in T, \quad g(s,t) \geqslant \underline{v}(g) - \varepsilon.$$

– On dit que $t \in T$ est une *stratégie ε-optimale* du joueur 2 si la stratégie t garantit $\overline{v}(g) + \varepsilon$:

$$\forall\, t \in T, \quad g(s,t) \leqslant \overline{v}(g) + \varepsilon.$$

– Un couple $(\overline{s}, \overline{t})$ est un point selle si :

$$\forall\,(s,t), \quad g(s,\overline{t}) \leqslant g(\overline{s},\overline{t}) \leqslant g(\overline{s},t).$$

On a les propriétés suivantes :

Proposition 2.4. — *– Il existe une stratégie optimale pour le joueur 1 (resp. le joueur 2) si et seulement si*

$$\underline{v}(g) = \max_{s\in S}\inf_{t\in T} g(s,t) \quad (\textit{resp. } \overline{v}(g) = \min_{t\in T}\sup_{s\in S} g(s,t)).$$

– S'il existe un point selle, alors le jeu a une valeur, les joueurs ont des stratégies optimales et on a :

$$v(g) = \max_{s\in S}\min_{t\in T} g(s,t) = \min_{t\in T}\max_{s\in S} g(s,t).$$

De plus si on note O^i *l'ensemble des stratégies optimales du joueur* $i = 1, 2$ *et* S *l'ensemble des points selles, on a* $S = O^1 \times O^2$.

2.2. Un théorème d'existence. — Le théorème principal d'existence de valeur dans les jeux à somme nulle est dû à Sion (1958) et généralise le célèbre théorème du minimax de Von Neumann (1928) (voir le paragraphe 4 ci-dessous).

Étant donné une fonction $f : X \to \mathbb{R}$ où X est un espace topologique, on dit que f est semi-continue supérieurement (scs) si pour tout réel c, l'ensemble $\{x \mid f(x) \geqslant c\}$ est fermé. On dit que f est semi-continue inférieurement (sci) si $-f$ est scs. On vérifie facilement que l'infimum d'une famille de fonctions scs est scs et que si f est scs pour toute suite $(x_n)_n$ qui converge vers x, $\overline{\lim}_n f(x_n) \leqslant f(x)$.

Lorsque X est un sous-ensemble convexe d'un espace vectoriel, on dit que f est quasi-concave si pour tout réel c, l'ensemble $\{x \mid f(x) \geqslant c\}$ est convexe. On dit que f est quasi-convexe si $-f$ est quasi-concave.

Théorème 2.5. — *Soit $G = (S, T, g)$ un jeu à somme nulle. On suppose :*

(i) *S et T sont des sous-ensembles convexes non-vides d'espaces vectoriels topologiques, l'un d'entre eux étant compact ;*

(ii) *pour tout $(s_0, t_0) \in S \times T$ et tout réel c, les ensembles $\{s \in S \mid g(s, t_0) \geqslant c\}$, $\{t \in T \mid g(s_0, t) \leqslant c\}$ sont convexes et fermés.*

Alors :

(a) $\underline{v}(g) = \overline{v}(g)$ *;*

(b) *si S (resp. T) est compact, le joueur 1 (resp. 2) a une stratégie optimale ;*

(c) *l'ensemble des stratégies ε-optimales ($\varepsilon \geqslant 0$) pour le joueur 1 (resp. 2) est convexe fermé.*

L'outil principal de la démonstration est le théorème de séparations des convexes. On utilise le lemme suivant :

Lemme 2.6. — *Soit $(F_i)_{i=1}^n$ une famille de sous-ensembles convexes compacts non-vides d'un espace vectoriel topologique localement convexe séparé (evtlcs), telle que $\bigcup_{i=1}^n F_i$ est convexe et pour tout j, $\bigcap_{i \neq j} F_i \neq \varnothing$. On a alors : $\bigcap_{i=1}^n F_i \neq \varnothing$.*

Démonstration. — Par récurrence sur n, le cas $n = 1$ étant clair, supposons le résultat vrai pour $n - 1$ et faux pour n. Prenons une famille $(F_i)_{i=1}^n$ vérifiant les hypothèses du lemme, posons $F = \bigcap_{i<n} F_i$

et supposons $F \cap F_n = \varnothing$. En utilisant le théorème de Hahn-Banach, il existe un hyperplan fermé H qui sépare F et F_n et comme les ensembles sont compacts on peut obtenir une séparation stricte : $F \cap H = \varnothing$ et $F_n \cap H = \varnothing$.

Nous pouvons appliquer l'hypothèse de récurrence à la famille $(F_i \cap H)_{i<n}$ car $\bigcup_{i<n}(F_i \cap H) = (\bigcup_{i=1}^{n} F_i) \cap H$ est convexe. Comme $\bigcap_{i<n}(F_i \cap H) = F \cap H = \varnothing$, d'après l'hypothèse de récurrence, il existe un indice $j < n$ tel que $\bigcap_{i<n, i\neq j}(F_i \cap H) = \varnothing$. Mais l'ensemble $G = \bigcap_{i<n, i\neq j} F_i$ contient F et $G \cap F_n \neq \varnothing$ par hypothèse, donc G doit rencontrer H d'où la contradiction. □

Nous pouvons alors démontrer le théorème de Sion pour des polytopes.

Lemme 2.7. — *Supposons que les hypothèses du théorème 2.5 sont satisfaites et qu'il existe deux ensembles finis S_0 et T_0 tels que S soit égal à l'enveloppe convexe $\operatorname{co} S_0$ de S_0 et T soit égal à $\operatorname{co} T_0$. Alors :*

$$\sup_{s\in S} \inf_{t\in T_0} g(s,t) \geqslant \inf_{t\in T} \sup_{s\in S_0} g(s,t)$$

Démonstration. — Remarquons que S et T peuvent s'identifier à des convexes compacts de dimension finie et donc être munis de l'unique topologie d'evtlcs correspondante. Celle-ci étant nécessairement plus fine que la topologie de départ, les hypothèses du théorème restent satisfaites. Supposons par l'absurde qu'il existe un réel c tel que :

$$\sup_{s\in S} \inf_{t\in T_0} g(s,t) < c < \inf_{t\in T} \sup_{s\in S_0} g(s,t)$$

Ceci revient à :

$$\forall\, s \in \operatorname{co} S_0,\ \exists\, t \in T_0, \quad g(s,t) < c$$
$$\forall\, t \in \operatorname{co} T_0,\ \exists\, s \in S_0, \quad g(s,t) > c.$$

Supposons de plus que (S_0, T_0) soit minimal (pour l'inclusion) parmi les paires d'ensembles non-vides vérifiant ces conditions. Posons alors pour tout $s \in S_0$, $T_s = \{t \in T \mid g(s,t) \leqslant c\}$, cet ensemble est convexe fermé. Les conditions ci-dessus impliquent $\bigcap_{s\in S_0} T_s = \varnothing$ et comme on a pris (S_0, T_0) minimal, $\bigcap_{s\in S_0, s\neq s_0} T_s \neq \varnothing$ pour tout $s_0 \in S_0$. Le lemme 2.6 implique que $\bigcup_{s\in S_0} T_s \neq T$ et donc il existe $t^* \in T$ tel que $\forall\, s \in S_0$, $g(s,t^*) > c$. Enfin, par convexité de l'ensemble $\{s \in S \mid g(s,t^*) \geqslant \min_{s'\in S_0} g(s',t^*)\}$, on a $g(s,t^*) > c$ pour tout $s \in S$. Échangeant les rôles des deux joueurs nous obtenons l'existence de s^* telle que $g(s^*, t) < c$ pour tout $t \in T$ et donc $g(s^*, t^*) < c < g(s^*, t^*)$, d'où la contradiction. $\square$

Démonstration du théorème 2.5. — Supposons par l'absurde que l'on ait :

$$\sup_{s \in S} \inf_{t \in T_0} g(s,t) < \inf_{t \in T} \sup_{s \in S_0} g(s,t)$$

On peut alors trouver $c \in \mathbb{R}$ tels que :

$$\begin{aligned} &\forall\, s \in S,\ \exists\, t \in T, \quad g(s,t) < c \\ &\forall\, t \in T,\ \exists\, s \in S, \quad g(s,t) > c \end{aligned}$$

Supposons S compact, les ensembles $S_t = \{s \in S \mid g(s,t) < c\}$ forment un recouvrement ouvert de S. Soit $(S_t)_{t \in T_0}$ un sous-recouvrement fini, on a :

$$\begin{aligned} \forall\, s \in S,\ \exists\, t \in T_0, &\quad g(s,t) < c \\ \forall\, t \in \operatorname{co} T_0,\ \exists\, s \in S, &\quad g(s,t) > c \end{aligned}$$

Remplaçant T par $\operatorname{co} T_0$ (qui est compact), nous pouvons échanger les rôles de S et T et obtenir un ensemble fini $S_0 \subset S$ tel que :

$$\begin{aligned} \forall\, s \in \operatorname{co} S_0,\ \exists\, t \in T_0, &\quad g(s,t) < c \\ \forall\, t \in \operatorname{co} T_0,\ \exists\, s \in S_0, &\quad g(s,t) > c \end{aligned}$$

et on conclut grâce au lemme 2.7.

Les points (b) et (c) du théorème 2.5 s'obtiennent directement par la semi-continuité supérieure et la quasi-concavité de $s \mapsto g(s,t)$ (et les conditions duales pour $t \mapsto g(s,t)$). $\square$

2.3. Continuité et dérivabilité de la valeur. — Fixons deux espaces compacts S et T. Soit E l'espace vectoriel des applications $g : S \times T \to \mathbb{R}$ bornées. On pose $\|g\| = \sup_{(s,t) \in S \times T} |g(s,t)|$, pour $g \in E$.

Proposition 2.8

(1) *Pour toutes applications $f, g \in E$:*

$$|\underline{v}(g) - \underline{v}(f)| \leqslant \|g - f\| \quad \text{et} \quad |\overline{v}(g) - \overline{v}(f)| \leqslant \|g - f\|\,.$$

(2) *Soit* $(f_\alpha)_\alpha$ *une suite généralisée de* E *telle que* $\|f_\alpha - f\| \to 0$ *avec* $f \in E$ *telle que pour tout* t*, l'application partielle* $f(\cdot, t)$ *est scs. Soit* $(s_\alpha)_\alpha$ *une suite généralisée de* S *telle que* s_α *est une stratégie* ε_α*-optimale du joueur 1 pour* f_α *avec* $\varepsilon_\alpha > 0$*,* $\varepsilon_\alpha \to 0$*. Alors, tout point d'accumulation de* $(s_\alpha)_\alpha$ *est une stratégie optimale du joueur 1 pour* f*.*

Démonstration

(1) Il est immédiat que pour toutes constantes $a > 0$ et $b \in \mathbb{R}$, si l'on pose $af + b$ l'application $(s,t) \mapsto af(s,t) + b$, on a : $\underline{v}(af+b) = a\underline{v}(f) + b$. De plus si pour tout (s,t), $f(s,t) \leqslant g(s,t)$, alors $\underline{v}(f) \leqslant \underline{v}(g)$. Comme $f(s,t) \leqslant g(s,t) + \|g - f\|$, on a $\underline{v}(f) \leqslant \underline{v}(g) + \|g - f\|$, d'où le résultat.

(2) Supposons que pour tout $t \in T$, $f_\alpha(s_\alpha, t) \geqslant \underline{v}(f_\alpha) - \varepsilon_\alpha$. D'après le point précédent $\underline{v}(f_\alpha) \to \underline{v}(f)$ et donc pour tout $t \in T$, $\underline{\lim} f_\alpha(s_\alpha, t) \geqslant \underline{v}(f)$. Comme f_α converge uniformément vers f, $\underline{\lim} f_\alpha(s_\alpha, t) = \underline{\lim} f(s_\alpha, t)$. Soit s un point d'accumulation de $(s_\alpha)_\alpha$, puisque $f(\cdot, t)$ est scs, $f(s,t) \geqslant \underline{\lim} f(s_\alpha, t) \geqslant \underline{v}(f)$. □

En utilisant ce résultat, nous obtenons une formule pour la dérivée directionnelle de la fonction valeur (Mills, 1956). Soient f, g dans E, toutes deux scs en s et sci en t. On suppose que pour tout $\varepsilon \geqslant 0$, $v(f + \varepsilon g)$ existe. Soit $S(f) = O^1(f) \times O^2(f)$ l'ensemble des points selles de f, on pose $v_f(g) = v(g|_{S(f)})$ (lorsque cette valeur existe).

Proposition 2.9. — *Avec ces hypothèses,* $v_f(g)$ *existe et on a :*

$$\lim_{\varepsilon \to 0^+} \frac{v(f + \varepsilon g) - v(f)}{\varepsilon} = v_f(g)$$

Démonstration. — Soit t une stratégie optimale du joueur 2 pour f ($t \in O^2(f)$), et s_ε une stratégie optimale du joueur 1 pour $f + \varepsilon g$. Il vient :

$$\begin{aligned} v(f + \varepsilon g) &\leqslant f(s_\varepsilon, t) + \varepsilon g(s_\varepsilon, t) \\ &\leqslant v(f) + \varepsilon g(s_\varepsilon, t) \end{aligned}$$

D'où $\frac{v(f+\varepsilon g) - v(f)}{\varepsilon} \leqslant g(s_\varepsilon, t)$, et comme ceci est vrai pour toute stratégie optimale du joueur 2 pour f, on obtient

$$\frac{v(f + \varepsilon g) - v(f)}{\varepsilon} \leqslant \inf_{t \in O^2(f)} g(s_\varepsilon, t).$$

Comme $g(\cdot, t)$ est scs, $\inf_{t\in O^2(f)} g(\cdot, t)$ l'est aussi. D'ou, pour tout point d'accumulation s de $(s_\varepsilon)_{\varepsilon\to 0}$, $\overline{\lim}_{\varepsilon\to 0^+} \inf_{t\in O^2(f)} g(s_\varepsilon, t) \leqslant \inf_{t\in O^2(f)} g(s,t)$, et $s \in O^1(f)$. Donc,

$$\overline{\lim}_{\varepsilon\to 0^+} \frac{v(f+\varepsilon g) - v(f)}{\varepsilon} \leqslant \sup_{s\in O^1(f)} \inf_{t\in O^2(f)} g(s,t)$$

On obtient l'égalité en échangeant les rôles des deux joueurs. □

3. Jeux à somme non nulle

La notion centrale de solution pour les jeux sous forme normale est l'équilibre de Nash (1950), qui généralise la notion de point selle. Soit $G = (N, (A^i)_{i\in N}, (g^i)_{i\in N})$ un jeu.

3.1. Équilibres de Nash

Définition 3.1. — Un équilibre de Nash du jeu G est un profil de stratégies $a = (a^i)_{i\in N}$ tel que :

$$\forall\, i \in N,\ \forall\, b^i \in A^i, \quad g^i(b^i, a^{-i}) \leqslant g^i(a)$$

Cette définition peut se reformuler de différentes façons. Par exemple, étant donné un profil de stratégies a, on dit que le joueur i a une *déviation profitable* par rapport à a, s'il existe b^i telle que $g^i(b^i, a^{-i}) > g^i(a)$. Un équilibre de Nash est un profil de stratégies pour lequel il n'existe pas de déviation profitable. C'est donc un point tel que, si tous les joueurs savent qu'on va jouer a, alors chacun a effectivement intérêt à le jouer. On peut remarquer que pour les jeux à somme nulle, équilibres de Nash et points selles coïncident.

Une autre reformulation va suggérer une méthode de calcul et de démonstration d'existence.

Définition 3.2

– Pour chaque joueur i et profil d'action de ses adversaires a^{-i}, on dit que a^i est *meilleure réponse* contre a^{-i} si :

$$\forall\, b^i \in A^i, \quad g^i(b^i, a^{-i}) \leqslant g^i(a^i, a^{-i}).$$

– On appelle *correspondance de meilleure réponse du joueur* i, l'application MR^i de A^{-i} dans les parties de A^i, qui à a^{-i} associe l'ensemble des meilleures réponses du joueur i.

– On appelle *correspondance de meilleure réponse* du jeu G, l'application $\mathrm{MR} : A \to 2^A$ définie par $\mathrm{MR}(a) = \prod_{i \in N} \mathrm{MR}^i(a^{-i})$.

On voit alors que a est un équilibre de Nash de G si et seulement si $a \in \mathrm{MR}(a)$. On dit que a est un *point fixe* de la correspondance de meilleure réponse. La procédure de calcul des équilibres est donc la suivante : tracer le graphe de la correspondance de meilleure réponse de chaque joueur, et chercher l'intersection des graphes.

Calcul des équilibres de Nash dans le jeu de Cournot. — On prend $G = (A^1, A^2, g^1, g^2)$ avec $A^1 = A^2 = \mathbb{R}_+$ et

$$g^i(a^1, a^2) = a^i(\alpha - \beta(a^1 + a^2))^+ - \gamma a^i \quad (\alpha, \beta, \gamma > 0,\ \alpha > \gamma).$$

On a :

$$\mathrm{MR}^1(a^2) = \begin{cases} \left\{ \frac{\alpha-\gamma}{2\beta} - \frac{a^2}{2} \right\} & \text{si } a^2 \leqslant \frac{\alpha-\gamma}{\beta}, \\ \{0\} & \text{sinon.} \end{cases}$$

La correspondance de meilleure réponse du joueur 2 est identique.

On voit alors facilement que ($a^1 \in \mathrm{MR}^1(a^2)$ et $a^2 \in \mathrm{MR}^2(a^1)$) si et seulement si, $a^1 = a^2 = \frac{\alpha-\gamma}{3\beta}$.

Lien avec l'élimination itérée des stratégies strictement dominées

Il découle des définitions qu'une stratégie strictement dominée n'est jamais meilleure réponse et donc n'est jamais utilisée dans un équilibre de Nash. Il s'ensuit que l'ensemble des équilibres de Nash est inchangé lors de l'élimination des stratégies strictement dominées. Pour trouver les équilibres d'un jeu on peut donc effectuer l'EISSD au préalable.

3.2. Un théorème d'existence. — La dernière reformulation suggère également une méthode de preuve d'existence d'équilibre : étant donné une correspondance de A dans A, c'est-à-dire une application F de A dans ses parties, on cherche un point fixe, c'est-à-dire un point a tel que $a \in F(a)$. Lorsque F est *univoque* : il existe une application $f : A \to A$ telle que, pour tout a, $F(a) = \{f(a)\}$, ceci revient à trouver un point fixe de l'application f. Les deux théorèmes de point fixes les plus utilisés sont le théorème de Brouwer et ses généralisations par Kakutani (1941) et Glicksberg (1952).

Théorème 3.3 **(Théorème de Brouwer).** — *Soit C un convexe compact non-vide de $\mathbb{R}^k$ et $f : C \to C$ continue, alors f admet un point fixe : il existe $c \in C$ tel que $c = f(c)$.*

Le théorème de Brouwer a été généralisé aux correspondances par Kakutani (1941) dans les espaces vectoriels de dimension finie et par Glicksberg (1952) dans les espaces vectoriels topologiques.

Théorème 3.4 **(Kakutani-Glicksberg).** — *Soit C un convexe compact non-vide d'un evtlcs et F une correspondance de C dans C telle que*

(i) *pour tout $c \in C$, $F(c)$ est un convexe compact non-vide,*
(ii) *le graphe de F, à savoir $\{(c, d) \in C \times C \mid c \in F(d)\}$, est fermé.*

Alors, il existe $c \in C$ tel que $c \in F(c)$.

On peut aussi formuler ce théorème de la façon suivante : Soit X un sous-ensemble fermé de $C \times C$ tel que pour tout $c \in C$, la section de X au-dessus de c ($i.e.$ $\{d \in C \mid (c, d) \in X\}$) est un convexe compact non-vide. Alors, X coupe la diagonale de $C \times C$.

Donnons une démonstration de ce résultat en dimension finie.

Démonstration. — On suppose que C est un convexe compact non-vide de $\mathbb{R}^k$. Pour tout entier n, il existe une famille finie de points $(x_i^n)_i$ de C telle que $C \subset \bigcup_i B(x_i^n, 1/n)$, ou $B(x_i^n, 1/n)$ désigne la boule ouverte de centre x_i^n et de rayon $1/n$. Notons $B^c(x_i^n, 1/n)$ le complémentaire dans $\mathbb{R}^k$ de cette boule et fixons, pour chaque x_i^n un élément $y_i^n \in F(x_i^n)$. Pour tout x dans C on pose :

$$f^n(x) = \sum_i \frac{d(x, B^c(x_i^n, 1/n))}{\sum_j d(x, B^c(x_j^n, 1/n))} \, y_i^n$$

Pour tout x dans C, il existe i tel que $x \in B(x_i^n, 1/n)$, d'où $\sum_j d(x, B^c(x_j^n, 1/n)) > 0$. L'application f^n est donc bien définie et continue sur C, et $f^n(x) \in C$ par convexité. D'après le théorème de Brouwer, il existe $x^n \in C$ tel que $x^n = f^n(x^n)$. Comme $d(x, B^c(x_i^n, 1/n)) > 0$ si et seulement si $d(x, x_i^n) < 1/n$, on a :

$$x^n = \sum_{i \,\mid\, d(x^n, x_i^n) < 1/n} \frac{d(x^n, B^c(x_i^n, 1/n))}{\sum_j d(x^n, B^c(x_j^n, 1/n))} \, y_i^n$$

Appliquons maintenant le théorème de Carathéodory : toute combinaison convexe des $(y_i^n)_i$ peut s'écrire comme combinaison d'au plus

$k+1$ points parmi cette famille. Cela fournit la conclusion suivante, pour tout n, il existe :

- $x^n, x_1^n, \dots, x_{k+1}^n$ éléments de C, tels que $d(x^n, x_i^n) < 1/n$ $(\forall\, i)$;
- $y_1^n, \dots, y_{k+1}^n$ avec $y_i^n \in F(x_i^n)$;
- $\lambda_1^n, \dots, \lambda_{k+1}^n$, $\lambda_i^n \geqslant 0$, $\sum_i \lambda_i^n = 1$;
- $x^n = \sum_{i=1}^{k+1} \lambda_i^n y_i^n$.

Par compacité, quitte à extraire une sous-suite, supposons que $x^n \to x$, $x_i^n \to x_i$, $y_i^n \to y_i$, $\lambda_i^n \to \lambda_i$. Alors, $x = x_i$ pour tout i, $y_i \in F(x)$ car le graphe de F est fermé et $x = \sum_{i=1}^{k+1} \lambda_i y_i$ appartient à $F(x)$ qui est convexe. □

On déduit de ce résultat un théorème d'existence d'équilibres (Glicksberg, 1952, généralisant Nash, 1950).

***Théorème 3.5* (Théorème de Glicksberg-Nash)**
Soit $G = (N, (A^i)_{i\in N}, (g^i)_{i\in N})$ un jeu tel que : pour tout $i \in N$, A^i est un convexe compact non-vide dans un evtlcs, $g^i : A \to \mathbb{R}$ est continue et pour tout $a^{-i} \in A^{-i}$, l'application partielle $a^i \mapsto g^i(a^i, a^{-i})$ est quasi-concave.

Alors, l'ensemble des équilibres de Nash de G est un fermé non-vide.

Démonstration. — On vérifie simplement que le théorème 3.4 s'applique. Les hypothèses de compacité, de continuité et quasi-concavité des fonctions de paiements assurent que $\mathrm{MR}^i(a^{-i})$ est convexe fermé non-vide $(\forall\, i, \forall\, a^{-i})$, $\mathrm{MR}(a)$ est donc bien convexe compact non-vide $(\forall\, a)$. Le graphe de la correspondance MR s'écrit :

$$\big\{(a,b) \in A \times A \mid \forall\, i \in N, \forall\, c^i \in A^i,\ g^i(a^i, b^{-i}) \geqslant g^i(c^i, b^{-i})\big\}$$

Les fonctions de paiement étant continues,

$$\big\{(a,b) \in A \times A \mid g^i(a^i, b^{-i}) \geqslant g^i(c^i, b^{-i})\big\}$$

est fermé $(\forall\, i \in N, \forall\, c^i \in A^i)$. Le graphe de la correspondance MR est donc une intersection de fermés. L'ensemble des équilibres de Nash est donc non vide et il est clairement fermé. □

4. Jeux finis et stratégies mixtes

On dit qu'un jeu $G = (N, (A^i)_{i\in N}, (g^i)_{i\in N})$ est fini lorsque les ensembles de stratégies sont tous des ensembles finis. Le jeu *matching pennies* (ci-dessus) n'admet pas de point selle ni de valeur. Toutefois, le but du joueur 1 étant de deviner l'action du joueur 2, et le but du joueur 2 étant de l'en empêcher, la meilleure « stratégie » du joueur 2 est de choisir son action aléatoirement et de façon équiprobable pour être le plus difficilement prédictible. On est donc naturellement conduit à élargir le modèle et à autoriser les joueurs à choisir leur action au hasard et avec les probabilités de leur choix, on parle de stratégies mixtes. Outre l'interprétation directe selon laquelle les joueurs utilisent un dispositif aléatoire pour choisir leur action, une stratégie mixte d'un joueur peut se voir comme la croyance qu'ont ses adversaires sur son action : le joueur lui, sait parfaitement quelle action il va jouer, les probabilités sont l'expression de l'incertitude des adversaires. On peut également donner une interprétation statistique en voyant les joueurs non pas comme des agents individuels mais comme des populations d'individus. La probabilité d'une stratégie s'interprète alors comme la proportion d'individus jouant cette stratégie. Cette interprétation est particulièrement fructueuse en biologie évolutionnaire où une stratégie représente un caractère génétique, et une stratégie mixte, la distribution statistique des gènes dans la population.

Soit E un ensemble fini. On note $\Delta(E)$ l'ensemble des distributions de probabilités sur E, que l'on identifiera avec des vecteurs de $\mathbb{R}^E$ à coordonnées positives de somme 1.

$$\Delta(E) = \left\{p \in \mathbb{R}^E \mid \forall\, e \in E,\, p(e) \geqslant 0,\, \textstyle\sum_{e\in E} p(e) = 1\right\}$$

Cet ensemble est l'enveloppe convexe de la base canonique de $\mathbb{R}^E$, c'est donc un convexe compact. L'ensemble E s'injecte naturellement dans $\Delta(E)$ en identifiant $e \in E$ à la masse de Dirac $\delta_e \in \Delta(E)$ qui vérifie $\delta_e(e') = 1$ si $e' = e$ et 0 sinon.

Définition 4.1. — Soit $G = (N, (A^i)_{i\in N}, (g^i)_{i\in N})$ un jeu fini.

– On appelle *stratégie mixte* du joueur i, une probabilité $\sigma^i \in \Delta(A^i)$ et *stratégie pure* une action $a^i \in A^i$. L'ensemble des stratégies mixtes du joueur i est $\Delta(A^i)$.

– Étant donné un profil de stratégies mixtes $\sigma = (\sigma^i)_{i\in N} \in \prod_{i\in N}\Delta(A^i)$, on appelle *paiement espéré* du joueur i la quantité :

$$g^i(\sigma) = \sum_{a\in A} \big(\textstyle\prod_{i\in N}\sigma^i(a^i)\big) g^i(a)$$

Ceci définit une extension de l'application g^i de $\prod_{i\in N} A^i$ à $\prod_{i\in N}\Delta(A^i)$ que l'on note encore g^i : on appelle cette application *extension multilinéaire* ou *extension mixte* de g^i.

– On appelle *extension mixte* du jeu G le jeu $(N, (\Delta(A^i))_{i\in N}, (g^i)_{i\in N})$.

– Un équilibre de Nash de l'extension mixte de G s'appellera équilibre de G en stratégies mixtes.

Les jeux finis joués en stratégies mixtes sont les premiers jeux pour lesquels ont été démontrés des résultats d'existence, Von Neumann (1928) pour les jeux à somme nulle et Nash (1950) pour les jeux à somme non-nulle.

Théorème 4.2 **(Théorème du MinMax, Von Neumann1928)**
Tout jeu fini (A^1, A^2, g) admet un point selle en stratégies mixtes. En particulier le jeu admet une valeur v et les deux joueurs ont des stratégies optimales. De plus,

$$v = \max_{\sigma^1\in\Delta(A^1)} \min_{a^2\in A^2} g(\sigma^1, a^2) = \min_{\sigma^2\in\Delta(A^2)} \max_{a^1\in A^1} g(a^1, \sigma^2)$$

Démonstration. — L'existence de valeur et de stratégies optimales pour les deux joueurs découle directement du théorème de Sion. Les ensembles de stratégies mixtes sont convexes et compacts et les extensions mixtes des fonctions de paiement sont multi-linéaires, donc continues et possèdent les propriétés requises de quasi-concavité.

De plus, si l'on fixe une stratégie mixte σ^1 du joueur 1, $g(\sigma^1, \sigma^2)$ est une application linéaire sur le polytope $\Delta(A^2)$. Elle atteint donc son minimum en un point extrême, c'est-à-dire en une stratégie pure. □

Théorème 4.3 **(Théorème de Nash, 1950).** — *Tout jeu fini $G = (N, (A^i)_{i\in N}, (g^i)_{i\in N})$ admet un équilibre de Nash en stratégies mixtes.*

Démonstration. — De la même façon, les ensembles de stratégies mixtes sont convexes et compacts et les extensions mixtes des fonctions de paiement sont multi-linéaires, donc continues et possèdent les propriétés requises de quasi-concavité. □

Extension : ensembles infinis de stratégies. — La notion de stratégie mixte peut s'étendre à des espaces mesurables d'actions, en posant des conditions de mesurabilité et intégrabilité sur les fonctions de paiements. Un cadre simple à traiter est le cas des espaces métriques compacts. Si E est un espace métrique compact, on le munit de la tribu borélienne et on pose $\Delta(E)$ l'ensemble des mesures probabilités boréliennes sur E. Muni de la topologie faible-$*$ (la plus petite topologie rendant continues les applications $\mu \mapsto \int_E f d\mu$ avec $f : E \to \mathbb{R}$ continue), $\Delta(E)$ est métrisable et compact.

On dit qu'un jeu $G = (N, (A^i)_{i\in N}, (g^i)_{i\in N})$ est *compact* si chaque A^i est métrique compact et chaque g^i est continue. L'extension mixte de g^i est définie sur $\prod_{i\in N} \Delta(A^i)$ par :

$$g^i(\sigma^1, \ldots, \sigma^n) = \int g^i d\sigma^1 \cdots d\sigma^n$$

On définit ainsi, comme dans le cas fini, l'extension mixte du jeu G. On obtient alors un résultat d'existence en stratégies mixtes.

Théorème 4.4. — *Tout jeu compact $G = (N, (A^i)_{i\in N}, (g^i)_{i\in N})$ admet un équilibre de Nash en stratégies mixtes.*

Démonstration. — Comme dans le cas fini, les espaces de stratégies mixtes sont des convexes compacts, les fonctions de paiements sont continues et multi-linéaires, l'aspect quasi-concave est donc garanti. □

Revenons aux jeux finis pour donner une caractérisation des équilibres de Nash en stratégies mixtes.

Théorème 4.5. — *Soit $G = (N, (A^i)_{i\in N}, (g^i)_{i\in N})$ un jeu fini. Un profil de stratégies mixtes $\sigma = (\sigma^i)_{i\in N}$ est un équilibre de Nash en stratégies mixtes de G si et seulement si :*

$$\forall\, i \in N,\ \forall\, a^i \in A^i, \quad (\sigma^i(a^i) > 0 \Rightarrow a^i \in \mathrm{MR}^i(\sigma^{-i}))$$

Autrement dit, toutes les stratégies pures jouées avec une probabilité non-nulle sont des meilleures réponses au profil de stratégies des autres joueurs, et en particulier elles donnent le même paiement. Ce résultat peut permettre un calcul assez simple des équilibres de Nash.

Dans le jeu de coordination :

	A	B
A	$1,1$	$0,0$
B	$0,0$	$1,1$

les profils (A, A) et (B, B) sont deux équilibres de Nash en stratégies pures. Pour déterminer les équilibres en stratégies mixtes, remarquons d'abord qu'il n'y a pas d'équilibre dans lequel un joueur joue une stratégie pure et l'autre joue une stratégie strictement mixte : les deux actions sont jouées avec probabilité strictement positive. En effet, dès qu'un joueur joue une stratégie pure, l'autre a une unique meilleure réponse qui est pure. On cherche donc les équilibres dans lesquels les deux joueurs jouent une stratégie strictement mixte. Chaque joueur doit alors avoir le même paiement espéré en jouant A ou B. Les équations d'égalisations des paiement donnent que chaque joueur joue la stratégie mixte $(\frac{1}{2}, \frac{1}{2})$. Ce jeu a donc exactement trois équilibres.

Démonstration. — Soit $\sigma = (\sigma^i)_{i \in N}$ un équilibre de G en stratégies mixtes et i un joueur. Par multi-linéarité :

$$g^i(\sigma^i, \sigma^{-i}) = \sum_{a^i \in A^i} \sigma^i(a^i)\, g^i(a^i, \sigma^{-i})$$

Comme σ est un équilibre, pour toute action a^i,

$$g^i(a^i, \sigma^{-i}) \leqslant g^i(\sigma^i, \sigma^{-i}) = \max_{\tau^i \in \Delta(A^i)} g^i(\tau^i, \sigma^{-i})$$

D'où, $g^i(a^i, \sigma^{-i}) = g^i(\sigma^i, \sigma^{-i})$ dès que $\sigma^i(a^i) > 0$. Réciproquement, si pour tout $i \in N$ et tout a^i tel que $\sigma^i(a^i) > 0$, on a $g^i(a^i, \sigma^{-i}) = \max_{\tau^i \in \Delta(A^i)} g^i(\tau^i, \sigma^{-i})$, alors

$$g^i(\sigma^i, \sigma^{-i}) = \sum_{a^i \in A^i} \sigma^i(a^i) g^i(a^i, \sigma^{-i}) = \max_{\tau^i \in \Delta(A^i)} g^i(\tau^i, \sigma^{-i})$$

et donc σ est un équilibre. $\square$

5. Jeux à information parfaite

Nous décrivons ici une classe de jeux contenant les jeux de plateaux traditionnels (échecs, dames, jeu de Go) : ce sont les *jeux à information parfaite* dans lesquels les joueurs jouent séquentiellement (et non pas simultanément) en ayant pleinement connaissance de l'état de la partie au moment de jouer. Ces interactions sont modélisées par des arbres de décision.

Définition 5.1. — On appelle *arbre* un ensemble de nœuds ou d'histoires H muni de :

– une racine ou histoire initiale $h_0 \in H$,

– une relation binaire sur H, $h \prec h'$ (h est le prédécesseur de h') telle que :

(i) tout $h \in H \setminus \{h_0\}$ a un unique prédécesseur $\pi(h)$ et il existe un entier $k \geqslant 1$ tel que $\pi^k(h) = h_0$,

(ii) h_0 n'a pas de prédécesseur.

Un arbre est un graphe connexe tel que pour toute paire d'histoires h, h', il existe un unique chemin de h à h'.

Partant d'un arbre, on définit un jeu en spécifiant quel joueur joue en chaque nœud, ce joueur choisit alors une branche issue de ce nœud. Posons $A(h) = \{h' : \ h \prec h'\}$ l'ensemble des successeurs de h. On dit que h est une histoire terminale si $A(h) = \varnothing$ et on pose H_T l'ensemble des histoires terminales.

Définition 5.2. — Un *jeu à information parfaite* est donné par un ensemble de joueurs N, un arbre $(H, h_0, \prec)$, une application $\iota : H \setminus H_T \to N$ et pour chaque joueur $i \in N$, une application $u^i : H_T \to \mathbb{R}$.

Le déroulement du jeu est le suivant. Le jeu commence en h_0. Au nœud $h \in H \setminus H_T$, le joueur $\iota(h)$ choisit un successeur h' de h dans l'ensemble $A(h)$, le jeu passe au nœud h'. Au nœud $h \in H_T$, le jeu est terminé et chaque joueur i reçoit le paiement $u^i(h)$.

Une stratégie du joueur i dans un tel jeu est le choix d'une action pour chaque nœud qui lui est attribué (le joueur i écrit un programme qui prévoit quoi jouer dans chaque cas possible). L'ensemble des stratégies du joueur i est donc : $S^i = \prod_{\{h:\ \iota(h)=i\}} A(h)$. Un profil de stratégies $s = (s^i)_{i \in N}$ induit une unique suite de nœuds $(h_0, h_1, \dots, h_t, \dots)$: si $i = \iota(h_t)$ alors, $h_{t+1} = s^i(h_t)$. Si il existe un rang t tel que $h_t \in H_T$, la suite s'arrête et on pose $h_T(s)$ le nœud terminal ainsi atteint. Lorsque l'arbre a un nombre fini de nœuds, on atteint toujours un nœud terminal.

Définition 5.3. — Soit $(N, (H, h_0, \prec), \iota, (u^i)_{i\in N})$ un jeu à information parfaite fini (N et H sont des ensembles finis), le jeu sous forme stratégique associé est le jeu $G = (N, (S^i)_{i\in N}, (g^i)_{i\in N})$ où $g^i(s) = u^i(h_T(s))$.

On a alors le théorème suivant prouvé par Zermelo (1912) pour le jeu d'échecs et basé sur l'algorithme de Kuhn.

Théorème 5.4. — *Tout jeu à information parfaite fini admet un équilibre de Nash en stratégies pures. En particulier, si le jeu est à somme nulle, le jeu a une valeur et les deux joueurs ont des stratégies optimales.*

Démonstration. — On procède par récurrence sur le nombre de nœuds de l'arbre. Si l'arbre a un seul nœud, l'assertion est évidente. Supposons qu'elle soit vraie pour tout jeu à information parfaite dont l'arbre a strictement moins que K nœuds, et soit un jeu à information parfaite dont l'arbre a K nœuds. Pour chaque successeur h de la racine h_0 on considère le sous-arbre issu de h. Ceci définit un *sous-jeu* à information parfaite qui a strictement moins que K nœuds et donc, par hypothèse de récurrence possède un équilibre en stratégies pures.

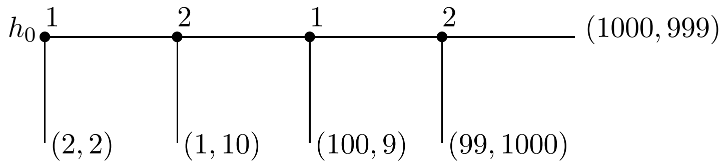
Fixons pour chaque $h \in A(h_0)$, un équilibre s_h du sous-jeu issu de h. Soit alors $i = \iota(h_0)$. On a un équilibre de Nash du jeu en définissant s tel que : au nœud h_0, i choisit h pour lequel $g^i(s_h)$ est maximal, et en tout autre nœud, on suit les stratégies s_h si h a été choisit par i en h_0. □

Cette démonstration décrit également un algorithme de résolution : on résout le jeu aux nœuds précédant les nœuds terminaux, et on remonte vers la racine (procédure de *backward induction* ou récurrence amont).

Ce théorème montre en particulier que le jeu d'échecs est parfaitement résoluble : soit les Blancs ont une stratégie gagnante, soit les Noirs ont une stratégie gagnante, soit les deux peuvent forcer la partie nulle. Seule la capacité de calcul empêche de décider dans quel cas on se trouve.

Donner quelques explications sur ce jeu ?

Exemple : Le jeu du Mille-Pattes de Rosenthal



La résolution de ce jeu par l'algorithme de Kuhn donne le paiement d'équilibre $(2, 2)$.

Exemple : Le Jeu de Gale. — Soit un échiquier de taille $n \times m$, n et m étant deux entiers finis supérieurs à 1. Deux joueurs (1 et 2) choisissent alternativement une case sur l'échiquier. Lorsqu'une case (i, j) est choisie, toutes les cases (i', j') situées au Nord-Est, c'est-à-dire telles que $i' \geqslant i$ et $j' \geqslant j$, sont éliminées de l'échiquier.

Le joueur 1 commence le jeu et est déclaré perdant le joueur contraint de choisir la case $(1, 1)$.

Alors :

(1) Le joueur 1 a une stratégie gagnante.

(2) On construit facilement une stratégie gagnante pour le joueur 1 dans le cas $n \times n$, et dans le cas $m \times 2$.

(3) On n'en connaît toujours pas dans le cas (n, m) général.

Démonstration de (1). — On sait que l'un des deux joueurs a une stratégie gagnante. Supposons que ce soit le joueur 2 et remarquons que, quel que soit le premier coup du joueur 1, la case (n, m) est effacée. Soit alors (i_0, j_0) le coup gagnant du joueur 2 qui suit le coup (n, m) du joueur 1. Ce coup gagnant du joueur 2 aurait pu être joué dès le départ par le joueur 1 d'où la contradiction. □

T. Tomala, Ceremade, Université Paris Dauphine
E-mail : `tomala@ceremade.dauphine.fr`

JEUX RÉPÉTÉS

par

Tristan Tomala

Table des matières

1. Modèle général

Dans un jeu répété, les joueurs interagissent à chaque date $t \in \mathbb{N}^*$. Chaque étape génère un paiement et le paiement global d'un joueur est fonction de la suite de ses paiements d'étapes. Nous allons tout d'abord décrire précisément ce modèle, préciser les notions de stratégies ainsi que les notions de solutions (valeur, équilibres). On rappelle que pour une famille d'ensembles $(E^i)_{i \in N}$, on note $E = \prod_{i \in N} E^i$.

1.1. Description du jeu. — Un *jeu répété* Γ est décrit par les données suivantes :

– Un ensemble de joueurs N ;
– Un ensemble d'états K ;
– Pour chaque joueur $i \in N$, un ensemble d'actions A^i et un ensemble de signaux U^i ;
– Une probabilité initiale p sur $K \times U$;

– Une probabilité de transition $Q : K \times A \to K \times U$;

– Pour chaque joueur $i \in N$, et chaque état $k \in K$, une fonction de paiement $g^{i,k} : A \to \mathbb{R}$.

Tous les ensembles ci-dessus sont finis et non vides.

Le déroulement du jeu est le suivant :

– Avant le début du jeu, la nature tire un état initial et des signaux $(k_1, (u_0^i)_{i\in N}) \in K \times U$ selon la probabilité p. Chaque joueur i observe le signal u_0^i.

– A chaque date $t \geqslant 1$, chaque joueur $i \in N$ choisit une action $a_t^i \in A^i$. Si l'état est $k_t \in K$, la nature tire un état et des signaux $(k_{t+1}, (u_t^i)_{i\in N}) \in K \times U$ selon $Q(\cdot|k_t, a_t)$. Le joueur i observe alors u_t^i et on passe dans l'état k_{t+1}.

La suite des paiements du joueur i est alors : $(g^{i,k_t}(a_t))_{t\geqslant 1}$.

Définition 1.1

– On appelle *jeu répété T fois*, et on note Γ_T, le jeu répété dont le déroulement est décrit ci-dessus et qui s'arrête à la date T. Le paiement final du joueur i est : $\frac{1}{T}\sum_{t=1}^{T} g^{i,k_t}(a_t)$.

– On appelle *jeu escompté au taux $\lambda \in]0,1]$*, et on note Γ_λ, le jeu répété une infinité de fois dont le déroulement est décrit ci-dessus. Le paiement final du joueur i est : $\sum_{t\geqslant 1} \lambda(1-\lambda)^{t-1} g^{i,k_t}(a_t)$.

– On appelle *jeu infiniment répété*, et on note Γ_∞, le jeu répété une infinité de fois dans lequel le paiement final du joueur i est : $\lim_T \frac{1}{T}\sum_{t=1}^{T} g^{i,k_t}(a_t)$ si cette limite existe.

Ce modèle rend compte de plusieurs aspects. Le jeu joué en date t est paramétré par l'état k et l'état évolue en fonctions des actions choisies (selon la probabilité de transition). Les joueurs feront donc face à un dilemme : jouer pour maximiser le paiement courant, ou pour induire des états favorables. Les signaux permettent de modéliser les imperfections et les asymétries d'information qu'ont les joueurs sur l'état et les actions passées de leurs adversaires. On distingue trois grandes classes de jeux répétés :

(1) *Les jeux répétés à information complète* appelés aussi super-jeux dans lesquels l'ensemble des états K est un singleton. Dans ce cas, le jeu répété est la répétition du jeu $G = (N, (A^i)_{i\in N}, (g^i)_{i\in N})$. On dit que le jeu répété est *à observation parfaite* si le signal de chaque

joueur révèle le profil d'actions : si a est le profil d'action joué, a est annoncé à chaque chaque joueur. C'est ce cas que nous étudierons dans la suite de ce texte.

(2) *Les jeux répétés à information incomplète.* Dans ce modèle les transitions sont telles que l'état reste constant égal à l'état initial. Cela revient à dire qu'on a une famille paramétrée de jeux $(G^k)_{k\in K}$, le paramètre du jeu k est tiré selon une probabilité initiale et le jeu G^k est répété. Les joueurs ont des information partielles et asymétriques sur sa valeur, données par les signaux initiaux u_0. Ce cas sera étudié dans le texte de Jérôme Renault (ce volume). On supposera toujours, comme dans le cas précédent, que les actions sont parfaitement observées.

(3) *Les jeux stochastiques.* On suppose dans ce cas que les signaux révèlent parfaitement les actions et les états. Le jeu est alors donné par une famille paramétrée de jeux $(G^k)_{k\in K}$, un état initial $k_1 \in K$ et une probabilité de transition qui détermine l'état suivant en fonction de l'état courant et des actions jouées.

1.2. Stratégies. — On appelle *histoire du jeu en date t*, la suite des états, profils d'actions et profils de signaux survenus avant la date t, c'est-à-dire une suite de la forme $h_t = (u_0, k_1, a_1, u_1, k_2 \dots, a_{t-1}, u_{t-1}, k_t)$. L'ensemble des histoires du jeu en date t est $H_t = U \times K \times (A \times U \times K)^{t-1}$ (par convention, nous supposerons qu'un ensemble à la puissance 0 est un singleton). On appelle *partie du jeu* une suite dans $H_\infty = U \times K \times (A \times U \times K)^{\mathbb{N}^*}$.

On appelle *histoire du joueur i* en date t une suite finie de la forme $h^i_t = (u^i_0, a^i_1, u^i_1, \dots, a^i_{t-1}, u^i_{t-1})$. L'ensemble de telles histoires est $H^i_t = U^i \times (A^i \times U^i)^{t-1}$.

Dans la suite nous allons travailler avec des ensembles produits. Donnons quelques définitions générales. Soit $(E_n, d_n)_n$ une famille dénombrable d'espaces métriques compacts. La *topologie produit* sur $E = \prod_n E_n$ est celle engendrée par les projections $E \to E_n$, $e = (e_m)_m \mapsto e_n$ (topologie de la convergence simple). Pour cette topologie, E est compact (théorème de Tychonoff) et métrisable (e.g. avec la distance $\sum_n 2^{-n} \max\{d_n, 1\}$). La *tribu produit* sur E est la tribu engendrée (également) par les projections. Elle coïncide avec la tribu borélienne associée à la topologie produit. Dans la suite, les produits de métriques compacts seront munis de cette topologie et de cette tribu.

Définissons les notions suivantes de stratégies.

Définition 1.2

– Une *stratégie pure* du joueur i est une famille d'applications $s^i = (s^i_t)_{t\geqslant 1}$, avec $s^i_t : H^i_t \to A^i$. On note S^i l'ensemble de ces stratégies.

– Une *stratégie mixte* du joueur i est une distribution de probabilité borélienne sur S^i.

– Une *stratégie de comportement* du joueur i est une famille d'applications $\sigma^i = (\sigma^i_t)_{t\geqslant 1}$, avec $\sigma^i_t : H^i_t \to \Delta(A^i)$. On note Σ^i l'ensemble de ces stratégies.

– Une *stratégie générale* du joueur i est une distribution de probabilité borélienne sur Σ^i.

Les ensembles de stratégies pures et de stratégies de comportement sont des produits dénombrables de métriques compacts. Les ensembles de stratégies mixtes et de stratégies générales sont donc également des métriques compacts (on prendra toujours la topologie faible-* sur l'ensemble des probabilités boréliennes sur un métrique compact). Grâce à l'injection naturelle de E dans $\Delta(E)$, on peut considérer que S^i est un sous-ensemble de $\Delta(S^i)$ et de Σ^i, et que $\Delta(S^i)$ et Σ^i sont des sous-ensembles de $\Delta(\Sigma^i)$. Les stratégies générales contiennent donc toutes les autres notions.

Lemme 1.3. — *Un profil de stratégies générales* $\alpha = (\alpha_i)_{i\in N} \in \prod_{i\in N}\Delta(\Sigma^i)$ *induit une unique mesure de probabilité* $\mathbb{P}_\alpha$ *sur* H_∞ *(muni de la tribu produit).*

Démonstration. — Commençons par montrer que tout profil de stratégies de comportement $\sigma = (\sigma^i)_{i\in N} \in \prod_{i\in N}\Sigma^i$ induit une unique mesure de probabilité $\mathbb{P}_\sigma$ sur H_∞. Pour chaque histoire $h_t = (u_0, k_1, a_1, u_1, k_2 \dots, a_{t-1}, u_{t-1}, k_t)$, on pose :

$$\begin{aligned}\mathbb{P}_\sigma(h_t) = p(u_0,k_1)\bigl(\textstyle\prod_i \sigma^i(h^i_1)[a^i_1]\bigr)Q(u_1,k_2|k_1,a_1)\cdots\\ \cdot\bigl(\textstyle\prod_i \sigma^i(h^i_{t-1})[a^i_{t-1}]\bigr)Q(u_{t-1},k_t|k_{t-1},a_{t-1}).\end{aligned}$$

Ceci définit la probabilité $\mathbb{P}_\sigma$ sur les cylindres $C(h_t)$ (l'ensemble des parties du jeu qui commencent par h_t), et celle-ci s'étend de manière unique sur H_∞, la tribu produit étant engendrée par les cylindres (par un théorème d'extension de Kolmogorov).

Soit maintenant un profil de stratégies générales $\alpha = (\alpha_i)_{i\in N} \in \prod_{i\in N} \Delta(\Sigma^i)$. On pose, pour tout ensemble mesurable B de H_∞, $\mathbb{P}_\alpha(B) = \int \mathbb{P}_\sigma(B)\Pi_i d\alpha^i(\sigma^i)$. $\square$

Le résultat suivant, dû à Kuhn (1953) pour le cas fini et généralisé par Aumann (1964), dit que pour analyser les jeux répétés, il suffit de considérer les stratégies de comportement.

Théorème 1.4

(1) *Toute stratégie de comportement du joueur i est équivalente à une stratégie mixte du joueur i dans le sens suivant : pour toute stratégie de comportement σ^i, il existe une stratégie mixte μ^i, telle que pour toute stratégie générale α^{-i} des autres joueurs, on ait $\mathbb{P}_{\sigma^i,\alpha^{-i}} = \mathbb{P}_{\mu^i,\alpha^{-i}}$.*

(2) *Toute stratégie mixte du joueur i est équivalente à une stratégie de comportement du joueur i.*

(3) *Les stratégies mixtes, de comportement, et générales, du joueur i sont équivalentes.*

Lors de constructions explicites de stratégies, il est souvent plus facile de définir une stratégie de comportement qu'une stratégie mixte. Dans le premier cas, on doit spécifier les lois de probabilités utilisées par le joueur pour choisir ses actions au moment de jouer, alors que dans le second cas, il faut définir une mesure de probabilité sur S^i, produit dénombrable d'ensembles finis. On verra également que dans certains cas, on est amené à choisir une stratégie de comportement au hasard, c'est-à-dire à construire une stratégie générale. Le théorème de Kuhn, nous donne donc la souplesse de passer d'une représentation à l'autre suivant les besoins.

Démonstration

(1) Soit σ^i une stratégie de comportement du joueur i. Pour toute stratégie pure s^i, posons $s^i(t)$ l'ensemble des stratégies pures du joueur i qui coïncident avec s^i jusqu'a la date t. Pour définir μ^i, il suffit de définir $\mu^i(s^i(t))$, $(\forall\, s^i,\ \forall\, t)$, la tribu produit sur S^i étant engendrée par les cylindres $s^i(t)$. On le définit alors comme la probabilité que le joueur i, tirant ses actions avec σ^i, joue ce qu'aurait

joué s^i jusqu'a la date t :

$$\mu^i(s^i(t)) = \prod_{h_t^i \in H_t^i} \prod_{r<t} \sigma_r^i(h_r^i)[s_r^i(h_r^i)]$$

On a alors clairement, $\mathbb{P}_{\sigma^i,\alpha^{-i}} = \mathbb{P}_{\mu^i,\alpha^{-i}}$.

(2) Réciproquement, soit μ^i une stratégie mixte du joueur i. Pour toute histoire h_t^i du joueur i, posons $S^i(h_t^i)$ l'ensemble des stratégies pures du joueur i compatibles avec h_t^i : si $h_t^i = (u_0^i, a_1^i, u_1^i, \ldots, a_{t-1}^i, u_{t-1}^i)$, $S^i(h_t^i)$ est l'ensemble des stratégies pures telles que pour tout $r < t-1$, $s_r^i(u_0^i, a_1^i, u_1^i, \ldots, a_{r-1}^i, u_{r-1}^i) = a_r^i$. On définit la probabilité de jouer a^i après h_t^i, comme la probabilité de l'ensemble des stratégies pures qui jouent a^i après h_t^i, conditionnellement au fait que la stratégie soit compatible avec h_t^i. Précisément, on pose $\sigma_t^i(h_t^i)[a^i] = \mu^i(\{s^i \mid s_t^i(h_t^i) = a^i\})/\mu^i(S^i(h_t^i))$, si $\mu^i(S^i(h_t^i)) > 0$, et on définit cette quantité arbitrairement si $\mu^i(S^i(h_t^i)) = 0$. Encore une fois, $\mathbb{P}_{\sigma^i,\alpha^{-i}} = \mathbb{P}_{\mu^i,\alpha^{-i}}$.

(3) Une stratégie générale α^i peut donc se voir comme une distribution de probabilité sur les stratégies mixtes qu'on peut réduire à une stratégie mixte μ^i en prenant l'espérance : si $E \subset S^i$ est un ensemble mesurable, on pose $\mu^i(E) = \int \mu(E) d\alpha^i(\mu)$. En termes géométriques, on peut voir une probabilité sur $\Delta(S^i)$ comme une combinaison convexe de stratégies mixtes et l'identifier à son barycentre. □

2. Équilibres

La problématique principale de la théorie des jeux répétés est d'étudier l'existence, voire la caractérisation, des solutions dans les jeux d'horizon long. Deux approches sont possibles : considérer les solutions de jeux à horizon fixé et étudier leur limite quand l'horizon tend vers l'infini, ou formuler un concept de solution directement sur le jeu d'horizon infini.

2.1. Équilibres : l'approche compacte

Définition 2.1

– La forme stratégique du jeu répété T fois Γ_T est donnée par :

$$(N, (S^i)_{i \in N}, \gamma_T^i)$$

avec $\gamma_T^i(s) = \mathbb{E}_{\mathbb{P}_s}[\frac{1}{T}\sum_{t=1}^{T} g^{i,k_t}(a_t)]$.

– La forme stratégique du jeu Γ_λ est donnée par :

$$(N, (S^i)_{i\in N}, \gamma^i_\lambda)$$

avec $\gamma^i_\lambda(s) = \mathbb{E}_{\mathbb{P}_s}[\sum_{t\geqslant 1} \lambda(1-\lambda)^{t-1} g^{i,k_t}(a_t)]$.

Remarque 2.2. — Grâce au théorème de Kuhn, nous pouvons définir les extension mixtes de ces jeux en utilisant les stratégies de comportement.

– L'extension mixte de Γ_T peut s'identifier au jeu sous forme stratégique :

$$(N, (\Sigma^i)_{i\in N}, \gamma^i_T)$$

avec $\gamma^i_T(\sigma) = \mathbb{E}_{\mathbb{P}_\sigma}[\frac{1}{T}\sum_{t=1}^T g^{i,k_t}(a_t)]$.

– L'extension mixte de Γ_λ peut s'identifier au jeu sous forme stratégique :

$$(N, (\Sigma^i)_{i\in N}, \gamma^i_\lambda)$$

avec $\gamma^i_\lambda(\sigma) = \mathbb{E}_{\mathbb{P}_\sigma}[\sum_{t\geqslant 1} \lambda(1-\lambda)^{t-1} g^{i,k_t}(a_t)]$.

On peut alors appliquer les théorèmes d'existence.

Théorème 2.3. — *Pour tout entier T et tout $\lambda \in\,]0,1]$, les jeux Γ_T et Γ_λ admettent des équilibres de Nash en stratégies de comportement. En particulier, dans le cas de jeu à somme nulle, Γ_T et Γ_λ ont chacun une valeur et les deux joueurs ont des stratégies optimales.*

Démonstration. — Les ensembles des stratégies pures sont compacts pour la topologie produit. De plus, les fonctions de paiements γ^i_T et γ^i_λ sont continues pour la topologie produit : on vérifie aisément que si une suite de stratégies $(s_q)_q$ converge simplement vers s alors les paiements associés convergent : c'est évident pour γ^i_T, et pour γ^i_λ, comme les paiements d'étapes sont uniformément bornés, on peut majorer la queue de la série par ε uniformément par rapport aux stratégies. D'après le théorème 4.4 du premier texte de ce volume, il existe un équilibre en stratégies mixtes, et grâce au théorème de Kuhn, c'est un équilibre en stratégies de comportement. $\square$

Notation

(1) On notera E_T l'ensemble des paiements d'équilibres du jeu Γ_T : l'ensemble des $x \in \mathbb{R}^N$ pour lesquels il existe un équilibre de Nash σ

de Γ_T tel que $\gamma^i_T(\sigma) = x^i$, $\forall\, i$. Dans le cas des jeux à somme nulle, on notera v_T la valeur de Γ_T.

(2) On notera E_λ l'ensemble des paiements d'équilibres du jeu Γ_λ et dans le cas des jeux à somme nulle, on notera v_λ la valeur de Γ_λ.

L'approche compacte consiste en l'étude de la limite de E_T (ou v_T) quand T tends vers l'infini et de E_λ (ou v_λ) quand λ tend vers 0.

2.2. Équilibres : l'approche uniforme. — On veut ici se placer directement dans le cas limite d'horizon infini : considérons le jeu Γ_∞ dans lequel le paiement final du joueur i induit par des stratégies de comportement $\sigma = (\sigma^i)_{i\in N}$ est $\lim_T \gamma^i_T(\sigma)$ si cette limite existe.

Pour pallier l'absence éventuelle de limite, on peut considérer la limite supérieure ou la limite inférieure (de ces suites bornées). On peut également avoir un critère de paiement linéaire en utilisant une *limite de Banach*, c'est-à-dire une forme linéaire L sur ℓ^∞ telle que pour toute suite réelle bornée $x = (x_n)_n$, $\underline{\lim}_n x_n \leqslant L(x) \leqslant \overline{\lim}_n x_n$. Une telle forme linéaire existe par application du théorème de Hahn-Banach à l'application sous-linéaire $\overline{\lim}$. Les limites de Banach sont les éléments du dual topologique de ℓ^∞ qui s'annulent sur l'espace des suites qui tendent vers 0, et valent 1 pour la suite constante égale à 1.

Toutefois, les applications $\sigma \mapsto \overline{\lim}_T \gamma^i_T(\sigma)$, $\sigma \mapsto \underline{\lim}_T \gamma^i_T(\sigma)$, $\sigma \mapsto L((\gamma^i_T(\sigma))_T)$ ne sont pas continues pour la topologie produit : changer le paiement en un nombre fini d'étapes ne change pas la limite de Césaro.

On adopte donc l'approche uniforme qui définit les solutions de Γ_∞ comme des solutions approchées de jeux finiment répétés arbitrairement longs. Nous donnons d'abord les définitions dans le cas de la somme nulle.

Définition 2.4. — Soit Γ_∞ un jeu infiniment répété à somme nulle, notons $g = g^1 = -g^2$.

– Le joueur 1 *garantit uniformément* le paiement $d \in \mathbb{R}$ si :

$$\forall\, \varepsilon > 0,\ \exists\, \sigma^1 \in \Sigma^1,\ \exists\, T_0,\ \forall\, \sigma^2 \in \Sigma^2,\ \forall\, T \geqslant T_0, \quad \gamma_T(\sigma^1, \sigma^2) \geqslant d - \varepsilon.$$

– Le joueur 2 *défend uniformément* le paiement $d \in \mathbb{R}$ si :

$$\forall\, \varepsilon > 0,\ \forall\, \sigma^1 \in \Sigma^1,\ \exists\, \sigma^2 \in \Sigma^2,\ \exists\, T_0,\ \forall\, T \geqslant T_0, \quad \gamma_T(\sigma^1, \sigma^2) \leqslant d + \varepsilon.$$

– Le maxmin uniforme $\underline{v}$ de Γ_∞, s'il existe, est tel que le joueur 1 garantit uniformément $\underline{v}$ et le joueur 2 défend uniformément $\underline{v}$. On définit le minmax uniforme $\overline{v}$ en échangeant les rôles des joueurs.

– Le jeu Γ_∞ a une valeur v si $\underline{v} = \overline{v} = v$.

– Lorsque le jeu a une valeur, une stratégie σ^1 du joueur 1 sera dite optimale si elle vérifie :

$$\forall\, \varepsilon > 0,\ \exists\, T_0,\ \forall\, \sigma^2 \in \Sigma^2,\ \forall\, T \geqslant T_0, \quad \gamma_T(\sigma^1, \sigma^2) \geqslant v - \varepsilon.$$

On définit similairement les stratégies optimales du joueur 2.

Pour les jeux à somme non nulle nous donnons une définition analogue :

Définition 2.5. — Soit Γ_∞ un jeu infiniment répété. Un profil de stratégies de comportement $\sigma = (\sigma^i)_{i \in N}$ est un *équilibre uniforme* de Γ_∞ si, pour tout $i \in N$, la suite $(\gamma_T^i(\sigma))_T$ converge vers une limite $\gamma_\infty^i(\sigma)$ quand T tend vers $+\infty$, et de plus,

$$\forall\, \varepsilon > 0,\ \exists\, T_0,\ \forall\, T \geqslant T_0,\ \forall\, i \in N,\ \forall\, \tau^i \in \Sigma^i, \quad \gamma_T^i(\tau^i, \sigma^{-i}) \leqslant \gamma_T^i(\sigma) + \varepsilon.$$

Un équilibre uniforme est donc un ε-équilibre dans tout jeu finiment répété suffisamment long. Nous noterons E_∞ l'ensemble des paiements d'équilibres uniformes de Γ_∞.

2.3. Principales questions. — Les questions auxquelles nous allons nous intéresser par la suite sont, pour la somme nulle et l'approche compacte : $\lim_T v_T$, $\lim_\lambda v_\lambda$ existent-elles, sont-elles égales, peut-on calculer ces limites ? Pour l'approche uniforme, le minmax et le maxmin uniformes existent-ils, existe-t-il une valeur, peut-on la calculer, quels sont les liens avec $\lim_T v_T$ et $\lim_\lambda v_\lambda$? Les mêmes questions se posent en somme non nulle pour les paiements d'équilibres.

La partie suivante traite du cas le plus simple de jeu répété : les jeux répétés à information complète et observation parfaite.

3. Jeux répétés à information complète et observation parfaite

Soit $G = (N, (A^i)_{i \in N}, (g^i)_{i \in N})$ un jeu fini. Considérons le jeu répété dans lequel le jeu G est joué à chaque étape $t \in \mathbb{N}^*$ et si $a_t \in A$ est le profil d'actions joué en date t, a_t est observé par tous les joueurs.

C'est un cas très simple de jeu répété dans lequel l'ensemble d'états est réduit à un singleton et les signaux des joueurs sont *parfaits*, c'est-à-dire qu'ils révèlent parfaitement les actions jouées.

Commençons par un lemme immédiat.

Lemme 3.1. — *Soit* $\alpha = (\alpha^i)_{i\in N} \in \prod_{i\in N} \Delta(A^i)$ *un équilibre de Nash du jeu* G. *Le profil de stratégies de comportement* σ *tel que pour tout* i, *pour tout* h^i_t, $\sigma^i(h^i_t) = \alpha^i$ *est un équilibre de Nash de* Γ_T, *de* Γ_λ *(pour tout* T *et tout* λ*) et un équilibre uniforme de* Γ_∞.

Il s'ensuit :

Corollaire 3.2. — *Si le jeu* G *est à somme nulle de valeur* v *alors, pour tout* T *et tout* λ, $v_T = v_\lambda = v$. *De plus, la valeur de* Γ_∞ *existe et vaut* v.

Dans le cas des jeux répétés à information complète, la répétition d'un jeu à somme nulle ne change pas la valeur. Examinons maintenant le cas de la somme quelconque.

Notations. — Posons $g : A \to \mathbb{R}$ l'application « vecteur de paiement », $g(a) = (g^i(a))_{i\in N}$. On appelle *ensemble réalisable* l'enveloppe convexe $\operatorname{co} g(A)$ des vecteurs $g(a)$, $a \in A$.

Pour chaque joueur i, posons

$$v^i = \min_{\alpha^{-i}\in\prod_{j\neq i}\Delta(A^j)} \max_{a^i\in A^i} g^i(a^i, \alpha^{-i})$$

et $\mathrm{IR} = \left\{x \in \mathbb{R}^n \mid \forall\, i \in N,\, x^i \geqslant v^i\right\}$. La quantité v^i s'appelle *niveau de rationalité individuelle* ou *niveau minmax* du joueur i : v^i est le plus petit paiement que le joueur i peut obtenir dans le jeu G en jouant une meilleure réponse contre un profil de stratégies mixtes de ses adversaires. C'est aussi le plus petit paiement que les joueurs $-i$ peuvent garantir s'ils jouent dans le but de minimiser le paiement du joueur i.

Posons enfin $E = \operatorname{co} g(A) \cap \mathrm{IR}$. Remarquons que si le jeu G est à somme nulle de valeur v, E se réduit au singleton $\{(v, -v)\}$.

Lemme 3.3. — *Les ensembles de paiements d'équilibres* E_T, E_λ *et* E_∞ *sont inclus dans* E *(pour tous* T, λ*).*

Démonstration. — Pour tout profil de stratégies σ, le vecteur de paiement appartient à chaque étape au convexe compact $\operatorname{co} g(A)$. En prenant la moyenne (arithmétique ou escomptée) puis l'espérance, le vecteur de paiement du jeu répété est également dans $\operatorname{co} g(A)$.

Pour l'inclusion dans IR, soit σ un équilibre de Γ_T (resp. de Γ_λ, resp. un équilibre uniforme) et soit i un joueur. Définissons une stratégie τ^i du joueur i qui à chaque étape t et après chaque histoire h_t joue une action qui maximise sur A^i la quantité $g^i(a^i, (\sigma^j(h_t))_{j\neq i})$. D'après la définition de v^i, $g^i(\tau^i(h_t), (\sigma^j(h_t))_{j\neq i}) \geqslant v^i$, pour tout t et pour toute histoire h_T. Il s'ensuit $\gamma_T^i(\tau^i, \sigma^{-i}) \geqslant v^i$ et $\gamma_\lambda^i(\tau^i, \sigma^{-i}) \geqslant v^i$, pour tous T et λ. Comme σ est un équilibre de Γ_T, $\gamma_T^i(\sigma) \geqslant \gamma_T^i(\tau^i, \sigma^{-i}) \geqslant v^i$. La conclusion est identique pour Γ_λ et pour les équilibres uniformes on conclut en remarquant que $\lim_T \gamma_T^i(\sigma) \geqslant \overline{\lim}_T \gamma_T^i(\tau^i, \sigma^{-i}) \geqslant v^i$. $\square$

3.1. Le jeu Γ_∞. — Le résultat principal dit que l'inclusion inverse est vraie : les paiements d'équilibres sont les paiements réalisables et individuellement rationnels. Ce résultat s'appelle communément le *Folk Théorème* car était informellement connu de bon nombre de chercheurs dès les années 60 mais est resté plusieurs années non publié. Les premières version publiées sont dues à Aumann et Shapley (1976) (ré-édité en 1994) et à Rubinstein (1977).

Ce résultat se formule pour le jeu Γ_∞.

Théorème 3.4. — *$E_\infty = E$.*

Démonstration. — Considérons un paiement x dans E. Comme x est réalisable, il existe une partie $h = (a_1, \dots, a_t, \dots)$ telle que pour tout joueur i, $\frac{1}{T}\sum_{t=1}^T g^i(a_t) \xrightarrow[T\to\infty]{} x^i$. Nous appellerons h le plan principal de la stratégie, et jouer selon h pour un joueur i en date t signifie jouer la i-ème composante de a_t. Pour chaque couple de joueurs distincts (i, j), fixons $\alpha^{i,j}$ dans $\Delta(A^j)$ de façon à ce que $(\alpha^{i,j})_{j\neq i}$ réalise le min dans l'expression de v^i. Fixons maintenant un joueur i dans N, et définissons une stratégie σ^i. σ^i commence en date 1 par jouer selon le plan principal, et continue de jouer selon h tant que tous les autres joueurs le font. Si à une certaine date $t \geqslant 1$, pour la première fois un joueur j ne joue pas selon le plan principal, alors σ^i joue à toutes les dates ultérieures la probabilité $\alpha^{j,i}$ (si pour la première fois à la même date plusieurs joueurs sortent du plan principal, on punit celui de ces joueurs qui est le plus petit).

Si tous les joueurs adoptent ces stratégies, la suite $(a_t)_{t\in\mathbb{N}^*}$ sera jouée et le paiement moyen limite réalisé est bien x. Supposons que le joueur i emploie une stratégie τ^i qui sort du plan principal pour la première fois en date t, alors que les autres jouent σ^{-i}. Le joueur i recevra au plus v^i à toutes les dates suivantes. Son paiement moyen en date T sera donc majoré par :

$$\begin{cases} \dfrac{1}{T}(\displaystyle\sum_{t=1}^{T-1} g^i(a_t) + M) & \text{si } t \geqslant T; \\ \dfrac{t-1}{T}\displaystyle\sum_{s=1}^{t-1} g^i(a_s) + \dfrac{1}{T}M + \dfrac{T-t}{T}v^i & \text{si } t < T, \end{cases}$$

où $M = \max_{i,a} \left|g^i(a)\right|$. Soit $\varepsilon > 0$, choisissons T_0 assez grand tel que pour tout $T \geqslant T_0$:

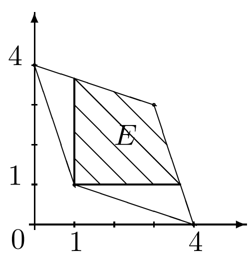
$$\frac{1}{T}M \leqslant \frac{\varepsilon}{2} \quad \text{et} \quad \Big(\frac{t-1}{T} > \frac{\varepsilon}{2M} \Longrightarrow \frac{1}{t-1}\sum_{s=1}^{t-1} g^i(a_s) \leqslant x^i + \frac{\varepsilon}{2}\Big).$$

Comme $v^i \leqslant x^i$, le paiement du joueur i est majoré par $x^i + \varepsilon$ et ce pour toute déviation. □

Le Dilemme du Prisonnier Répété. — Reprenons le jeu du Dilemme du Prisonnier :

	C	D
C	$3,3$	$0,4$
D	$4,0$	$1,1$

Dans ce cas, l'ensemble E est le suivant :



Bien que l'action C soit strictement dominée pour chaque joueur, on construit un équilibre uniforme de paiement $(3,3)$: jouer C si l'autre a toujours fait de même dans le passé, sinon jouer D.

3.2. Le jeu Γ_T. — Nous allons maintenant voir si l'approche uniforme et l'approche compacte donnent les mêmes résultats : la limite de E_T est-elle égale à E ? La convergence d'ensemble est au sens de la distance de Hausdorff. Étant donnés deux compacts A, B de $\mathbb{R}^n$, on pose $d(A,B) = \max\{\max_{a\in A} d(a,B), \max_{b\in B} d(b,A)\}$. Comme $E_T \subset E$, $E_T \to E$ revient à :

$$\forall\,\varepsilon > 0,\ \exists\, T_0,\ \forall\, T \geqslant T_0,\ \forall\, x \in E,\ \exists\, y \in E_T : \quad \|x-y\| \leqslant \varepsilon$$

On a la propriété :

Proposition 3.5. — *Pour tous entiers T et T', on a*

$$TE_T + T'E_{T'} \subset (T+T')E_{T+T'}.$$

Ici, $\alpha A + \beta B$ désigne l'ensemble des $\alpha a + \beta b$, avec $a \in A$, $b \in B$. Ce résultat s'obtient simplement en remarquant que jouer un équilibre de E_T puis un équilibre de $E_{T'}$ est un équilibre de $E_{T+T'}$. $(E_T)_T$ se comporte comme une suite croissante d'ensembles : si elle converge c'est forcément vers $\cup_T E_T$.

Considérons le dilemme du prisonnier encore une fois. Il est clair que pour tout équilibre de Nash du jeu Γ_T, les joueurs doivent jouer D à la dernière étape, quelle que soit l'histoire passée : on ne peut en effet pas les menacer de représailles. Le jeu est donc équivalent à un jeu de longueur $T-1$, les actions d'étapes T étant fixées. Mais alors les joueurs ne peuvent pas non plus se menacer mutuellement en étape $T-1$, puisqu'il joueront D ensuite quoiqu'il arrive. Ils joueront donc D également en $T-1$. Par récurrence on obtient : $E_T = \{(3,3)\}$ pour tout T. Plus généralement, posons $v = (v^i)_{i\in N}$ le vecteur des niveaux de rationalité individuelle ou *point de menace*, on a le résultat du à Sorin (1986) :

Proposition 3.6. — *Si $E_1 = \{v\}$ alors $E_T = \{v\}$ pour tout T.*

La convergence de E_T vers E nécessite donc une condition sur les paiements du jeu. Dans un équilibre d'un jeu finiment répété, un équilibre de Nash du jeu statique doit être joué à la dernière étape. Pour que les menaces soient dissuasives, il faut que les paiements de punitions, *i.e.* les niveaux de rationalité individuels, soient strictement inférieurs à un paiement d'équilibre de Nash de G : c'est précisément la condition qui manque pour le Dilemme du Prisonnier. On a alors le théorème (Benoit et Krishna, 1987) :

Théorème 3.7. — *Supposons que pour tout joueur i, il existe $e(i) \in E_1$ tel que $e^i(i) > v^i$. Alors $E_T \to_{T\to+\infty} E$.*

Démonstration. — Soit $\varepsilon > 0$ et $x \in E$. Prenons une histoire $h = (a_1, \dots, a_L)$ telle que $y = \frac{1}{L}\sum_{t=1}^{L} g(a_t)$ est à distance inférieure à ε de x. Soit la suite d'actions consistant à jouer K fois de suite l'histoire h, pour un certain entier K.

La stratégie consiste à jouer cette suite puis à jouer R fois les équilibres de G correspondant aux paiements $(e(1), \dots, e(N))$, si aucune déviation n'apparaît au cours des LK premières étapes (on note N le nombre de joueurs). Si un joueur i dévie à une de ces étapes, il est puni au niveau v^i jusqu'a la fin du jeu. Ceci définit un profil de stratégies pour le jeu répété $T = LK + RN$ fois. Vérifions que c'est un équilibre dont le paiement est proche de x pour K et R bien choisis.

Supposons que le joueur i dévie. Une déviation dans les RN dernières étapes n'est pas profitable. S'il dévie pendant les LK premières étapes, son paiement total augmente d'au plus $2M$ (on a encore posé $M = \max_{i,a} \left|g^i(a)\right|$), alors que le fait d'être puni à la fin lui fait perdre $R(e^i(i) - v^i)$. Comme $e^i(i) - v^i > 0$, pour R assez grand, $R(e^i(i) - v^i) > 2M$ pour tout i.

Le paiement moyen si tous les joueurs utilisent ces stratégies est

$$\frac{LK}{LK + RN}\, y + \frac{R}{LK + RN} \sum_i g(e(i)),$$

dont la distance à y est inférieure à

$$\frac{2MN}{L} \cdot \frac{R}{K}$$

et peut être rendue inférieure à ε en choisissant K grand devant R.

Enfin, si le nombre de répétitions T n'est pas de la forme $LK + RN$, il existe K tel que $LK + RN < T < L(K + 1) + RN$. On prend alors la stratégie définie ci-dessus pour K et on la complète en jouant un équilibre de G (par exemple celui de paiement $e(1)$) aux dernières étapes du jeu. $\square$

3.3. Le jeu Γ_λ. — A la différence du jeu finiment répété, le jeu escompté n'a pas de fin, il est donc possible, à n'importe quelle étape de menacer les joueurs de représailles. Toutefois, considérons le jeu à trois joueurs suivant (Forges, Mertens et Neyman, 1986), dans lequel le joueur 3 n'a qu'une action.

	P	F
P	$1,-1,0$	$-1,1,0$
F	$-1,1,0$	$1,-1,1$

Il s'agit d'un jeu à somme nulle entre les joueurs 1 et 2 (c'est Matching Pennies) dont la valeur est 0 et chaque joueur (1 ou 2) a une unique stratégie optimale : le jeu à un unique équilibre de Nash dans lequel chaque joueur joue l'action mixte $(\frac{1}{2}, \frac{1}{2})$. Dans le jeu Γ_λ (ou dans Γ_T), l'unique équilibre consiste à jouer $(\frac{1}{2}, \frac{1}{2})$ à chaque étape, indépendamment du passé. Le seul paiement d'équilibre de Γ_λ (resp. de Γ_T) est donc $(0, 0, \frac{1}{4})$ alors que $(0, 0, \frac{1}{2})$ est dans E. Comme dans le jeu finiment répété, il faut introduire une condition pour garantir $E_\lambda \to {}_{\lambda\to 0}E$.

***Théorème 3.8* (Sorin, 1986).** — *Si le jeu est à deux joueurs ou qu'il existe $x \in E$ tel que $x^i > v^i$ pour tout i, alors $E_\lambda \to {}_{\lambda\to 0}E$.*

Démonstration

(1) Supposons que l'ensemble des x de E tels que $x^i > v^i$ pour tout i est non vide et prenons $\varepsilon > 0$. Il existe un entier n tel que pour tout $y \in E$, on peut trouver dans la boule de centre y et de rayon ε un x_n vérifiant $x_i^n > v^i + \frac{\varepsilon}{2}$ pour tout i et tel que x_n est une combinaison convexe des $g(a)$ avec des coefficients rationnels de la forme n_a/n.

Construisons alors une suite n-périodique de profils d'actions telle que chaque profil $a \in A$ apparaît n_a fois dans une période et notons $(a_t^*)_t$ cette suite. Ainsi, $x^n = \frac{1}{n}\sum_{t=1}^{n} g(a_t^*)$. Soit σ le profil de stratégies de comportement qui consiste à jouer a_t^* en date t si cette consigne a été respectée par tous les joueurs dans le passé, et à punir pour toujours le joueur i (au niveau v^i) si celui-ci dévie. Montrons maintenant que pour λ suffisamment proche de 0, σ est un équilibre qui induit un paiement proche de y.

Comme la suite d'actions est périodique, pour $0 < \lambda < 1$,

$$\gamma_\lambda(\sigma) = \sum_{t=1}^{n} \frac{\lambda(1-\lambda)^{t-1}}{1-(1-\lambda)^n}\, g(a_t^*)$$

Pour tout t,

$$\lim_{\lambda\to 0} \frac{\lambda(1-\lambda)^{t-1}}{1-(1-\lambda)^n} = \frac{1}{n},$$

donc $\lim_{\lambda\to 0}\gamma_\lambda(\sigma) = x^n$, de plus la convergence est uniforme par rapport au point x^n considéré. Il existe donc λ_0 tel que pour tout $\lambda < \lambda_0$, on ait $\|\gamma_\lambda(\sigma) - y\| \leqslant 2\varepsilon$, λ_0 étant uniforme par rapport au choix de y dans E.

Supposons maintenant que le joueur i dévie. Grâce à la périodicité, supposons sans perte de généralité qu'il joue pour la première fois une action a^i différente de a_T^{*i} au cours de la première période en date T, $1 \leqslant T \leqslant n$. Posons encore une fois $M = \max_{i,a} \left|g^i(a)\right|$ et majorons le paiement du joueur i aux T premières étapes par M. Le paiement global est alors majoré par :

$$(1-(1-\lambda)^T)M + (1-\lambda)^T v^i < (1-(1-\lambda)^T)M + (1-\lambda)^T\left(x_n^i - \frac{\varepsilon}{2}\right)$$

Pour avoir un équilibre on veut donc

$$(1-(1-\lambda)^T)M + (1-\lambda)^T\left(x_n^i - \frac{\varepsilon}{2}\right) \leqslant x_n^i$$

or pour cela il suffit d'avoir $(1-(1-\lambda)^T)2M \leqslant (1-\lambda)^T\varepsilon/2$ ce qui est vrai pour λ suffisamment petit, uniformément par rapport au point x_n. On a donc bien la convergence de E_λ vers E.

(2) Supposons que le jeu soit à deux joueurs. Si le cas précédent ne s'applique pas alors soit, $E = \{v\}$ mais alors v est le paiement d'un équilibre du jeu G et il suffit de répéter cet équilibre : $E_\lambda = E$. Soit, pour tout x dans E, $x^1 = v^1$ et il existe x dans E tel que $x^2 > v^2$ (ou la condition symétrique en échangeant les deux joueurs). Ceci implique que v^1 est le paiement maximal du joueur 1. On construit donc un équilibre comme dans le cas précédent mais sans tenir compte des déviations de ce joueur (qui n'y a jamais intérêt). □

3.4. Équilibres sous-jeux parfaits. — Certains équilibres construits dans les paragraphes précédents sont critiquables en tant que solutions rationnelles car reposant sur des menaces non crédibles : étant donné qu'un joueur a dévié, rien n'assure que ses adversaires auront intérêt à le punir. On introduit alors la notion d'*équilibre sous-jeu parfait*. Un profil de stratégies σ est un équilibre sous-jeu parfait si pour toute histoire h, σ induit un équilibre dans le jeu répété qui reste à jouer après h. On note E'_∞ l'ensemble des paiements d'équilibres sous-jeu parfaits de Γ_∞ (de même E'_T et E'_λ). Cette notion avait déjà été considérée par Aumann-Shapley et Rubinstein et on a :

Théorème 3.9. — $E'_\infty = E_\infty = E$.

La démonstration est simple, il suffit d'adapter la construction du Folk théorème : si un joueur dévie en date t, il est puni jusqu'à la date t^2, après quoi les joueurs oublient la déviation et se remettent à jouer la suite prévue. La longueur de la punition est suffisante pour faire perdre tout intérêt à dévier, les $t^2 - t$ étapes de punitions étant prépondérantes devant les t premières étapes. De plus un joueur n'hésitera pas à appliquer la punition puisqu'un nombre fini de dates ne compte pas sur son paiement limite.

Pour les jeux finiment répétés et les jeux escomptés, il faut ajouter encore des conditions.

***Théorème 3.10* (Benoit et Krishna, 1985, Gossner, 1995)**
Supposons que pour tout i, il existe $e(i), f(i) \in E_1$ tels que $e^i(i) > f^i(i)$. Alors $E'_T \to_{T\to+\infty} E$.

***Théorème 3.11* (Fudenberg et Maskin, 1986).** — *Si E est d'intérieur non vide alors, $E'_\lambda \to_{\lambda\to 0} E$.*

L'intuition est la même pour ces deux résultats. On construit une suite d'actions qui donne le bon paiement et que les joueurs doivent suivre. En cas de déviation, on punit pendant un nombre fini d'étapes et la phase de punition est suivie d'une phase de récompense : si le joueur j a correctement puni le joueur i, il reçoit la récompense, sinon il est puni à son tour. Dans le cas des jeux finiment répétés, comme on doit finir le jeu par des équilibres statiques, on récompense i en lui donnant le paiement $e^i(i)$ (et $f^i(i)$ s'il n'a pas appliqué les punitions). Dans le cas des jeux escomptés, l'hypothèse d'intérieur non vide assure que, si le vecteur de paiement à atteindre est à l'intérieur de E, il est possible d'induire des paiements escomptés futurs dans n'importe quelle direction, et donc d'augmenter ou de diminuer le paiement de n'importe quel joueur.

3.5. Extensions : jeux avec signaux. — Nous n'avons traité ici que du cas d'observation parfaite. Cette hypothèse est cruciale pour le Folk théorème et la construction d'équilibre correspondante. Dans le modèle général avec des signaux quelconques, les phénomènes suivants (notamment) apparaissent :

(1) Il peut y avoir des déviations indétectables : un joueur peut choisir une autre action que celle prescrite sans modifier les signaux des autres joueurs. Un profil d'actions a pour lequel un joueur i peut dévier profitablement – jouer b^i tel que $g^i(b^i, a^{-i}) > g^i(a)$ – sans changer les signaux des autres joueurs, ne peut pas être joué dans un équilibre du jeu répété.

(2) Lorsqu'une déviation est détectée, plusieurs joueurs peuvent être suspectés : les signaux peuvent être compatibles avec des déviations de différents joueurs. On ne peut donc pas toujours construire des équilibres en punissant *un* joueur à son niveau minmax.

(3) La finesse des signaux pouvant différer d'un joueur à l'autre, certains joueurs peuvent détecter une déviation et pas d'autres. Les joueurs les mieux informés peuvent essayer de communiquer leurs informations aux autres, cette communication devant se faire au travers des signaux et devant être robuste aux déviations unilatérales.

Les deux branches principales de cette littérature suivent la dichotomie que nous avons introduite : approche uniforme *vs* approche compacte. Dans l'approche uniforme les premières caractérisations de paiements d'équilibres dans les jeux avec signaux sont dues à E. Lehrer (1989, 1992). Des avancées récentes sont dues à Renault et Tomala (1998, 2004a) et à Gossner et Tomala (2006a, 2006b).

Pour l'approche compacte, la majorité des articles étudient le jeux escomptés et les équilibres sous-jeux parfaits. Les articles de référence sont Abreu, Pearce et Stachetti (1990), Fudenberg et Levine (1994) et Fudenberg, Levine et Maskin (1994).

La recherche est actuellement très active dans ces deux branches.

Bibliographie

Abreu (D.), Pearce (D.) & Stacchetti (E.)

[1990] Toward a theory of discounted repeated games with imperfect monitoring, *Econometrica*, 58 (1990), p. 1041–1063.

Aumann (R.J.)

[1964] Mixed and behaviour strategies in infinite extensive games, dans Dresher, Shapley & Tucker, éd., *Advances in Game Theory*, Annals of Mathematics Studies, vol. 52, Princeton University Press, 1964, p. 627–650.

Aumann (R.J.) & Maschler (M.)

[1995] *Repeated games with incomplete information*, M.I.T. Press, 1995.

Aumann (R.J.) & Shapley (L.S.)

[1994] Long-term competition–A game theoretic analysis, dans Megiddo (N.), éd., *Essays on game theory*, New-York : Springer-Verlag, 1994, p. 1–15.

Benoit (J-P.) & Krishna (V.)

[1985] Finitely repeated games, *Econometrica*, 53 (1985), p. 905–922.

[1987] Nash equilibria of finitely repeated games, *International Journal of Game Theory*, 16 (1987), p. 197–204.

Blackwell (D.)

[1956] An analog of the minmax theorem for vector payoffs, *Pacific Journal of Mathematics*, 65 (1956), p. 1–8.

Borel (É.)

[1921] La théorie du jeu et les équations intégrales à noyau symétrique gauche, *C. R. Acad. Sci. Paris*, 173 (1921), p. 1304–1308.

Forges (F.), Mertens (J.-F.) & Neyman (A.)

[1986] A counterexample to the Folk theorem with discounting, *Economic Letters*, 20 (1986), p. 7.

Forges (F.), Renault (J.), Sorin (S.) & Vieille (N.)

[2006] Théorie des jeux : le prix Nobel pour les travaux de R.J. Aumann, *MATAPLI, Bulletin de liaison de la SMAI*, 79 (2006), p. 47–70.

Fudenberg (D.) & Maskin (E.)

[1986] The folk theorem in repeated games with discounting or with incomplete information, *Econometrica*, 54 (1986), p. 533–554.

Fudenberg (D.) & Levine (D.K.)

[1994] Efficiency and observability with long-run and short-run players, *Journal of Economic Theory*, 62 (1994), p. 103–135.

Fudenberg (D.), Levine (D.K.) & Maskin (E.)

[1994] The folk theorem with imperfect public information, *Econometrica*, 62 (1994), p. 997–1039.

Glicksberg (I.)

[1952] A further generalization of the Kakutani fixed point theorem, with applications to Nash equilibrium points, *Proceedings of the American Mathematical Society*, 3 (1952), p. 170–174.

Gossner (O.)

[1995] The folk theorem for finitely repeated games with mixed strategies, *International Journal of Game Theory*, 24 (1995), p. 95–107.

Gossner (O.) & Tomala (T.)

[2006] Empirical distributions of beliefs under imperfect monitoring, *Mathematics of Operations Research*, (2006).

[à paraître] Secret correlation in repeated games with signals, *Mathematics of Operations Research*, (à paraître).

Kakutani (S.)

[1941] A generalization of Brouwer's fixed point theorem, *Duke Mathematical Journal*, 8 (1941), p. 416–427.

Kohlberg (E.)

[1975] Optimal strategies in repeated games with incomplete information, *International Journal of Game Theory*, 4 (1975), p. 7–24.

Kuhn (H.W.)

[1953] Extensive games and the problem of information, dans Kuhn & Tucker, éd., *Contributions to the Theory of Games, vol. II*, Annals of Mathematical Studies, vol. 28, Princeton University Press, 1953, p. 193–216.

Lehrer (E.)

[1989] Nash equilibria of n player repeated games with semi-standard information, *International Journal of Game Theory*, 19 (1989), p. 191–217.

[1992] Correlated Equilibria in two-Player Repeated Games with non-Observable Actions, *Mathematics of Operations Research*, 17 (1992), p. 175–199.

Mertens (J.-F.)

[1986] Repeated Games, dans *Proceedings of the International Congress of Mathematicians (Berkeley)*, Providence, RI : American Mathematical Society, 1986, p. 1528–1577.

Myerson (R.)

[1991] *Game Theory*, Harvard University Press, 1991.

Mertens (J.-F.), Sorin (S.) & Zamir (S)

[1994] Repeated games, 1994, p. 9420–9422 ; CORE discussion paper.

Mills (H.D.)

[1956] Marginal value of matrix games and linear programs, dans Kuhn & Tucker, éd., *Linear Inequalities and Related Systems*, Annals of Mathematical Studies, vol. 38, Princeton University Press, 1956, p. 183–193.

NASH (J.)
[1950] Equilibrium points in n-person games, *Proceedings of the National Academy of Sciences*, 36 (1950), p. 48–49.

OSBORNE (M.J.) & RUBINSTEIN (A.)
[1994] *A course in Game Theory*, M.I.T. Press, 1994.

RENAULT (J.) & TOMALA (T.)
[1998] Repeated proximity games, *International Journal of Game Theory*, 27 (1998), p. 539–559.
[2004] Communication equilibrium payoffs of repeated games with imperfect monitoring, *Games and Economic Behavior*, 49 (2004), p. 313–344.

RUBINSTEIN (A.)
[1977] *Equilibrium in supergames*, Research Memorandum, vol. 25, Center for Research in Mathematical Economics and Game Theory, 1977.

SHAPLEY (L.S.)
[1953] Stochastic games, *Proceedings of the National Academy of Sciences of the U.S.A.*, 39 (1953), p. 1095–1100.

SION (M.)
[1958] On General Minimax Theorems, *Pacific Journal of Mathematics*, 8 (1958), p. 171–176.

SORIN (S.)
[1986] On Repeated Games with Complete Information, *Mathematics of Operations Research*, 11 (1986), p. 147–160.
[1992] Repeated Games with Complete Information, dans Aumann (R.J.) & Hart (S.), éd., *Handbook of Game Theory, vol. I*, Elsevier Science Publishers, 1992, p. 71–107.
[2002] *A First Course on Zero-Sum Repeated Games*, Mathématiques et Applications, Springer, 2002.

VAN DAMME (E.)
[1987] *Stability and Perfection of Nash Equilibria*, Springer, 1987.

VON NEUMANN (J.)
[1928] Zur Theorie der Gesellschaftsspiele, *Mathematische Annalen*, 100 (1928), p. 295–320.

VON NEUMANN (J.) & MORGENSTERN (O.)
[1944] *Games and Economic Behavior*, Princeton University Press, 1944.

ZERMELO (E.)
[1912] Über eine Anwendung der Mengenlehrer auf die Theorie des Schachspiels, dans *Proceedings of the Fifth International Congress of Mathematicians (Cambridge), vol. II*, 1912, p. 501.

T. Tomala, Ceremade, Université Paris Dauphine
E-mail : `tomala@ceremade.dauphine.fr`

JEUX RÉPÉTÉS À INFORMATION INCOMPLÈTE

par

Jérôme Renault

Table des matières

Dans un jeu répété à information incomplète, les joueurs jouent à chaque étape le même jeu de base, mais celui-ci est imparfaitement connu. On expose ici d'abord le cas le plus simple appelé modèle standard à manque d'information d'un seul côté. Ensuite on présentera quelques variantes et extensions importantes.

Si S est un ensemble fini, on note $|S|$ son cardinal et $\Delta(S)$ l'ensemble des probabilités sur S, identifié à

$$\left\{p = (p_s)_{s\in S} \in \mathbb{R}^S \mid \forall\, s \in S\, p_s \geqslant 0 \text{ et } \textstyle\sum_{s\in S} p_s = 1\right\}.$$

Merci à Dinah Rosenberg pour ses remarques.

Pour $p = (p_s)_{s\in S}$ et $q = (q_s)_{s\in S}$ dans $\mathbb{R}^S$, $\|p - q\|$ désigne, sauf précision contraire, $\sum_{s\in S} |p_s - q_s|$. Pour $C \subset \mathbb{R}^S$, on note $\mathrm{int}(C)$ l'intérieur de C, ∂C la frontière de C et $\mathrm{conv}(C)$ l'enveloppe convexe de C. Si $G = (G(i,j))_{(i,j)\in I\times J}$ est une matrice de paiement de format $I \times J$, si $x \in \Delta(I)$ et $y \in \Delta(J)$, $G(x,y)$ désigne le paiement espéré

$$\sum_{\substack{i\in I\\ j\in J}} x_i y_j G(i,j).$$

1. Le modèle standard à manque d'information d'un seul côté

Ce modèle a été introduit et étudié par Aumann et Maschler dans les années 1966-68, aux cours de travaux réédités en 1995. On se place dans le cadre des jeux à deux joueurs et à somme nulle, et on fait l'hypothèse que l'un des deux joueurs a toute l'information. Le manque d'information est donc du côté d'un seul joueur. Formellement, on a une famille $(G^k)_{k\in K}$ de jeux matriciels de même taille $I \times J$, et une probabilité p sur K. Le jeu $\Gamma_\infty(p)$ est joué ainsi :

– initialement, un état de la nature k est tiré, une fois pour toutes, selon p. Le joueur 1 apprend k, pas le joueur 2.

– à chaque étape $t = 1, 2, \dots$, simultanément le joueur 1 choisit une action i_t dans I et le joueur 2 choisit une action j_t dans J. Le paiement d'étape du joueur 1 est alors $G^k(i_t, j_t)$, celui du joueur 2 est $-G^k(i_t, j_t)$, mais tout ce que les joueurs apprennent avant de passer à l'étape $t+1$ est le couple (i_t, j_t).

Le joueur 1 est aussi appelé joueur informé, alors que le joueur 2 est dit non informé. Remarquons que le joueur 1, connaissant k, est capable de déduire son paiement d'étape t ; ce n'est pas le cas du joueur 2 qui n'observe pas son paiement mais uniquement le couple (i_t, j_t). L'ensemble I est appelé l'ensemble d'actions du joueur 1, l'ensemble J celui du joueur 2, et l'ensemble K celui des états de la nature. Ces trois ensembles seront supposés finis et non vides dans tout ce qui suit.

Une stratégie du joueur 1 est un élément $\sigma = (\sigma_t)_{t\geqslant 1}$, avec σ_t : $K \times (I \times J)^{t-1} \to \Delta(I)$ pour tout t (par convention, $(I \times J)^0$ est un singleton). Une stratégie du joueur 2 est un élément $\tau(\tau_t)_{t\geqslant 1}$, avec

$\tau_t : (I \times J)^{t-1} \to \Delta(J)$ pour tout t. On note respectivement Σ et $\mathcal{T}$ les ensembles de stratégies des joueurs 1 et 2. Un couple de stratégies (σ, τ) induit, avec p, une probabilité sur l'ensemble des parties $K \times (I \times J)^\infty$ muni de la tribu produit, et pour tout entier strictement positif T on définit le paiement moyen jusqu'à l'étape T par :

$$\gamma_T^p(\sigma, \tau) = \mathbb{E}_{p,\sigma,\tau}\left(\frac{1}{T}\sum_{t=1}^{T} G^{\widetilde{k}}(\widetilde{\imath}_t, \widetilde{\jmath}_t)\right) = \sum_{k \in K} p^k \gamma_T^k(\sigma, \tau).$$

(les « tilde » indiquent des variables aléatoires, et par ailleurs on assimile k et la mesure de Dirac sur k). Le jeu répété T fois est le jeu sous forme stratégique $(\Sigma, \mathcal{T}, \gamma_T^p)$. Par des arguments standards (théorème 2.3 du texte de T. Tomala sur les jeux répétés dans ce volume) il a une valeur notée $v_T(p)$ et les deux joueurs y ont des stratégies optimales. En ce qui concerne la valeur du jeu infiniment répété $\Gamma_\infty(p)$, rappelons la définition suivante :

Définition 1.1. — Soit v un réel.

– Le joueur 1 garantit v dans $\Gamma_\infty(p)$ si :

$$\forall \varepsilon > 0,\ \exists \sigma \in \Sigma,\ \exists T_0,\ \forall T \geqslant T_0,\ \forall \tau \in \mathcal{T}, \quad \gamma_T^p(\sigma, \tau) \geqslant v - \varepsilon.$$

– Le joueur 2 garantit v dans $\Gamma_\infty(p)$ si :

$$\forall \varepsilon > 0,\ \exists \tau \in \mathcal{T},\ \exists T_0,\ \forall T \geqslant T_0,\ \forall \sigma \in \Sigma, \quad \gamma_T^p(\sigma, \tau) \leqslant v + \varepsilon.$$

– v est la valeur du jeu $\Gamma_\infty(p)$ si les deux joueurs y garantissent v.

La valeur du jeu, quand elle existe, est nécessairement unique. Quand le jeu $\Gamma_\infty(p)$ a une valeur v, une stratégie σ du joueur 1 est dite optimale si :

$$\forall \varepsilon > 0,\ \exists T_0,\ \forall T \geqslant T_0,\ \forall \tau \in \mathcal{T}, \quad \gamma_T^p(\sigma, \tau) \geqslant v - \varepsilon,$$

et similairement, une stratégie τ du joueur 2 est dite optimale si :

$$\forall \varepsilon > 0,\ \exists T_0,\ \forall T \geqslant T_0,\ \forall \sigma \in \Sigma, \quad \gamma_T^p(\sigma, \tau) \leqslant v + \varepsilon.$$

1.1. Exemples élémentaires : stratégies complètement révélatrices, non révélatrices et partiellement révélatrices. — Les exemples suivants sont classiques (voir Aumann et Maschler 1995, Zamir 1992). À chaque fois, il y a deux états : $K = \{a, b\}$, et $p = (1/2, 1/2)$.

Exemple 1.2. — $G^a = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$ et $G^b = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$.

Cet exemple est trivial. Afin de maximiser son paiement, le joueur 1 n'a qu'à jouer, à chaque étape, l'action H(aut) si l'état est a et l'action B(as) si l'état est b. Ainsi $v_T(p) = 0 = v_\infty(p) = 0$.

Exemple 1.3. — $G^a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ et $G^b = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

Une stratégie « naïve » du joueur 1 jouerait à l'étape 1 l'action H si l'état est a, et l'action B si l'état est b. Une telle stratégie est dite *complètement révélatrice*, ou CR, car elle permet au joueur 2 de déduire l'état sélectionné après avoir observé les actions du joueur 1. Cette stratégie est optimale ici dans le jeu à une étape, et $v_1(p) = 1/2$. Mais elle est très mauvaise quand le jeu est répété, et ne garantit rien de plus que 0 dans $\Gamma_\infty(p)$. À l'inverse, le joueur 1 peut toujours ne pas tenir compte de son information, et jouer une stratégie *non révélatrice*, ou NR, *i.e.* jouer indépendamment de l'état. Il considère alors la matrice moyenne

$$\frac{1}{2}G^a + \frac{1}{2}G^b = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix},$$

et peut y jouer à chaque étape une stratégie optimale, qui est ici unique et vaut $\frac{1}{2}H + \frac{1}{2}B$. Ainsi a-t-on : $v_T(p) \geqslant 1/4$ pour tout T. Nous verrons plus tard que cette façon de jouer est ici optimale pour le joueur 1 dans $\Gamma_\infty(p)$.

Exemple 1.4. — $G^a = \begin{pmatrix} 4 & 0 & 2 \\ 4 & 0 & -2 \end{pmatrix}$ et $G^b = \begin{pmatrix} 0 & 4 & -2 \\ 0 & 4 & 2 \end{pmatrix}$.

Jouer une stratégie CR ne garantit que 0 pour le joueur 1, car le joueur 2 pourra finalement jouer l'action M(du milieu) si l'état est a, et l'action G(auche) si l'état est b. Mais jouer NR revient à se placer dans le jeu

$$\frac{1}{2}G^a + \frac{1}{2}G^b = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \end{pmatrix},$$

et donc ne garantit que 0. Nous prouverons plus tard qu'il est ici optimal pour le joueur 1 de jouer la stratégie σ présentée maintenant.

Le joueur 1 choisit aléatoirement, une fois pour toutes, un élément s dans $\{H, B\}$ de la façon suivante : si $k = a$, alors $s = H$ avec probabilité 3/4, et donc $s = B$ avec probabilité 1/4 ; et si $k = b$, alors

$s = H$ avec probabilité $1/4$, et $s = B$ avec probabilité $3/4$. Ensuite le joueur 1 joue l'action s à chaque étape, indépendamment des coups du joueur 2.

Les probabilités conditionnelles vérifient : $P(k = a|s = H) = 3/4$, et $P(k = a|s = B) = 1/4$. Donc à la fin de l'étape 1, le joueur 2 ayant observé le premier coup du joueur 1 aura appris quelque chose sur l'état de la nature : sa croyance sera passée de $\frac{1}{2}a + \frac{1}{2}b$ à $\frac{3}{4}a + \frac{1}{4}b$ ou à $\frac{1}{4}a + \frac{3}{4}b$. Mais il ne connaît toujours pas l'état avec probabilité 1 : on parle de *révélation partielle d'information.*

1.2. Utilisation de l'information : éclatement et concavification

$p = (p^k)_{k \in K}$ est la probabilité initiale, ou l'*a priori* du joueur 2 sur l'état de la nature. Le joueur 1 choisit sa première action (ou plus généralement un message ou signal s au sein d'un ensemble fini S), en fonction de l'état de la nature k sélectionné. Notons $x = (x^k)_{k \in K} \in \Delta(S)^K$ la probabilité de transition utilisée : si l'état est k, le joueur 1 choisit le signal s avec probabilité $x^k(s)$.

La probabilité totale que s arrive est notée : $\lambda_s = \sum_{k \in K} p^k x^k(s)$, et si $\lambda_s > 0$, la probabilité conditionnelle sur K (ou *a posteriori*) sachant s vaut : $\widehat{p}(x, s) = \left(\frac{p^k x^k(s)}{\lambda_s}\right)_{k \in K}$. Il est clair que :

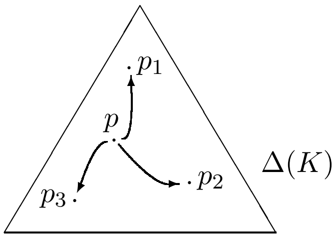
$$\sum_{s \in S} \lambda_s \widehat{p}(x, s) = p. \tag{1}$$

Les *a posteriori* contiennent donc l'*a priori* p dans leur enveloppe convexe. Le lemme suivant est fondamental et exprime une sorte de réciproque : on dit que le joueur 1 peut amener n'importe quels *a posteriori* contenant la probabilité initiale dans leur enveloppe convexe.

***Lemme 1.5* (éclatement).** — *Supposons que $p = \sum_{s \in S} \lambda_s p_s$, avec S fini, pour tout s de S $\lambda_s > 0$, $p_s \in \Delta(K)$, et $\sum_{s \in S} \lambda_s = 1$. Alors il existe une probabilité de transition $x \in \Delta(S)^K$ telle que :*

$$\forall\, s \in S, \quad \lambda_s = \sum_{k \in K} p^k x^k(s) \quad \textit{et} \quad \widehat{p}(x, s) = p_s.$$

Démonstration. — On pose $x^k(s) = \lambda_s p_s^k / p^k$ si $p^k > 0$. □



Soit f une application semi-continue supérieurement de $\Delta(K)$ dans $\mathbb{R}$. La plus petite fonction concave de $\Delta(K)$ dans $\mathbb{R}$ partout plus grande que f, notée $\operatorname{cav} f$, est continue et, pour tout p dans $\Delta(K)$,

$$\operatorname{cav} f(p) = \max\Big\{ \textstyle\sum_{s\in S} \lambda_s f(p_s) \mid S \text{ fini}, \forall\, s \in S\ \lambda_s \geqslant 0, \\ p_s \in \Delta(K),\ \sum_{s\in S} \lambda_s = 1,\ \sum_{s\in S} \lambda_s p_s = p \Big\}.$$

Nous allons voir que le lemme 1.5 a une conséquence importante du point de vue de ce qui peut être garanti par le joueur 1. Le résultat suivant est valable dans n'importe quel jeu finiment ou infiniment répété (en fait dans tout jeu à somme nulle où le joueur 1 a toute l'information). On dit que le joueur 1 garantit f si pour chaque valeur de p, le joueur 1 garantit $f(p)$ dans le jeu de probabilité initiale p.

Lemme 1.6. — *Si le joueur 1 garantit f, alors le joueur 1 garantit* $\operatorname{cav} f$.

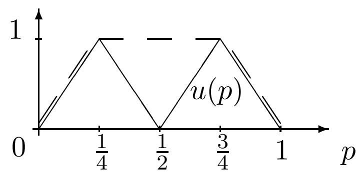
Démonstration. — Fixons p, et considérons l'égalité $\operatorname{cav} f(p) = \sum_{s\in S} \lambda_s f(p_s)$, avec les notations précédentes. Le joueur 1 observe l'état k, puis choisit s dans S selon $x^k(s)$, où $x = (x^k)_{k\in K}$ est donné par le splitting (lemme 1.5). Puis il joue de façon à garantir $f(p_s)$ dans le jeu de probabilité initiale p_s. Par le théorème de Kuhn sur l'équivalence entre les stratégies générales et les stratégies de comportement (voir le théorème 1.4 du texte de T. Tomala sur les jeux répétés dans ce volume), la stratégie du joueur 1 ainsi définie peut se voir comme une stratégie dans Σ. Ainsi le joueur 1 garantit-il $\sum_{s\in S} \lambda_s f(p_s)$ dans le jeu de probabilité initiale p. $\square$

Remarquons que le fait que le joueur 2 observe ou non, apprenne ou pas, l'élément s n'a aucune influence sur la preuve ci-dessus. Dans un jeu à somme nulle où le joueur 1 (maximisateur) a toute l'information, le lemme 1.6 donne : si le jeu a une valeur $v(p)$ pour toute probabilité initiale p, alors v est nécessairement concave. On peut voir cela comme le fait que la valeur de l'information est positive dans un jeu à somme nulle.

1.3. Le joueur 1 garantit $\mathrm{cav}\, u$**.** — Notons, pour toute probabilité p dans $\Delta(K)$, $u(p)$ la valeur du jeu matriciel $G(p) = \sum_{k\in K} p^k G^k$. Le joueur 1 peut garantir $u(p)$ dans $\Gamma_\infty(p)$ en jouant à chaque étape une stratégie optimale dans $G(p)$, indépendamment de l'état et des coups passés. u est la valeur du jeu répété où le joueur 1 joue indépendamment de l'état, et $G(p)$ s'appelle le jeu non révélateur à p. Clairement le joueur 1 garantit u, donc on a, dans un jeu à T étapes ou dans le jeu infiniment répété :

Corollaire 1.7. — *Le joueur 1 garantit* $\mathrm{cav}\, u$.

Revenons sur les exemples précédents. Dans l'exemple 1.2, on a $u(p) = -p(1-p)$ pour tout p (en faisant l'abus de notation : p probabilité de l'état a). u est convexe, et $\mathrm{cav}\, u(p) = 0$ pour tout p. Dans l'exemple 1.3, on a $u(p) = p(1-p)$ pour tout p, donc u est concave et $u = \mathrm{cav}\, u$. Concernant l'exemple 1.4, on a représenté u, en plein, et $\mathrm{cav}\, u$, en pointillés, sur la figure suivante.



Reprenons la stratégie partiellement révélatrice du joueur 1 présentée précédemment. Avec probabilité $1/2$, l'*a posteriori* vaudra $\frac{3}{4}a + \frac{1}{4}b$, et le joueur 1 jouera H qui est optimale dans

$$\frac{3}{4}G^a + \frac{1}{4}G^b = \begin{pmatrix} 3 & 1 & 1 \\ 3 & 1 & -1 \end{pmatrix}.$$

De même avec probabilité $1/2$ l'*a posteriori* vaudra $\frac{1}{4}a + \frac{3}{4}b$ et le joueur 1 jouera une stratégie optimale dans $\frac{1}{4}G^a + \frac{3}{4}G^b$. Cette stratégie garantit donc :

$$\frac{1}{2}\, u(3/4) + \frac{1}{2}\, u(1/4) = \mathrm{cav}\, u(1/2).$$

1.4. Le joueur 2 garantit $\lim_T v_T$**.** — T étant fixé, le jeu finiment répété à T étapes $\Gamma_T(p)$ a une valeur $v_T(p)$ et les deux joueurs y ont des stratégies optimales. v_T est concave d'après le lemme 1.6, et lipschitzienne de constante

$$M = \max\{|G^k(i,j)| \mid k \in K,\, i \in I,\, j \in J\}$$

indépendante de T.

Définition 1.8. — Pour p dans $\Delta(K)$, on pose $v^*(p) = \inf_{T \geqslant 1} v_T(p)$.

v^* est concave et M-Lipschitz.

Lemme 1.9. — *La suite de fonctions $(v_T)_T$ converge uniformément sur $\Delta(K)$ vers v^*, et le joueur 2 garantit v^* dans le jeu infiniment répété.*

Démonstration. — Le joueur 2 peut se comporter ainsi dans le jeu $\Gamma_\infty(p)$: jouer une stratégie optimale de $\Gamma_T(p)$, puis tout oublier, et recommencer à jouer une stratégie optimale de $\Gamma_T(p)$, puis tout oublier, et recommencer etc. Ainsi le joueur 2 garantit $v_T(p)$ dans $\Gamma_\infty(p)$. Donc le joueur 2 garantit $v^*(p)$. D'après la définition 1.1, cela implique que : $\limsup_T v_T(p) \leqslant v^*(p)$. Ce qui entraîne que v_T converge simplement vers v^*. La convergence est uniforme car toutes ces fonctions sont lipschitziennes de même constante. □

Indiquons également que l'on a pour tous T, $T' \geqslant 1$, et p dans $\Delta(K)$:

$$(T + T')\, v_{T+T'}(p) \leqslant T\, v_T(p) + T'\, v_{T'}(p).$$

En effet, le joueur 2 peut jouer une stratégie optimale dans $\Gamma_T(p)$, puis tout oublier et jouer une stratégie optimale dans $\Gamma_{T'}(p)$.

1.5. Martingale des *a posteriori*. — La définition suivante est essentielle.

Définition 1.10. — Soit (σ, τ) dans $\Sigma \times \mathcal{T}$ un profil de stratégies dans $\Gamma_\infty(p)$. Pour t dans $\mathbb{N}$ et $h_t = (i_1, j_1, \ldots, i_t, j_t) \in (I \times J)^t$, on définit l'*a posteriori* du joueur 2 après h_t comme :

$$p_t(\sigma, \tau, h_t) = (p_t^k(\sigma, \tau, h_t))_{k \in K} \in \Delta(K),$$

avec $p_t^k(\sigma, \tau, h_t) = \mathbb{P}_{p,\sigma,\tau}(\widetilde{k} = k | h_t)$ pour tout k.

$p_t(\sigma, \tau, h_t)$ est la croyance du joueur 2 sur l'état de la nature à la fin de l'étape t si h_t a été précédemment joué et que le joueur 1 utilise σ. On a facilement que $p_t(\sigma, \tau, h_t)$ ne dépend ni de τ, ni de l'action jouée par le joueur 2 en date t, pourvu que $\mathbb{P}_{p,\sigma,\tau}(h_t) > 0$. On note donc aussi $p_t(\sigma, h_t)$ pour $p_t(\sigma, \tau, h_t)$, dès qu'il existe τ telle que $\mathbb{P}_{p,\sigma,\tau}(h_t) > 0$ (et on peut définir arbitrairement $p_t(\sigma, \tau, h_t)$ et $p_t(\sigma, h_t)$ dans $\Delta(K)$ sinon). Notons $H_t = (I \times J)^t$ et $\mathcal{H}_t$ la tribu sur $K \times (I \times J)^\infty$ engendrée par la projection sur H_t donnant les actions jouées aux t premières étapes. $p_t(\sigma)$ est vue comme une application mesurable de $(K \times (I \times J)^\infty, \mathcal{H}_t)$ dans $\Delta(K)$, donc est une variable aléatoire à valeurs dans $\Delta(K)$. $p_0(\sigma)$ est la probabilité initiale p. La propriété suivante est cruciale.

Propriété 1.11. — Par rapport à $\mathbb{P}_{p,\sigma,\tau}$, la suite $(p_t(\sigma))_{t\geqslant 0}$ est une $(\mathcal{H}_t)_{t\geqslant 0}$- martingale à valeurs dans $\Delta(K)$.

Cette propriété est très générale et concerne l'apprentissage bayésien d'un paramètre inconnu : l'espérance de ce que je saurai demain est ce que je sais aujourd'hui. On peut facilement en donner une démonstration analytique de façon analogue à l'obtention de l'équation (1).

Fixons maintenant une stratégie σ du joueur 1. L'idée des calculs suivants est la suivante. La martingale $(p_t(\sigma))_{t\geqslant 0}$ étant bornée, elle converge p.s. et on a une borne uniforme sur sa variation L^1 (voir le lemme 1.12). Cela implique qu'au bout d'un moment, la martingale sera essentiellement constante, et donc qu'à partir d'une certaine étape le joueur 1 jouera approximativement de façon non révélatrice, et ne pourra pas garantir beaucoup plus que $u(q)$, où q est « un *a posteriori* limite ». Les *a posteriori* étant liés à la probabilité initiale, le joueur 1 ne pourra pas garantir plus, sur le long terme, que

$$\max\Big\{ \textstyle\sum_{s\in S} \lambda_s u(p_s) \mid S \text{ fini}, \forall\, s \in S\, \lambda_s \geqslant 0, \\ p_s \in \Delta(K), \sum_{s\in S} \lambda_s = 1, \sum_{s\in S} \lambda_s p_s = p\Big\},$$

autrement dit que $\operatorname{cav} u(p)$. Passons maintenant à la preuve formelle.

σ étant fixée, on définit la stratégie $\tau = (\tau_t)_{t\geqslant 1}$ du joueur 2 de la façon suivante : jouer à chaque étape une stratégie optimale dans le jeu matriciel $G(p_t)$, où p_t est l'a posteriori courant.

Supposons que (σ,τ) soit joué dans le jeu répété $\Gamma_\infty(p)$. Afin d'alléger les notations, on note dans les calculs suivants $\mathbb{P} = \mathbb{P}_{p,\sigma,\tau}$, on note $\mathbb{E}$ l'espérance par rapport à $\mathbb{P}$, et $p_t(h_t)$ pour $p_t(\sigma, h_t)$. Toutes les normes indiquées sont des normes 1. L'inégalité suivante est juste due au fait que $(p_t)_t$ est une martingale à valeurs dans $\Delta(K)$ et d'espérance p.

Lemme 1.12

$$\forall\, T \geqslant 1, \quad \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}(\|p_{t+1} - p_t\|) \leqslant \frac{\sum_{k\in K}\sqrt{p^k(1-p^k)}}{\sqrt{T}}.$$

Démonstration. — On a pour tous k dans K et $t \geqslant 0$:

$$\mathbb{E}\big((p^k_{t+1} - p^k_t)^2\big) = \mathbb{E}\Big(\mathbb{E}\big((p^k_{t+1} - p^k_t)^2|\mathcal{H}_t\big)\Big) = \mathbb{E}\big((p^k_{t+1})^2\big) - \mathbb{E}\big((p^k_t)^2\big).$$

Donc

$$\mathbb{E}\bigg(\sum_{t=0}^{T-1}(p^k_{t+1} - p^k_t)^2\bigg) = \mathbb{E}\big((p^k_T)^2\big) - (p^k)^2 \leqslant p^k(1-p^k).$$

Comme par Cauchy-Schwartz

$$\mathbb{E}\bigg(\frac{1}{T}\sum_{t=0}^{T-1}\Big|p^k_{t+1} - p^k_t\Big|\bigg) \leqslant \sqrt{\frac{1}{T}\mathbb{E}\bigg(\sum_{t=0}^{T-1}(p^k_{t+1} - p^k_t)^2\bigg)}$$

pour chaque k, on a le résultat voulu. □

Pour h_t dans H_t, on note $\sigma^k(h_t) = \sigma_{t+1}(k, h_t) \in \Delta(I)$ l'action mixte jouée par le joueur 1 après h_t si l'état est k, et $\overline{\sigma}(h_t)$ la loi de l'action de date $t+1$ jouée par le joueur 1 après h_t :

$$\overline{\sigma}(h_t) = \sum_{k\in K} p^k_t(\sigma, h_t)\sigma^k(h_t) \in \Delta(I).$$

$\overline{\sigma}(h_t)$ peut se voir comme la stratégie moyenne après h_t, et servira d'approximation non révélatrice à $(\sigma^k(h_t))_k$. Le lemme suivant exprime le lien entre, d'une part, la variation de la martingale $(p_t(\sigma))_{t\geqslant 0}$, *i.e.* l'information révélée par le joueur 1, et d'autre part, la dépendance par rapport à l'état de l'action jouée par le joueur 1 en date $t+1$, *i.e.* l'information employée par le joueur 1.

Lemme 1.13. — $\forall\, t \geqslant 0$, $\forall\, h_t \in H_t$,

$$\mathbb{E}\left(\|p_{t+1} - p_t\|\,|h_t\right) = \mathbb{E}\Big(\big\|\sigma^{\widetilde{k}}(h_t) - \overline{\sigma}(h_t)\big\||h_t\Big).$$

Démonstration. — Fixons $t \geqslant 0$ et h_t dans H_t tel que $\mathbb{P}_{p,\sigma,\tau}(h_t) > 0$. Pour (i_{t+1}, j_{t+1}) dans $I \times J$, on a :

$$\begin{aligned} p_{t+1}^k(h_t, i_{t+1}, j_{t+1}) &= \mathbb{P}(\widetilde{k} = k | h_t, i_{t+1}) \\ &= \frac{\mathbb{P}(\widetilde{k} = k | h_t)\mathbb{P}(i_{t+1} | k, h_t)}{\mathbb{P}(i_{t+1} | h_t)} \\ &= \frac{p_t^k(h_t)\sigma^k(h_t)(i_{t+1})}{\overline{\sigma}(h_t)(i_{t+1})}. \end{aligned}$$

Donc :

$$\begin{aligned} \mathbb{E}(\|p_{t+1} - p_t\| | h_t) &= \sum_{i_{t+1} \in I} \overline{\sigma}(h_t)(i_{t+1}) \sum_{k \in K} |p_{t+1}^k(h_t, i_{t+1}) - p_t^k(h_t)| \\ &= \sum_{i_{t+1} \in I} \sum_{k \in K} |p_t^k(h_t)\sigma^k(h_t)(i_{t+1}) - \overline{\sigma}(h_t)(i_{t+1}) p_t^k(h_t)| \\ &= \sum_{k \in K} p_t^k(h_t) \|\sigma^k(h_t) - \overline{\sigma}(h_t)\| \\ &= \mathbb{E}\big(\|\sigma^{\widetilde{k}}(h_t) - \overline{\sigma}(h_t)\| | h_t\big). \qquad \square \end{aligned}$$

On peut maintenant majorer les paiements. Pour $t \geqslant 0$ et h_t dans H_t :

$$\begin{aligned} \mathbb{E}\left(G^{\widetilde{k}}(\widetilde{i}_{t+1}, \widetilde{j}_{t+1}) | h_t\right) &= \sum_{k \in K} p_t^k(h_t) G^k(\sigma^k(h_t), \tau_{t+1}(h_t)) \\ &\leqslant \sum_{k \in K} p_t^k(h_t) G^k(\overline{\sigma}(h_t), \tau_{t+1}(h_t)) \\ &\qquad + M \sum_{k \in K} p_t^k(h_t) \|\sigma^k(h_t) - \overline{\sigma}(h_t)\| \\ &\leqslant u(p_t(h_t)) + M \sum_{k \in K} p_t^k(h_t) \|\sigma^k(h_t) - \overline{\sigma}(h_t)\|, \end{aligned}$$

où $u(p_t(h_t))$ vient de la définition de τ. Par le lemme 1.13, on obtient :

$$\mathbb{E}\left(G^{\widetilde{k}}(\widetilde{i}_{t+1}, \widetilde{j}_{t+1}) | h_t\right) \leqslant u(p_t(h_t)) + M\mathbb{E}\left(\|p_{t+1} - p_t\| | h_t\right).$$

En appliquant l'inégalité de Jensen, il vient :

$$\mathbb{E}\left(G^{\widetilde{k}}(\widetilde{i}_{t+1}, \widetilde{j}_{t+1})\right) \leqslant \operatorname{cav} u(p) + M\mathbb{E}\left(\|p_{t+1} - p_t\|\right).$$

On applique maintenant le lemme 1.12, et on obtient :

$$\begin{aligned}\gamma_T^p(\sigma,\tau) &= \mathbb{E}\bigg(\frac{1}{T}\sum_{t=0}^{T-1} G^{\widetilde{k}}(\widetilde{i}_{t+1},\widetilde{j}_{t+1})\bigg)\\ &\leqslant \operatorname{cav} u(p) + \frac{M}{\sqrt{T}}\sum_{k\in K}\sqrt{p^k(1-p^k)}.\end{aligned}$$

Il reste à considérer le cas d'une stratégie σ optimale dans $\Gamma_T(p)$ et on a prouvé le résultat suivant.

Proposition 1.14. — *Pour p dans $\Delta(K)$ et $T \geqslant 1$,*

$$v_T(p) \leqslant \operatorname{cav} u(p) + \frac{M'}{\sqrt{T}}, \quad \textit{avec} \quad M' = M\sum_{k\in K}\sqrt{p^k(1-p^k)}.$$

1.6. Valeur de $\Gamma_\infty(p)$. — D'après le corollaire 1.7, pour tout T le joueur 1 garantit $\operatorname{cav} u(p)$ dans $\Gamma_T(p)$, donc on a $v_T(p) \geqslant \operatorname{cav} u(p)$, et donc $v^* \geqslant \operatorname{cav} u$. En passant à la limite en T dans la proposition 1.14, on trouve $v^* \leqslant \operatorname{cav} u$, donc $v^* = \operatorname{cav} u$. Avec le corollaire 1.7 et le lemme 1.9, on obtient :

***Théorème 1.15* (Aumann et Maschler).** — *Le jeu $\Gamma_\infty(p)$ a une valeur qui vaut* $\operatorname{cav} u(p)$.

1.7. Formule de récurrence. — On montre facilement la formule de récurrence suivante, où $\max_{x\in\Delta(I)^K}$ et $\min_{y\in\Delta(J)}$ commutent par le théorème de Sion.

$$\begin{aligned}&v_{T+1}(p)\\ &\quad= \frac{1}{T+1}\max_{x\in\Delta(I)^K}\min_{y\in\Delta(J)}\bigg(G(p,x,y) + T\sum_{i\in I} x(p)(i) v_T(\widehat{p}(x,i))\bigg),\end{aligned}$$

où, avec des notations similaires au paragraphe 1.2 :

– $x = (x^k(i))_{i\in I,k\in K}$, avec x^k la probabilité utilisée par le joueur 1 à l'étape 1 si l'état est k,

– $y = (y(j))_{j\in J}$ est la probabilité utilisée par le joueur 2 à l'étape 1,

– $G(p,x,y) = \sum_k p^k G^k(x^k,y)$ est le paiement espéré d'étape 1,

– $x(p)(i) = \sum_{k\in K} p^k x^k(i)$ est la probabilité que i soit jouée en date 1,

– $\widehat{p}(x,i)$ est la probabilité conditionnelle sur K sachant i.

La propriété suivante s'interprète facilement : plus on joue longtemps, moins l'information initiale n'est importante.

Lemme 1.16. — $(v_T(p))_T$ *est décroissante.*

Démonstration. — Montrons $v_{T+1} \leqslant v_T$ par récurrence sur T.

$$v_2(p) = \frac{1}{2} \max_{x \in \Delta(I)^K} \min_{y \in \Delta(J)} \Big(G(p,x,y) + T \sum_{i \in I} x(p)(i) v_1(\widehat{p}(x,i)) \Big)$$

Comme v_1 est concave,

$$v_2(p) \leqslant \frac{1}{2} \max_{x \in \Delta(I)^K} \min_{y \in \Delta(J)} (G(p,x,y) + v_1(p)) = v_1(p).$$

Soit $T \geqslant 2$ tel que : $\forall\, p,\ v_T(p) \leqslant v_{T-1}(p)$. Alors,

$$\begin{aligned}(T+1)v_{T+1}(p) &= \max_{x \in \Delta(I)^K} \min_{y \in \Delta(J)} \Big(G(p,x,y) \\ &\qquad + (T-1) \sum_{i \in I} x(p)(i) v_T(\widehat{p}(x,i)) + \sum_{i \in I} x(p)(i) v_T(\widehat{p}(x,i) \Big) \\ &\leqslant \max_{x \in \Delta(I)^K} \min_{y \in \Delta(J)} \Big(G(p,x,y) \\ &\qquad\qquad + (T-1) \sum_{i \in I} x(p)(i) v_{T-1}(\widehat{p}(x,i)) + v_T(p) \Big) \\ &= T v_T(p) + v_T(p) \\ &= (T+1) v_T(p). \end{aligned}$$

□

2. Jeux à paiements vectoriels et approchabilité

Le modèle suivant a été introduit par D. Blackwell (1956). Il est intéressant en tant que tel, mais permettra aussi d'expliciter, en 2.4, la construction d'une stratégie optimale pour le joueur non informé dans le jeu $\Gamma_\infty(p)$ de la section précédente. On a ici aussi une famille de matrices $(G^k)_{k \in K}$ de même taille $I \times J$. À chaque étape t, le joueur 1 choisit $i_t \in I$, simultanément le joueur 2 choisit $j_t \in J$, et le « paiement » d'étape t est alors le vecteur $G(i_t, j_t) = (G^k(i_t, j_t))_{k \in K}$ dans $\mathbb{R}^K$. Précisons qu'il n'y a pas de vrai état k ici, ni de probabilité a priori sur K, et les deux joueurs ont un rôle symétrique. On suppose qu'après chaque étape chaque joueur observe le vecteur de paiement. Les actions des joueurs ne sont en principe pas observées (on pourra remarquer plus tard que cela ne changerait en fait rien aux résultats, qui nécessitent juste de supposer que les joueurs observent *au moins* le vecteur de paiement). L'approchabilité vise à répondre aux questions du genre : dans quels ensembles le joueur 1 (par exemple) peut-il amener le paiement moyen de long terme ?

Dans toute cette section sur l'approchabilité, on considère norme et distance euclidiennes. La présentation s'est inspirée de Sorin (2002). Notons $F = \{(G^k(i,j))_{k\in K} \mid i \in I,\, j \in J\}$ l'ensemble fini des paiements d'étapes possibles, et M une constante telle que $\|u\| \leqslant M$ pour tout u de F. Une stratégie du joueur 1 est un élément $\sigma = (\sigma_t)_{t\geqslant 1}$, où σ_t est une application de F^{t-1} dans $\Delta(I)$ pour tout t. De même pour le joueur 2 en remplaçant $\Delta(I)$ par $\Delta(J)$. On note respectivement Σ et $\mathcal{T}$ les espaces de stratégies des joueurs 1 et 2. Un profil de stratégies (σ,τ) dans $\Sigma\times\mathcal{T}$ induit « naturellement » une unique probabilité sur $(I \times J \times F)^\infty$ notée $\mathbb{P}_{\sigma,\tau}$. On note g_t la variable aléatoire, à valeurs dans F, du paiement d'étape t, et $\overline{g}_t = \frac{1}{t}\sum_{t'=1}^{t} g_{t'} \in \mathrm{conv}(F)$. Soit $C \subset \mathbb{R}^K$ un ensemble « cible », que l'on supposera toujours sans perte de généralité fermé. On note $d_t = d(\overline{g}_t, C)$ la v.a. de la distance euclidienne de $\overline{g}_t$ à C.

Définition 2.1

– C est approchable par le joueur 1 si :

$$\forall\, \varepsilon > 0,\ \exists\, \sigma \in \Sigma,\ \exists\, T,\ \forall\, \tau \in \mathcal{T},\ \forall\, t \geqslant T, \quad \mathbb{E}_{\sigma,\tau}(d_t) \leqslant \varepsilon.$$

– C est approchable par le joueur 2 si :

$$\forall\, \varepsilon > 0,\ \exists\, \tau \in \mathcal{T},\ \exists\, T,\ \forall\, \sigma \in \Sigma,\ \forall\, t \geqslant T, \quad \mathbb{E}_{\sigma,\tau}(d_t) \leqslant \varepsilon.$$

– C est évitable (on dit aussi repoussable) par le joueur 1, resp. joueur 2, s'il existe $\delta > 0$ tel que $\{z \in \mathbb{R}^K, d(z,C) \geqslant \delta\}$ soit approchable par le joueur 1, resp. joueur 2.

C est approchable par le joueur 1 si pour tout $\varepsilon > 0$, le joueur 1 peut s'assurer qu'au bout d'un certain temps on aura $\mathbb{E}(d_t) \leqslant \varepsilon$, donc on sera en espérance proche de C à ε près. Un ensemble ne peut être à la fois approchable par le joueur 1 et évitable par le joueur 2. Par exemple si K est un singleton, on est en dimension 1 et le théorème du minmax implique que pour tout réel t, l'ensemble $[t, +\infty[$ est soit approchable par le joueur 1, soit évitable par le joueur 2 (selon la position de t par rapport à $\max_{x\in\Delta(I)} \min_{y\in\Delta(J)} G(x,y)$).

2.1. Conditions suffisantes et nécessaires d'approchabilité

Pour x dans $\Delta(I)$, on note

$$xG = \{G(x,y) \mid y \in \Delta(J)\} = \mathrm{conv}\{\textstyle\sum_{i\in I} x_i G(i,j) \mid j \in J\}.$$

xG est l'ensemble des paiements vectoriels espérés possibles quand le joueur 1 joue l'action mixte x. De même, pour y dans $\Delta(J)$ on note

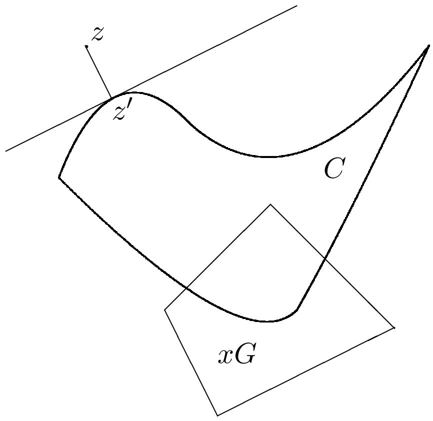
$$Gy = \{G(x,y) \mid x \in \Delta(I)\}.$$

Dans la définition suivante, la lettre B fait référence à Blackwell.

Définition 2.2. — C est un B-ensemble pour le joueur 1 si pour tout $z \notin C$, il existe $z' \in C$ et $x \in \Delta(I)$ satisfaisant les deux conditions suivantes :

(i) $\|z' - z\| = d(z, C)$,

(ii) L'hyperplan affine passant par z' et orthogonal à $[z, z']$ sépare z de xG.



Par exemple, si x est dans $\Delta(I)$, l'ensemble convexe xG est approchable par le joueur 1. Étant donné un B-ensemble C pour le joueur 1, on définit une stratégie σ adaptée à C de la façon suivante. Pour tout t de $\mathbb{N}$, en date $t+1$ le joueur 1 considère le paiement moyen courant $\overline{g}_t$. Si $\overline{g}_t \in C$ (ou si $t = 0$), σ joue arbitrairement en date $t+1$. Si $\overline{g}_t \notin C$ (et $t \geqslant 1$), σ joue en date $t+1$ une action mixte $x \in \Delta(I)$ vérifiant la définition précédente pour $z = \overline{g}_t$.

Théorème 2.3. — *Soit C un B-ensemble pour le joueur 1. Une stratégie σ adaptée à C vérifie :*

$$\forall\, \tau \in \mathcal{T},\, \forall\, t \geqslant 1 \quad \mathbb{E}_{\sigma,\tau}(d_t) \leqslant \frac{2M}{\sqrt{t}} \quad \textit{et} \quad d_t \underset{t\to\infty}{\longrightarrow} 0 \quad \mathbb{P}_{\sigma,\tau}\, p.s.$$

En dimension 1 et pour $C = \{0\}$, ce théorème implique en particulier qu'une suite de réels $(x_t)_t$ bornée, telle que le produit

$$x_{T+1}\left(\frac{1}{T}\sum_{t=1}^{T} x_T\right)$$

soit négatif pour tout T, converge en moyenne de Césaro vers 0.

Démonstration. — Supposons que le joueur 1 joue σ adaptée à C, alors que le joueur 2 joue une stratégie quelconque τ. Soit $t \geqslant 1$, et supposons que $\overline{g}_t \notin C$. On note $z' \in C$ et $x \in \Delta(I)$ qui satisfont (i) et (ii) de la définition 2.2 pour $z = \overline{g}_t$. On a :

$$\begin{aligned} d_{t+1}^2 = d(\overline{g}_{t+1}, C)^2 \leqslant \|\overline{g}_{t+1} - z'\|^2 & \\ &= \left\|\frac{1}{t+1}\sum_{l=1}^{t+1} g_l - z'\right\|^2 \\ &= \left\|\frac{1}{t+1}(g_{t+1} - z') + \frac{t}{t+1}(\overline{g}_t - z')\right\|^2 \\ &= \left(\frac{1}{t+1}\right)^2 \|g_{t+1} - z'\|^2 + \left(\frac{t}{t+1}\right)^2 {d_t}^2 \\ &\qquad + \frac{2t}{(t+1)^2}\langle g_{t+1} - z', \overline{g}_t - z'\rangle. \end{aligned}$$

Par hypothèse, l'espérance, sachant les t premiers coups $h_t \in (I\times J)^t$, du produit scalaire ci-dessus est négatif, donc

$$\mathbb{E}\left(d_{t+1}^2 | h_t\right) \leqslant \left(\frac{t}{t+1}\right)^2 {d_t}^2 + \left(\frac{1}{t+1}\right)^2 \mathbb{E}\left(\|g_{t+1} - z'\|^2 | h_t\right).$$

Or

$$\mathbb{E}\left(\|g_{t+1} - z'\|^2 | h_t\right) \leqslant \mathbb{E}\left(\|g_{t+1} - \overline{g}_t\|^2 | h_t\right) \leqslant (2M)^2,$$

donc :

$$\mathbb{E}\left(d_{t+1}^2 | h_t\right) \leqslant \left(\frac{t}{t+1}\right)^2 {d_t}^2 + \left(\frac{1}{t+1}\right)^2 4M^2. \tag{2}$$

En prenant l'espérance, on obtient, que $\overline{g}_t \notin C$ ou pas :

$$\forall\, t \geqslant 1, \quad \mathbb{E}\left(d_{t+1}^2\right) \leqslant \left(\frac{t}{t+1}\right)^2 \mathbb{E}({d_t}^2) + \left(\frac{1}{t+1}\right)^2 4M^2.$$

Donc on a par récurrence, pour tout $t \geqslant 1$, $\mathbb{E}(d_t^2) \leqslant 4M^2/t$, puis

$$\mathbb{E}(d_t) \leqslant \frac{2M}{\sqrt{t}}.$$

Posons maintenant, comme dans Sorin (2002),

$$e_t = d_t^2 + \sum_{t'>t} \frac{4M^2}{t'^2}.$$

L'inégalité (2) donne :

$$\mathbb{E}(e_{t+1}|h_t) \leqslant e_t,$$

donc (e_t) est une surmartingale positive dont l'espérance tend vers zéro. Donc $e_t \to_{t\to\infty} 0$ $\mathbb{P}_{\sigma,\tau}$ p.s., et enfin $d_t \to_{t\to\infty} 0$ $\mathbb{P}_{\sigma,\tau}$ p.s. □

Le théorème 2.3 implique que tout B-ensemble pour le joueur 1 est approchable par le joueur 1. La réciproque est vraie dans le cas des ensembles convexes.

Théorème 2.4. — *Soit C un convexe fermé de $\mathbb{R}^K$.*

(i) *C est un B-ensemble pour le joueur 1,*

$\iff$ (ii) $\forall\, y \in \Delta(J), Gy \cap C \neq \varnothing$,

$\iff$ (iii) *C est approchable par le joueur 1,*

$\iff$ (iv) $\forall\, q \in \mathbb{R}^K,\ \max_{x\in\Delta(I)} \min_{y\in\Delta(J)} \sum_{k\in K} q^k G^k(x,y) \geqslant \inf_{c\in C} \langle q, c\rangle.$

Démonstration. — (i) ⇒ (iii) vient du théorème précédent.

Montrons (iii) ⇒ (ii). Supposons qu'il existe $y \in \Delta(J)$ tel que $Gy \cap C = \varnothing$. Comme Gy est approchable par le joueur 2, alors C est évitable par le joueur 2 et donc pas approchable par le joueur 1.

Montrons (ii) ⇒ (i). Supposons que $Gy \cap C \neq \varnothing$ pour tout $y \in \Delta(J)$. Soient $z \notin C$ et z' sa projection sur C. Considérons le jeu matriciel où les paiements de G sont projetés dans la direction $z' - z$, *i.e.* le jeu matriciel $\sum_{k\in K}(z'^k - z^k)G^k$. Par hypothèse, on a :

$$\forall\, y \in \Delta(J),\ \exists\, x \in \Delta(I), \quad G(x,y) \in C,$$

donc :

$$\langle z' - z, G(x,y)\rangle \geqslant \min_{c\in C}\{\langle z' - z, c\rangle, \langle z' - z, z'\rangle\}.$$

Donc

$$\min_{y\in\Delta(J)} \max_{x\in\Delta(I)} \langle z' - z, G(x,y)\rangle \geqslant \langle z' - z, z'\rangle.$$

Par le théorème du minmax, il existe x dans $\Delta(I)$ tel que, pour tout $y \in \Delta(J)$,

$$\langle z' - z, G(x,y)\rangle \geqslant \langle z' - z, z'\rangle,$$

soit

$$\langle z' - z, z' - G(x,y)\rangle \leqslant 0.$$

(iv) signifie que tout demi-espace contenant C est approchable par le joueur 1. (iii) $\Rightarrow$ (iv) est donc clair. (iv) $\Rightarrow$ (i) est similaire à (ii) $\Rightarrow$ (i). □

Les théorèmes précédents 2.3 et 2.4 sont, à des différences mineures de formulation près, dues à Blackwell (1956). Indiquons que X. Spinat (2002) a montré récemment le résultat suivant :

Théorème 2.5. — *Un ensemble est approchable par le joueur 1 si et seulement si il contient un B-ensemble pour le joueur 1.*

Cela implique notamment qu'ajouter la condition

$$d_t \underset{t\to\infty}{\longrightarrow} 0 \quad \mathbb{P}_{\sigma,\tau}\, p.s.$$

dans la définition d'approchabilité ne change pas la notion.

2.2. Approchabilité par le joueur 1 versus évitabilité par le joueur 2. — Le résultat suivant est un corollaire du théorème 2.4, voir (ii).

Corollaire 2.6. — *Tout sous-ensemble convexe de $\mathbb{R}^K$ est soit approchable par le joueur 1, soit évitable par le joueur 2.*

Remarque 2.7. — Cas de la dimension 1. On montre que si K est un singleton, alors tout ensemble est soit approchable par le joueur 1, soit évitable par le joueur 2.

***Exemple 2.8* (Existence d'un ensemble ni approchable par le joueur 1, ni évitable par le joueur 2 en dimension 2 (voir Sorin, 2002))**

Posons

$$G = \begin{pmatrix} (0,0) & (0,0) \\ (1,0) & (1,1) \end{pmatrix},$$

$$C = \{(1/2, v) \mid 0 \leqslant v \leqslant 1/4\} \cup \{(1,v), 1/4 \leqslant v \leqslant 1\}.$$

Considérons la stratégie suivante σ du joueur 1 dans un jeu à $2T$ étapes : σ joue B(as) pendant les T premières étapes, puis de deux choses l'une : si la deuxième coordonnée des paiements est en moyenne pendant les T premières étapes supérieure (resp. strictement inférieure) à $1/2$, σ joue B (resp. Haut) les T étapes suivantes. Quelque que soit la stratégie du joueur 2, on aura $\overline{g}_{2T} \in C$. Donc C n'est pas évitable par le joueur 2.

On peut montrer que C n'est pas non plus approchable par le joueur 1 en considérant des stratégies du joueur 2 qui joue D(roite) pendant longtemps, induisant un paiement sur la première diagonale, puis G(auche) toujours.

2.3. Approchabilité faible. — On peut affaiblir la notion d'approchabilité en abandonnant l'uniformité de la stratégie par rapport au temps.

Définition 2.9

– C est faiblement approchable par le joueur 1 si :

$$\forall\,\varepsilon > 0,\ \exists\,T,\ \forall\,t \geqslant T,\ \exists\,\sigma \in \Sigma,\ \forall\,\tau \in \mathcal{T},\quad \mathbb{E}_{\sigma,\tau}(d_t) \leqslant \varepsilon.$$

– C est faiblement approchable par le joueur 2 si :

$$\forall\,\varepsilon > 0,\ \exists\,T,\ \forall\,t \geqslant T,\ \exists\,\tau \in \mathcal{T},\ \forall\,\sigma \in \Sigma,\quad \mathbb{E}_{\sigma,\tau}(d_t) \leqslant \varepsilon.$$

– C est faiblement évitable par le joueur 1, resp. joueur 2, s'il existe $\delta > 0$ tel que $\{z \in \mathbb{R}^K \mid d(z, C) \geqslant \delta\}$ soit faiblement approchable par le joueur 1, resp. joueur 2.

L'ensemble C de l'exemple 2.8 est faiblement approchable par le joueur 1. N. Vieille (1992) a prouvé le résultat suivant *via* l'introduction de certains jeux différentiels.

Théorème 2.10. — *Tout sous-ensemble de $\mathbb{R}^K$ est faiblement approchable par le joueur 1 ou faiblement repoussable par le joueur 2.*

2.4. Retour au modèle standard à manque d'information d'un seul côté : stratégie optimale explicite pour le joueur non informé. — Revenons ici au formalisme de la section 1. On a une famille de matrices $(G^k)_{k \in K}$, et une probabilité initiale p sur $\Delta(K)$. D'après le théorème 1.15, le jeu $\Gamma_\infty(p)$ a une valeur qui vaut $\operatorname{cav} u(p)$. En considérant un hyperplan tangent à $\operatorname{cav} u$ au point p, on peut trouver un vecteur $\ell \in \mathbb{R}^K$ tel que

$$\langle \ell, p \rangle = \operatorname{cav} u(p) \quad \text{et} \quad \forall\, q \in \Delta(K),\ \langle \ell, q \rangle \geqslant \operatorname{cav} u(q) \geqslant u(q).$$

Considérons maintenant l'orthant $C = \{z \in \mathbb{R}^K \mid \forall\, k \in K,\ z^k \leqslant \ell^k\}$. Soit $q = (q^k)_k$ dans $\mathbb{R}^K$.

S'il existe k tel que $q^k > 0$, on a :

$$\inf_{c \in C} \langle q, c\rangle = -\infty \leqslant \max_{y \in \Delta(J)} \min_{x \in \Delta(I)} \sum_{k \in K} q^k G^k(x,y).$$

Supposons maintenant que $q^k \leqslant 0$ pour tout k, avec $q \neq 0$. Posons $q' = -q$ et $s = \sum_k q'^k$.

$$\begin{aligned}\inf_{c \in C} \langle q, c\rangle &= \sum_{k \in K} q^k \ell^k = -s \sum_{k \in K} \frac{q'^k}{s} \ell^k = -s \langle \ell, q'/s\rangle \\ &\leqslant -s\, u(q'/s) \\ &\leqslant -s \max_{x \in \Delta(I)} \min_{y \in \Delta(J)} \sum_{k \in K} \frac{q'^k}{s} G^k(x,y) \\ &= \max_{y \in \Delta(J)} \min_{x \in \Delta(I)} \sum_{k \in K} q^k G^k(x,y)\end{aligned}$$

Donc l'analogue, pour le joueur 2, de la condition (iv) du théorème 2.4 est vérifiée. C est un B-ensemble pour le joueur 2, et est donc approchable par le joueur 2. Une stratégie τ adaptée à C pour le joueur 2 vérifie, par le théorème 2.3 : $\forall\, \sigma \in \Sigma$, $\forall\, k \in K$,

$$\mathbb{E}_{\sigma,\tau}\biggl(\frac{1}{T}\sum_{t=1}^{T} G^k(\widetilde{i}_t, \widetilde{j}_t) - \ell^k\biggr) \leqslant \mathbb{E}_{\sigma,\tau}\biggl(d\biggl(\frac{1}{T}\sum_{t=1}^{T} G^k(\widetilde{i}_t, \widetilde{j}_t), C\biggr)\biggr) \leqslant \frac{2M}{\sqrt{T}},$$

où M est ici un majorant de la norme euclidienne des vecteurs $(G^k(i,j))_{k \in K}$, avec $i \in I$ et $j \in J$. Donc

$$\begin{aligned}\gamma_T^p(\sigma,\tau) &= \sum_{k \in K} p^k \biggl(\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{\sigma,\tau}(G^k(\widetilde{i}_t, \widetilde{j}_t))\biggr) \\ &\leqslant \langle p, \ell\rangle + \frac{2M}{\sqrt{T}} = \operatorname{cav} u(p) + \frac{2M}{\sqrt{T}}.\end{aligned}$$

La stratégie « d'approchabilité » τ est donc optimale pour le joueur 2 dans $\Gamma_\infty(p)$.

3. Manque d'information des deux côtés

On ne suppose plus ici qu'un des deux joueurs a toute l'information. On se donne un ensemble produit $K \times L$ et une famille $(G^{k,l})_{(k,l) \in K \times L}$

de matrices de même taille $I \times J$, ainsi que des probabilités initiales p sur K et q sur L. K, L, I et J sont supposés finis non vides. Le jeu $\Gamma_\infty(p,q)$ est le suivant :

– un état de la nature (k,l) est tiré selon la probabilité produit $p \otimes q$, puis k est annoncé uniquement au joueur 1 et l est annoncé uniquement au joueur 2.

– les joueurs répètent ensuite le jeu matriciel $G^{k,l}$, et observent après chaque étape les actions jouées dans $I \times J$.

k (resp. l) représente donc l'information initiale du joueur 1 (resp. du joueur 2). On s'est placé pour simplifier dans le cas d'informations initiales indépendantes, voir la remarque 3.6 en fin de section pour plus de généralité. Le modèle de manque d'information des deux côtés fut également introduit par Aumann et Maschler dans les années soixante (réédition en 1995). On peut consulter avec profit Zamir, 1992, et Sorin, 2002.

Une stratégie du joueur 1 est un élément $\sigma = (\sigma_t)_{t\geqslant 1}$, avec $\sigma_t : K \times (I \times J)^{t-1} \to \Delta(I)$ pour tout t. Une stratégie du joueur 2 est un élément $\tau = (\tau_t)_{t\geqslant 1}$, avec $\tau_t : L \times (I \times J)^{t-1} \to \Delta(J)$ pour tout t. On note respectivement Σ et $\mathcal{T}$ les ensembles de stratégies des joueurs 1 et 2. (p,q,σ,τ) induit une probabilité sur l'ensemble des parties $K \times L \times (I \times J)^\infty$ muni de la tribu produit, et pour tout entier strictement positif T on pose :

$$\gamma_T^{p,q}(\sigma,\tau) = \mathbb{E}_{\sigma,\tau}^{p,q}\Big(\frac{1}{T}\sum_{t=1}^{T} G^{\widetilde{k},\widetilde{l}}(\widetilde{\imath}_t,\widetilde{\jmath}_t)\Big).$$

Là encore les « tilde » indiquent des variables aléatoires. $\gamma_T^{p,q}$ est la fonction de paiement du jeu finiment répété à T étapes. Il a une valeur notée $v_T(p,q)$. On note aussi :

$$u(p,q) = \max_{x\in\Delta(I)} \min_{y\in\Delta(J)} \sum_{k,l} p^k q^l G^{k,l}(x,y). \tag{3}$$

$u(p,q)$ est la valeur du jeu matriciel $\sum_{k,l} p^k q^l G^{k,l}$, appelé « jeu non révélateur à (p,q) », au sens où aucun des joueurs n'utilise son information.

Pour une application continue $f : \Delta(K)\times\Delta(L) \to \mathbb{R}$, on note $\mathrm{cav_I}\, f$ la concavification de f par rapport à la première variable : pour tout (p,q) dans $\Delta(K)\times\Delta(L)$, $\mathrm{cav_I}\, f(p,q)$ est la valeur en p de la plus petite fonction concave de $\Delta(K)$ dans $\mathbb{R}$ qui soit supérieure à $f(.,q)$. De même, on note $\mathrm{vex}_{\mathrm{II}}\, f$ la convexification de f par rapport à la seconde variable. On montre que $\mathrm{cav}_{\mathrm{I}}\, f$ et $\mathrm{vex}_{\mathrm{II}}\, f$ sont continues, et on peut former les composées $\mathrm{cav}_{\mathrm{I}}\, \mathrm{vex}_{\mathrm{II}}\, f$ et $\mathrm{vex}_{\mathrm{II}}\, \mathrm{cav}_{\mathrm{I}}\; f$. Ce sont des fonctions concaves par rapport à la première variable, convexes par rapport à la seconde, et on a toujours $\mathrm{cav}_{\mathrm{I}}\, \mathrm{vex}_{\mathrm{II}}\, f(p,q) \leqslant \mathrm{vex}_{\mathrm{II}}\, \mathrm{cav}_{\mathrm{I}}\; f(p,q)$.

Étant donnés (σ,τ) dans $\Sigma \times \mathcal{T}$, on définit pour toute histoire $h_t = (i_1, j_1, \ldots, i_t, j_t) \in (I \times J)^t$ les *a posteriori* :

$$p_t(\sigma,\tau,h_t) = (p_t^k(\sigma,\tau,h_t))_{k\in K} = \left(\mathbb{P}^{p,q}_{\sigma,\tau}(\widetilde{k} = k|h_t)\right)_{k\in K}.$$

$$q_t(\sigma,\tau,h_t) = (q_t^l(\sigma,\tau,h_t))_{l\in L} = \left(\mathbb{P}^{p,q}_{\sigma,\tau}(\widetilde{l} = l|h_t)\right)_{l\in L}.$$

On procède comme dans le modèle standard à manque d'information d'un seul côté (voir la partie 1.5) : $p_t(\sigma,\tau,h_t) = p_t(\sigma,h_t)$ et $q_t(\sigma,\tau,h_t) = q_t(\tau,h_t)$. Par indépendance, on a pour tout (k,l) :

$$\mathbb{P}^{p,q}_{\sigma,\tau}\left((\widetilde{k},\widetilde{l}) = (k,l)|h_t\right) = p_t^k(\sigma,h_t)q_t^l(\tau,h_t).$$

Par rapport à $\mathbb{P}^{p,q}_{\sigma,\tau}$, $(p_t(\sigma))_{t\geqslant 0}$ et $(q_t(\tau))_{t\geqslant 0}$ sont des martingales. Rappelons que l'on a pour tout T, comme dans la preuve du lemme 1.12.

$$\mathbb{E}^{p,q}_{\sigma,\tau}\left(\sum_{t=0}^{T-1}(p^k_{t+1}(\sigma) - p^k_t(\sigma))^2\right) = \mathbb{E}^{p,q}_{\sigma,\tau}\left((p^k_T(\sigma))^2\right) - (p^k)^2 \leqslant p^k(1-p^k).$$

Définition 3.1. — Le contenu informationnel d'une stratégie σ du joueur 1 est :

$$I(\sigma) = \sup_{\tau\in\mathcal{T}} \mathbb{E}^{p,q}_{\sigma,\tau}\left(\sum_{k\in K}\sum_{t=0}^{\infty}\left(p^k_{t+1}(\sigma) - p^k_t(\sigma)\right)^2\right).$$

Par linéarité de l'espérance par rapport à τ, le sup peut être pris sur les stratégies τ du joueur 2 qui sont à la fois pures et indépendantes de l'état $l \in L$.

3.1. Maxmin et Minmax de $\Gamma_\infty(p,q)$. — Rappelons par définition que le joueur 1 garantit le réel v si :

$$\forall\,\varepsilon > 0,\ \exists\,\sigma,\ \exists\,T,\ \forall\,t \geqslant T,\ \forall\,\tau,\quad \gamma^{p,q}_t(\sigma,\tau) \geqslant v - \varepsilon.$$

On dit que le joueur 2 défend v si :

$$\forall\,\varepsilon > 0,\ \forall\,\sigma,\ \exists\,T,\ \exists\,\tau,\ \forall\,t \geqslant T,\quad \gamma^{p,q}_t(\sigma,\tau) \leqslant v + \varepsilon.$$

Si à la fois le joueur 1 garantit v et le joueur 2 défend v, on dit que v est le maxmin du jeu répété $\Gamma_\infty(p,q)$. En particulier c'est le plus grand réel qui peut être garanti par le joueur 1. On définit de même le minmax du jeu répété : s'il existe, c'est l'unique réel qui peut être à la fois garanti par le joueur 2 et défendu par le joueur 1. Le jeu $\Gamma_\infty(p,q)$ a une valeur si et seulement si il existe un réel qui peut être garanti par les deux joueurs. C'est équivalent à l'existence et à l'égalité du maxmin et du minmax. On a ici :

***Théorème 3.2* (Aumann Maschler Stearns).** — *Dans le jeu répété $\Gamma_\infty(p,q)$, le maxmin existe et vaut* $\mathrm{cav}_{\mathrm{I}}\,\mathrm{vex}_{\mathrm{II}}\,u(p,q)$*, le minmax existe et vaut* $\mathrm{vex}_{\mathrm{II}}\,\mathrm{cav}_{\mathrm{I}}\,u(p,q)$.

Démonstration. — On se place dans le jeu répété $\Gamma_\infty(p,q)$.

(1) Tout d'abord, le joueur 2 peut jouer sans tenir compte de son information. On se trouve alors dans un jeu à manque d'information d'un seul côté comme dans la section 1. Le joueur 2 garantit donc $\mathrm{cav}\, v_q(p)$, où pour tout p' dans $\Delta(K)$,

$$v_q(p') = \max_{x\in\Delta(I)} \min_{y\in\Delta(J)} \sum_{k\in K} p'^k \Big(\sum_{l\in L} q^l G^{k,l}\Big)(x,y) = u(p',q).$$

Donc le joueur 2 garantit $\mathrm{cav}_{\mathrm{I}}\, u(p,q)$.

Le joueur 2 peut utiliser son information, et en utilisant le lemme 1.6 avec le joueur 2 comme joueur informé (qui minimise les paiements, donc en changeant cav en vex car $-\mathrm{cav}(-f) = \mathrm{vex}(f)$), on obtient que le joueur 2 garantit ici $\mathrm{vex}_{\mathrm{II}}\,\mathrm{cav}_{\mathrm{I}}\,u(p,q)$. Symétriquement, le joueur 1 garantit $\mathrm{cav}_{\mathrm{I}}\,\mathrm{vex}_{\mathrm{II}}\,u(p,q)$.

(2) Pour conclure, il suffit de montrer que le joueur 2 défend $\mathrm{cav}_{\mathrm{I}}\,\mathrm{vex}_{\mathrm{II}}\,u(p,q)$. Fixons une stratégie σ du joueur 1, et soit $\varepsilon \in (0,1)$.

Soit τ^* une stratégie du joueur 2 qui soit indépendante de l, et soit une date N tel que

$$\mathbb{E}^{p,q}_{\sigma,\tau^*}\bigg(\sum_{k\in K}\sum_{t=0}^{N-1}\big(p^k_{t+1}(\sigma) - p^k_t(\sigma)\big)^2\bigg) \geqslant I(\sigma) - \varepsilon.$$

On définit la stratégie τ du joueur 2 de la façon suivante. Pendant les N premières étapes, τ joue selon τ^*, de façon à extraire quasiment toute l'information contenue dans σ. Supposons que $h_N \in (I\times J)^N$ soit jouée lors des N premiers coups. Le joueur 2 calcule alors l'a posteriori $p_N(\sigma, h_N) \in \Delta(K)$, et joue une stratégie optimale dans le *jeu répété à manque d'information d'un seul côté* où :

– Le joueur 2 est informé de $l \in L$, initialement tiré selon q.

– Le joueur 1 est non informé, et si $l \in L$ est l'état sélectionné, le jeu matriciel joué à chaque étape est $\sum_{k\in K} p_N^k(\sigma, h_N) G^{k,l}$.

Par le théorème 1.15, ce jeu a une valeur qui vaut $\text{vex}_{\text{II}}\, u(p_N(\sigma, h_N), q)$, et les joueurs y ont des stratégies optimales. À partir de la date $N+1$, la stratégie τ joue une stratégie optimale dans ce jeu à manque d'information d'un seul côté.

Toutes les probabilités et espérances qui suivent sont prises par rapport à $\mathbb{P}^{p,q}_{\sigma,\tau}$. On a pour toute date $T \geqslant N$:

$$\begin{aligned}
\mathbb{E}\Big(\sum_k \big(p_T^k(\sigma) - p_N^k(\sigma)\big)^2\Big) &= \mathbb{E}\Big(\sum_k \big(p_T^k(\sigma)\big)^2 - \big(p_N^k(\sigma)\big)^2\Big) \\
&= \mathbb{E}\Big(\sum_k \sum_{t=N}^{T-1} \big(p_{t+1}^k(\sigma)\big)^2 - \big(p_t^k(\sigma)\big)^2\Big) \\
&= \mathbb{E}\Big(\sum_k \sum_{t=N}^{T-1} \big(p_{t+1}^k(\sigma) - p_t^k(\sigma)\big)^2\Big) \\
&\leqslant \varepsilon \qquad\qquad \text{par définition de } \tau^*.
\end{aligned}$$

L'inégalité de Cauchy-Schwartz,

$$\mathbb{E}\left(\|p_T(\sigma) - p_N(\sigma)\|_1\right) \leqslant \sqrt{|K|}\,\sqrt{\sum_k \big(\mathbb{E}(p_T^k(\sigma) - p_N^k(\sigma))\big)^2}$$

donne alors :

$$\mathbb{E}\left(\|p_T(\sigma) - p_N(\sigma)\|_1\right) \leqslant \sqrt{|K|}\sqrt{\varepsilon} \tag{4}$$

et l'erreur commise en supposant que l'information sur $\widetilde{k}$ n'évolue plus après l'étape N est faible.

Calculons maintenant les paiements. Soit $t \geqslant N$, et $h_t \in (I \times J)^t$.

$$\begin{aligned}
\mathbb{E}\big(G^{\widetilde{k},\widetilde{l}}(\widetilde{i}_{t+1},\widetilde{j}_{t+1}) \mid h_t\big) \\
&= \sum_{k,l} p_t^k(\sigma, h_t) q_t^l(\tau, h_t) G^{k,l}\left(\sigma_{t+1}(k, h_t), \tau_{t+1}(l, h_t)\right) \\
&\leqslant \sum_{k,l} p_t^k(\sigma, h_t) q_t^l(\tau, h_t) G^{k,l}\left(\overline{\sigma}(h_t), \tau_{t+1}(l, h_t)\right) \\
&\quad + \sum_k \sum_l p_t^k(\sigma, h_t) q_t^l(\tau, h_t) M \|\sigma_{t+1}(k, h_t) - \overline{\sigma}(h_t)\| \\
&= \sum_{k,l} p_t^k(\sigma, h_t) q_t^l(\tau, h_t) G^{k,l}\left(\overline{\sigma}(h_t), \tau_{t+1}(l, h_t)\right) \\
&\qquad\qquad + M\mathbb{E}\left(\|p_{t+1}(\sigma) - p_t(\sigma)\| | h_t\right)
\end{aligned}$$

où comme dans le cas de manque d'information d'un seul côté, on pose $\overline{\sigma}(h_t) = \sum_{k \in K} p_t^k(\sigma, h_t)\sigma_{t+1}(k, h_t)$, la constante M majore tous les paiements en valeur absolue et on peut appliquer le lemme 1.13. On introduit maintenant $p_N = p_N(\sigma, h_N)$, où $h_N \in (I \times J)^N$ correspond aux premières étapes de h_t. Notons

$$\xi_{t+1}(\sigma, \tau)(h_t) = \sum_k \sum_l p_N^k(\sigma, h_N) q_t^l(\tau, h_t) G^{k,l}\left(\overline{\sigma}(h_t), \tau_{t+1}(l, h_t)\right).$$

Alors :

$$\begin{aligned}
\mathbb{E}\left(G^{\widetilde{k},\widetilde{l}}(\widetilde{i}_{t+1}, \widetilde{j}_{t+1}) | h_t\right) &\leqslant \xi_{t+1}(\sigma, \tau)(h_t) + M\|p_N(\sigma, h_N) - p_t(\sigma, h_t)\| \\
&\quad + M\mathbb{E}\left(\|p_{t+1}(\sigma) - p_t(\sigma)\| | h_t\right)
\end{aligned}$$

Or τ joue indépendamment de l pendant les N premières étapes. Donc $q_N(\tau, h_N) = q$, et par le théorème 1.15 on peut trouver T_0 tel que pour tout $T \geqslant T_0$:

$$\mathbb{E}\left(\frac{1}{T} \sum_{t=N}^{N+T-1} \xi_{t+1}(\sigma, \tau)(h_t) | h_N\right) \leqslant \mathrm{vex}_{\mathrm{II}}\, u(p_N(\sigma, h_N), q) + \varepsilon.$$

(on pourrait même avoir

$$\mathbb{E}\left(\frac{1}{T} \sum_{t=N}^{N+T-1} \xi_{t+1}(\sigma, \tau)(h_t) | h_N\right) \leqslant \mathrm{vex}_{\mathrm{II}}\, u(p_N(\sigma, h_N), q)$$

pour tout T en reprenant les preuves de la sous-section 1.3). Nous obtenons ici :

$$\begin{aligned}
&\mathbb{E}\bigg(\frac{1}{T}\sum_{t=N}^{N+T-1} G^{\widetilde{k},\widetilde{l}}(\widetilde{i}_{t+1},\widetilde{j}_{t+1})|h_N\bigg)\\
&\qquad\leqslant \mathrm{vex}_{\mathrm{II}}\, u(p_N(\sigma,h_N),q)+\varepsilon+\frac{M}{T}\sum_{t=N}^{N+T-1}\mathbb{E}\left(\|p_N(\sigma,h_N)-p_t(\sigma)\||h_N\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad+\frac{M}{T}\sum_{t=N}^{N+T-1}\mathbb{E}\left(\|p_{t+1}(\sigma)-p_t(\sigma)\||h_N\right),\\
&\qquad\leqslant \mathrm{vex}_{\mathrm{II}}\, u(p_N(\sigma,h_N),q)+\varepsilon+\frac{M}{T}\sum_{t=N}^{N+T-1}\mathbb{E}\left(\|p_N(\sigma,h_N)-p_t(\sigma)\||h_N\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad+\frac{M}{\sqrt{T}}\sum_{k}\sqrt{p_N^k(1-p_N^k)},
\end{aligned}$$

par le lemme 1.12. Comme

$$\mathrm{vex}_{\mathrm{II}}\, u(p_N(\sigma,h_N),q)\leqslant \mathrm{cav}_{\mathrm{I}}\,\mathrm{vex}_{\mathrm{II}}\, u(p_N(\sigma,h_N),q),$$

l'inégalité de Jensen ainsi que l'inégalité précédente (4) donnent finalement :

$$\begin{aligned}
&\mathbb{E}\bigg(\frac{1}{T}\sum_{t=N}^{N+T-1} G^{\widetilde{k},\widetilde{l}}(\widetilde{i}_{t+1},\widetilde{j}_{t+1})\bigg)\\
&\qquad\qquad\qquad\leqslant \mathrm{cav}_{\mathrm{I}}\,\mathrm{vex}_{\mathrm{II}}\, u(p,q)+\varepsilon+M\sqrt{|K|}\sqrt{\varepsilon}+\frac{M|K|}{\sqrt{T}},
\end{aligned}$$

et donc pour T assez grand,

$$\mathbb{E}\bigg(\frac{1}{T}\sum_{t=1}^{T} G^{\widetilde{k},\widetilde{l}}(\widetilde{i}_t,\widetilde{j}_t)\bigg)\leqslant \mathrm{cav}_{\mathrm{I}}\,\mathrm{vex}_{\mathrm{II}}\, u(p,q)+(2+M)\sqrt{|K|}\sqrt{\varepsilon}. \qquad \square$$

On a toujours $\mathrm{cav}_{\mathrm{I}}\,\mathrm{vex}_{\mathrm{II}}\, u(p,q)\leqslant \mathrm{vex}_{\mathrm{II}}\,\mathrm{cav}_{\mathrm{I}}\, u(p,q)$, et un corollaire immédiat du théorème 3.2 est que le jeu $\Gamma_\infty(p,q)$ a une valeur si et seulement si $\mathrm{cav}_{\mathrm{I}}\,\mathrm{vex}_{\mathrm{II}}\, u(p,q)\,\mathrm{vex}_{\mathrm{II}}\,\mathrm{cav}_{\mathrm{I}}\, u(p,q)$. Ce n'est pas toujours le cas, et il y a des contre-exemples à l'existence de la valeur (les premiers étant dûs à Aumann et Maschler).

Exemple 3.3. — Prenons $K = \{a, a'\}$, et $L = \{b, b'\}$, avec p et q uniformes.

$$G^{a,b} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -1 & 1 & 1 & -1 \end{pmatrix} \qquad G^{a,b'} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$G^{a',b} = \begin{pmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad G^{a',b'} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

Mertens et Zamir (1971) ont montré qu'ici

$$\mathrm{cav_I}\,\mathrm{vex_{II}}\,u(p,q) = -\frac{1}{4} < 0 = \mathrm{vex_{II}}\,\mathrm{cav_I}\,u(p,q).$$

3.2. Les jeux compacts. — v_T et v_λ sont des applications M-Lipschitz, concaves par rapport à la première variable et convexes par rapport à la seconde. Rappelons que u est la valeur du jeu non révélateur. Le résultat suivant, que nous nous contentons d'énoncer, est dû à Mertens et Zamir (1971). On se place dans l'ensemble $\mathcal{C}$ des applications continues de $\Delta(K) \times \Delta(L)$ dans $\mathbb{R}$.

Théorème 3.4. — *$(v_T)_T$ et $(v_\lambda)_\lambda$ convergent uniformément vers l'unique solution f du système suivant :*

$$\begin{cases} f = \mathrm{vex_{II}} \max\{u, f\} \\ f = \mathrm{cav_I} \min\{u, f\} \end{cases}$$

L'étude du système ci-dessus peut se faire sans référence aux jeux répétés (voir Mertens et Zamir 1977, Sorin 1984, Laraki 2001a, Laraki 2001b).

Remarque 3.5. — Structure des fonctions non révélatrices u. Au sein de $\mathcal{C}$, on considère le sous-ensemble U constitué des valeurs d'une famille de matrices, au sens de l'équation (3). Il est possible de montrer que U est un sous-espace vectoriel de $\mathcal{C}$, contient les fonctions affines, est stable par produit, par passage au sup et à l'inf, et est dense dans $\mathcal{C}$ pour la topologie de la convergence uniforme.

Remarque 3.6. — Le modèle de manque d'information des deux côtés se généralise au cas où les informations initiales des joueurs ne sont plus indépendantes. Notons le nouvel ensemble d'états R (au lieu de $K \times L$ précédemment). L'état r dans R est tiré selon une probabilité connue $p = (p^r)_{r \in R}$, puis chaque joueur observe un signal déterministe dépendant de r. Cela revient à considérer, pour chaque joueur i, une partition R^i de R et à supposer que le joueur i observe de façon privée l'élément de sa partition qui contient l'état sélectionné.

À la première étape, le joueur 1 va jouer une action $x = (x^r)_{r\in R}$ mesurable par rapport à R^1, *i.e.* telle que $(r \to x^r)$ soit constante sur chaque élément de R^1. Après l'observation du premier coup du joueur 1 dans I, on peut calculer la probabilité conditionnelle sur R. Celle-ci appartiendra à l'ensemble suivant :

$$\Pi^{\mathrm{I}}(p) = \big\{(\alpha^r p^r)_{r\in R} \mid \forall\, r\ \alpha^r \geqslant 0,\\ \textstyle\sum_r \alpha^r p^r = 1 \text{ et } (\alpha^r)_r \text{ est } R^1\text{-mesurable}\big\}.$$

$\Pi^{\mathrm{I}}(p)$ contient p, est convexe compact dans $\Delta(R)$, et on dit qu'une application f de $\Delta(R)$ dans $\mathbb{R}$ est I-concave si pour tout p dans $\Delta(R)$ sa restriction à $\Pi^{\mathrm{I}}(p)$ est concave. Pour $g : \Delta(R) \to \mathbb{R}$ majorée, on définit $\mathrm{cav}_{\mathrm{I}}\, g$ comme la plus petite fonction I-concave supérieure à g. On définit de façon analogue l'ensemble $\Pi^{\mathrm{II}}(p)$, la notion de II- convexité et la II-convexifiée $\mathrm{vex}_{\mathrm{II}}\, g$. Ces définitions généralisent celles du cas d'informations initiales indépendantes, et les résultats des théorèmes 3.2 et 3.4 s'étendent parfaitement (voir Mertens et Zamir, 1971).

4. Somme non nulle et manque d'information d'un seul côté

Ici, deux joueurs vont répéter indéfiniment un même jeu *bimatriciel* tiré au départ selon une probabilité connue, seul le joueur 1 prenant connaissance de la réalisation du tirage. Formellement, on se donne deux familles $(A^k)_{k\in K}$ et $(B^k)_{k\in K}$ de matrices de même taille $I \times J$, et une probabilité p sur K. K, I, J, sont des ensembles finis non vides. On suppose que chaque joueur a au moins deux actions, $|I| \geqslant 2$ et $|J| \geqslant 2$, et on suppose $p^k > 0$ pour tout k de K. Le jeu $\Gamma_\infty(p)$ se déroule ainsi :

– initialement, un état de la nature k est tiré, une fois pour toutes, selon p. Le joueur 1 apprend k, pas le joueur 2.

– à chaque étape $t = 1, 2, \ldots$, simultanément le joueur 1 choisit une action $i_t \in I$ et le joueur 2 choisit une action j_t dans J. Le paiement d'étape du joueur 1 est alors $A^k(i_t, j_t)$, celui du joueur 2 est $B^k(i_t, j_t)$, et tout ce que les joueurs apprennent avant de passer à l'étape $t+1$ est le couple (i_t, j_t).

Lorsque $B^k = -A^k$ pour tout k, le jeu est à somme nulle et on est dans le cadre de la section 1. Les ensembles de stratégies des joueurs sont définis comme dans cette section, et une paire de stratégies (σ, τ) dans $\Sigma \times \mathcal{T}$ induit une probabilité sur l'ensemble des parties $K \times (I \times J)^\infty$ muni de la tribu produit. Les paiements moyens espérés sont notés :

$$\alpha_T^p(\sigma,\tau) = \mathbb{E}_{p,\sigma,\tau}\left(\frac{1}{T}\sum_{t=1}^{T} A^{\widetilde{k}}(\widetilde{i}_t,\widetilde{j}_t)\right) = \sum_{k\in K} p^k \alpha_T^k(\sigma,\tau),$$

$$\beta_T^p(\sigma,\tau) = \mathbb{E}_{p,\sigma,\tau}\left(\frac{1}{T}\sum_{t=1}^{T} B^{\widetilde{k}}(\widetilde{i}_t,\widetilde{j}_t)\right) = \sum_{k\in K} p^k \beta_T^k(\sigma,\tau).$$

Il est pratique d'utiliser ici la définition suivante.

Définition 4.1. — (σ^*, τ^*) est un équilibre du jeu répété à information incomplète $\Gamma_\infty(p)$ si :

(i) pour tout $k \in K$, $(\alpha_T^k(\sigma^*, \tau^*))_T$ et $(\beta_T^p(\sigma^*, \tau^*))_T$ convergent vers des limites respectivement notées $\alpha^k(\sigma^*, \tau^*)$ et $\beta^p(\sigma^*, \tau^*)$,

(ii)

$$\forall\,\varepsilon > 0,\ \exists\, T_0,\ \forall\, T \geqslant T_0,\ \forall\, k \in K,\ \forall\, \sigma \in \Sigma, \quad \alpha_T^k(\sigma, \tau^*) \leqslant \alpha_T^k(\sigma^*, \tau^*) + \varepsilon$$

et

$$\forall\,\varepsilon > 0,\ \exists\, T_0,\ \forall\, T \geqslant T_0,\ \forall\, \tau \in \mathcal{T}, \quad \beta_T^p(\sigma^*, \tau) \leqslant \beta_T^p(\sigma^*, \tau^*) + \varepsilon.$$

$\big((\alpha^k(\sigma^*, \tau^*))_{k\in K}, (\beta^p(\sigma^*, \tau^*))\big) \in \mathbb{R}^K \times \mathbb{R}$ est alors appelé paiement d'équilibre de $\Gamma_\infty(p)$.

Comme $p \in \mathrm{int}(\Delta(K))$, la première ligne de (ii) équivaut à :

$$\forall\,\varepsilon > 0,\ \exists\, T_0,\ \forall\, T \geqslant T_0,\ \forall\, \sigma \in \Sigma, \quad \alpha_T^p(\sigma, \tau^*) \leqslant \alpha_T^p(\sigma^*, \tau^*) + \varepsilon.$$

Remarquons que cette définition d'équilibre est tout de même légèrement plus forte que la définition usuelle d'équilibre uniforme (voir le texte de T. Tomala sur les jeux répétés dans ce volume) : il est ici pratique d'imposer la convergence de $(\alpha_T^k(\sigma^*, \tau^*))_T$ pour chaque valeur de k, et non pas seulement la convergence de $(\alpha_T^p(\sigma^*, \tau^*))_T$. En quelque sorte, le joueur 1 est vu comme $|K|$ différents types possibles ayant chacun une fonction de paiement spécifique, et on veut que le paiement de chaque type converge. Si l'on est dans le cas de la somme nulle ($A^k = -B^k$ pour tout k), l'existence d'un tel équilibre implique l'existence de la valeur et de stratégies optimales pour chaque joueur.

La question de l'existence d'équilibre a été posée par Aumann, Maschler et Stearns dans les années soixante. Sorin (1983) a prouvé l'existence pour deux états de la nature, et le cas général a été résolu en 1995 par Simon, Spież et Toruńczyk (voir les théorèmes suivants 4.6 et 4.7). Hart (1985) a donné une caractérisation des paiements d'équilibres, qui n'a toutefois pas entraîné de preuve d'existence. On peut consulter Forges (1992) pour un survey.

On note, pour toute probabilité q dans $\Delta(K)$,

$$A(q) = \sum_k q^k A^k, \qquad u(q) = \max_{x\in\Delta(I)} \min_{y\in\Delta(J)} A(q)(x,y),$$

$$B(q) = \sum_k q^k B^k, \qquad v(q) = \max_{y\in\Delta(J)} \min_{x\in\Delta(I)} B(q)(x,y).$$

Si $\gamma = (\gamma(i,j))_{(i,j)\in I\times J} \in \Delta(I\times J)$,

$$A(q)(\gamma) = \sum_{(i,j)\in I\times J} \gamma(i,j) A(q)(i,j)$$

et de même on pose

$$B(q)(\gamma) = \sum_{(i,j)\in I\times J} \gamma(i,j) B(q)(i,j).$$

4.1. Existence d'un équilibre. — Exactement comme dans le cas de la somme nulle, une paire de stratégies (σ,τ) induit une suite d'*a posteriori* $(p_t(\sigma))_{t\geqslant 0}$ qui est une $\mathbb{P}_{p,\sigma,\tau}$- martingale à valeurs dans $\Delta(K)$. Du point de vue de l'existence d'un équilibre, on va se restreindre aux cas où cette martingale « bouge » au plus une fois.

Définition 4.2. — Un plan joint dans $\Gamma_\infty(p)$ est un triplet (S,λ,γ), où :

– S est un ensemble (« de messages ») fini non vide,

– $\lambda = (\lambda^k)_{k\in K}$ (« stratégie de signalling ») avec pour tout k, $\lambda^k \in \Delta(S)$ et pour tout s, $\lambda_s = \sum_{k\in K} p^k \lambda^k_s > 0$,

– $\gamma = (\gamma_s)_{s\in S}$ (« contrat ») avec pour tout s, $\gamma_s \in \Delta(I\times J)$.

L'idée, due à Aumann, Maschler et Stearns, est la suivante. Le joueur 1 observe k, puis choisit $s \in S$ selon λ^k et annonce s au joueur 2. Ensuite les joueurs jouent des actions pures correspondant aux fréquences $\gamma_s(i,j)$, pour i dans I et j dans J. Étant donné un plan joint (S,λ,γ), on définit :

– $\forall\, s \in S$, $p_s = (p_s^k)_{k\in K} \in \Delta(K)$, avec $p_s^k = p^k \lambda_s^k/\lambda_s$ pour tout k. p_s est l'*a posteriori* sur K sachant s.

– $\varphi = (\varphi^k)_{k\in K} \in \mathbb{R}^K$, avec pour tout k, $\varphi^k = \max_{s\in S} A^k(\gamma_s)$.

– $\forall\, s \in S$, $\psi_s = B(p_s)(\gamma_s)$ et

$$\psi = \sum_{k\in K} p^k \sum_{s\in S} \lambda_s^k B^k(\gamma_s) = \sum_{s\in S} \lambda_s \psi_s.$$

Définition 4.3. — Un plan joint (S, λ, γ) est un *plan joint si :

(i) $\forall\, s \in S$, $\psi_s \geqslant \operatorname{vex} v(p_s)$.

(ii) $\forall\, k \in K$, $\forall\, s \in S$ tel que $p_s^k > 0$, $A^k(\gamma_s) = \varphi^k$ (« incitation du joueur 1 à choisir s selon λ^k »).

(iii) $\forall\, q \in \Delta(K)$, $\langle \varphi, q\rangle \geqslant u(q)$.

Étant donné un *plan joint, on définit un couple de stratégies (σ^*, τ^*) adapté au plan joint. Pour tout message s de S, fixons tout d'abord une suite $(i_t^s, j_t^s)_{t\geqslant 1}$ d'éléments de $I \times J$ telle que pour tout couple (i, j), la suite des fréquences empiriques converge vers la probabilité correspondante :

$$\frac{1}{T}\bigl|\{t \mid 1 \leqslant t \leqslant T,\ (i_t^s, j_t^s) = (i_t, j_t)\}\bigr| \underset{T\to\infty}{\longrightarrow} \gamma_s(i, j).$$

On se donne également un entier ℓ et une application injective $f : S \to I^\ell$ correspondant à un code entre les joueurs pour annoncer un élément de S.

On définit σ^* de la façon suivante. Le joueur 1 observe l'état k sélectionné, puis choisit s selon la probabilité λ^k, et annonce s au joueur 2 en jouant $f(s)$ pendant les ℓ premiers coups. Enfin, σ^* joue i_t^s à chaque date $t > \ell$ tant que le joueur 2 joue j_t^s. Si à une date $t > \ell$ le joueur 2 ne joue pas j_t^s alors le joueur 1 se met à jouer une stratégie de punition du joueur 2 dans le jeu de probabilité initiale p_s, *i.e.* le joueur 1 joue une stratégie optimale dans le jeu à somme nulle de probabilité initiale p_s où les paiements du joueur 1 sont $(-B^k)_{k\in K}$.

On définit maintenant τ^*. Le joueur 2 joue arbitrairement au début puis à la fin de l'étape ℓ il déduit le message s des coups du joueur 1. Il joue ensuite à chaque date $t > \ell$ l'action j_t^s tant que le joueur 1 joue i_t^s. Si à une date $t > \ell$, le joueur 1 ne joue pas i_t^s, ou si les ℓ premiers coups du joueur 1 ne correspondent à aucun message, alors le joueur 2 se met à jouer une stratégie de punition $\overline{\tau}$ telle que :

$$\forall\,\varepsilon>0,\ \exists\,T_0,\ \forall\,T\geqslant T_0,\ \forall\,\sigma\in\Sigma,\ \forall\,k\in K,\quad \alpha_T^k(\sigma,\overline{\tau})\leqslant\varphi^k+\varepsilon.$$

Une telle stratégie $\overline{\tau}$ existe en raison de (iii) : c'est une stratégie d'approchabilité par le joueur 2 de l'orthant

$$\{x\in\mathbb{R}^K \mid \forall\,k\in K,\ x^k\leqslant\varphi^k\}$$

(voir section 2, 2.4).

***Lemme 4.4* (Sorin, 1983).** — *Un couple de stratégies (σ^*,τ^*) adapté à un *plan joint est un équilibre de $\Gamma_\infty(p)$.*

Démonstration. — Pour tout k,

$$\alpha^k(\sigma^*,\tau^*)=\sum_{s\in S}\lambda_s^k A^k(\gamma_s)=\sum_{s\in S}\lambda_s^k\varphi^k\varphi^k,$$

d'après (ii), et

$$\beta(\sigma^*,\tau^*)=\sum_{k\in K}p^k\sum_{s\in S}\lambda_s^k B^k(\gamma_s)=\psi.$$

Supposons que le joueur 2 joue τ^*. L'existence de $\overline{\tau}$ fait qu'aucune déviation détectable du joueur 1 n'est profitable, et donc que si l'état est k, le joueur 1 ne pourra gagner plus que $\max_{s'\in S}A^k(\gamma_{s'})$. Or ceci vaut $\varphi^k=\alpha^k(\sigma^*,\tau^*)$. La preuve peut être uniforme en σ et on obtient :

$$\forall\,\varepsilon>0,\ \exists\,T_0,\ \forall\,T\geqslant T_0,\ \forall\,k\in K,\ \forall\,\sigma\in\Sigma,\quad \alpha_T^k(\sigma,\tau^*)\leqslant\alpha^k(\sigma^*,\tau^*)+\varepsilon.$$

Supposons enfin que le joueur 1 joue σ^*. La condition (i) implique que le joueur 2 qui joue τ^* gagne, si le message vaut s, au moins $\operatorname{vex} v(p_s)$. Comme $\operatorname{vex} v(p_s)$ $(=-\operatorname{cav}(-v(p_s)))$ est la valeur, pour le joueur 2 de paiements $(B^k)_k$, du jeu à somme nulle de probabilité initiale p_s, le joueur 2 craint la punition du joueur 1, et

$$\forall\,\varepsilon>0,\ \exists\,T_0,\ \forall\,T\geqslant T_0,\ \forall\,\tau\in\mathcal{T},\quad \beta_T^p(\sigma^*,\tau)\leqslant\sum_{s\in S}\lambda_s\psi_s+\varepsilon=\psi+\varepsilon.$$

$\square$

Afin de prouver l'existence d'équilibres dans $\Gamma_\infty(p)$, on se retrouve à chercher des *plans joints. L'idée est tout d'abord de considérer, pour chaque probabilité r sur K, l'ensemble des vecteurs φ possibles s'il y a un plan joint dont r fait partie des a posteriori. Ceci amène à considérer la correspondance[1] suivante.

$$\begin{aligned}\Phi : \Delta(K) &\rightrightarrows \mathbb{R}^K\\ r &\longmapsto \{(A^k(\gamma))_{k\in K} \mid \gamma \in \Delta(I\times J), B(r)(\gamma) \geqslant \operatorname{vex} v(r)\}\end{aligned}$$

Il est facile de voir que Φ a un graphe compact et des valeurs convexes non vides, et vérifie :

$$\forall\, r \in \Delta(K), \forall\, q \in \Delta(K), \exists\, \varphi \in \Phi(r), \quad \langle \varphi, q\rangle \geqslant u(q).$$

Supposons maintenant que l'on trouve un ensemble fini $(p_s)_{s\in S}$ d'éléments de $\Delta(K)$, ainsi que des vecteurs de $\mathbb{R}^K$ φ et φ_s pour tout s tels que :

– $p \in \operatorname{conv}\{p_s \mid s \in S\}$,
– $\forall\, q \in \Delta(K)$, $\langle \varphi, q\rangle \geqslant u(q)$,
– $\forall\, s \in S$, $\varphi_s \in \Phi(p_s)$,
– $\forall\, s \in S$, $\forall\, k \in K$, $\varphi_s^k \leqslant \varphi^k$ avec égalité si $p_s^k > 0$.

Alors il est facile de construire un *plan joint. On se retrouve donc à essayer de démontrer le résultat suivant :

Proposition 4.5. — *Soient $p \in \operatorname{int}(\Delta(K))$, $u : \Delta(K) \to \mathbb{R}$ une application continue, et $\Phi : \Delta(K) \rightrightarrows \mathbb{R}^K$ une correspondance de graphe compact à valeurs convexes non vides tels que :*

$$\forall\, r \in \Delta(K), \forall\, q \in \Delta(K), \exists\, \varphi \in \Phi(r), \quad \langle \varphi, q\rangle \geqslant u(q).$$

Alors il existe un ensemble fini S, une famille $(p_s)_{s\in S}$ d'éléments de $\Delta(K)$, ainsi que des vecteurs de $\mathbb{R}^K$ φ et φ_s pour tout s de S tels que :

– *$p \in \operatorname{conv}\{p_s \mid s \in S\}$,*
– *$\forall\, q \in \Delta(K)$, $\langle \varphi, q\rangle \geqslant u(q)$,*
– *$\forall\, s \in S$, $\varphi_s \in \Phi(p_s)$,*
– *$\forall\, s \in S$, $\forall\, k \in K$, $\varphi_s^k \leqslant \varphi^k$ avec égalité si $p_s^k > 0$.*

[1] Rappelons qu'une correspondance F d'un ensemble X dans un ensemble Y est une application de X dans l'ensemble des parties de Y. Le graphe de la correspondance F est alors définie comme $\{(x,y) \in X \times Y \mid y \in F(x)\}$.

La preuve de ce résultat repose sur un théorème de point fixe de type Borsuk-Ulam[(2)] démontré par Simon, Spież et Toruńczyk (1995) *via* des outils de topologie algébrique (on peut voir Renault, 2000, ou Simon 2002 pour le passage à la proposition 4.5). Une version simplifiée de ce résultat de type point fixe est donnée maintenant :

***Théorème 4.6* (Simon, Spież et Toruńczyk, 1995).** — *Soient C un compact d'un espace euclidien de dimension n, $x \in \operatorname{int}(C)$, et Y une union finie de sous-espaces affines de dimension $n-1$ d'un espace euclidien. Soit F une correspondance de C dans Y de graphe compact à valeurs convexes non vides. Alors il existe $L \subset \partial C$ et $y \in Y$ tels que :*

$$\forall\, l \in L, \quad y \in F(l) \quad \textit{et} \quad x \in \operatorname{conv}(L).$$

Remarquons que pour $n = 1$ (qui correspond à deux états possibles), l'image par F de la composante connexe de C contenant x est nécessairement un singleton, donc le résultat est clair. Tous comptes faits, on aboutit donc à :

***Théorème 4.7* (Simon, Spież et Toruńczyk, 1995).** — *Il existe un *plan joint. Donc il existe un équilibre dans le jeu répété $\Gamma_\infty(p)$.*

4.2. Caractérisation des paiements d'équilibre. — On présente ici la caractérisation des paiements d'équilibre due à S. Hart (1985). Notons dans cette partie $p_0 \in \operatorname{int}(\Delta(K))$ la probabilité initiale. Soit un équilibre (σ^*, τ^*) de $\Gamma_\infty(p_0)$ de paiement $(a, \beta) \in \mathbb{R}^K \times \mathbb{R}$. On a d'après la définition 4.1 :

$$\forall\, k \in K, \forall\, \varepsilon > 0, \exists\, T_0, \forall\, T \geqslant T_0, \forall\, \sigma \in \Sigma, \quad \alpha_T^k(\sigma, \tau^*) \leqslant a^k + \varepsilon.$$

Donc l'orthant $\{x \in \mathbb{R}^K \mid \forall\, k \in K,\ x^k \leqslant a^k\}$ est approchable par le joueur 2, et on montre avec le théorème 2.4 (voir aussi la partie 2.4) :

$$\forall\, q \in \Delta(K), \quad \langle a, q \rangle \geqslant u(q) \tag{5}$$

La propriété (5) s'appelle la condition de rationalité individuelle du joueur 1, et ne dépend pas de la probabilité initiale dans $\operatorname{int}(\Delta(K))$. En ce qui concerne le joueur 2, on a :

$$\forall\, \varepsilon > 0, \exists\, T_0, \forall\, T \geqslant T_0, \forall\, \tau \in \mathcal{T}, \quad \beta_T^{p_0}(\sigma^*, \tau) \leqslant \beta + \varepsilon,$$

[(2)] Pour toute application continue de la sphère unité de $\mathbb{R}^{n+1}$ dans $\mathbb{R}^n$ il existe deux points diamétralement opposés ayant la même image.

donc par le théorème 1.15 :

$$\beta \geqslant \operatorname{vex} v(p_0). \tag{6}$$

La propriété (6) s'appelle la condition de rationalité individuelle du joueur 2 : à l'équilibre, ce joueur doit avoir au moins la valeur du jeu où les paiements du joueur 1 sont opposés aux siens.

Supposons un instant que σ^* soit une stratégie non révélatrice du joueur 1, au sens où le joueur 1 joue indépendamment de l'état k sélectionné. Supposons également que les joueurs jouent des actions dont les fréquences empiriques convergent vers les probabilités d'une distribution $\pi = (\pi_{i,j})_{(i,j)\in I\times J} \in \Delta(I \times J)$. On aura alors :

$$\forall\, k \in K, \quad a^k \sum_{i,j} \pi_{i,j} A^k(i,j) \quad \text{et} \quad \beta = \sum_k p_0^k \sum_{i,j} \pi_{i,j} B^k(i,j),$$

et si les conditions de rationalité individuelle sont vérifiées, alors aucune déviation détectable d'un joueur n'est profitable. Ceci amène à définir l'ensemble suivant, où M est une constante fixée égale à $\max\{|A^k(i,j)|, |B^k(i,j)|, (i,j) \in I \times J\}$, et où $\mathbb{R}_M = [-M, M]$.

Définition 4.8. — Soit G l'ensemble des triplets $(a, \beta, p) \in \mathbb{R}_M^K \times \mathbb{R}_M \times \Delta(K)$ tels que :

(1) $\forall\, q \in \Delta(K)$, $\langle a, q\rangle \geqslant u(q)$,

(2) $\beta \geqslant \operatorname{vex} v(p)$,

(3) il existe $\pi \in \Delta(I \times J)$ tel que $\beta = \sum_k p^k \sum_{i,j} \pi_{i,j} B^k(i,j)$ et pour tout $k \in K$, $a^k \geqslant \sum_{i,j} \pi_{i,j} A^k(i,j)$ avec égalité si $p^k > 0$.

On est amené à considérer toutes les probabilités initiales possibles $p \in \Delta(K)$ car la variable d'état importante du modèle est, là encore, la martingale des *a posteriori* du joueur 2 sur l'état de la nature. Pour $p \in \operatorname{int}(\Delta(K))$, $\{(a, \beta) \mid (a, \beta, p) \in G\}$ est l'ensemble des paiements d'équilibre non révélateurs de $\Gamma_\infty(p)$. La définition suivante est essentielle.

Définition 4.9. — On définit l'ensemble G^* comme l'ensemble des éléments $g = (a, \beta, p) \in \mathbb{R}_M^K \times \mathbb{R}_M \times \Delta(K)$ tels qu'il existe un espace probabilisé $(\Omega, \mathcal{A}, Q)$, une suite croissante $(\mathcal{F}_n)_{n\geqslant 1}$ de sous-tribus finies de $\mathcal{A}$, une suite de v.a. $(g_n)_{n\geqslant 1}(a_n, \beta_n, p_n)_{n\geqslant 1}$ définies sur $(\Omega, \mathcal{A})$ à valeurs dans $\mathbb{R}_M^K \times \mathbb{R}_M \times \Delta(K)$ satisfaisant :

(i) $g_1 = (a, \beta, p)$ p.s.,

(ii) $(g_n)_{n\geqslant 1}$ est une $(\mathcal{F}_n)_{n\geqslant 1}$ martingale,
(iii) $\forall\, n \geqslant 1$, $a_{n+1} = a_n$ p.s. ou $p_{n+1} = p_n$ p.s., et
(iv) $(g_n)_n$ converge p.s. vers une v.a. g_∞ à valeurs dans G.

Oublions dans un premier temps la composante du paiement du joueur 2. Un processus $(g_n)_n$ vérifiant (ii) et (iii) s'appelle une bi-martingale, c'est une martingale telle qu'à chaque étape, il existe une des deux composantes qui n'évolue p.s. pas. G^* peut donc se voir comme l'ensemble des points de départ des bi-martingales qui convergent dans G. L'importance ici de l'ensemble G^* vient du résultat suivant.

***Théorème 4.10* (Hart, 1985).** — *Soit* $(a,\beta) \in \mathbb{R}^K \times \mathbb{R}$.

(a,β) *est un paiement d'équilibre de* $\Gamma_\infty(p_0) \Longleftrightarrow (a,\beta,p_0) \in G^*$.

Donnons maintenant dans les deux paragraphes suivants, non pas une démonstration, mais une idée approximative de la preuve du théorème 4.10.

Commençons par l'implication $\Rightarrow$. Fixons un équilibre (σ^*,τ^*) de $\Gamma_\infty(p)$. La suite des *a posteriori* $(p_t(\sigma))_{t\geqslant 0}$ est une $\mathbb{P}_{p,\sigma,\tau}$- martingale. Modifions légèrement la structure temporelle de telle sorte qu'à chaque étape le joueur 1 joue d'abord, puis que le joueur 2 joue sans avoir pris connaissance du coup du joueur 1. À chaque demi-étape où le joueur 2 joue, l'*a posteriori* reste constant. À chaque demi-étape où le joueur 1 joue, l'espérance du paiement futur du joueur 1 (qui reste à définir proprement, à l'aide notamment d'une limite de Banach) reste constante. D'où, de façon heuristique, l'apparition de la bimartingale. Enfin, par convergence des martingales bornées, au bout d'un moment tout sera fixé et on jouera alors approximativement un équilibre non révélateur pour un *a posteriori* limite, donc on convergera vers des éléments de G.

Passons maintenant à l'implication $\Leftarrow$. Soit (a,β) tel que $(a,\beta,p_0) \in G^*$, et supposons pour simplifier que la bi-martingale associée (a_n,β_n,p_n) converge en un nombre fixé N d'étapes :

$$\forall\, n \geqslant N, \quad (a_n,\beta_n,p_n) = (a_N,\beta_N,p_N) \in G.$$

On peut construire un équilibre (σ^*,τ^*) de $\Gamma_\infty(p_0)$ de paiement (a,β) de la façon suivante. À chaque fois, (a_n,β_n) sera un paiement d'équilibre du jeu de probabilité initiale p_n. Après une certaine étape, le joueur 1 jouera de façon indépendante de l'état de la nature, l'*a posteriori* du joueur 2 sera p_N, et on jouera jusqu'à la fin des temps un équilibre non révélateur de $\Gamma_\infty(p_N)$ de paiement (a_N, β_N). Comment arriver jusque là ? Souvenons-nous que, de par la définition même des paiements dans le jeu infiniment répété, le poids d'un nombre fini d'étapes est nul. Les joueurs peuvent donc passer un grand nombre d'étapes à communiquer, sans que cela n'influe sur les paiements. Soit un indice $n < N$ tel que $a_{n+1} = a_n$. Pour passer de (a_n, β_n, p_n) à $(a_n, \beta_{n+1}, p_{n+1})$, le joueur 1 peut utiliser le lemme 1.5 de splitting afin de « signaller » une partie de l'information au joueur 2. Soit maintenant un indice $n < N$ tel que $p_{n+1} = p_n$. On souhaite passer de (a_n, β_n, p_n) à $(a_{n+1}, \beta_{n+1}, p_n)$. Le joueur 1 va jouer indépendamment de l'état, et les deux joueurs vont agir de façon à convexifier leurs paiements futurs. Ceci peut se faire au moyen de procédures appelées *loteries conjointement contrôlées*, et introduites dès les années soixante (Aumann Maschler, 1995). L'idée est la suivante. Imaginons que les deux joueurs doivent décider avec probabilités égales de jouer l'équilibre $E1$ de paiement (a^1, β^1) ou de jouer l'équilibre $E2$ de paiement (a^2, β^2). Les joueurs ne sont pas nécessairement indifférents entre les deux équilibres, il est possible par exemple que le joueur 1 préfère $E1$ alors que le joueur 2 préfère $E2$. On peut procéder ainsi. Notons i et i', respectivement j et j', deux actions différentes du joueur 1, resp. joueur 2. Simultanément et indépendamment, le joueur 1 va jouer i ou i' à probabilités égales, et le joueur 2 va jouer j ou j' à probabilités égales.

$$\begin{array}{cc} & \begin{array}{cc} j & j' \end{array} \\ \begin{array}{c} i \\ i' \end{array} & \begin{pmatrix} \times & \\ & \times \end{pmatrix} \end{array}.$$

Puis les joueurs décident conjointement de jouer $E1$ si on est sur la diagonale, *i.e.* si (i, j) ou (i', j') est joué, et de jouer $E2$ sinon. Cette procédure est robuste aux déviations unilatérales : aucun des joueurs ne peut dévier de façon à empêcher que $E1$ et $E2$ soient choisis avec probabilités égales. Plus généralement, les loteries conjointement contrôlées permettent de sélectionner une alternative au sein d'un ensemble fini selon une probabilité voulue (penser aux expansions binaires), et ceci de façon robuste aux déviations de la part d'un seul joueur. S. Hart a montré qu'en combinant des étapes de « signalling » et des loteries conjointement contrôlées, il était possible de construire un équilibre de $\Gamma_\infty(p_0)$ de paiement (a, β).

4.3. Biconvexité et bimartingales. — L'analyse précédente incite à définir et étudier certaines propriétés générales dites de biconvexité. La référence ici est l'article de Aumann et Hart (1986).

Soient X et Y des convexes compacts d'espaces euclidiens, et $(\Omega, \mathcal{F}, \mathcal{P})$ un espace probabilisé sans atome.

Définition 4.11. — Un sous-ensemble B de $X \times Y$ est biconvexe si pour tous x de X et y de Y, les sections $B_{x.} = \{y' \in Y \mid (x, y') \in B\}$ et $B_{.y} = \{x' \in X \mid (x', y) \in B\}$ sont convexes. Pour B biconvexe, une application $f : B \to \mathbb{R}$ est dite biconvexe si pour tout $(x, y) \in X \times Y$, $f(., y)$ et $f(x, .)$ sont convexes.

On a, comme pour le cas classique de convexité, que si f est biconvexe, alors pour tout réel α, l'ensemble $\{(x, y) \in B \mid f(x, y) \leqslant \alpha\}$ est un ensemble biconvexe.

Définition 4.12. — Une suite $Z_n = (X_n, Y_n)_{n\geqslant 1}$ de v.a. à valeurs dans $X \times Y$ est une bimartingale si :

(1) il existe une suite croissante $(\mathcal{F}_n)_{n\geqslant 1}$ de sous-tribus finies de $\mathcal{F}$ telle que $(Z_n)_n$ soit une $(\mathcal{F}_n)_{n\geqslant 1}$-martingale.

(2) $\forall\, n \geqslant 1$, $X_n = X_{n+1}$ p.s. ou $Y_n = Y_{n+1}$ p.s.

(3) Z_1 est p.s. constante.

$(Z_n)_{n\geqslant 1}$ étant une martingale bornée, elle converge presque sûrement vers une limite Z_∞.

Définition 4.13. — Soit A un sous-ensemble mesurable de $X \times Y$. On note A^* l'ensemble des $z \in X \times Y$ pour lesquels il existe une bimartingale $(Z_n)_{n\geqslant 1}$ avec $Z_1 = z$ p.s. et convergeant vers Z_∞ avec $Z_\infty \in A$ p.s.

On peut montrer que tout espace probabilisé $(\Omega, \mathcal{F}, \mathcal{P})$ sans atome, ou encore tout produit de convexes compacts $X \times Y$ contenant A, induisent le même ensemble A^*. On peut aussi remplacer la condition (2) par :

$$\forall\, n \geqslant 1, \quad (X_n = X_{n+1} \text{ ou } Y_n = Y_{n+1}) \text{ p.s.}$$

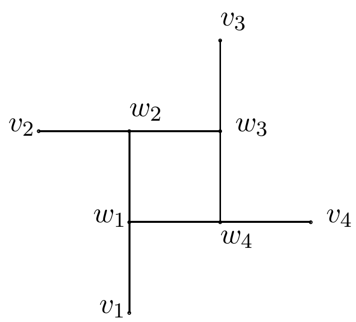
Remarquons que, sans cette condition (2) de bi-martingale, l'ensemble A^* serait seulement l'enveloppe convexe de A, et on a toujours $A \subset A^* \subset \mathrm{conv}(A)$. Ces inclusions peuvent être strictes. Par exemple, si $X = Y = [0,1]$ et $A = \{(0,0),(1,0),(0,1)\}$, on montre que

$$A^* = \{(x,y) \in [0,1] \times [0,1] \mid x = 0 \text{ ou } y = 0\}.$$

A^* est toujours biconvexe, et contient donc $\mathrm{biconv}(A)$, le plus petit ensemble biconvexe qui contient A. L'inclusion $\mathrm{biconv}(A) \subset A^*$ peut également être stricte, comme le montre l'exemple suivant :

Exemple 4.14. — On pose

$X = Y = [0,1]$,
$v_1 = (1/3, 0),\ v_2 = (0, 2/3),\ v_3 = (2/3, 1),\ v_4 = (1, 1/3)$,
$w_1 = (1/3, 1/3),\ w_2 = (1/3, 2/3),\ w_3 = (2/3, 2/3),\ w_4 = (2/3, 1/3)$,
$A = \{v_1, v_2, v_3, v_4\}$.



A est biconvexe, donc $A = \mathrm{biconv}(A)$. Soit maintenant le processus markovien $(Z_n)_{n\geqslant 1}$ suivant : $Z_1 = w_1$. Si $Z_n \in A$, alors $Z_{n+1} = Z_n$. Si $Z_n = w_i$ pour un i, alors $Z_{n+1} = w_{i+1(mod\,4)}$ avec probabilité $1/2$, et $Z_{n+1} = v_i$ avec probabilité $1/2$. $(Z_n)_n$ est une bimartingale qui converge p.s. vers un point de A, donc $w_1 \in A^* \setminus \mathrm{biconv}(A)$.

Donnons maintenant une caractérisation géométrique de l'ensemble A^*. On suppose ici A fermé. Pour chaque sous-ensemble biconvexe B de $X \times Y$ qui contient A, on note $\mathrm{nsc}(B)$ l'ensemble des points de B qui ne peuvent être séparés de A par une fonction biconvexe bornée continue sur A. Plus précisément,

$$\mathrm{nsc}(B) = \big\{z \in B \mid \forall\, f : B \to \mathbb{R} \text{ bornée biconvexe continue sur } A,$$
$$f(z) \leqslant \sup\{f(z'), z' \in A\}\big\}.$$

***Théorème 4.15* (Aumann et Hart, 1986).** — *A^* est le plus grand ensemble biconvexe B contenant A tel que* $\mathrm{nsc}(B) = B$.

Revenons au contexte des jeux et aux notations de la partie 4.2. Pour être précis, il faut tenir compte de la composante paiement du joueur 2, donc modifier très légèrement les définitions. G est fermé dans $\mathbb{R}_M^K \times \mathbb{R}_M \times \Delta(K)$. Pour $B \subset \mathbb{R}_M^K \times \mathbb{R}_M \times \Delta(K)$, B est biconvexe si pour tous a dans $\mathbb{R}_M^K$ et pour tout p dans $\Delta(K)$, les sections $\{(\beta, p'), (a, \beta, p') \in B\}$ et $\{(a', \beta), (a', \beta, p) \in B\}$ sont convexes. Une fonction réelle f définie sur un ensemble biconvexe B est dite biconvexe si pour tous a et p, $f(a, ., .)$ et $f(., ., p)$ sont convexes.

***Théorème 4.16* (Aumann et Hart, 1986).** — *G^* est le plus grand ensemble biconvexe B contenant G tel que : $\forall\, z \in B$, $\forall\, f : B \to \mathbb{R}$ biconvexe bornée continue sur A, $f(z) \leqslant \sup\{f(z'), z' \in G\}$.*

5. Extensions, divers

Concernant la modélisation, les fondements des jeux à information incomplète sont étudiés par Harsanyi (1967) et par Mertens et Zamir dans (1985). Par ailleurs, les résultats présentés précédemment ne constituent que la base des jeux répétés à information incomplète, et il existe de nombreuses extensions et variantes. La présentation suivante est sûrement imparfaite et ne prétend pas à l'exhaustivité. On n'aborde notamment pas ici les liens entre jeux répétés à information incomplète et phénomènes de réputation, « merging » de probabilités, l'apprentissage, le cheap-talk,...

5.1. Signaux. — Les modèles définis dans ce texte se généralisent au cas d'observation imparfaite. On se donne des ensembles (finis) de signaux U et V, et une application $\ell : K \times I \times J \to \Delta(U \times V)$. Après chaque étape t, si l'état est k et que (i_t, j_t) a été joué, on tire (u_t, v_t) selon $\ell(k, i_t, j_t)$. Le joueur 1, resp. joueur 2, apprend alors uniquement u_t, resp. v_t, avant de passer à l'étape $t+1$. Quand ℓ ne dépend pas de l'état k on dit que les signaux sont indépendants de l'état. Le cas où $U = V = I \times J$ et $\ell(k,i,j)$ est la mesure de Dirac sur $((i,j),(i,j))$ pour tous k, i, j, correspond aux modèles précédents dits d'observation parfaite.

Aumann et Maschler ont généralisé le théorème 1.15 au cadre général de signaux. Pour une action $x \in \Delta(I)$, une action j dans J et un état k, notons xQ_j^k la marginale sur V de la distribution $\sum_{i\in I} x_i \;\; \ell(k,i,j)$ des signaux reçus par le joueur 2. On définit l'ensemble des stratégies non révélatrices du joueur 1 comme :

$$\begin{aligned}&\mathrm{NR}(p) = \{x = (x^k)_{k\in K} \in \Delta(I)^K \mid \\ &\qquad \forall\, k \in K,\, \forall\, k' \in K \text{ tels que } p^k p^{k'} > 0,\, \forall\, j \in J,\, x^k Q_j^k = x^{k'} Q_j^{k'}\}.\end{aligned}$$

Si la probabilité initiale est p et que le joueur 1 joue selon une stratégie dans $\mathrm{NR}(p)$, l'*a posteriori* du joueur 2 restera presque sûrement égal à l'*a priori* p. La valeur du jeu NR à p devient ici :

$$\begin{aligned}u(p) &= \max_{x\in\mathrm{NR}(p)} \min_{y\in\Delta(J)} \sum_{k\in K} p^k G^k(x^k, y)\\ &= \min_{y\in\Delta(J)} \max_{x\in\mathrm{NR}(p)} \sum_{k\in K} p^k G^k(x^k, y),\end{aligned}$$

avec la convention $u(p) = -\infty$ si $\mathrm{NR}(p) = \varnothing$. Avec ces notations, on a le même énoncé que le théorème 1.15 : la valeur du jeu répété de probabilité initiale p existe et vaut $\operatorname{cav} u(p)$. Kohlberg (1975) et Mertens, Sorin et Zamir (1994) (partie B, chapitre V, 3.d.) ont généralisé dans ce cadre la construction d'une stratégie optimale explicite de type approchabilité pour le joueur 2.

En ce qui concerne les jeux à somme nulle et manque d'information des deux côtés, Mertens (1972) et Mertens et Zamir (1971 et 1977), ont généralisé l'étude de la partie 3 au cas de signaux indépendants de l'état. Dans les jeux à manque d'information d'un seul côté et somme non nulle, l'existence d'équilibre a été généralisée pour des signaux indépendants de l'état dans Renault, 2000 (voir aussi Simon, Spież et Toruńczyk 2002).

5.2. Deux joueurs somme nulle. — Indiquons tout d'abord qu'il est crucial, pour la validité du théorème 1.15, que le joueur 1 connaisse l'*a priori* p du joueur 2 sur l'état de la nature (voir Sorin et Zamir 1985, pour un jeu répété à manque d'information « d'un côté et demi » sans valeur).

Dans l'exemple 1.3, Mayberry (1967) a étudié la valeur v_λ du jeu escompté et a montré que pour $2/3 < \lambda < 1$, v_λ a une dérivée discontinue en tout point rationnel p (voir aussi Zamir, 1992 ou Sorin, 2002).

Dans le cas général, de nombreux travaux étudient les liens entre les valeurs des jeux finiment répétés, ou escomptés, et obtiennent notamment des propriétés sur la convergence de $(v_T)_T$ ou de $(v_\lambda)_\lambda$: par exemple Zamir, 1971, Zamir 1973, Mertens et Zamir 1976, de Meyer 1996a, 1996b, Laraki 2001a. On trouve de nombreuses généralisations des théorèmes principaux. Laraki (2001b) étudie des jeux dits de « splitting ». B. de Meyer a introduit la notion de « jeu dual » (voir de Meyer 1996b, de Meyer 1998, Rosenberg 1998, de Meyer et Rosenberg 1999, Laraki 2002).

Donnons juste une idée de ce jeu dual dans le cadre du modèle standard de la section 1. Soit z un paramètre dans $\mathbb{R}^K$. Dans le jeu dual $\Gamma_T^*(z)$, le joueur 1 commence par choisir secrètement l'état k. Puis à chaque étape $t \leqslant T$, les joueurs choisissent classiquement des actions i_t et j_t qui sont annoncées avant de passer à l'étape suivante. Le paiement du joueur 1 est finalement $\frac{1}{T}\sum_{t=1}^{T} G^k(i_t, j_t) - z^k$. Ce joueur peut donc maintenant choisir l'état k, mais doit le payer au prix z^k. On montre que $\Gamma_T^*(z)$ a une valeur $w_T(z)$. w_T est convexe, et liée à la valeur du jeu primal $\Gamma_T(p)$ par les formules de conjugaison :

$$\begin{aligned} w_T(z) &= \max_{p\in\Delta(K)} (v_T(p) - \langle p, z\rangle), \\ v_T(p) &= \inf_{z\in\mathbb{R}^K} (w_T(z) + \langle p, z\rangle). \end{aligned}$$

Certains travaux privilégient une approche fonctionnelle *via* l'étude d'opérateurs (Rosenberg et Sorin 2001, voir aussi Sorin, 2002). Cette approche s'applique à la fois aux jeux répétés à information incomplète et aux jeux stochastiques.

Laraki (2004) étudie l'opérateur de convexification et s'intéresse à la préservation de la continuité et du caractère Lipschitz.

Par ailleurs, il est possible de généraliser le modèle standard de la partie 1 et de prouver l'existence de la valeur dans le cas où l'état n'est plus tiré aléatoirement une fois pour toutes au départ, mais évolue selon une chaîne de Markov observée uniquement par le joueur 1 (Renault, 2003).

Enfin, de Meyer et Moussa Saley (2003) se sont intéressés à l'origine des mouvements browniens dans les modèles financiers. Ils ont introduit un modèle de jeu de marché basé sur un jeu répété à manque d'information d'un seul côté, et prouvent l'apparition endogène d'un mouvement brownien.

5.3. Somme non nulle. — Dans le cadre de la partie 4.2 des jeux à manque d'information d'un seul côté et somme non nulle, on peut étudier le nombre d'étapes de communication nécessaires à la réalisation d'équilibres, lié à la convergence des bimartingales (Aumann et Maschler 1995, Aumann et Hart 1986, Forges 1984, Forges 1990). Indiquons que F. Forges (1988) a aussi donné une caractérisation des paiements d'équilibres, pour une notion plus générale d'équilibre appelée équilibre en communication.

Par ailleurs, on peut étudier le sous cas où chaque joueur connaît ses propres paiements. Lorsqu'il y a manque d'information d'un seul côté, cela correspond à supposer que la matrice des paiements du joueur 2 est indépendante de k. On montre (Shalev, 1994) que tout paiement d'équilibre s'obtient alors comme paiement d'équilibre complètement révélateur. Ce résultat peut se généraliser au cas de manque d'information des deux côtés et somme non nulle (voir l'article non publié de Koren, 1992), et il peut ne pas exister d'équilibre même quand les deux joueurs connaissent leurs paiements.

Un autre modèle traite du cas d'information dit symétrique. Les deux joueurs ont alors une information incomplète, mais identique, sur l'état de la nature. Ils reçoivent après chaque étape le même signal, dépendant notamment de cet état. A. Neyman et S. Sorin (1998) ont montré l'existence de *paiements* d'équilibres dans le cas de deux joueurs (pour la somme nulle, voir Forges, 1982).

Très peu d'études ont concerné le cas d'au moins 3 joueurs. On trouve un résultat partiel (pour deux états de la nature) d'existence d'équilibre de type plan joint dans Renault (2001a). Enfin, pour des modèles de jeux répétés à n joueurs à information incomplète et avec signaux, on trouve des résultats d'existence d'équilibres particuliers (complètement révélateurs) chez Renault et Tomala, 2004b (voir aussi Renault, 2001b), où la transmission stratégique d'information est étudiée indépendamment des paiements.

Bibliographie

AUMANN (R.J.) & HART (S.)

[1986] Bi-convexity and bi-martingales, *Israel Journal of Mathematics*, 54 (1986), p. 159–180.

AUMANN (R.J.) & MASCHLER (M.)

[1995] *Repeated games with incomplete information*, M.I.T. Press, 1995 ; avec la collaboration de R. Stearns (contient une réédition de travaux de 1966,67,68).

BLACKWELL (D.)

[1956] An analog of the minmax theorem for vector payoffs, *Pacific Journal of Mathematics*, 65 (1956), p. 1–8.

DE MEYER (B.)

[1996a] Repeated games and partial differential equations, *Mathematics of Operations Research*, 21 (1996), p. 209–236.

[1996b] Repeated games, duality and the central limit theorem, *Mathematics of Operations Research*, 21 (1996), p. 237–251.

[1998] The maximal variation of a bounded martingale and the central limit theorem, *Annales de l'Institut Henri Poincaré, Probabilités et statistiques*, 34 (1998), p. 49–59.

DE MEYER (B.) & MOUSSA SALEY (H.)

[2003] On the strategic origin of Brownian motion in finance, *International Journal of Game Theory*, 31 (2003), p. 285–319.

DE MEYER (B.) & ROSENBERG (D.)

[1999] « *Cavu* » and the dual game, *Mathematics of Operations Research*, 24 (1999), p. 619–626.

FORGES (F.)

[1982] Infinitely repeated games of incomplete information : symmetric case with random signals, *International Journal of Game Theory*, 11 (1982), p. 203–213.

[1984] A note on Nash equilibria in repeated games with incomplete information, *International Journal of Game Theory*, 13 (1984), p. 179–187.

[1988] Communication equilibria in repeated games with incomplete information, *Mathematics of Operations Research*, 13 (1988), p. 191–231.

[1990] Equilibria with communication in a job market example, *Quarterly Journal of Economics*, 105 (1990), p. 375–398.

[1992] Repeated Games of Incomplete Information : Non-zero sum, dans Aumann (R.J.) & Hart (S.), éd., *Handbook of Game Theory, I*, Elsevier Science Publishers, 1992, p. 155–177.

HARSANYI (J.)

[1967-68] Games with incomplete information played by 'Bayesian'

players, parts I-III, *Management Science*, 8 (1967-68), p. 159–182, 320–334, 486–502.

Hart (S.)

[1985] Nonzero-sum two-person repeated games with incomplete information, *Mathematics of Operations Research*, 10 (1985), p. 117–153.

Kohlberg (E.)

[1975] Optimal strategies in repeated games with incomplete information, *International Journal of Game Theory*, 4 (1975), p. 7–24.

Koren (G.)

[avril 1992] Two-person repeated games where players know their own payoffs, avril 1992 ; document de travail basé sur une « master thesis » à l'Université de Tel-Aviv, 50 pages, `http://www.ma.huji.ac.il/hart/papers/koren.pdf`.

Laraki (R.)

[2001a] Variational inequalities, system of functional equations and incomplete information repeated games, *SIAM Journal on control and optimization*, 40 (2001), p. 516–524.

[2001b] The splitting game and applications, *International Journal of Game Theory*, 30 (2001), p. 359–376.

[2002] Repeated games with lack of information on one side : the dual differential approach, *Mathematics of Operations Research*, 27 (2002), p. 419–440.

[2004] On the regularity of the convexification operator on a compact set, *Journal of Convex Analysis*, 11 (2004), p. 209–234.

Mayberry (J.-P.)

[1967] Discounted repeated games with incomplete information, dans *Report of the U.S. Arms control and disarmament agency*, vol. ST116, chapter V, Princeton : Mathematica, 1967, p. 435–461.

Mertens (J.-F.)

[1972] The value of two-person zero-sum repeated games : the extensive case, *International Journal of Game Theory*, 1 (1972), p. 217–227.

Mertens (J.-F.), Sorin (S.) & Zamir (S.)

[1994] Repeated games, dans *CORE discussion paper*, 1994, p. 9420–9422.

Mertens (J.-F.) & Zamir (S.)

[1971] The value of two-person zero-sum repeated games with lack of information on both sides, *International Journal of Game Theory*, 1 (1971), p. 39–64.

[1976] The normal distribution and repeated games, *International Journal of Game Theory*, 5 (1976), p. 187–197.

[1977] A duality theorem on a pair of simultaneous functional equations, *Journal of Mathematical Analysis and Applications*, 60

(1977), p. 550–558.
[1985] Formulation of Bayesian analysis for games with incomplete information, *International Journal of Game Theory*, 14 (1985), p. 1–29.

Neyman (A.) & Sorin (S.)
[1998] Equilibria in Repeated Games with Incomplete Information : The General Symmetric Case, *International Journal of Game Theory*, 27 (1998), p. 201–210.

Renault (J.)
[2000] 2-player repeated games with lack of information on one side and state independent signalling, *Mathematics of Operations Research*, 4 (2000), p. 552–572.
[2001a] 3-player repeated games with lack of information on one side, *International Journal of Game Theory*, 30 (2001), p. 221–246.
[2001b] Learning sets in state dependent signalling game forms : a characterization, *Mathematics of Operations Research*, 26 (2001), p. 832–850.
[à paraître] The value of Markov chain games with lack of information on one side, *Mathematics of Operations Research*, (à paraître).

Renault (J.) & Tomala (T.)
[2004] Learning the state of nature in repeated games with incomplete information and signals, *Games and Economic Behavior*, 47 (2004), p. 124–156.

Rosenberg (D.)
[1998] Duality and Markovian strategies, *International Journal of Game Theory*, 27 (1998), p. 577–597.

Rosenberg (D.) & Sorin (S.)
[2001] An operator approach to zero- sum repeated games, *Israel Journal of Mathematics*, 121 (2001), p. 221–246.

Shalev (J.)
[1994] Nonzero-Sum Two-Person Repeated Games with Incomplete Information and Known-Own Payoffs, *Games and Economic Behavior*, 7 (1994), p. 246–259.

Simon (R.S.)
[2002] Separation of joint plan equilibrium payoffs from the min-max functions, *Games and Economic Behavior*, 1 (2002), p. 79–102.

Simon (R.S.), Spież (S.) & Toruńczyk (H.)
[1995] The existence of equilibria in certain games, separation for families of convex functions and a theorem of Borsuk-Ulam type, *Israel Journal of Mathematics*, 92 (1995), p. 1–21.
[2002] Equilibrium existence and topology in some repeated games with incomplete information, *Transactions of the AMS*, 354 (2002), p. 5005–5026.

SORIN (S.)

[1983] Some results on the existence of Nash equilibria for non- zero sum games with incomplete information, *International Journal of Game Theory*, 12 (1983), p. 193–205.

[1984] On a pair of simultaneous functional equations, *Journal of Mathematical Analysis and Applications*, 98 (1984), p. 296–303.

SORIN (S.) & ZAMIR (S.)

[1985] A 2-person game with lack of information on 1 and 1/2 sides, *Mathematics of Operations Research*, 10 (1985), p. 17–23.

SORIN (S.)

[2002] *A first course on zero-sum repeated games*, Mathématiques et Applications, Springer, 2002.

SPINAT (X.)

[2002] A necessary and sufficient condition for approchability, *Mathematics of Operations Research*, 27 (2002), p. 31–44.

VIEILLE (N.)

[1992] Weak approachability, *Mathematics of Operations Research*, 17 (1992), p. 781–791.

ZAMIR (S.)

[1971] On the relation between finitely and infinitely repeated games with incomplete information, *International Journal of Game Theory*, 1 (1971), p. 179–198.

[1973] On repeated games with general information function, *International Journal of Game Theory*, 21 (1973), p. 215–229.

[1992] Repeated Games of Incomplete Information : zero-sum, dans Aumann (R.J.) & Hart (S.), éd., *Handbook of Game Theory, I*, Elsevier Science Publishers, 1992, p. 109–154.

J. RENAULT, Ceremade, Université Paris Dauphine
E-mail : renault@ceremade.dauphine.fr

JEUX STOCHASTIQUES

par

Rida Laraki

Table des matières

1. Introduction

Les jeux stochastiques modélisent l'interaction entre des décideurs pouvant influencer leur environnement. Ces jeux ont d'abord été introduits et étudiés par Loyd Shapley (1953). Depuis, la littérature n'a cessé de croître.

Dans un jeu stochastique, les joueurs font face à des buts potentiellement différents. Ils doivent assurer un bon paiement aujourd'hui tout en maintenant une espérance de paiement élevée pour demain.

Les jeux stochastiques utilisent des outils mathématiques très variés. Nous allons présenter ici quelques résultats classiques, principalement pour les jeux à somme nulle. Plus précisément, les sections 2 à 4 présentent le modèle. Les sections (5, 6 puis 12) sont dédiées aux jeux à n joueurs et somme non nulle. Les sections 7 à 11 sont dédiées aux jeux à somme nulle.

Ce texte s'est inspiré principalement du cours sur les jeux répétés à somme nulle de Sorin (2002), du cours NATO sur les jeux stochastiques et leurs applications édité par Neyman et Sorin (2003), du polycopié de cours de DEA — non publié — sur les jeux stochastiques par Solan (2006) et, enfin, d'un article sur l'étude asymptotique des jeux absorbants à somme nulle par Laraki (2006).

2. Déroulement

Nous considérons un espace d'états (ou d'environnements) Ω fini. Nous avons un ensemble fini de joueurs noté $N = \{1, \ldots, |N|\}$. Dans chaque état ω le joueur i aura un ensemble d'actions (par étape) $A^i(\omega)$ considéré lui aussi fini (et non vide). $A(\omega) = \Pi_i A^i(\omega)$ est donc l'ensemble de tous les profils d'actions admissibles en une étape donnée à l'état ω. Nous notons l'ensemble des couples (état, profil d'actions) par :

$$\Omega A = \{(\omega, a) \mid a \in A(\omega)\}.$$

Donnons nous aussi une famille de probabilités de transition $q : \Omega A \to \Delta(\Omega)$ où $\Delta(X)$ est l'ensemble des probabilités sur X, et un état initial ω_1. Enfin, soit $g^i : \Omega A \to \mathbb{R}$ la fonction de paiement d'étape du joueur i.

Le jeu se déroule comme suit :

– *Étape 1* : l'état initial est noté ω_1. De manière simultanée et indépendante, chaque joueur choisit une action dans son ensemble d'actions admissibles en ω_1. Si le profil $a_1 = (a_1^i)_{i \in N} \in A(\omega)$ a été choisi, chaque joueur i reçoit pour l'étape 1 le paiement $g_1^i = g^i(\omega_1, a_1)$. Un état ω_2 est alors tiré selon la distribution de probabilité $q(\omega_1, a_1)$.

Ensuite, le couple (état, profil d'actions) (a_1, ω_2) est annoncé publiquement à tous les joueurs.

Le déroulement est maintenant défini par induction.

– *Étape* $t \geqslant 2$: connaissant l'histoire passée h_t,

$$h_t = (\omega_1, a_1, \ldots, \omega_{t-1,} a_{t-1}, \omega_t),$$

simultanément, chaque joueur choisit de façon indépendante aux autres une action dans son ensemble d'actions admissibles en ω_t. Si le profil $a_t = (a_t^i)_{i \in N}$ est choisi, chaque joueur i reçoit pour l'étape t le paiement $g_t^i = g^i(\omega_t, a_t)$. Un état ω_{t+1} est alors tiré selon la distribution de probabilité $q(\omega_t, a_t)$. Enfin, le couple (profil d'actions,état) (a_t, ω_{t+1}) est annoncé publiquement à tous les joueurs.

À titre d'exemple, la pêche est une industrie importante au Royaume-Uni et en Islande. Les deux pays partagent le même territoire dans l'Atlantique. Chaque année ils doivent fixer des quotas pour leurs pêcheurs respectifs. La décision des quotas est déterminée chaque année par rapport au nombre moyen z de kg de poisson par km^2. Celui-ci est mesuré chaque année à la fin de septembre. Les pêcheurs pêchent généralement l'intégralité de leurs quotas. Le taux de croissance des poissons est supposé être de 2% $(1 - \exp(-cz))$ où c est une constante fixée. Le gain pour le Royaume-Uni est mesuré par le nombre x de kg de poissons pêchés par ses pêcheurs par km^2. Le gain pour l'Islande est mesuré par le nombre y de kg de poissons pêchés par ses pêcheurs par km^2. En supposant que le Royaume-Uni a un pouvoir de négociation par rapport à l'Islande égal à $\alpha \in [0, 1]$, cette description peut s'écrire comme un jeu stochastique (avec un espace d'état et des ensembles d'actions compacts mais qui peuvent tous être discrétisés). La variable d'état serait alors $\omega = z$. Quand l'état est z, le Royaume-Uni peut choisir x dans l'intervalle $A^1(z) = [0, \alpha z]$ et l'Islande peut choisir y dans l'intervalle $A^1(z) = [0, (1 - \alpha)z]$. Si l'état actuel est z_t et que x_t et y_t ont été sélectionnés l'état demain sera,

$$z_{t+1} = (z_t - x_t - y_t) + (z_t - x_t - y_t) \times 2\% \, (1 - \exp(-c(z_t - x_t - y_t)) \, .$$

La loi de transition est donc déterministe. Enfin la fonction de paiement du Royaume-Uni est $g^1(\omega, x, y) = x$ et celle de l'Islande est $g^2(\omega, x, y) = y$.

3. Stratégies

C'est essentiellement la même définition que dans le texte sur les jeux répétés (ce volume). Ici nous l'adaptons à notre contexte. Pour tout entier (ou étape) t, l'*ensemble de toutes les histoires* possibles jusqu'à la date t est noté :

$$H_t = (\Omega A)^{t-1} \times \Omega.$$

Un élément de H_t sera noté h_t et la dernière composante est notée ω_t. H_1 est identifié avec l'espace d'état Ω. La première histoire n'est autre que ω_1. L'ensemble de toutes *les histoires de longueur finie* est noté

$$H = \bigcup_{t \geqslant 1} H_t.$$

Enfin, l'espace de toutes les *histoires d'une longueur infinie* (appelé *l'ensemble des parties*) est noté :

$$H_\infty = (\Omega A)^{\mathbb{N}^*}.$$

Pour chaque date t, H_t définit une partition (une algèbre ou un cylindre) de H_∞ : à chaque histoire de longueur finie $h_t \in H_t$ est associée l'ensemble des histoires infinies qui coïncident avec h_t jusqu'à l'étape t. Nous notons cette algèbre par $\mathcal{H}_t$ et notons par $\mathcal{H}$ la σ-algèbre générée par tous les cylindres.

Une stratégie de comportement σ^i pour le joueur i est une fonction qui associe à chaque histoire de longueur finie une action mixte dans $\Delta(A^i)$. L'ensemble des stratégies de comportement du joueur i est noté Σ^i.

Une stratégie de comportement est *pure* si pour chaque histoire finie h_t, $\sigma^i(h_t)$ est pure (est une masse de dirac).

Une stratégie *mixte* est une distribution de probabilité sur les stratégies pures.

Une stratégie de comportement est dite *stationnaire* si, pour chaque couple d'histoires de longueur finie $h_t = (\omega_1, a_1, \ldots, \omega_{t-1,} a_{t-1}, \omega_t)$ et $\widehat{h}_{\widehat{t}} = (\widehat{\omega}_1, \widehat{a}_1, \ldots, \widehat{\omega}_{\widehat{t}-1,} \widehat{a}_{\widehat{t}-1}, \widehat{\omega}_{\widehat{t}})$,

$$\omega_t = \widehat{\omega}_{\widehat{t}} \Longrightarrow \sigma(h_t) = \sigma(\widehat{h}_{\widehat{t}}).$$

Une stratégie stationnaire pour le joueur i sera notée x^i et un profil sera noté par $x = (x^1, \ldots, x^{|N|})$. L'ensemble des stratégies stationnaires du joueur i sera noté X^i, qui peut être identifié à

$\prod_{\omega\in\Omega}\Delta\left(A^i(\omega)\right)$. Ainsi, le nombre de stratégies stationnaires pures du joueur i est $\prod_{\omega\in\Omega}\left|A^i(\omega)\right|$ (où $|F|$ désigne le cardinal de l'ensemble F).

Une stratégie est dite *markovienne* si elle dépend seulement de l'état en cours et du nombre d'étapes écoulées. Mathématiquement, une stratégie de comportement est markovienne si, pour chaque couple d'histoires de même longueur $h_t=(\omega_1,a_1,\ldots,\omega_{t-1,}a_{t-1},\omega_t)$ et $\widehat{h}_t=(\widehat{\omega}_1,\widehat{a}_1,\ldots,\widehat{\omega}_{t-1,}\widehat{a}_{t-1},\widehat{\omega}_t)$,

$$\omega_t=\widehat{\omega}_t\Longrightarrow\sigma(h_t)=\sigma(\widehat{h}_t).$$

Chaque profil de stratégies σ et chaque état initial ω_1 définissent une unique distribution de probabilité sur H_∞ (voir le texte sur les jeux répétés (ce volume)). Cette probabilité sera notée $\mathbb{P}_{\omega_1,\sigma}$ et l'espérance mathématique associée sera noté $\mathbb{E}_{\omega_1,\sigma}$.

4. Objectifs

Ce sont essentiellement les mêmes définitions que celles du le texte sur les jeux répétés (ce volume). Ici, nous les adaptons à notre contexte et les reprenons pour préserver une certaine indépendance entre les textes de ce volume.

4.1. Approche compacte. — Dans cette approche on cherche une modélisation qui permet l'application directe des théorèmes standards d'existence de l'équilibre. On suppose donc que chaque joueur i cherche à maximiser une fonction de paiement total. Une telle fonction ϕ^i associe à chaque suite de paiements d'étapes $\left\{g^i(\omega_t,a_t)\right\}_{t=1,\ldots}$ une valeur dans l'intervalle $[-M,M]$ que l'on peut interpréter comme le paiement moyen par étape, où

$$M=\max_{(i,\omega,a)\in I\times\Omega A}\left|g^i(\omega,a)\right|.$$

Rappelons que H_∞ est toujours munie de la topologie naturelle induite par les cylindres. On supposera dans l'approche compacte que pour chaque profil de stratégies pures des autres joueurs σ^{-i}, la fonction $\sigma^i\mapsto\phi^i\left(\sigma^i,\sigma^{-i}\right)$ où σ^i parcourt l'ensemble des stratégies pures du joueur i, est continue pour la topologie naturelle induite par les cylindres.

L'ensemble de stratégies pures étant clairement compact pour la topologie naturelle, il est possible d'appliquer un théorème standard et d'en déduire l'existence d'un équilibre en stratégies mixtes. Le théorème de Kuhn (voir le texte sur les jeux répétés (ce volume)), nous permet ensuite de déduire l'existence d'un équilibre en stratégies de comportement.

Pour prouver l'existence d'un équilibre ayant des propriétés plus spécifiques (stationnaire ou markovien) il nous faut plus de structure dans les paiements. Les deux types de jeux qui vont suivre nous intéressent tout particulièrement du fait de leur récursivité et de l'intérêt qu'ils suscitent dans les applications.

Le jeu répété fini Γ_T dure T étapes. Dans un tel jeu, on suppose que le paiement (moyen) du joueur i pour une partie $h_\infty = (\omega_1, a_1, \ldots, \omega_t, a_t, \ldots) \in H_\infty$ est

$$g_T^i(h_\infty) = \sum_{t=1}^{T} \frac{1}{T}\, g^i\left(\omega_t, a_t\right).$$

Dans *le jeu escompté* Γ_λ, avec $\lambda \in]0, 1]$, on suppose que le paiement du joueur i est

$$g_\lambda^i(h_\infty) = \lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} g^i(\omega_t, a_t).$$

Il est facile de voir que ces deux jeux appartiennent à la famille des jeux compacts et donc admettent des équilibres en stratégies de comportement. Deux questions restent à résoudre : trouver des équilibres particuliers (stationnaires ou markoviens) mais aussi étudier le comportement asymptotique quand les joueurs sont de plus en plus patients ($T \to \infty$ ou $\lambda \to 0$).

Dans le cadre des jeux à deux joueurs et à somme nulle, on notera dans la suite $v_T(\omega_1)$ (resp. $v_\lambda(\omega_1)$) *la valeur* du jeu fini T fois (resp. du jeu λ-escompté).

4.2. Approche uniforme. — Nous allons montrer dans le section Big-Match que l'unique stratégie optimale peut dépendre explicitement de la durée T du jeu (respectivement du taux d'escompte λ). On montrera aussi que la stratégie limite, quand $T \to \infty$ (resp. $\lambda \to 0$) peut converger vers une mauvaise stratégie. Ceci justifie l'approche uniforme, dans laquelle on cherche des stratégies *uniformément* *bonnes* : c'est-à-dire presque optimales pour tout jeu fini de durée T assez grande (resp. tout jeu escompté avec un taux d'escompte λ assez petit).

Dans un jeu à somme nulle, on dit que la valeur uniforme $v(\omega_1)$ existe si pour tout $\varepsilon > 0$ il existe un profil de stratégies de comportement $(\sigma, \tau) \in \Sigma^1 \times \Sigma^2$ et il existe $T(\varepsilon)$ tels que pour tout $T \geqslant T(\varepsilon)$,

$$(1) \qquad \mathbb{E}_{\omega_1,\sigma,\widetilde{\tau}}\left(\sum_{t=1}^{T} \frac{1}{T} g^1(\omega_t, a_t)\right) \geqslant v(\omega_1) - \varepsilon \qquad \forall\, \widetilde{\tau} \in \Sigma^2,$$

$$(2) \qquad \mathbb{E}_{\omega_1,\widetilde{\sigma},\tau}\left(\sum_{t=1}^{T} \frac{1}{T} g^2(\omega_t, a_t)\right) \leqslant v(\omega_1) + \varepsilon \qquad \forall\, \widetilde{\sigma} \in \Sigma^1.$$

L'équation (1) (resp. l'équation (2)) sera interprétée : le joueur 1 (resp. 2) peut *garantir uniformément* $v(\omega_1)$.

La somme d'Abel pour une suite bornée de réels $\{z_t\}_{t\in\mathbb{N}^*}$ peut s'écrire comme une combinaison convexe infinie des sommes de Césaro :

$$\lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} z_t = \sum_{T=1}^{\infty} T\lambda^2 (1-\lambda)^{T-1} \Big(\frac{1}{T} \sum_{t=1}^{T} z_t\Big).$$

avec

$$\sum_{T=1}^{\infty} T\lambda^2(1-\lambda)^{T-1} = 1.$$

Ainsi, si la moyenne de Césaro $(\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} z_t)$ existe, alors la moyenne d'Abel $(\lim_{\lambda\to 0} \lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} z_t)$ existe aussi.

Ceci implique en particulier que si la valeur uniforme existe, alors pour tout $\varepsilon > 0$, il existe un profil de stratégies de comportement $(\sigma, \tau) \in \Sigma^1 \times \Sigma^2$ et il existe $\lambda(\varepsilon)$ tels que pour tout $0 < \lambda \leqslant \lambda(\varepsilon)$ et tout $(\widetilde{\sigma}, \widetilde{\tau}) \in \Sigma^1 \times \Sigma^2$,

$$\mathbb{E}_{\omega_1,\sigma,\widetilde{\tau}}\left(\lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} g^1(\omega_t, a_t)\right) \geqslant v_\infty(\omega_1) - \varepsilon,$$

$$\mathbb{E}_{\omega_1,\widetilde{\sigma},\tau}\left(\lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} g^2(\omega_t, a_t)\right) \leqslant v_\infty(\omega_1) + \varepsilon.$$

Donc si la valeur uniforme existe, alors nous avons :

$$v = \lim_{\lambda\to 0} v_\lambda = \lim_{T\to\infty} v_T.$$

Pour la définition d'un équilibre uniforme en somme non nulle, consulter le texte sur les jeux répétés (ce volume). Cependant, même dans les jeux à somme nulle, un équilibre uniforme n'existe pas toujours — par exemple dans le Big Match plus bas — d'où le recours à la valeur dans le cas à somme nulle et à des paiements uniformes (les limites de paiements d'ε-équilibres uniformes).

La définition exacte de l'ensemble des paiements d'équilibres uniformes E_0 est similaire à celle de la valeur uniforme (voir Mertens, Sorin et Zamir, p. 403) : on définit E_ε comme l'ensemble des paiements $v = (v^1, v^2, \ldots)$ tels qu'on trouve un profil de stratégies et une date T qui vérifient : dans tout jeu fini à au moins T étapes, d'une part le paiement du joueur i est au moins $v^i - \varepsilon$ et d'autre part en déviant, aucun des joueurs i ne peut gagner plus que $v^i + \varepsilon$. Puis E_0 est l'intersection des E_ε pour $\varepsilon > 0$.

Dans la suite nous allons présenter plusieurs techniques proposées dans la littérature pour montrer l'existence et/ou la caractérisation de $\lim_{\lambda\to 0} v_\lambda$, ainsi que l'égalité entre $\lim_{\lambda\to 0} v_\lambda$ et $\lim_{T\to\infty} v_T$. Enfin, nous développerons la preuve difficile concernant l'existence de la valeur uniforme (valable seulement dans notre cadre où tous les ensembles sont finis et où les joueurs se rappellent de tout le passé, observent les états et les actions passées).

4.3. Approche infinie. — Cette approche consiste à définir pour chaque joueur i, une fonction mesurable et bornée par M, définie directement sur l'ensemble des parties H_∞. Les exemples de l'approche compacte y sont inclus mais nous avons aussi les exemples suivants :

– $\liminf_{T\to\infty} \sum_{t=1}^{T} \frac{1}{T}\, g^i(\omega_t, a_t)$;
– $\limsup_{T\to\infty} \sum_{t=1}^{T} \frac{1}{T}\, g^i(\omega_t, a_t)$;
– $\liminf_{\lambda\to 0} \lambda \sum_{t=1}^{\infty} (1-\lambda)^t g^i(\omega_t, a_t)$;
– $\limsup_{\lambda\to 0} \lambda \sum_{t=1}^{\infty} (1-\lambda)^t g^i(\omega_t, a_t)$.

D'après la relation précédente entre les sommes d'Abel et de Césaro nous déduisons que

$$\lim_{T\to\infty} \inf \frac{1}{T} \sum_{t=1}^{T} z_t \leqslant \lim \inf_{\lambda\to 0} \lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} z_t.$$

$\lim_{T\to\infty} \inf \frac{1}{T} \sum_{t=1}^{T} z_t$ est donc l'évaluation la plus pessimiste de la suite des paiements d'étapes des quatre : si un joueur garantit un montant pour cette évaluation, il en est de même pour les trois autres évaluations.

5. Équilibre markovien

Théorème 5.1. — *Tout jeu stochastique fini Γ_T à N joueurs admet un équilibre en stratégies markoviennes.*

Démonstration. — Nous montrons ce résultat par récurrence sur la longueur du jeu T. Pour cela nous utilisons un argument de programmation dynamique.

Sans perte de généralité, nous supposerons que les joueurs maximisent la somme des paiements d'étapes (au lieu de la somme divisée par la longueur du jeu T).

Pour $T = 1$ c'est trivial.

Supposons que le résultat est vrai pour un jeu qui dure T étapes.

Soit le jeu de longueur $T+1$ et d'état initial ω_1. Quand les joueurs ont joué une fois, on arrive à un état ω_2 et il reste à jouer T étapes. En utilisant l'hypothèse de récurrence, nous sélectionnons pour chaque état futur possible ω_2 un équilibre markovien dans le jeu qui dure T étapes. Chaque joueur décide alors de jouer, à partir de l'étape 2, l'équilibre markovien du jeu fini qui dure T étapes.

Soit $f^i(\omega_2)$ le paiement global d'un tel équilibre (en sommant les paiements d'étapes à partir de l'étape 2). Ainsi, à l'étape 1, on peut considérer que les joueurs font face au jeu statique suivant :

– L'ensemble des joueurs est N ;
– L'ensemble d'actions du joueur i est $A^i(\omega_1)$;
– La fonction de paiement du joueur i est :

$$r^i(\omega_1, a) = g^i(\omega_1, a) + \sum_{\omega_2 \in \Omega} q(\omega_1, a)(\omega_2) f^i(\omega_2).$$

Ce jeu fini admet un équilibre en stratégies mixtes par application du théorème de Nash. Supposons alors que dans le jeu fini de $T+1$ étapes, les joueurs commencent par jouer un équilibre du jeu statique puis jouent l'équilibre markovien sélectionné dans le jeu fini qui dure T étapes. Ceci définit un équilibre markovien du jeu fini en $T+1$ étapes. □

6. Équilibre stationnaire

Shapley (1953), l'inventeur du modèle des jeux stochastiques, a montré l'existence de la valeur et des stratégies optimales stationnaires pour les jeux escomptés à deux joueurs et à somme nulle. Nous montrons ici le résultat de Fink (1964) et Takahashi (1964) qui généralisent Shapley aux jeux à n joueurs.

On commence par proposer une méthodologie générale pour calculer explicitement le paiement λ-escompté $g^i_\lambda(\omega, x)$ pour un profil de stratégies stationnaires $x = (x^i)_{i \in N}$.

Proposition 6.1. — *$\left\{g^i_\lambda(\omega, x)\right\}_{\omega \in \Omega}$ est l'unique solution au système linéaire de $|\Omega|$-équations :*

$$g^i_\lambda(\omega, x) = \sum_{a \in A(\omega)} \Big(\prod_{i \in N} x^i(a^i)\Big) \begin{pmatrix} \lambda g^i(\omega, a) \\ + (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, a)(\omega') g^i_\lambda(\omega', x) \end{pmatrix}.$$

Démonstration. — Puisque les joueurs jouent des stratégies stationnaires, l'espérance de paiement dépend seulement de l'état courant. Donc $\left\{g^i_\lambda(\omega, x)\right\}_{\omega \in \Omega}$ satisfait nécessairement ce système d'équations. Pour montrer que c'est l'unique solution, nous utilisons un principe de maximum. Supposons qu'il y ait deux solutions, $\{\alpha(\omega)\}_{\omega \in \Omega}$ et $\{\beta(\omega)\}_{\omega \in \Omega}$, au système. Soit

$$\omega_0 \in \arg\max_{\omega \in \Omega} \alpha(\omega) - \beta(\omega),$$

et supposons sans perte de généralité que

$$\alpha(\omega_0) - \beta(\omega_0) = k \geqslant 0.$$

Alors,

$$\begin{aligned} k &= \alpha(\omega_0) - \beta(\omega_0) \\ &= \sum_{a \in A(\omega_0)} \Big(\prod_{i \in N} x^i(a^i)\Big) \Big((1-\lambda) \sum_{\omega' \in \Omega} q(\omega_0, a)(\omega')\Big) \big(\alpha(\omega') - \beta(\omega')\big) \\ &\leqslant (1-\lambda) k. \end{aligned}$$

D'où $k = 0$. □

Nous en déduisons les propriétés suivantes.

Proposition 6.2. — *Pour tout joueur i et tout état initial ω, la fonction $(\lambda, x) \mapsto g^i_\lambda(\omega, x)$ est continue et une fraction rationnelle en λ.*

Démonstration. — Puisque le système est linéaire en λ, la solution est nécessairement une fraction rationnelle en λ. La continuité découle du fait que le système linéaire admet une unique solution. □

Nous allons maintenant utiliser la continuité pour montrer que le jeu Γ_λ admet une stationnarité qui permet l'existence d'un équilibre stationnaire. Pour cela, on définit une famille de jeux auxiliaires statiques similaire à celle de la section précédente. Pour chaque $|N|$-uplet de fonctions $f^1, \ldots, f^{|N|}$ de $\Omega \to \mathbb{R}$ bornées par M et chaque $\omega_1 \in \Omega$, on considère un jeu en un coup $G_{\omega_1}(f^1, \ldots, f^{|N|})$ défini comme suit :

– l'ensemble des joueurs est N ;
– l'ensemble d'actions du joueur i est $A^i(\omega)$;
– la fonction de paiement du joueur i est :

$$r^i(\omega_1, a) = \lambda g^i(\omega_1, a) + (1 - \lambda) \sum_{\omega_2 \in \Omega} q(\omega_1, a)(\omega_2) f^i(\omega_2).$$

***Théorème 6.3* (Fink (1964), Takahashi (1964)).** — *Tout jeu stochastique escompté à N joueurs admet un équilibre en stratégies stationnaires.*

Démonstration

Étape 1 : nous appliquons le théorème de Kakutani. Considérons l'ensemble

$$XF := \prod_{i \in N} \prod_{\omega \in \Omega} \Delta\left(A^i(\omega)\right) \times [-M, M]^{|N| \times |\Omega|}.$$

C'est un convexe compact d'un espace euclidien de dimension finie.

Un élément de XF sera noté

$$(x, f) = \left(\left\{x^i(\omega)\right\}^{i \in N}_{\omega \in \Omega}, \left\{f^i(\omega)\right\}^{i \in N}_{\omega \in \Omega}\right).$$

$x^i(\omega)$ doit être interprétée comme la stratégie jouée par le joueur i si l'état aujourd'hui est ω et $f^i(\omega)$ comme étant le paiement de continuation du joueur i si le nouvel état est ω.

Nous allons définir une correspondance $W : XF \to XF$ comme suit. Si les coordonnées de W sont notées

$$W = \left(W^{i,\omega}_X, W^{i,\omega}_F\right)^{i \in N}_{\omega \in \Omega},$$

alors

$$W_X^{i,\omega}(x,f)$$
$$= \arg \max_{y^i \in \Delta(A^i(\omega))} \left\{ \begin{array}{c} \lambda g^i(\omega, y^i, x^{-i}(\omega)) \\ +(1-\lambda)\sum_{\omega' \in \Omega} q(\omega, y^i, x^{-i}(\omega))(\omega') f^i(\omega') \end{array} \right\},$$

et

$$W_F^{i,\omega}(x,f) = \lambda g^i(\omega, x(\omega)) + (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, x)(\omega') f^i(\omega').$$

où la notation (y^i, x^{-i}) veut dire que le joueur i utilise la stratégie stationnaire y^i et les autres joueurs le profil x^{-i}. La motivation est la suivante. Considérons le jeu en un coup $G_\omega(f^1, \ldots, f^{|N|})$. Alors, $W_F^{i,\omega}(x,f)$ est le paiement espéré du joueur i si les joueurs utilisent le profil de stratégies stationnaires x et $W_X^{i,\omega}(x,f)$ est l'ensemble de toutes les meilleures réponses possibles du joueur i face à x^{-i}.

Il est facile de vérifier que cette correspondance satisfait aux hypothèses du théorème de Kakutani (pour la semi-continuité supérieure, on utilise la propriété de continuité dans la proposition 6.2). Nous concluons alors à l'existence d'un point fixe que nous notons (x,f).

Étape 2 : nous montrons que $g_\lambda^i(\omega,x) = f^i(\omega)$.

Ceci résulte de la proposition 6.1. En effet, puisque (x,f) est un point fixe de W, nous avons que pour tout i et tout ω :

$$f^i(\omega) = \lambda g^i(\omega, x(\omega)) + (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, x(\omega))(\omega') f^i(\omega').$$

Étape 3 : nous prouvons que pour toute stratégie σ^i du joueur i, $g_\lambda^i(\omega, \sigma^i, x^{-i}) \leqslant g_\lambda^i(\omega, x)$.

Par définition de F, nous avons que pour tout ω, $x^i(\omega)$ est une meilleure réponse du joueur i contre $x^{-i}(\omega)$ dans le jeu $G_\omega(f^1, \ldots, f^{|N|})$. D'où, pour tout ω et tout $y^i(\omega) \in \Delta\left(A^i(\omega)\right)$,

$$\begin{aligned} \lambda g^i(\omega, y^i, x^{-i}(\omega)) &+ (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, y^i, x^{-i}(\omega))(\omega') f^i(\omega') \\ &\leqslant \lambda g^i(\omega, x(\omega)) + (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, x(\omega))(\omega') f^i(\omega') \\ &= f^i(\omega) \\ &= g_\lambda^i(\omega, x). \end{aligned}$$

Soit $h_t = (\omega_1, a_1, \ldots, \omega_{t-1}, a_{t-1}, \omega_t)$ une histoire partielle de durée t et soit σ^i une stratégie quelconque du joueur i (qui peut dépendre de toute l'histoire du jeu). La dernière inégalité permet alors de déduire que :

$$\mathbb{E}_{\omega_1,\sigma^i,x^{-i}}\left(\begin{array}{l}\lambda g^i(\omega_t,\sigma^i(h_t),x^{-i}(\omega_t)) \\ +(1-\lambda)\sum\limits_{\omega_{t+1}\in\Omega} q(\omega_t,\sigma^i(h_t),x^{-i}(\omega_t))(\omega_{t+1})g^i_\lambda(\omega_{t+1},x)\,|h_t\end{array}\right) \leqslant g^i_\lambda(\omega_t,x).$$

Ceci implique, après une sommation, que

$$\begin{aligned} g^i_\lambda(\omega_1,\sigma^i,x^{-i}) &= \mathbb{E}_{\omega_1,\sigma^i,x^{-i}}\left(\lambda\sum_{t=1}^{\infty}(1-\lambda)^{t-1}g^i(\omega_t,a_t)\right) \\ &\leqslant g^i_\lambda(\omega_1,x). \end{aligned}$$

Nous avons donc bien un équilibre de Nash stationnaire. □

7. Opérateur de Shapley

À partir de maintenant et sauf mention explicite, nous nous focaliserons sur les jeux à deux joueurs et à somme nulle.

Puisque l'état du jeu est connu des deux joueurs, chaque joueur peut écrire le principe de programmation dynamique pour calculer sa stratégie optimale. En fait, v_λ et v_T peuvent être calculés à l'aide d'un même opérateur appelé l'opérateur de Shapley. Celui-ci étend le principe de programmation dynamique de Bellman. Ce principe a été publié par Shapley avant et indépendamment de Bellman. De plus Shapley traite le cas de deux joueurs alors que Bellman considère seulement celui d'un seul joueur.

L'opérateur de Shapley n'est autre que l'opérateur valeur pour un jeu statique, similaire à celui introduit dans la preuve du théorème 5.1. Pour chaque état possible ω_1 et chaque fonction $f : \Omega \to \mathbb{R}$ bornée par M, on introduit le jeu statique suivant à deux joueurs et à somme nulle :

– l'ensemble de stratégies pures du joueur 1 est $A^1(\omega_1)$;
– l'ensemble de stratégies pures du joueur 2 est $A^2(\omega_1)$;

– la fonction de paiement du joueur 1 est :

$$r^1(a,\omega_1) = g^1(a,\omega_1) + \sum_{\omega_2\in\Omega} q(\omega_1,a)(\omega_2)f(\omega_2).$$

Ce jeu admet une valeur en stratégies mixtes, noté $\Psi(f)(\omega_1)$. L'opérateur de Shapley Ψ est défini sur l'espace $\mathcal{F}$ des fonctions $f:\Omega\to\mathbb{R}$ qui sont bornées par M.

Pour tout $\lambda\in\,]0,1[$ nous définissons l'opérateur Φ comme suit :

$$\Phi(\lambda,f) = \lambda\Psi\Big(\frac{1-\lambda}{\lambda}f\Big).$$

$\Phi(\lambda,f)(\omega_1)$ correspond à la valeur en stratégies mixtes du jeu statique ayant comme paiement :

$$\lambda g^1(a,\omega_1) + (1-\lambda)\sum_{\omega_2\in\Omega} q(\omega_1,a)(\omega_2)f(\omega_2).$$

L'analyse de la section précédente montre le résultat suivant.

Théorème 7.1. — *v_λ est l'unique élément de $\mathcal{F}$ qui vérifie*

$$v_\lambda = \Phi(\lambda,v_\lambda).$$

La suite $\{v_T\}_{T\in\mathbb{N}^*}$ est l'unique suite dans $\mathcal{F}$ solution de :

$$v_{T+1} = \Phi\Big(\frac{1}{T+1},v_T\Big);$$
$$v_0 = 0.$$

Il est facile de voir que l'opérateur $\Phi(\lambda,\cdot)$ est contractant et donc admet un unique point fixe. Nous n'avons donc pas besoin d'utiliser le théorème de point fixe de Kakutani pour l'existence de v_λ ou de v_T mais seulement du théorème du minmax pour montrer que Φ est bien défini, et de l'argument de contraction de Picard.

Remarquons que dans Γ_λ et Γ_T, les joueurs ont une stratégie optimale markovienne et n'ont donc besoin d'observer que l'état courant pour bien jouer. Nous allons voir que, pour bien jouer uniformément, il est nécessaire que les joueurs observent leurs paiements. Sans cette observabilité, la valeur uniforme n'existerait pas. Ceci montre une différence fondamentale et intuitive entre l'approche compacte et l'approche uniforme : pour bien jouer uniformément, un joueur joue d'une manière élaborée.

Nous nous intéressons à l'étude de v_λ et v_T quand les joueurs sont de plus en plus patients ($\lambda \to 0$ et $T \to \infty$) et à la caractérisation de leur limite si possible. Nous allons commencer par un cas (plus) simple : celui des jeux absorbants.

8. Jeux absorbants

Dans cette section nous étudions une classe de jeu, introduite formellement par Kohlberg (1974) et qui va nous servir pour illustrer certains des résultats exposés et expliquer une partie des difficultés qui peuvent être rencontrées dans les jeux stochastiques.

Un état $\omega \in \Omega$ est dit absorbant si, une fois atteint, les joueurs ne peuvent jamais en sortir. Mathématiquement, cela veut dire que pour tout profil $a \in A(\omega)$, on a $q\,(\omega, a)\,(\omega) = 1$. Un jeu est absorbant s'il admet seulement un unique état non absorbant.

Une fois qu'un état absorbant est atteint, le jeu est réduit à un jeu répété à information parfaite (déjà analysé dans le premier texte de ce volume). Nous savons alors que, partant d'un tel état, un équilibre existe. Si nous nous intéressons à l'analyse des équilibres, on peut supposer, sans perte de généralité, qu'une fois qu'un état absorbant est atteint, la suite des paiements est constante et égale à un paiement d'équilibre (que nous avons préalablement sélectionné dans le jeu répété).

En résumé, on va supposer dans toute la suite et sans perte de généralité, qu'à tout état absorbant dans un jeu stochastique est associé un paiement absorbant (un paiement d'étape que les joueurs reçoivent à toutes les étapes suivantes du jeu). Nous supposerons que l'état initial ω_1 d'un jeu absorbant est l'état non absorbant (sinon le jeu serait trivial et sans enjeu). Dès que l'on quitte cet état, le jeu est essentiellement terminé (il n'y a plus de difficulté mathématique liée à l'aspect stochastique). Il n'est donc plus nécessaire de spécifier l'état de départ dans un jeu absorbant.

Ainsi, un jeu absorbant à somme nulle peut être décrit d'une manière compacte comme suit. Il y a deux joueurs, 1 et 2. Le jeu est donné par deux ensembles finis d'actions, A^1 pour le joueur 1 et A^2 pour le joueur 2. Nous avons par ailleurs besoin de deux fonctions de paiements $\widetilde{g} : A^1 \times A^2 \to \mathbb{R}$ et $g^* : A^1 \times A^2 \to \mathbb{R}$. Enfin, nous avons besoin d'une famille de probabilités de transition $p^* : A^1 \times A^2 \to [0,1]$. Le jeu est joué comme suit. À l'étape $t = 1, 2, \ldots$, le joueur 1 choisit une action $a_t^1 \in A^1$ et simultanément le joueur 2 une action $a_t^2 \in A^2$, puis

(i) avec probabilité $p^*\left(a_t^1, a_t^2\right)$ le jeu est absorbé et le joueur 1 reçoit le paiement $g^*\left(a_t^1, a_t^2\right)$ à chacune des étapes restantes.

(ii) avec probabilité $1 - p^*\left(a_t^1, a_t^2\right) = \widetilde{p}\left(a_t^1, a_t^2\right)$ le jeu n'est pas absorbé et le paiement de l'étape t du joueur 1 est $\widetilde{g}\left(a_t^1, a_t^2\right)$.

Exemple d'un jeu d'arrêt. — Soit le jeu absorbant suivant à deux joueurs et à somme nulle.

	C	A
C	0	1^*
A	1^*	0^*

Dans ce jeu, il y a deux paiements absorbants : 1 et 0 (ils sont marqués par $*$). Le paiement absorbant 1^* est atteint si un des profils d'actions (A,C) ou (C,A) est joué (et dans ce cas le joueur gagne 1 par étape, et ce jusqu'à la fin des temps, et le joueur 2 gagne l'opposé, soit -1). Le paiement absorbant 0^* est atteint si (A,A) est joué. Le jeu reste dans l'état non absorbant si et seulement si les joueurs jouent (C,C) (c'est la seule case de la matrice sans $*$).

Dans un *jeu d'arrêt* général chaque joueur a deux options : A (arrêter) ou C (continuer). Dès qu'un joueur choisit A, le jeu est absorbé (s'arrête) et tant que le jeu continue les joueurs reçoivent un paiement de 0 par étape.

Considérons le profil suivant de stratégies stationnaires dans l'exemple du jeu d'arrêt : le joueur 1 joue $(xH, (1-x)B)$ et le joueur 2 joue $(yG, (1-y)D)$. Calculons maintenant le paiement du joueur 1 $g_\lambda(x,y)$ dans le jeu escompté Γ_λ :

$$g_\lambda(x,y) = xy\left(\lambda \times 0 + (1-\lambda)\, g_\lambda(x,y)\right) + \left((1-x)y + (1-y)x\right),$$

d'où

$$g_\lambda(x,y) = \frac{x + y - 2xy}{1 - xy(1-\lambda)}.$$

Dans cet exemple, la valeur $v_\lambda \in [0,1]$ satisfait :

$$\begin{aligned} v_\lambda &= \text{valeur}\left(\begin{array}{c|c|c|} & \text{C} & \text{A} \\ \cline{2-3} \text{C} & (1-\lambda)v_\lambda & 1 \\ \cline{2-3} \text{A} & 1 & 0 \\ \cline{2-3} \end{array}\right) \\ &= \max_{x\in[0,1]} \min_{y\in[0,1]} \left[xy(1-\lambda)v_\lambda + x(1-y) + y(1-x)\right] \\ &= \min_{y\in[0,1]} \max_{x\in[0,1]} \left[xy(1-\lambda)v_\lambda + x(1-y) + y(1-x)\right]. \end{aligned}$$

On vérifie alors que les joueurs n'ont pas de stratégies optimales pures. Si $x_\lambda \in]0,1[$ (resp. $y_\lambda \in]0,1[$) est la stratégie optimale du joueur 1 (resp. du joueur 2) alors, en utilisant le fait que chaque joueur est indifférent entre ces deux actions (les deux ont la même espérance de paiement) nous trouvons que :

$$v_\lambda = x_\lambda(1-\lambda)v_\lambda + (1-x_\lambda) = x_\lambda = y_\lambda.$$

Donc il existe des stratégies optimales uniques, elles sont stationnaires et l'on a :

$$v_\lambda = x_\lambda = y_\lambda = \frac{1-\sqrt{\lambda}}{1-\lambda}.$$

Remarquons que v_λ n'est pas une fraction rationnelle de λ. Ceci est une différence fondamentale avec les jeux à un seul joueur (programmation dynamique). Dans le cas d'un joueur, il existe toujours une stratégie optimale qui est stationnaire et pure pour tout λ. Ceci implique en particulier que v_λ est une fraction rationnelle de λ (car c'est un maximum sur un ensemble fini de fonctions rationnelles en λ) mais aussi qu'il existe un $\lambda_0 > 0$ et une même stratégie pure qui est optimale pour tout $\lambda < \lambda_0$ (une stratégie uniformément bonne, Blackwell (1962)). Enfin, si le jeu d'arrêt se jouait en temps continu (escompté ou non), le joueur 1 pourrait arrêter seul le jeu avec probabilité 1 et garantir ainsi un paiement proche de 1 : il lui suffit de tirer un instant uniformément entre le début du jeu et un temps très proche du début et d'arrêter le jeu à cet instant. Ceci montre une différence fondamentale entre le comportement optimal en temps discret et celui en temps continu. En effet, en temps discret, le joueur 1 ne peut garantir d'arrêter le jeu seul.

Dans cet exemple, $\lim v_\lambda$ existe et est égale à 1. Dans la section suivante nous allons montrer, par l'utilisation d'une approche semi-algébrique, que la valeur asymptotique existe pour tout jeu stochastique à somme nulle. Dans le cas des jeux absorbants, la convergence peut être démontrée plus simplement, en utilisant une approche variationnelle (Laraki 2006). La méthode s'inspire de l'utilisation des solutions de viscosité pour montrer l'existence et la caractérisation de la valeur dans les jeux différentiels à somme nulle et découle d'une méthodologie plus générale (voir Laraki 2001, 2002). Cette approche dans les jeux absorbants permet une caractérisation explicite de $\lim v_\lambda$ comme la valeur d'un jeu.

Nous avons besoin des *notations suivantes spécifiques à cette section* :

– $\mathbb{R}_+^{A^1} = \{z = (z_{a^1})_{a^1\in A^1} \mid z_{a^1} \in \mathbb{R}_+\}$ est l'orthant positif associé à l'ensemble fini d'actions du joueur 1 A^1 (qu'on peut identifier à l'ensemble des mesures positives sur A^1).

– Pour $z \in \mathbb{R}_+^{A^1}$, $x \in \Delta(A^1)$ et $a^2 \in A^2$,

- x_{a^1} est la probabilité avec laquelle le joueur 1 joue l'action a^1.
- $\widetilde{g}(x, a^2) = \sum_{a^1\in A^1} x_{a^1}(1 - p^*(a^1, a^2))\widetilde{g}(a^1, a^2)$ est le paiement non absorbant d'étape si le joueur 1 joue x et le joueur 2 joue j ;
- $g^*(z, a^2) = \sum_{a^1\in A^1} z_{a^1} p^*(a^1, a^2)g^*(a^1, a^1)$ est l'extension linéaire de $\widetilde{g}(\cdot, a^2)$ à $\mathbb{R}_+^{A^1}$;
- $p^*(z, a^2) = \sum_{a^1\in A^1} z_{a^1} p^*\left(a^1, a^2\right)$ est la probabilité d'absorption étendue linéairement à $\mathbb{R}_+^{A^1}$;
- $\widetilde{p}(x, a^2) = 1 - p^*(x, a^2)$ est la probabilité de continuation.

L'opérateur de Shapley implique que v_λ existe et est l'unique réel dans $[-M, +M]$ qui satisfait

$$v_\lambda = \max_{x\in\Delta(A^1)} \min_{a^2\in A^2} \left[\lambda\widetilde{g}(x, a^2) + (1 - \lambda)\,\widetilde{p}(x, a^2)v_\lambda + g^*(x, a^2)\right]. \tag{3}$$

Proposition 8.1. — *v_λ converge vers v quand λ tend vers zéro, où v est donné par la formule suivante :*

$$v = \sup_{z\in\mathbb{R}_+^{A^1}} \sup_{x\in\Delta(A^1)} \min_{a^2\in A^2} \begin{pmatrix} \frac{g^*(x,a^2)}{p^*(x,a^2)} 1_{\{p^*(x,a^2)>0\}} \\ +\frac{\widetilde{g}(x,a^2)+g^*(z,a^2)}{\widetilde{p}(x,a^2)+p^*(z,a^2)} 1_{\{p^*(x,a^2)=0\}} \end{pmatrix}.$$

Démonstration. — Soit w un point d'accumulation de v_λ quand λ tend vers zéro : $w = \lim_{n\to\infty} v_{\lambda_n}$ où $\lambda_n \to 0$.

Étape 1 : nous allons montrer que $w \leqslant v$. L'idée est de considérer une stratégie stationnaire optimale $x\,(\lambda_n)$ pour le joueur 1 puis d'aller à la limite dans l'opérateur de Shapley. Ce procédé est bien connu dans la théorie du contrôle optimal.

De (3), nous déduisons qu'il existe $x\,(\lambda_n) \in \Delta(A^2)$ tel que,

$$v_{\lambda_n} = \min_{a^2\in A^2} \begin{bmatrix} \lambda_n \widetilde{g}(x(\lambda_n), a^2) + (1-\lambda_n)\,\widetilde{p}(x(\lambda_n), a^2) v_{\lambda_n} \\ +g^*(x(\lambda_n), a^2) \end{bmatrix}.$$

D'où, pour tout $a^2 \in A^2$,

$$v_{\lambda_n} \leqslant \frac{\lambda_n \widetilde{g}(x(\lambda_n), a^2) + g^*(x(\lambda_n), a^2)}{\lambda_n \widetilde{p}(x(\lambda_n), a^2) + p^*(x(\lambda_n), a^2)}. \tag{4}$$

Par compacité de $\Delta(A^1)$, et en considérant si nécessaire une sous-suite de $\{\lambda_n\}$, on peut supposer sans perte de généralité que $x\,(\lambda_n) \to x \in \Delta(A^1)$.

Si $p^*(x, a^2) > 0$, alors en faisant tendre λ_n vers zéro, on trouve que :

$$w = \lim v\,(\lambda_n) \leqslant \frac{g^*(x, a^2)}{p^*(x, a^2)}.$$

Supposons maintenant que $p^*(x, a^2) = \sum_{a^1\in A^1} x_{a^1} p^*(a^1, a^2) = 0$, et donc que pour tout a^1 tel que $x_{a^1} > 0$, nous avons $p^*(a^1, a^2) = 0$. Soit alors $z(\lambda_n) = (x_{a^1}\,(\lambda_n)/\lambda_n)_{a^1\in A^1} \in \mathbb{R}_+^{A^1}$. L'équation (4) devient, après la division par λ_n,

$$\begin{aligned} v\,(\lambda_n) &\leqslant \frac{\widetilde{g}(x(\lambda_n), a^2) + g^*(z(\lambda_n), a^2)}{\widetilde{p}(x(\lambda_n), a^2) + p^*(z(\lambda_n), a^2)} \\ &\triangleq r\left(\lambda_n, a^2\right). \end{aligned}$$

Puisque A^2 est fini, en considérant si nécessaire une sous-suite de $\{\lambda_n\}$, on peut supposer que $r\left(\lambda_n, a^2\right)$ converge pour tout a^2. Puisque $\widetilde{p}(x, a^2) = \lim_{n\to\infty} \widetilde{p}(x(\lambda_n), a^2) = 1$ et $\widetilde{g}(x, a^2) = \lim_{n\to\infty} \widetilde{g}(x(\lambda_n), a^2)$, on en déduit que

$$w \leqslant \lim_{n\to\infty} \frac{\widetilde{g}(x, a^2) + g^*(z(\lambda_n), a^2)}{\widetilde{p}(x, a^2) + p^*(z(\lambda_n), a^2)}.$$

Soit $\varepsilon > 0$. Puisque A^2 est fini, on en déduit l'existence d'un entier $N(\varepsilon)$ tel que, pour tout $a^2 \in A^2$,

– si $p^*(x,a^2)>0$ alors $w \leqslant \dfrac{g^*(x,a^2)}{p^*(x,a^2)}$.

– si $p^*(x,a^2)=0$ alors $w \leqslant \dfrac{\widetilde{g}(x,a^2)+g^*(z(\lambda_{N(\varepsilon)}),a^2)}{\widetilde{p}(x,a^2)+p^*(z(\lambda_{N(\varepsilon)}),a^2)}+\varepsilon$.

En conséquence, $w \leqslant v$.

Étape 2 : nous montrons que $w \geqslant v$. L'idée est de construire une stratégie du joueur 1 dans le jeu λ_n-escompté qui lui garantit approximativement v.

Rappelons que λ_n converge vers 0, et que $w=\lim v_{\lambda_n}$.

Soit $(z,x)\in \mathbb{R}_+^{A^1}\times\Delta(A^1)$ ε-optimal pour le joueur 1 dans l'expression de v. Supposons que λ_n soit assez petit. Soit

$$A^1(x) := \{a^1 \in A^1 \mid x_{a^1}=0\}$$

et définissons $x(\lambda_n)\in\Delta(A^1)$ comme suit :

– si $A^1(x)=\varnothing$ alors $x(\lambda_n)=x$.

– si $a^1\in A^1(x)$ alors $x_{a^1}(\lambda_n)=z_{a^1}\times\lambda_n$.

– si $a^1\notin A^1(x)$ et $A^1(x)\neq\varnothing$ alors $x_{a^1}(\lambda_n)=x_{a^1}-\dfrac{\sum_{a^1\in A^1(x)}z_{a^1}}{|A^1(x)|}\lambda_n$.

Ainsi, nous avons $v(\lambda_n)\geqslant r(\lambda_n)-\varepsilon$, où $r(\lambda_n)$ est l'unique réel dans $[-M,M]$ qui satisfait,

$$r(\lambda_n)=\min_{a^2\in A^2}\begin{bmatrix}\lambda_n\left[\widetilde{g}(x(\lambda_n),a^2)\right]\\ +(1-\lambda_n)\left(\widetilde{p}(x(\lambda_n),a^2)\right)r(\lambda_n)\\ +g^*(x(\lambda_n),a^2)\end{bmatrix}. \tag{5}$$

En effet, $r(\lambda_n)$ est ce que le joueur 1 s'assure de gagner dans le jeu λ_n-escompté s'il joue la stratégie stationnaire $x(\lambda_n)$. Soit $a^2_{\lambda_n}\in A^2$ une stratégie optimale stationnaire pure pour le joueur 2 contre $x(\lambda_n)$ dans le jeu λ_n-escompté (un élément du arg min dans (5)). Puisque la suite $(a^2_{\lambda_n})_{n\geqslant1}$ appartient à un ensemble fini A^2, il y a un nombre fini de sous-suites de $\{\lambda_n\}$ pour lesquelles $a^2_{\lambda_n}$ est constant pour n grand. Pour chacune de ces sous-suites, (que l'on continuera à appeler λ_n) on peut supposer que $r(\lambda_n)$ converge vers un certain r. En reprenant exactement le calcul de l'étape 1, nous en déduisons que $w\geqslant v-\varepsilon$ pour tout $\varepsilon>0$. $\square$

Remarquez qu'on aurait pu obtenir une autre formule en considérant dans la preuve une stratégie optimale du joueur 2 (au lieu du joueur 1).

Coulomb (2001), obtient une formule explicite pour le maxmin uniforme des jeux absorbants avec signaux. Si l'on suppose l'observation parfaite des actions et en utilisant le fait que dans ce cas le maxmin uniforme coïncide avec la valeur uniforme et donc avec $\lim v_\lambda$, nous aboutissons à une autre formule pour $\lim v_\lambda$.

Cette analyse variationnelle et la formule qui en résulte peuvent être étendues aux cas où les ensembles d'actions sont compacts et les fonctions $\widetilde{g}$, g^* et p^* continues. Ceci ne peut être fait dans le cadre de Coulomb (2001) qui utilise une approche semi-algébrique. Cependant, Coulomb étudie le cas uniforme qui nécessite des outils plus complexes et des stratégies plus élaborées...

9. Approche semi-algébrique

Ici, nous montrons l'existence de $\lim_{\lambda\to 0} v_\lambda$ pour tout jeu stochastique à somme nulle en utilisant une approche semi-algébrique, initiée par Bewley et Kohlberg (1976 a et b). Cela permet de montrer en particulier que v_λ est à variation bornée. Cette propriété impliquera que $\lim_{T\to\infty} v_T$ existe et est égale à $\lim_{\lambda\to 0} v_\lambda$. Ici on suit Sorin (2002).

Un ensemble dans $\mathbb{R}^m$ est semi-algébrique s'il peut s'écrire comme union finie d'ensembles A_k de la forme :

$$A_k = \{x \in \mathbb{R}^m \mid p_k(x) > 0\} \quad \text{ou} \quad A_k = \{x \in \mathbb{R}^m \mid p_k(x) = 0\}.$$

où p_k est un polynôme de $\mathbb{R}^m$.

En utilisant le fait que v_λ est l'unique point fixe de $\Phi(\lambda, \cdot)$, on en déduit le résultat suivant.

Proposition 9.1. — *L'ensemble des $(\lambda, v_\lambda, x_\lambda)$ tel que λ parcourt $]0,1]$, v_λ est la valeur du jeu λ-escompté, x_λ^i est une stratégie optimale stationnaire du joueur i est semi-algébrique.*

Ce résultat peut facilement être étendu aux jeux à somme non nulle en remplaçant valeur par paiement d'équilibre et stratégie optimale par profil d'équilibre de Nash (pour plus de détails sur l'approche semi-algébrique, consulter le chapitre 6 par Neyman dans Neyman et Sorin 2003).

En utilisant l'élimination de Tarski-Seidenberg (Benedetti et Risler 1990, théorème 2.21, p. 54) on en déduit qu'il existe une sélection semi-algébrique par rapport à $\lambda \in \,]0,1[$. Ceci implique l'existence d'un développement en série de Puiseux au voisinage de 0 pour chaque élément $z_\lambda \in \{x_\lambda, y_\lambda, v_\lambda\}$.

***Théorème 9.2* (Bewley and Kohlberg 1976).** — *Étant donné un jeu stochastique (où tout est fini) à somme nulle, pour chaque $z_\lambda \in \{x_\lambda, y_\lambda, v_\lambda\}$, il existe $\lambda_0 > 0$, $k \in \mathbb{N}^*$, $r_n : \Omega \to \mathbb{R}$, $n = 0, \dots,$ tels que :*

$$z_\lambda(\omega) = \sum_{n=0}^{\infty} r_n(\omega)\, \lambda^{\frac{n}{k}}.$$

pour tout $\lambda \in \left]0, \lambda_0\right[$ et tout $\omega \in \Omega$.

Démonstration. — On applique le résultat de Forster 1981, théorème 8.14, p. 58, au cas où les paiements sont uniformément bornés. □

Ainsi dans l'exemple du jeu d'arrêt nous avons :

$$\begin{aligned} v_\lambda = x_\lambda = y_\lambda &= \frac{1-\sqrt{\lambda}}{1-\lambda} \\ &= \left(1-\sqrt{\lambda}\right)\left(1+\lambda+\lambda^2+\cdots\right). \end{aligned}$$

Dans le cadre des jeux avec plus de deux joueurs ou à somme non nulle, on peut montrer l'existence d'une sélection de stratégies et de paiements d'équilibre ayant un développement en série de Puiseux.

La fonction $\lambda \to f_\lambda$ où $f_\lambda \in \mathcal{F}$ et $\lambda \in \left]0,1\right[$ sera dite *à variation bornée* si pour toute suite $\{\lambda_n\}_{n\in\mathbb{N}}$ dans $\left]0,1\right[$ décroissante vers 0, nous avons :

$$\sum_{n=1}^{\infty} \left\| f_{\lambda_{n+1}} - f_{\lambda_n} \right\|_\infty < \infty.$$

Corollaire 9.3. — *v_λ est une fonction à variation bornée au voisinage de 0 et admet une limite quand $\lambda \to 0$.*

Démonstration. — La convergence est une conséquence du développement de Puiseux. La variation bornée est une conséquence du fait que pour tout $\lambda \in (0, \lambda_0)$ nous avons :

$$\left\| \frac{dv_\lambda}{d\lambda} \right\|_\infty \leqslant C\lambda^{(1-k)/k}.$$ □

Corollaire 9.4. — *$\lim_{T\to\infty} v_T$ existe et est égale à $\lim_{\lambda\to 0} v_\lambda$.*

Ceci est une conséquence directe d'un résultat général (voir le chapitre 26 par Neyman dans Neyman et Sorin 2003).

Soit Θ un opérateur d'un espace de Banach $\mathcal{Z}$ dans lui-même qu'on supposera non dilatant, *i.e.* pour tout z et z' dans $\mathcal{Z}$, $\|\Theta(z) - \Theta(z')\| \leqslant \|z - z'\|$. Il est facile de voir alors que, pour tout z et z' dans $\mathcal{Z}$,

$$\left\|\lambda\Theta\left(\frac{1-\lambda}{\lambda}z\right) - \lambda\Theta\left(\frac{1-\lambda}{\lambda}z'\right)\right\| \leqslant (1-\lambda)\left\|z - z'\right\|.$$

de sorte que $w \to \lambda\Theta\left(\frac{1-\lambda}{\lambda}w\right)$ admet un point fixe qui est nécessairement unique (par contraction de Picard). Soit w_λ ce point fixe.

Définissons la suite $\{w_T\}_{T=0,1,\ldots}$ dans $\mathcal{Z}$ par récurrence comme suit : $w_0 = 0$ et $w_{T+1} = \frac{1}{T+1}\Theta(Tw_T)$, celle-ci pouvant s'écrire aussi $w_T = \frac{\Theta^T(0)}{T}$.

***Théorème 9.5* (Neyman 2003).** — *Si w_λ est à variation bornée alors $\lim_{\lambda\to 0} w_\lambda = \lim_{T\to\infty} w_T$.*

Pour montrer le dernier corollaire il suffit de prendre $\Theta = \Psi$ et $\mathcal{Z} = \mathcal{F}$.

Démonstration. — Remarquons que si w_λ est à variation bornée, elle converge nécessairement vers un certain w. Supposons que $\lambda_t = 1/t$. Dans ce cas,

$$\begin{aligned}\left\|w_{t+1} - w_{\lambda_{t+1}}\right\| &= \left\|\frac{1}{t+1}\Theta(tw_t) - \frac{1}{t+1}\Theta\left(tw_{\lambda_{t+1}}\right)\right\| \\ &\leqslant \frac{t}{t+1}\left\|w_t - w_{\lambda_{t+1}}\right\| \\ &\leqslant \frac{t}{t+1}\left(\|w_t - w_{\lambda_t}\| + \left\|w_{\lambda_t} - w_{\lambda_{t+1}}\right\|\right),\end{aligned}$$

soit

$$(t+1)\left\|w_{t+1} - w_{\lambda_{t+1}}\right\| \leqslant t\left\|w_t - w_{\lambda_t}\right\| + t\left\|w_{\lambda_t} - w_{\lambda_{t+1}}\right\|.$$

En sommant on trouve

$$(T+1)\left\|w_{T+1} - w_{\lambda_{T+1}}\right\| \leqslant \|w_1 - w_{\lambda_1}\| + \sum_{t=1}^{T} t\left\|w_{\lambda_t} - w_{\lambda_{t+1}}\right\|.$$

Puisque la somme $\sum_{t=1}^{T}\left\|w_{\lambda_t} - w_{\lambda_{t+1}}\right\|$ est bornée, nous obtenons que $\frac{1}{T+1}\sum_{t=1}^{T} t\left\|w_{\lambda_t} - w_{\lambda_{t+1}}\right\| \to 0$, d'où le résultat. □

10. Big-Match

Avant d'étudier l'approche uniforme dans les jeux à somme nulle en général, nous étudions le Big-Match. C'est le premier exemple résolu explicitement dans le cadre uniforme par Blackwell et Ferguson (1968). C'est le jeu suivant :

	$\overline{0}$	$\overline{1}$
$\overline{0}$	1	0
$\overline{1}$	0^*	1^*

L'histoire qu'ils ont utilisée pour le présenter est la suivante :

« Chaque jour le joueur 2 choisit un nombre $\overline{0}$ ou $\overline{1}$, et le joueur 1 essaye de prédire le choix du joueur 2, gagnant un point à chaque fois qu'il a raison. Cela continue tant que le joueur 1 prédit $\overline{0}$. Si un jour il prédit $\overline{1}$, tous les choix futurs pour les deux joueurs sont contraints à être les mêmes que le choix de ce jour : si le joueur 1 avait raison ce jour là, il gagne 1 *point chaque jour suivant ; s'il s'est trompé ce jour là, il gagne* 0 *tous les jours suivants. »*

Seul le joueur 1 peut arrêter le jeu et ceci arrive dès qu'il choisit l'option $\overline{1}$. La formule du théorème 7.1 implique que :

$$v_{T+1} = \frac{1}{T+1} \text{ valeur} \left(\begin{array}{c|c|c} & \overline{0} & \overline{1} \\ \hline \overline{0} & 1+\mathrm{T}v_T & \mathrm{T}v_T \\ \hline \overline{1} & 0 & 1 \end{array} \right).$$

Par récurrence, on montre que $v_T = 1/2$. La stratégie $(\frac{1}{2}\overline{0}, \frac{1}{2}\overline{1})$ du joueur 2 lui garantit une espérance de gain de 1/2 par étape et donc 1/2 uniformément. L'unique stratégie optimale du joueur 1 est markovienne et consiste de jouer $\overline{1}$ avec probabilité $x_m = \frac{1}{m+1}$, s'il reste m étapes à jouer (en particulier il n'existe pas de stratégie stationnaire optimale pour le joueur 1). On calcule facilement que $v_\lambda = 1/2$ et que l'unique stratégie optimale du joueur 1 est évidemment stationnaire. Elle consiste à jouer $\overline{1}$ avec probabilité $x_\lambda = \frac{\lambda}{1+\lambda}$. À la limite quand $\lambda \to 0$, le joueur 1 joue $\overline{0}$ avec probabilité 1 à chaque étape. Cette stratégie est mauvaise car elle ne permet d'assurer que zéro par étape (face à cette stratégie, le joueur 2 pourrait jouer l'option $\overline{1}$ à chaque étape).

Nous allons montrer que le joueur 1 peut garantir uniformément tout paiement strictement moins que $1/2$ (ce qui implique que la valeur uniforme de ce jeu existe et est égale à $1/2$). Cependant, il assez facile de montrer qu'il n'a pas de stratégie uniforme qui garantit exactement $1/2$. Pour le montrer il suffit de remarquer que le joueur 1 doit assurer tout le temps un paiement de continuation de $1/2$ ce qui l'oblige à continuer le jeu avec probabilité 1 à chaque étape (en jouant tout le temps l'option $\overline{0}$). Cette stratégie ne lui garantit qu'un paiement de 0. Le Big-Match n'admet donc pas d'équilibre uniforme mais seulement un paiement d'équilibre uniforme.

L'idée de la construction est la suivante : tant que le jeu n'est pas absorbé, si le joueur 1 a obtenu dans le passé un bon paiement (plus de 1 que de 0), il joue avec une probabilité relativement grande la stratégie $\overline{0}$, sinon il augmente sa probabilité de jouer $\overline{1}$.

***Théorème 10.1* (Blackwell et Ferguson 1968).** — *Pour tout entier $K > 0$ le joueur 1 a une stratégie qui garantit uniformément le paiement $K/2(K+1)$.*

Démonstration. — Fixons un entier $K > 0$. Pour tout nombre d'étapes T, définissons trois entiers :

– 0_T le nombre d'étapes avant T où le joueur 2 joue $\overline{0}$;
– 1_T le nombre d'étapes avant T où le joueur 2 joue $\overline{1}$;
– $k_T = 0_T - 1_T$.

On remarque que $0_T + 1_T = T - 1$, ce qui implique :

$$k_T = 2l_T - T - 1.$$

Si le joueur 2 a souvent joué $\overline{0}$, k_T est grand (en particulier positif). Dans ce cas, le joueur 1 augmentera sa probabilité de jouer $\overline{0}$. Dans le cas contraire, k_T est petit (en particulier négatif) et le joueur 1 augmentera sa probabilité de jouer $\overline{1}$.

Nous définissons la stratégie du joueur 1, σ_K, comme suit : à l'étape T, si le jeu n'a pas été absorbé, jouer $\overline{1}$ avec probabilité $1/(k_T + K + 1)^2$. On note que si $k_T = -K$ alors le jeu est absorbé avec probabilité 1. En particulier, tout choix suivant des joueurs n'affecte plus les paiements. Nous allons montrer que pour toute suite d'actions $a^2 = \left(a_1^2, \ldots, a_t^2, \ldots\right)$ du joueur 2, nous avons :

$$\mathbb{E}_{\sigma_K^1, a^2}\left(\sum_{t=1}^{T} \frac{1}{T}\, g^1\left(\omega_t, a_t\right)\right) \geqslant \frac{K}{2(K+1)} - \frac{K+1}{2T}.$$

Ceci impliquerait que la même propriété est vraie pour toute stratégie pure du joueur 2 et donc aussi pour toute stratégie mixte du joueur 2. Nous déduirons alors que le joueur 1 peut garantir uniformément $\frac{K}{2(K+1)} - \varepsilon$ pour tout ε.

Fixons donc une suite d'actions $a^2 = \left(a_1^2, \ldots, a_t^2, \ldots\right)$ pour le joueur 2. Remarquons alors que :

– si $a_1^2 = \overline{1}$, alors σ_K^1 coïncide avec σ_{K-1}^1 à partir de l'étape 2,
– si $a_1^2 = \overline{0}$, alors σ_K^1 coïncide avec σ_{K+1}^1 à partir de l'étape 2.

Soit t_* la première étape où le jeu est absorbé (le joueur 1 joue $\overline{1}$). Définissons la variable aléatoire X_T comme suit :

$$X_T = \begin{cases} 1/2 & \text{si } t_* > T \\ 1 & \text{si } t_* \leqslant T, \quad a_{t_*}^2 = \overline{1} \\ 0 & \text{si } t_* \leqslant T, \quad a_{t_*}^2 = \overline{0} \end{cases}$$

Si le jeu s'est arrêté avant ou à l'étape T, X_T représente le paiement d'absorption du joueur 1. Sinon, X_T vaut $1/2$ (l'objectif approximatif du joueur 1). Nous allons montrer par récurrence sur T que

$$\mathbb{E}_{\sigma_K^1, a^2}\left(X_T\right) \geqslant \frac{K}{2(K+1)}.$$

Pour $T = 1$:

– Si $a_1^2 = \overline{1}$, alors

$$\begin{aligned} \mathbb{E}_{\sigma_K^1, a^2}\left(X_T\right) &= \left(1 - \frac{1}{(K+1)^2}\right)\frac{1}{2} + \frac{1}{(K+1)^2} \\ &> \frac{1}{2} \\ &> \frac{K}{2(K+1)}. \end{aligned}$$

– Si $a_1^2 = \overline{0}$, alors

$$\begin{aligned}\mathbb{E}_{\sigma_K^1,a^2}(X_T) &= \left(1 - \frac{1}{(K+1)^2}\right)\frac{1}{2}\\ &= \frac{K(K+2)}{2(K+1)^2}\\ &> \frac{K}{2(K+1)}.\end{aligned}$$

Supposons maintenant que la dernière inégalité est satisfaite pour $T = t_0$ et montrons qu'elle est vraie pour $T = t_0 + 1$.

– Si $a_1^2 = \overline{1}$, alors

$$\begin{aligned}\mathbb{E}_{\sigma_K^1,a^2}(X_{t_0+1}) &= \left(1 - \frac{1}{(K+1)^2}\right) E_{\sigma_{K-1}^1,a^2}(X_{t_0}) + \frac{1}{(K+1)^2}\\ &> \left(1 - \frac{1}{(K+1)^2}\right)\frac{K-1}{2K} + \frac{1}{(K+1)^2}\\ &= \frac{K}{2\,(K+1)}.\end{aligned}$$

– Si $a_1^2 = \overline{0}$, alors

$$\begin{aligned}\mathbb{E}_{\sigma_K^1,a^2}(X_{t_0+1}) &= \left(1 - \frac{1}{(K+1)^2}\right) E_{\sigma_{K+1}^1,a^2}(X_{t_0})\\ &> \left(1 - \frac{1}{(K+1)^2}\right)\frac{K+1}{2(K+2)}\\ &= \frac{K}{2(K+1)}.\end{aligned}$$

Soit $t^* := \min\{t_*, T\}$: c'est l'étape courante si le jeu n'a pas encore été absorbé, sinon c'est l'étape d'absorption. Rappelons que si $k_T = -K$, alors le jeu est absorbé avec probabilité 1. Donc $k_T \geqslant -K$ d'où

$0_T \geqslant (T-K-1)/2$, ce qui implique que :

$$\begin{aligned}
\mathbb{E}_{\sigma^1_K,a^2}\left(\sum_{t=1}^{T}\frac{1}{T}\,g^1\left(\omega_t,a_t\right)\right) &= \mathbb{E}_{\sigma^1_K,a^2}\left(\frac{0_{t^*}+\left(T-t^*\right)1_{a^2_{t^*}=\bar{1}}}{T}\right)\\
&\geqslant \mathbb{E}_{\sigma^1_K,a^2}\left(\frac{\frac{t^*-K-1}{2}+\left(T-t^*\right)1_{a^2_{t^*}=\bar{1}}}{T}\right)\\
&= \mathbb{E}_{\sigma^1_K,a^2}\left(\frac{\frac{t^*}{2}+\left(T-t^*\right)1_{a^2_{t^*}=\bar{1}}}{T}\right)-\frac{K+1}{2T}\\
&= \frac{1}{T}\mathbb{E}_{\sigma^1_K,a^2}\left(\sum_{t=1}^{T}X_t\right)-\frac{K+1}{2T}\\
&\geqslant \frac{K}{2(K+1)}-\frac{K+1}{2T}. \qquad \square
\end{aligned}$$

La stratégie utilisée dans la preuve nécessite l'observation par les joueurs des paiements et de l'état. Sans l'observation des paiements, la valeur uniforme du Big-Match n'existe pas (Coulomb (1992)). Cependant, dans le cas d'un joueur qui n'observe rien (ni état, ni paiements), il peut être démontré qu'une stratégie uniforme existe toujours (Rosenberg, Solan et Vieille (2002)) mais elle n'est pas aussi simple (stationnaire) que la stratégie uniforme de Blackwell (1962) en observation parfaite de l'état.

11. Valeur uniforme

Ici nous étendons le résultat précédent, à savoir l'existence de la valeur uniforme, à tout jeu stochastique à somme nulle.

***Théorème 11.1* (Mertens et Neyman (1981)).** — *Tout jeu stochastique à deux joueurs et à somme nulle admet une valeur uniforme.*

Démonstration. — Comme cela a déjà été montré précédemment, si la valeur uniforme existe, elle est nécessairement égale à $\lim_{\lambda\to 0} v_\lambda(\omega_1) := v\left(\omega_1\right)$. Nous allons montrer que pour tout $\varepsilon > 0$, le joueur 1 peut s'assurer uniformément $v\left(\omega_1\right)-\varepsilon$. Par symétrie, le joueur 2 peut aussi garantir uniformément $v\left(\omega_1\right)+\varepsilon$, d'où le résultat.

Par la section sur l'opérateur de Shapley, nous avons que pour tout taux d'escompte λ, il existe une stratégie optimale stationnaire x^1_λ

pour le joueur 1 (le joueur qui maximise). Donc pour tout $\omega \in \Omega$ et tout $y^2_\lambda \in \Delta\left(A^2(\omega)\right)$ nous avons

$$\lambda g^1(\omega, x^1_\lambda(\omega), y^2_\lambda(\omega)) \\ + (1-\lambda) \sum_{\omega' \in \Omega} q(\omega, x^1_\lambda(\omega), y^2_\lambda(\omega))(\omega') v_\lambda(\omega') \geqslant v_\lambda(\omega).$$

Nous allons construire explicitement une stratégie uniforme ε-optimale pour le joueur 1 comme suit. À chaque étape t, le joueur 1 va jouer la stratégie $x^1_{\lambda_t}(\omega_t)$, où λ_t est une fonction du paiement reçu dans le passé et ω_t est l'état courant. Si le paiement moyen passé est bon, le joueur 1 va augmenter le poids des opportunités futures ($\lambda_{t+1} < \lambda_t$). Dans le cas contraire, il doit essayer d'assurer un bon paiement aujourd'hui ($\lambda_{t+1} > \lambda_t$).

En effet, le taux d'escompte (comme le taux d'intérêt) mesure l'importance du paiement de court terme relativement au paiement de long terme. Dans l'étude de la valeur uniforme, le taux d'escompte peut être arbitrairement petit de telle sorte que les opportunités futures ont un poids très élevé relativement aux opportunités présentes. Cela dit, si le joueur a tout le temps des paiements faibles, il ne va jamais pouvoir utiliser les opportunités futures. Le mécanisme d'adaptation du taux d'escompte dynamiquement en fonction des performances passées permet au joueur de faire proprement son arbitrage entre le paiement aujourd'hui et le paiement demain.

D'après le corollaire 9.3, il existe λ_0 tel que pour tout $0 < \lambda < \lambda' \leqslant \lambda_0$, il existe une fonction positive et intégrable ψ sur $[0, \lambda_0]$ telle que

$$\max_{\omega \in \Omega} \left| v_\lambda(\omega) - v_{\lambda'}(\omega) \right| = \int_\lambda^{\lambda'} \psi(s) ds.$$

Soit $\varepsilon \in]0, M[$ fixé. On définit une fonction $D : [0, \lambda_0] \to \mathbb{R}$ comme suit :

$$D(y) = \frac{12M}{\varepsilon} \int_y^{\lambda_0} \frac{\psi(s)}{s} ds + \frac{1}{\sqrt{y}}$$

Nous allons commencer par montrer quelques propriétés sur D avant de prouver le théorème.

Propriété 1. — *D est décroissante et intégrable ($\int_0^{\lambda_0} D(y) dy < +\infty$).*

Démonstration. — La décroissance est évidente. Le second terme, $\frac{1}{\sqrt{y}}$ est clairement intégrable. Pour le deuxième terme, on intervertit les intégrales ce qui permet d'obtenir :

$$\begin{aligned}\int_{y=0}^{\lambda_0}\int_{s=y}^{\lambda_0}\frac{\psi(s)}{s}dsdy &= \int_{s=y}^{\lambda_0}\int_{y=0}^{\lambda_0}\frac{\psi(s)}{s}dsdy\\ &= \int_{s=0}^{\lambda_0}\psi(s)ds \leqslant 2M.\end{aligned}$$ □

Propriété 2

$$\lim_{y\to 0} D\left(1-\frac{\varepsilon}{6M}y\right)-D(y) = \lim_{y\to 0} D\left(1\right)-D(y+\frac{\varepsilon}{6M}y) = +\infty.$$

Démonstration. — Nous avons :

$$\lim_{y\to 0} D\left(1-\frac{\varepsilon}{6M}y\right)-D(y) = \lim_{y\to 0}\int_{1-\frac{\varepsilon}{6M}y}^{y}\frac{\psi(s)}{s}ds + \frac{1}{\sqrt{1-\frac{\varepsilon}{6M}y}}-\frac{1}{\sqrt{y}}.$$

Le premier terme est positif et le second tend vers $+\infty$. L'autre limite se démontre de la même façon. □

Nous définissons maintenant une deuxième fonction φ sur $[0,\lambda_0]$ comme suit :

$$\varphi(\lambda) = \int_{D(\lambda)}^{+\infty} D^{-1}(y)dy = \int_0^{\lambda} D(y)dy - \lambda D(\lambda)$$

L'égalité plus haut se déduit géométriquement.

Définissons aussi pour tout entier t la variable aléatoire Z_t comme suit :

$$Z_t = v_{\lambda_t}(\omega_t) - \varphi(\lambda_t).$$

Puisque D est intégrable, quand $\lambda_t \to 0$ la quantité $\varphi(\lambda_t)$ approche 0 et donc Z_t approche $v_0(\omega_t)$. Notre but est de construire une stratégie du joueur 1 telle que Z_t soit une sous-martingale. Puisque la stratégie du joueur 1 va être de jouer la stratégie optimale du jeu λ_t-escompté, il nous faut d'abord construire les variables aléatoires λ_t. Nous allons les construire inductivement.

Nous choisissons λ_1 assez petit pour que

(i) $\varphi(\lambda_1) < \varepsilon$;
(ii) $\forall\, y > D(\lambda_1)$, $D\left(1-\frac{\varepsilon}{6M}y\right) - D(y) > 6M$;
(iii) $\forall\, y > D(\lambda_1)$, $D\left(y\right) - D(1+\frac{\varepsilon}{6M}y) > 6M$.

Ceci est possible grâce aux propriétés 1 et 2. Nous définissons $\lambda_{t+1} = D^{-1}(d_{t+1})$ où la suite $\{d_t\}$ est définie comme suit :

$$\begin{aligned} d_1 &= D(\lambda_1) \\ d_{t+1} &= \max\left\{D(\lambda_1), d_t + g^1(\omega_t, a_t) - v_{\lambda_t}(\omega_{t+1}) + 4\varepsilon\right\}. \end{aligned}$$

Puisque $d_t \geqslant D(\lambda_1)$ pour $t \geqslant 1$, nous concluons que $\lambda_t \leqslant \lambda_1$.

Nous allons commencer par étudier quelques propriétés utiles des deux suites.

Propriété 3. — $|d_{t+1} - d_t| \leqslant 6M$.

Démonstration. — Cela découle de sa définition puisque $\left|g^1(\omega_t, a_t)\right|$, $|v_{\lambda_t}(\omega_{t+1})|$ et ε sont bornés par M. □

Propriété 4. — $|\lambda_{t+1} - \lambda_t| \leqslant \frac{\varepsilon\lambda_t}{6M}$.

Démonstration. — Étudions le cas $\lambda_{t+1} \leqslant \lambda_t$ (l'autre cas se montre de la même façon). Donc $d_t \leqslant d_{t+1}$ ce qui implique d'après la propriété 3 que $d_{t+1} \leqslant d_t + 6M$.

Supposons maintenant que $\lambda_{t+1} < \lambda_t - \frac{\varepsilon\lambda_t}{6M}$. En utilisant (ii) nous aurons alors que

$$d_{t+1} - d_t \geqslant D(\lambda_t - \frac{\varepsilon\lambda_t}{6M}) - D(\lambda_t) > 6M,$$

ce qui est une contradiction. □

Nous définissons maintenant la stratégie du joueur 1 comme suit : à l'étape t, connaissant l'état actuel ω_t et observant le paiement de l'étape précédente, le joueur actualise le taux d'escompte à λ_t puis à l'étape t il joue la stratégie optimale $x^1_{\lambda_t}(\omega_t)$ dans le jeu λ_t escompté. Appelons cette stratégie σ^1.

On définit

$$\begin{aligned} C_1 &= \varphi(\lambda_t) - \varphi(\lambda_{t+1}); \\ C_2 &= v_{\lambda_{t+1}}(\omega_{t+1}) - v_{\lambda_t}(\omega_{t+1}); \\ C_3 &= \lambda_t\left(g^1\left(\omega_t, a_t\right) - v_{\lambda_t}(\omega_{t+1})\right). \end{aligned}$$

Rappelons que $Z_t := v_{\lambda_t}(\omega_t) - \varphi(\lambda_t)$ et que

$$\varphi(\lambda) = \int_0^\lambda D(y)dy - \lambda D(\lambda).$$

Propriété 5. — *Pour tout t, nous avons :*

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[Z_{t+1} - Z_t \,|\mathcal{H}_t\right] \geqslant \mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[C_1 + C_2 - C_3 \,|\mathcal{H}_t\right].$$

Démonstration. — Nous avons, par définition de σ^1

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\lambda_t g^1(\omega_t, a_t) + (1-\lambda_t) v_{\lambda_{t+1}}(\omega_{t+1}) \,|\mathcal{H}_t\right] \geqslant v_{\lambda_t}(\omega_t),$$

d'où

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\begin{array}{l}\lambda_t\left(g^1(\omega_t, a_t) - v_{\lambda_t}(\omega_t)\right)\\ +(1-\lambda_t)\left(v_{\lambda_{t+1}}(\omega_{t+1}) - v_{\lambda_t}(\omega_t)\right)|\mathcal{H}_t\end{array}\right] \geqslant 0.$$

Donc

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\begin{array}{l}\lambda_t\left(g^1(\omega_t, a_t) - v_{\lambda_t}(\omega_t)\right)\\ +(1-\lambda_t)\left(v_{\lambda_{t+1}}(\omega_{t+1}) - v_{\lambda_t}(\omega_t)\right)\\ +v_{\lambda_{t+1}}(\omega_{t+1}) - v_{\lambda_{t+1}}(\omega_{t+1})\\ +\varphi(\lambda_t) - \varphi(\lambda_t)\\ +\varphi(\lambda_{t+1}) - \varphi(\lambda_{t+1})\,|\mathcal{H}_t\end{array}\right] \geqslant 0.$$

Nous en déduisons donc que

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[C_3 - C_2 - C_1 + Z_{t+1} - Z_t \,|\mathcal{H}_t\right] \geqslant 0. \qquad \square$$

Nous allons maintenant borner les termes C_1, C_2 et C_3.

Propriété 6. — $C_1 \geqslant \lambda_t \left(d_{t+1} - d_t\right) - \varepsilon\lambda_t$.

Démonstration. — Considérons le cas $\lambda_{t+1} < \lambda_t$ (l'autre cas se traite de la même façon). Dans ce cas on observe géométriquement, en utilisant la définition de φ et la propriété 1, que

$$\varphi(\lambda_t) - \varphi(\lambda_{t+1}) \geqslant \lambda_t \left(d_{t+1} - d_t\right) - \left(\lambda_{t+1} - \lambda_t\right)\left(d_{t+1} - d_t\right),$$

puis nous utilisons les propriétés 3 et 4. $\square$

Propriété 7. — $|C_2| \leqslant \varepsilon\lambda_t$.

Démonstration. — Nous avons :

$$\begin{aligned}|C_2| &= \left|v_{\lambda_{t+1}}(\omega_{t+1}) - v_{\lambda_t}(\omega_{t+1})\right|\\ &\leqslant \left|\int_{\lambda_t}^{\lambda_{t+1}} \psi(y)dy\right|.\end{aligned}$$

La propriété 4 implique alors que pour tout y entre λ_t et λ_{t+1} nous avons,

$$\frac{\lambda_t}{y} \geqslant \frac{1}{2},$$

d'où

$$\begin{aligned}|C_2| &\leqslant \left|\int_{\lambda_t}^{\lambda_{t+1}} \frac{\psi(y)}{y} dy\right| 2\lambda_t \\ &= \lambda_t \left|\frac{\varepsilon}{6M}(d_t - d_{t+1}) + \frac{1}{\sqrt{\lambda_t}} - \frac{1}{\sqrt{\lambda_{t+1}}}\right|.\end{aligned}$$

On observe par ailleurs que l'intégrale $\int_{\lambda_t}^{\lambda_{t+1}} \frac{\psi(y)}{y} dy$ est positive si et seulement si $\lambda_t < \lambda_{t+1}$. Ceci est vrai si et seulement si $\frac{1}{\sqrt{\lambda_t}} > \frac{1}{\sqrt{\lambda_{t+1}}}$. Nous déduisons donc que $|C_2| \leqslant \frac{\varepsilon}{6M}|d_t - d_{t+1}|\lambda_t$. Il nous suffit donc d'appliquer la propriété 3. □

Propriété 8. — $C_3 \leqslant \lambda_t (d_{t+1} - d_t) - 4\varepsilon\lambda_t$.

Démonstration. — Par définition de d_{t+1} nous avons

$$d_{t+1} - d_t \geqslant g^1(\omega, a_t) - v_{\lambda_t}(\omega_{t+1}) + 4\varepsilon,$$

d'où l'inégalité souhaitée. □

Nous en déduisons maintenant la propriété souhaitée.

Propriété 9. — *Z_t est une sous-martingale qui satisfait*

$$\begin{aligned}\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}[Z_t] &\geqslant 2\varepsilon\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\sum_{t=1}^{T}\lambda_t\right] + Z_1 \\ &\geqslant Z_1.\end{aligned}$$

Démonstration. — En utilisant les propriétés 5 à 8 nous obtenons que

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}[Z_{t+1} - Z_t \,|\mathcal{H}_t] \geqslant 2\varepsilon\lambda_t.$$

Puis en sommant nous en déduisons l'inégalité recherchée. □

Nous finissons maintenant la preuve du théorème. Pour cela nous allons montrer que pour toute stratégie σ^2 du joueur 2 nous avons

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left(\sum_{t=1}^{T}\frac{1}{T}g^1(\omega_t, a_t)\right) \geqslant v_0(\omega_1) - 7\varepsilon,$$

pour tout $T \geqslant T_0$ que nous choisirons plus tard.

En remplaçant Z_t par sa valeur $v_{\lambda_t}(\omega_t) - \varphi(\lambda_t)$ dans la propriété 9 (la dernière inégalité) nous obtenons que

$$\begin{aligned}\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[v_{\lambda_t}(\omega_t)\right] &\geqslant v_{\lambda_1}(\omega_1) + \varphi(\lambda_t) - \varphi(\lambda_1)\\ &\geqslant v_{\lambda_1}(\omega_1) - \varphi(\lambda_1)\\ &\geqslant v_{\lambda_1}(\omega_1) - \varepsilon,\end{aligned}$$

où la dernière inégalité utilise (i). Nous en déduisons donc que l'espérance du paiement λ_t-escompté en ω_t est élevée.

Par (i) et par définition de Z_t nous obtenons que $|Z_t| \leqslant 2M$. Ainsi, utilisant la propriété 9, nous en déduisons que

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\sum_{t=1}^{+\infty}\lambda_t\right] \leqslant \frac{M}{\varepsilon},$$

d'où

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\sum_{t=1}^{+\infty}\mathbf{1}_{\{\lambda_t=\lambda_1\}}\right] \leqslant \frac{M}{\varepsilon}.$$

Par définition de d_{t+1} et en utilisant la propriété 3 nous obtenons :

$$d_{t+1} - d_t \leqslant g^1(\omega_t, a_t) - v_{\lambda_t}(\omega_{t+1}) + 4\varepsilon + 6M\mathbf{1}_{\{\lambda_{t+1}=\lambda_1\}},$$

en sommant nous obtenons :

$$\begin{aligned}\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\frac{1}{T}\sum_{t=1}^{T} g^1(\omega_t, a_t)\right] &\geqslant \mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\frac{1}{T}\sum_{t=1}^{T} v_{\lambda_t}(\omega_{t+1})\right]\\ &\quad + \frac{1}{T}\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[d_{T+1} - d_1\right] - 4\varepsilon\\ &\quad - \frac{6M}{T}\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\sum_{t=1}^{+\infty}\mathbf{1}_{\{\lambda_t=\lambda_1\}}\right]\\ &\geqslant v_{\lambda_1}(\omega_1) - \varepsilon - \frac{d_1}{T} - 4\varepsilon - \frac{6M^2}{T\varepsilon},\end{aligned}$$

car $d_{T+1} \geqslant 0$. Il nous suffit maintenant pour conclure de choisir $T \geqslant T_0$ avec

$$T_0 = \max\left\{\frac{d_1}{\varepsilon}, \frac{6M^2}{\varepsilon^2}\right\}. \qquad \square$$

Remarque. — Nous avons prouvé par la même occasion que

$$\mathbb{E}_{\omega_1,\sigma^1,\sigma^2}\left[\liminf \frac{1}{T}\sum_{t=1}^{T} g^1(\omega_t, a_t)\right] \geqslant v_{\lambda_1}(\omega_1) - 7\varepsilon.$$

ce qui montre que tous les exemples de l'approche infinie ont une valeur qui coïncide avec la valeur uniforme.

12. Paris Match

Ce jeu a été introduit et étudié explicitement par Sorin (1986) et il appartient à la famille du Big-Match. L'exemple est célèbre car il montre que dans les jeux stochastiques à somme non nulle, il n'y a aucun lien entre l'approche asymptotique (l'ensemble des paiements d'équilibres de Nash quand le taux d'escompte tend vers zéro) et l'approche uniforme (l'ensemble des paiements d'équilibres de Nash uniformes). Ceci montre aussi la différence fondamentale entre les jeux répétés sans aléas (section 3 du le texte sur les jeux répétés (ce volume)) et les jeux stochastiques. L'exemple de Sorin (1986) est le suivant :

	G	D
H	(1,0)	(0,1)
B	(0,2)*	(1,0)*

Ainsi, en gardant seulement les paiements du joueur 1, nous retrouvons le Big-Match étudié par Blackwell et Ferguson (1968).

Soit E_λ l'ensemble des paiements d'équilibres de Nash du Paris-Match λ-escompté, E_T l'ensemble des paiements d'équilibres de Nash du Paris-Match répété T fois et enfin E_0 l'ensemble des paiements d'équilibres uniformes.

***Théorème 12.1* (Sorin (1986)).** — *E_λ et E_T sont disjoints de E_0. Plus précisément, pour tout λ et T nous avons :*

(a) $E_\lambda = \{(\frac{1}{2}, \frac{2}{3})\}$;
(b) $E_T = \{(\frac{1}{2}, \frac{2}{3})\}$;
(c) $E_0 = \{\alpha(\frac{1}{2}, 1) + (1-\alpha)(\frac{2}{3}, \frac{2}{3}); \alpha \in [0,1]\}$.

Démonstration. — Nous donnons seulement l'idée de la preuve en suivant Thuijsman au chapitre 12 dans Neyman et Sorin (2003). Pour une preuve complète veuillez consulter Sorin (1986).

Étape 1. — Nous prouvons (a), (b) se prouve de la même manière.

Remarquons que tout paiement d'équilibre doit donner au moins $\frac{1}{2}$ au joueur 1 et doit donner au moins $\frac{2}{3}$ au joueur 2. En effet, chaque joueur peut garantir au moins son montant (le joueur 2 en jouant

$\frac{1}{3}G + \frac{2}{3}D$ à chaque étape et le joueur 1 en jouant comme dans le Big-Match escompté $\frac{\lambda}{1+\lambda}H + \frac{1}{1+\lambda}B$).

Soit w le paiement d'équilibre le plus favorable au joueur 2 dans E_λ. Nous allons prouver que $w \leqslant \frac{2}{3}$ (et donc est égal à $\frac{2}{3}$).

Soit $\left(\sigma^1, \sigma^2\right)$ un équilibre dans Γ_λ tel que $g_\lambda^2\left(\sigma^1, \sigma^2\right) = w$. Soit w_1 (resp. w_2) le paiement de continuation sous $\left(\sigma^1, \sigma^2\right)$ si à l'étape 1 la paire (H, G) a été jouée (resp. (H, D) a été jouée).

Remarquez que w_1 et w_2 doivent aussi correspondre à des paiements d'équilibres dans Γ_λ pour le joueur 2 (car sinon le joueur 2 pourrait dévier). Donc on a :

$$w_1 \leqslant w \text{ et } \ w_2 \leqslant w.$$

Soit maintenant p la probabilité avec laquelle le joueur 1 joue B à la première étape sous σ^1 et soit q la probabilité avec laquelle le joueur 2 joue D sous σ^2. On vérifie facilement que si $\left(\sigma^1, \sigma^2\right)$ est un équilibre alors p et q sont dans $]0, 1[$. Nous en déduisons alors que le joueur 2 est indifférent entre ces deux actions à la première étape, d'où

$$\begin{aligned} w &= 2p + (1-p)(1-\lambda)w_1 \\ &= 0p + (1-p)(\lambda + (1-\lambda)w_2) \end{aligned}$$

La ligne 2 (resp. 3) correspond à l'espérance de gain si G est jouée (resp. si D est jouée). Puisque $w_1 \leqslant w$ et $w_2 \leqslant w$ nous obtenons :

$$\begin{aligned} &w \leqslant 2p + (1-p)(1-\lambda)w \\ &w \leqslant (1-p)(\lambda + (1-\lambda)w). \end{aligned}$$

La deuxième inégalité devient

$$2 - w \geqslant (1-p)(2 - (1-\lambda)w),$$

et en utilisant la première inégalité nous trouvons

$$(2-w)\,(1-p)(\lambda + (1-\lambda)w) \geqslant (1-p)(2 - (1-\lambda)w)w$$

qui est équivalente à $w \leqslant \frac{2}{3}$.

Un calcul similaire montre que le paiement maximal possible pour le joueur 1 n'excède pas $\frac{1}{2}$.

Étape 2. — Pour prouver que $E_0 \subset \{\alpha(\frac{1}{2}, 1) + (1-\alpha)(\frac{2}{3}, \frac{2}{3}) \mid \alpha \in [0,1]\}$ Sorin (1986) donne l'argument suivant :

L'idée de la preuve est très simple : si la probabilité d'obtenir un paiement absorbant sur le chemin d'équilibre est moins que 1, alors après un certain temps le joueur 1 joue essentiellement l'action H; les paiements de continuations correspondants à partir de cette étape ne sont pas individuellement rationnels, d'où la contradiction.

Le lecteur est prié de consulter Sorin (1986) pour la traduction mathématique de cet argument.

Étape 3. — Nous prouvons que $\{\alpha(\frac{1}{2}, 1)+(1-\alpha)(\frac{2}{3}, \frac{2}{3}) \mid \alpha \in [0,1]\} \subset E_0$.

L'argument se comprend facilement à l'aide d'un exemple. Prenons par exemple le couple de paiements $\left(\frac{7}{12}, \frac{10}{12}\right)$ et expliquons la méthode de Sorin pour construire un ε-équilibre uniforme. Considérons le jeu auxiliaire à somme nulle suivant :

	G	D
H	$\frac{7}{12}$	$-\frac{5}{12}$
B	$-\frac{7}{12}{}^*$	$\frac{5}{12}{}^*$

Soit σ_ε^1 une stratégie uniformément ε-optimale pour le joueur 1 dans le jeu auxiliaire, similaire à celle de Blackwell et Ferguson 1968. Soit x^2 la stratégie stationnaire du joueur 2 qui joue $\frac{5}{12}G + \frac{7}{12}D$ à chaque étape. Sous $(\sigma_\varepsilon^1, x^2)$ le jeu est absorbé avec probabilité 1 et induit donc un vecteur de paiements dans le vrai jeu exactement égal à $\left(\frac{7}{12}, \frac{10}{12}\right)$. Donc, σ_ε^1 est une meilleure réponse du joueur 1 dans le Paris-Match face à x^2 car sans absorption (en jouant H) le joueur 1 gagne seulement $\frac{5}{12}$. En utilisant la définition de la stratégie du joueur 1, il est possible de montrer que x^2 est uniformément ε-optimale face à σ_ε^1 dans le Paris Match (voir Sorin (1986)). $\square$

13. Extensions

La construction d'une stratégie maxmin uniforme en cas d'observation imparfaite des actions a été faite d'une manière indépendante par Coulomb (2003) et Rosenberg, Solan et Vieille (2003).

Vieille (2000 a,b) a généralisé remarquablement le résultat de Mertens et Neyman (1982) aux jeux à deux joueurs et à somme non nulle en prouvant l'existence d'un paiement d'équilibre uniforme. Sa preuve, en deux étapes, consiste à réduire d'abord le problème d'existence à la classe des jeux récursifs (jeux dans lesquels les joueurs ont une fonction de paiement nulle dans tous les états non absorbants). Les jeux d'arrêts présentent la particularité d'être à la fois récursifs et absorbants d'où leur intérêt.

Exception faite de quelques cas particuliers, comme les jeux absorbants à trois joueurs résolus positivement par Solan (1999) qui prouve l'existence d'un paiement d'équilibre uniforme, le problème pour les jeux à plus de deux joueurs reste entièrement ouvert. Solan et Vieille (2003) ont exhibé un jeu d'arrêt à quatre joueurs qui prouve que pour montrer l'existence d'un paiement d'équilibre uniforme, il faut nécessairement inventer une nouvelle classe de stratégies, différentes de celles utilisées habituellement (comme celle dans la preuve de Mertens-Neyman 1981).

Plusieurs personnes conjecturent la non-existence d'un paiement d'équilibre uniforme pour tout jeu stochastique à n joueurs. Pour prouver cela, il y a deux possibilités : trouver un contre exemple ou prouver que l'existence induit un paradoxe mathématique. La non-existence a été prouvée en temps continu. Laraki, Solan et Vieille (2004) ont exhibé un jeu d'arrêt simple à trois joueurs qui n'admet pas d'équilibre approché, que ça soit au sens faible (escompté) ou fort (uniforme). Le même exemple en temps discret admet un équilibre stationnaire pour le jeu escompté (par Fink (1964)) et un paiement d'équilibre uniforme (par Solan (1999)).

La classe la plus simple à étudier est certainement celle des jeux d'arrêts. Dans un jeu d'arrêt, chaque joueur a le choix entre A (arrêter) ou C (continuer). Tant que le jeu n'est pas arrêté, le gain d'étape est zéro pour chaque joueur. La première fois qu'un ou plusieurs joueurs choisissent l'action A le jeu est absorbé. Le paiement d'absorption de chaque joueur dépend de la coalition des joueurs S qui ont arrêté le jour de l'absorption. À part une sous classe résolue positivement par Solan et Vieille (2001), le problème de l'existence d'un paiement d'équilibre uniforme reste entièrement ouvert (peut être pourriez vous le résoudre ?).

L'existence de la valeur dans les jeux escomptés a été généralisée à des espaces d'états et d'actions plus généraux par Maitra et Parthasarathy (1970) et d'autres auteurs jusqu'au résultat très général de Nowak (1985).

Une façon très naturelle d'étudier v_λ et v_T consiste à étudier les propriétés mathématiques de l'opérateur de Shapley et de sa dérivée pour en déduire la convergence et une caractérisation de la limite. Cette approche a été utilisée pour la première fois par Kohlberg (1974) dans le cadre restreint des jeux absorbants et a été définie et étudiée d'une manière générale pour les jeux stochastiques à somme nulle par Rosenberg et Sorin (2001). Elle a été appliquée par les deux auteurs pour étudier les jeux absorbants avec des espaces d'actions compacts mais aussi pour l'étude des jeux répétés à information incomplète des deux côtés (*cf.* texte sur les jeux répétés (ce volume)). La même approche permet à Rosenberg (2000) de montrer l'existence et la caractérisation variationnelle de la valeur asymptotique des jeux absorbants et information incomplète d'un côté.

Nous invitons les lecteurs intéressés par d'autres résultats ou extensions à consulter le livre sur les jeux stochastiques édités par Neyman et Sorin (2003), le livre de Sorin (2002) ou encore l'ouvrage de référence sur les jeux répétés par Mertens Sorin et Zamir (1994). Pour une étude détaillée de l'approche infinie (avec des fonctions de paiements mesurables définies sur l'ensemble des parties comme les liminf et limsup des sommes de Césaro mais aussi d'autres objectifs comme atteindre ou éviter un sous ensemble mesurable de H_∞) vous pouvez consulter l'ouvrage de référence de Maitra et Sudderth (1996).

Bibliographie

Blackwell (D.)
[1962] Discrete dynamic programming, *Annals of Mathematical Statistics*, 33 (1962), p. 719–726.

Blackwell (D.) & Ferguson (T.)
[1968] The Big Match, *Annals of Mathematical Statistics*, 33 (1968), p. 882–886.

Benedetti (R.) & Risler (J.-J.)
[1990] *Real algebraic and semi-algebraic sets*, Paris : Hermann, 1990.

Bewley (T.) & Kohlberg (E.)

[1976a] The asymptotic theory of stochastic games, *Mathematics of Operation Research*, 1 (1976), p. 197–208.

[1976b] The asymptotic solution of a recursion equation occuring in stochastic games, *Mathematics of Operation Research*, 1 (1976), p. 321–336.

Coulomb (J.M.)

[1992] Repeated games with absorbing states and no signals, *International Journal of Game Theory*, 21 (1992), p. 161–174.

[2001] Repeated games with absorbing states and signaling structure, *Mathematics of Operation Research*, 26 (2001), p. 286–303.

[2003] Stochastic games without perfect monitoring, *International Journal of Game Theory*, 32 (2003), p. 723–96.

Fink (A.M.)

[1964] Equilibrium in a stochastic n-person game, *J. Sci. Hiroshima Univ.*, 28 (1964), p. 89–93.

Foster (O.)

[1981] *Lectures on Riemann surfaces*, Springer, 1981.

Kohlberg (E.)

[1974] Repeated games with absorbing states, *Annals of Statistics*, 2 (1974), p. 724–738.

Laraki (R.)

[2001] Variational inequalities, system of functional equations, and incomplete information repeated games, *SIAM Journal of Control and Optimization*, 40(2) (2001), p. 516–524.

[2002] Repeated games with lack information on one side : the dual differential approach, *Mathematics of Operation Research*, 27(2) (2002), p. 419–440.

Laraki (R.), Solan (E.) & Vieille (N.)

[2005] Continuous-time games of timing, *Journal of Economic Theory*, 120 (2005), p. 206–238.

Laraki (R.)

[2006] A variational approach in zero-sum absorbing games, 2006 ; preprint.

Maitra (A.) & Parthasarathy (T.)

[1970] On stochastic games, *Journal of Optimization Theory and Applications*, 5 (1970), p. 289–300.

Maitra (A.) & Sudderth (W.)

[1996] *Discrete gambling and stochastic games*, Springer, 1996.

Mertens (J.-F.) & Neyman (A.)

[1981] Stochastic games, *International Journal of Game Theory*, 10 (1981), p. 53–66.

Mertens (J.-F.), Sorin (S.) & Zamir (S.)

[1994] Repeated games, CORE D.P., 1994, p. 9420–21–22.

Neyman (A.) & Sorin (S.)
[2003] *Stochastic games and applications*, NATO Science Series, Kluwer Academic Publishers, 2003.

Nowak (A.S.)
[1985] Universally measurable strategies in zero-sum stochastic games, *Annals of Probability*, 13 (1985), p. 269–287.

Rosenberg (D.)
[2000] Zero-sum absorbing games with incomplete information on one side : asymptotic analysis, *SIAM Journal of Control and Optimization*, 39 (2000), p. 208–225.

Rosenberg (D.) & Sorin (S.)
[2001] An operator approach to zero-sum repeated games, *Israël Journal of Mathematics*, 121 (2001), p. 221–246.

Rosenberg (D.), Solan (E.) & Vieille. (N.)
[2002a] Blackwell Optimality in Markov Decision Processes with Partial Observation, *Annals of Statistics*, 30 (2002), p. 1178–1193.
[2002b] The MaxMin value of stochastic games with imperfect monitoring, *International Journal of Game Theory*, 32 (2002), p. 133–150.

Solan (E.)
[1999] Three-Player absorbing games, *Mathematics of Operation Research*, 24 (1999), p. 669–698.

Solan (E.) & Vieille (N.)
[2001] Quitting games, *Mathematics of Operation Research*, 26 (2001), p. 265–285.
[2003] Quitting games – an example, *International Journal of Game Theory*, 31 (2003), p. 365–381.

Solan (E.)
[2006] Stochastic games, 2006 ; a non-published master course.

Sorin (S.)
[1986] Asymptotic properties of a non zero-sum stochastic game, *International Journal of Game Theory*, 15 (1986), p. 101–107.
[2002] *A First course on zero-sum repeated games*, Springer-Verlag, 2002.

Shapley (L.S.)
[1953] Stochastic games, *Proceedings of the National Academy of Sciences of the U. S. A.*, 39 (1953), p. 1095–1100.

Takahashi (M.)
[1964] Equilibrium points of stochastic non-cooperative n-person game, *J. Sci. Hiroshima Univ.*, 28 (1964), p. 95–99.

Vieille (N.)
[2000a] Two-player stochastic games I : a reduction, *Israël Journal of Mathematics*, 119 (2000), p. 55–91.

[2000b] Two-player stochastic games II : the case of recursive games, *Israël Journal of Mathematics*, 119 (2000), p. 93–126.

R. Laraki, CNRS, Laboratoire d'Économétrie, École polytechnique
E-mail : Rida.Laraki@shs.polytechnique.fr