Exposé de première année : Marches Aléatoires Contrôlées

Nicolas Coudray   Aymeric Dieuleveut

Encadrant : Gilles Stoltz

Table des matières

1 Stratégies déterministes

Avant tout, donnons quelques définitions :

Nous considérons ici un jeu entre deux joueurs. On a deux ensembles finis d'actions $\mathcal{A}$ pour le joueur 1 et $\mathcal{B}$ pour le joueur 2. A chaque tour, les joueurs choisissent un élément de $\Delta(\mathcal{A})$ et $\Delta(\mathcal{B})$ (ensembles des lois de probabilité sur $\mathcal{A}$ et $\mathcal{B}$, identifiés aux ensembles des vecteurs de $\mathbb{R}^{\mathcal{A}}$ et $\mathbb{R}^{\mathcal{B}}$ à coordonnées positives et dont la somme des coordonnées vaut 1). La fonction de gain du joueur 1 est $m : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$ (vue comme une matrice, $[m(i,j)]_{i,j}$), prolongée linéairement à $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$ par la formule $m(p,q) = \displaystyle\sum_{i,j} p_i q_j m(i,j)$.
(Par abus de notation, on notera parfois $m(i,q) := m(\delta_i, q) := \sum_j q_j m(i,j)$ et $m(p,j) := m(p, \delta_j)$.)

Par exemple, si on considère le jeu "pierre - feuille - ciseaux", la matrice des gains sera :

	Pierre	Feuille	Ciseaux
Pierre	0	-1	1
Feuille	1	0	-1
Ciseaux	-1	1	0

On a $\mathcal{A} = \mathcal{B} = \{Pierre, Feuille, Ciseaux\}$ et le gain du joueur 1 s'il joue "Feuille" et que son adversaire 2 joue "Pierre" est 1.
Dans le cas des stratégies déterministes, le joueur 1 peut choisir de jouer le vecteur $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ et dans ce cas il est assuré que son gain sera nul.

Si on autorise au joueur 1 le "puits", qui bat les ciseaux et la pierre , la matrice devient alors :

	Pierre	Feuille	Ciseaux
Pierre	0	-1	1
Feuille	1	0	-1
Ciseaux	-1	1	0
Puits	1	-1	1

Les gains que l'on considère par la suite sont, contrairement à l'exemple, des vecteurs de $\mathbb{R}^d$.

Définition 1. *Dans le cas des stratégies déterministes, les joueurs choisissent $(p_t, q_t) \in \Delta(\mathcal{A}) \times \Delta(\mathcal{B})$. Le gain reçu est $m(p_t, q_t)$. A la fin de chaque tour, q_t est révélé au joueur 1.*

On appelle stratégie (pour le joueur 1) une suite de fonctions $f_T : (\Delta(\mathcal{A}) \times \Delta(\mathcal{B}))^{(T-1)} \mapsto \Delta(\mathcal{A})$, qui à partir des choix passés (p_t, q_t pour $t \leqslant T-1$) donne le coup suivant, p_T pour le joueur 1 (ou $q_T \in \Delta(\mathcal{B})$ pour le joueur 2).

On pose alors D_T distance entre le gain moyen à l'instant $T (\in \mathbb{N})$ et un ensemble C :

$$D_T := \inf_{c \in C} \left\| c - \frac{1}{T} \sum_{t=1}^{T} m(p_t, q_t) \right\|_2 .$$

Définition 2. *Avec ces notations, on dira qu'un ensemble $C \subseteq \mathbb{R}^d$ est m-approchable par des stratégies déterministes s'il existe une stratégie pour le joueur 1 telle que pour toute stratégie du joueur 2,*

$$\lim_{T \to \infty} D_T = 0 \ .$$

Remarquons qu'un ensemble est approchable si et seulement si son adhérence l'est ($d(x, C) = d(x, \bar{C})$ car la distance d'un point x à un ensemble C est atteinte sur l'adhérence de ce dernier). Aussi, nous ne nous intéresserons dans la suite qu'à l'approchabilité des ensembles fermés.

Nous allons maintenant démontrer le théorème suivant (dû à Blackwell) :

Théorème 1.
Soit C un convexe fermé. Alors, en reprenant les notations précédentes, C est approchable si et seulement si

$$\forall q \in \Delta(\mathcal{B}), \exists p \in \Delta(\mathcal{A}), \qquad m(p,q) \in C.$$

Une stratégie possible d'approchabilité est la suivante : on choisit p_1 arbitrairement, puis à chaque tour, le joueur 1 joue un coup p_T qui résout le problème suivant :

$$\min_{p\in\Delta(\mathcal{A})} \max_{q\in\Delta(\mathcal{B})} < \overline{m}_T - \pi_T, m(p,q) > ,$$

avec :

- $\overline{m}_T := \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} m(p_t, q_t)$,
- *π_T la projection orthogonale de $\overline{m}_T$ sur C ($\pi_T = \overline{m}_T$ si $\overline{m}_T \in C$),*
- *$< .,. >$ le produit scalaire euclidien sur $\mathbb{R}^d$.*

De plus, si l'on note $M = \sup_{(i,j)\in\mathcal{A}\times\mathcal{B}} \|m(i,j)\|_2$, alors on a une estimation de la vitesse de convergence en suivant cette stratégie :

$$D_T \leqslant \frac{2M}{\sqrt{T}}.$$

Avant de donner la démonstration, on donne une interprétation simple et des exemples dans le cas ou les gains sont des réels et non des vecteurs.

Le théorème signifie que si $d = 1$, il existe deux valeurs limites $Mm = \max_{q\in\mathcal{B}} \min_{p\in\mathcal{A}} m(p,q)$ et $mM = \min_{q\in\mathcal{B}} \max_{p\in\mathcal{A}} m(p,q)$ telles qu'un convexe $C := [v; V]$ de $\Re$ est appprochable si et seulement si $mM \geqslant v$ et $Mm \leqslant V$.

En effet, on peut vérifier simplement que la condition est nécessaire, même sans le théorème : en reprenant les notations ci-dessus,

1. si on a $Mm > V$ alors $\exists q_0 \in \mathcal{B}, \quad \min_{p\in\mathcal{A}} m(p,q) > V + \varepsilon$ c'est-à-dire $\exists q_0 \in \mathcal{B}, \forall p \in \mathcal{A}, \quad m(p,q) > V + \varepsilon$, donc si le joueur 2 joue la stratégie qui consiste à , quel que soit le passé, jouer q_0, le gain moyen sera strictement supérieur à $V + \varepsilon$.
2. et de même si $mM < V$ alors $\exists q_0 \in \mathcal{B}, \quad \max_{p\in\mathcal{A}} m(p,q) < V - \varepsilon$ c'est-à-dire $\exists q_0 \in \mathcal{B}, \forall p \in \mathcal{A}, \quad m(p,q) < V - \varepsilon$, donc si le joueur 2 joue la stratégie qui consiste à , quel que soit le passé, jouer q_0, le gain moyen sera strictement inférieur à $v - \varepsilon$.

Dans les deux cas, C n'est pas approchable.

Réciproquement, si on a cette condition, on a la condition du théorème : en effet on peut remarquer que dans un cadre tout à fait général (indépendant de d), la convexité de C permet de se contenter de l'hypothèse :

$$\forall q \in \mathcal{B}, \exists p \in \Delta(\mathcal{A}), \qquad m(p,q) \in C,$$

et on a avec notre hypothèse que $\forall q \in \mathcal{B}, \qquad \min_{p\in\mathcal{A}} m(p,q) \leqslant V$ et $\max_{p\in\mathcal{A}} m(p,q) \geqslant v$.
Donc $\forall q \in \mathcal{B}, \exists p \in \Delta(\mathcal{A}), \qquad m(p,q) \in C$.

On peux donner briévement deux exemples à partir de jeux à deux actions :
Cas 1 :

	p_1	p_2
q_1	0	1
q_2	4	3

Ici, $Mm = \max(0,3) = 3$ et $mM = \min(1,4) = 1$ un convexe C est donc approchable si et seulement si : $[1;3] \subset C$.

Cas 2 :

	p_1	p_2
q_1	0	5
q_2	4	3

Ici, $Mm = \max(0,3) = 3$ et $mM = \min(5,4) = 4$ un convexe $C = [v;V]$ est donc approchable si et seulement si : $v \leqslant 4$ et $3 \leqslant V$, c'est à dire si et seulement s'il contient un élément de [3 ;4].

On passe maintenant à la démonstration du théorème, qui va se faire par double implication.

Démonstration. L'idée de la démonstration du sens indirect provient essentielement de l'article de Blackwell [1] (p.6). En remarquant que si

$$\exists q_0 \in \Delta(\mathcal{B}), \forall p, \qquad m(p,q_0) \notin C,$$

alors en considérant

$$R(q_0) := \{m(p,q_0) \ / \ p \in \Delta(\mathcal{A})\} = \left\{ \sum_i p_i m(i,q_0) \ / \ p = (p_i)_{i\in\mathcal{A}} \in \Delta(\mathcal{A}) \right\}$$

l'enveloppe convexe des $m(i,q_0)$, alors pour le jeu associé à la matrice ${}^t m$, on a

$$\forall p, \exists q (= q_0), \qquad {}^t m(q,p) \in R(q_0),$$

$R(q_0)$ étant un convexe fermé d'intersection vide avec C. Si l'on démontre le sens direct, on aura alors $R(q_0)$ approchable par le joueur 2. Or cela signifie par définition qu'il existe une stratégie du joueur 2 telle que pour toute stratégie du joueur 1, $d\left(\frac{1}{T}\sum_{t=1}^T m(p_t,q_t), R(q_0)\right) \to 0$. On utilise ensuite le fait que $d\left(R(q_0), C\right) := \inf_{x\in R(q_0)} d\left(x, C\right) = \inf_{x\in R(q_0), y\in C} d\left(x,y\right) > 0$ (car $R(q_0)$ et C sont disjoints, $R(q_0)$ est compact et C fermé) et par inégalité triangulaire, si $a \in C$ et $b \in R(q_0)$, on sait que

$$d\left(\frac{1}{T}\sum_{t=1}^T m(p_t,q_t), a\right) + d\left(\frac{1}{T}\sum_{t=1}^T m(p_t,q_t), b\right) \geqslant d(a,b),$$

et donc en passant à l'inf sur C et $R(q_0)$,

$$d\left(\frac{1}{T}\sum_{t=1}^T m(p_t,q_t), C\right) + d\left(\frac{1}{T}\sum_{t=1}^T m(p_t,q_t), R(q_0)\right) \geqslant d\left(C, R(q_0)\right),$$

d'où

$$\liminf_{T\to\infty} d\left(\frac{1}{T}\sum_{t=1}^{T} m(p_t,q_t), C\right) \geqslant d(R(q_0),C) > 0,$$

et donc C n'est pas approchable pour le joueur 1.

Il ne reste donc plus qu'à montrer le sens direct. On utilise la stratégie donnée dans l'énoncé du théorème (la demonstration qui suit provient de l'article [4]).

Par le théorème du minimax (démontré en annexe) appliqué à la matrice $A := [< \overline{m}_T - \pi_T, m(i,j) >]_{i,j}$, on sait que le problème "p_T résout $\min\limits_{p\in\Delta(\mathcal{A})} \max\limits_{q\in\Delta(\mathcal{B})} < \overline{m}_T - \pi_T, m(p,q) >$" proposé pour construire une stratégie admet une solution et que :

$$\min_{p\in\Delta(\mathcal{A})}\max_{q\in\Delta(\mathcal{B})} < \overline{m}_T - \pi_T, m(p,q) >= \max_{q\in\Delta(\mathcal{B})}\min_{p\in\Delta(\mathcal{A})} < \overline{m}_T - \pi_T, m(p,q) > .$$

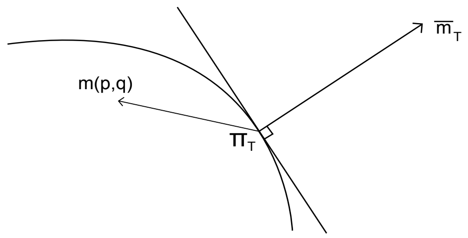
Or on a :

$$\forall q, \exists p, \qquad m(p,q) \in C \ ,$$

et donc, par les propriétés de projection sur les convexes,

$$\forall q, \exists p, \qquad < \overline{m}_T - \pi_T, m(p,q) - \pi_T > \leqslant \ 0$$

(voir par exemple le théorème 6.32 de [7] dans le cas d'un produit scalaire réel).



On a donc prouvé :

$$\min_{p\in\Delta(\mathcal{A})}\max_{q\in\Delta(\mathcal{B})} < \overline{m}_T - \pi_T, m(p,q) - \pi_T >= \max_{q\in\Delta(\mathcal{B})}\min_{p\in\Delta(\mathcal{A})} < \overline{m}_T - \pi_T, m(p,q) - \pi_T > \leqslant \ 0 \ .$$

En particulier puisque p_{T+1} est défini comme solution du problème $\min\limits_{p\in\Delta(\mathcal{A})}\max\limits_{q\in\Delta(\mathcal{B})} < \overline{m}_T - \pi_T, m(p,q) >$ on a la relation :

$$\forall q, \quad < \overline{m}_T - \pi_T, m(p_{T+1},q) - \pi_T >\leqslant 0 \ ,$$

et donc en particulier :

$$< \overline{m}_T - \pi_T, m(p_{T+1},q_{T+1}) - \pi_T >\leqslant 0 \ , \tag{1}$$

(qui reste vrai dans le cas $\overline{m}_T \in C$).

Un calcul algébrique donne alors :

$$\begin{aligned}
\overline{m}_{T+1} &= \frac{1}{T+1}\sum_{\tau=1}^{T+1} m(p_\tau,q_\tau) \\
&= \frac{1}{T+1} m(p_{T+1},q_{T+1}) + \frac{1}{T}\sum_{\tau=1}^{T} m(p_\tau,q_\tau) - \frac{1}{T(T+1)}\sum_{\tau=1}^{T} m(p_\tau,q_\tau) \\
&= \overline{m}_T + \frac{1}{T+1}(m(p_{T+1},q_{T+1}) - \overline{m}_T) \ .
\end{aligned}$$

Cela permet de déterminer une inégalité sur D_T :

$$\begin{aligned}
D_{T+1}^2 &= \inf_{c\in C}\left\|c-\frac{1}{T+1}\sum_{t=1}^{T+1}m(p_t,q_t)\right\|_2^2 \\
&= \|\pi_{T+1}-\overline{m}_{T+1}\|_2^2 \\
&\leqslant \|\pi_T-\overline{m}_{T+1}\|_2^2 \\
&= \left\|(\pi_T-\overline{m}_T)-\frac{1}{T+1}(m(p_{T+1},q_{T+1})-\overline{m}_T)\right\|_2^2,
\end{aligned}$$

où l'on a utilisé pour l'inégalité le fait que $\pi_T\in C$. Puis en développant :

$$D_{T+1}^2 \leqslant \|\pi_T-\overline{m}_T\|_2^2+\frac{2}{T+1}<\overline{m}_T-\pi_T, m(p_{T+1},q_{T+1})-\overline{m}_T> \\ +\frac{\|(m(p_{T+1},q_{T+1})-\overline{m}_T)\|_2^2}{(T+1)^2},$$

et en ajoutant et soustrayant $\frac{2}{T+1}<\overline{m}_T-\pi_T, m(p_{T+1},q_{T+1})-\pi_T>$ on fait apparaître $D_T=\|\pi_T-\overline{m}_T\|_2^2$:

$$D_{T+1}^2 \leqslant \|\pi_T-\overline{m}_T\|_2^2+\frac{2}{T+1}<\overline{m}_T-\pi_T, \pi_T-\overline{m}_T> \\ +\frac{2}{T+1}<\overline{m}_T-\pi_T, m(p_{T+1},q_{T+1})-\pi_T>+\frac{\|(m(p_{T+1},q_{T+1})-\overline{m}_T)\|_2^2}{(T+1)^2}$$

$$D_{T+1}^2 \leqslant \|\pi_T-\overline{m}_T\|_2^2-\frac{2}{T+1}\|\overline{m}_T-\pi_T\|_2^2 \\ +\frac{2}{T+1}<\overline{m}_T-\pi_T, m(p_{T+1},q_{T+1})-\pi_T>+\frac{\|(m(p_{T+1},q_{T+1})-\overline{m}_T)\|_2^2}{(T+1)^2}$$

$$D_{T+1}^2 \leqslant (1-\frac{2}{T+1})D_T^2+\frac{2}{T+1}<\overline{m}_T-\pi_T, m(p_{T+1},q_{T+1})-\pi_T> \\ +\frac{\|(m(p_{T+1},q_{T+1})-\overline{m}_T)\|_2^2}{(T+1)^2}.$$

Or le produit scalaire est négatif d'après (1). De plus, par inégalité triangulaire,

$$\|m(p_{T+1},q_{T+1})-\overline{m}_T\|_2\leqslant 2M.$$

Finalement on a :

$$D_{T+1}^2\leqslant(1-\frac{2}{T+1})D_T^2+\frac{4M^2}{(T+1)^2}.$$

On démontre alors par récurrence que $D_T^2\leqslant\frac{4M^2}{T}$ pour tout T strictement positif.

Initialisation : On sait que π_1 réalise $\min_{x\in C}\|x-m(p_1,q_1)\|_2^2$, et donc si l'on considère p_1' un élément quelconque tel que $m(p_1',q_1)\in C$ (qui existe par hypothèse)

$$D_1^2:=\|\pi_1-m(p_1,q_1)\|_2^2\leqslant\|m(p_1',q_1)-m(p_1,q_1)\|_2^2\leqslant 4M^2\ .$$

Hérédité :

$$\begin{aligned}
\left(1-\frac{2}{T+1}\right)\frac{1}{T}+\frac{1}{(T+1)^2} &= \frac{(T+1)^2-2(T+1)+T}{T(T+1)^2}\\
&= \frac{T^2-1+T}{T(T+1)^2}\\
&\leqslant \frac{T^2+T}{T(T+1)^2}\\
&\leqslant \frac{1}{T+1}\ ,
\end{aligned}$$

et le résultat en découle alors immédiatement, ce qui achève la preuve. $\square$

2 Stratégies aléatoires

On se place à présent dans le cadre d'un jeu où un facteur aléatoire va apparaître : dans le cas des stratégies aléatoires, les joueurs 1 et 2 choisissent à chaque instant t une loi de probabilité p_t et q_t dans les ensembles $\Delta(\mathcal{A})$ et $\Delta(\mathcal{B})$, mais cette fois-ci ils tirent leurs actions $I_t \in \mathcal{A}$ (respectivement $J_t \in \mathcal{B}$) selon les lois p_t (respectivement q_t). A la fin du tour, l'action J_t du joueur 2, et donc le gain du joueur 1 $m(I_t, J_t)$ sont révélés au joueur 1.

Dans cette situation, I_t (resp. J_t) est une variable aléatoire à valeurs dans $\mathcal{A}$ (resp. $\mathcal{B}$) de loi conditionnelle à $\mathcal{F}_{t-1}$ p_t (resp. q_t).

Dans toute la suite on notera $\mathcal{F}_T := \sigma\left(\{I_t, J_t \ / \ 1 \leqslant t \leqslant T\} \bigcup \{J_{T+1}\}\right)$, la tribu par rapport à laquelle p_t, I_t, J_t pour $t \leqslant T$ et p_{T+1}, J_{T+1} sont mesurables. En effet, si $1 \leqslant t \leqslant T+1$, alors p_t est une fonction de I_τ et J_τ pour $1 \leqslant \tau \leqslant t-1$.

Définition 3. *On appelle stratégie (pour le joueur 1) une suite de fonctions $f_T : (\mathcal{A} \times \mathcal{B})^{(T-1)} \mapsto \Delta(\mathcal{A})$, qui à partir des tirages passés (I_t, J_t pour $t \leqslant T-1$) donne le coup suivant, p_T pour le joueur 1 (ou $q_T \in \Delta(\mathcal{B})$ pour le joueur 2).*

On note alors D_T la distance du gain moyen à l'instant T à l'ensemble C, c'est-à-dire

$$D_T := \inf_{c \in C} \left\| c - \frac{1}{T} \sum_{t=1}^{T} m(I_t, J_t) \right\|_2 .$$

D_T est donc une variable aléatoire à valeurs réelles, $\mathcal{F}_T$-mesurable.

Définition 4. *On dira qu'un ensemble $C \subseteq \mathbb{R}^d$ est m-approchable par des stratégies aléatoires s'il existe une stratégie pour le joueur 1 telle que pour toute stratégie du joueur 2,*

$$\lim_{T \to \infty} D_T = 0 \quad ps.$$

Cette définition est la même que pour les stratégies déterministes, mais avec l'introduction du caractère probabiliste.

La démonstration du théorème de Blackwell utilise la démonstration du cas déterministe, qui va permettre de contrôler le comportement de la moyenne. Cependant, un argument de déviation (qui décrit l'écartement à la moyenne) est nécessaire. On utilisera pour cela l'inégalité d'Hoeffding-Azuma, démontrée en annexe.

Théorème 2 (Inégalité d'Hoeffding-Azuma)**.**
Soit $(\mathcal{F}_n)_{n \geqslant 0}$ une filtration, (c_n) et (ℓ_n) deux suites réelles et $(Y_n)_{n \geqslant 1}$ une suite de variables aléatoires telles que $\mathbb{E}[Y_n | \mathcal{F}_{n-1}] = 0$ et $c_n \leqslant Y_n \leqslant c_n + \ell_n$ ps. Alors pour tout $\lambda \geqslant 0$,

$$\mathbb{P}\left(\left| \sum_{i=1}^{n} Y_i \right| \geqslant \lambda \right) \leqslant 2 \exp\left(-\frac{2\lambda^2}{\sum_{i=1}^{n} \ell_i^2} \right).$$

En particulier, si $|Y_n| \leqslant c$ ps, et si $0 < \delta < 1$, on obtient qu'avec probabilité au moins $1 - \delta$,

$$\left| \sum_{i=1}^{n} Y_i \right| \leqslant c \sqrt{2n \ln\left(\frac{2}{\delta}\right)}.$$

Théorème 3 (Blackwell)**.**
Soit C un convexe fermé. Alors C est approchable si et seulement si

$$\forall q \in \Delta(\mathcal{B}), \exists p \in \Delta(\mathcal{A}), \qquad m(p,q) \in C$$

Une stratégie d'approchabilité possible est de jouer p_1 arbitrairement, puis de jouer à chaque tour p_T résolvant le problème suivant :

$$\min_{p\in\Delta(\mathcal{A})} \max_{q\in\Delta(\mathcal{B})} < \tilde{m}_T - \pi_T, m(p,q) > ,$$

avec :

- $\tilde{m}_T := \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} m(p_t, J_t)$,
- π_T *la projection orthogonale de* $\tilde{m}_T$ *sur* C *(*$\pi_T = \tilde{m}_T$ *si* $\tilde{m}_T \in C$*),*
- $< .,. >$ *le produit scalaire euclidien sur* $\mathbb{R}^d$.

De plus, si l'on note $M = \sup\limits_{(i,j)\in\mathcal{A}\times\mathcal{B}} \|m(i,j)\|_2$*, alors avec probabilité au moins* $1-\delta$ *:*

$$D_T \leqslant \frac{2M}{\sqrt{T}}\left(1 + \sqrt{2d\ln\left(\frac{2d}{\delta}\right)}\right).$$

Cas particulier : si le joueur 2 joue selon une stratégie fixe connue (par exemple iid) (cas qui ne rentre pas dans le cadre du théorème), ou si la loi du gain ne dépend pas du coup de 2, alors le résultat d'approchabilité découle de la loi des grands nombres.
En effet le joueur 1 peut choisir un $p_0 \in \Delta(\mathcal{A})$ tel que pour un $q_0 \in \Delta(\mathcal{B})$ on aît $m(p_0, q_0) \in C$ et donc en jouant la stratégie constante p_0 on aura $m(I_T, J_T) = m(I_0, J_0)$ en loi. Et $\mathbb{E}\left(m(I_T, J_T)\right) = m(p_0, q_0) \in C$. Les gains sont donc une suite de variables indépendantes et de même loi, bornés donc L^1. Par la loi forte des grands nombres, on aura :

$$\frac{1}{T}\sum_{t=1}^{T} m(I_t, J_t) \to m(p_0, q_0) \text{ ps.}$$

C'est-à-dire $D_T \to 0$ ps.

Ce cas particulier recouvre par exemple le cas de l'exemple de "Pierre-Feuille-Ciseaux", dans lequel quelles que soient les stratégies choisies le gain suit une loi répartie uniformément sur $\{-1, 0, 1\}$. Dans ce cas, on vérifie facilement qu'un convexe fermé est approchable si et seulement s'il contient 0.

Si la norme de la matrice vaut 1 en dimension 2, et que la condition nécessaire et suffisante est vérifiée, le théorème donne par exemple qu'avec au moins 99% de chance, on aura que au bout de 1000 coups la distance sera inférieure à $\frac{1}{2}$ et pour savoir qu'avec au moins 99,99% de chance on a $D_T \leqslant \frac{1}{2}$ il faut attendre 1600 coups environ. Une très bonne certitude n'est pas couteuse.

Démonstration. La démonstration du sens indirect se déduit du sens direct exactement de la même manière que dans le cas déterministe.

Pour le sens direct, on reprend à nouveau une démonstration de l'article [4]. Dans le cas des stratégies déterministes, la distance correspond à une distance réelle de la moyenne au convexe, ou encore à l'espérance conditionnelle de la distance dans le cas des stratégies aléatoires : en effet pour tout temps t, en utilisant :

1. le caractère $\mathcal{F}_T$-mesurable de J_{T+1},
2. le fait que la loi conditonelle de I_{T+1} sachant $\mathcal{F}_T$ est p_{T+1},

on a :

$$\mathbb{E}\left[m(I_{T+1}, J_{T+1})|\mathcal{F}_T\right] = m(p_{T+1}, J_{T+1}). \tag{2}$$

Cette égalité prouve que si on pose

$$Y_T := m(I_T, J_T) - m(p_T, J_T) \ ,$$

alors pour tout k tel que $1 \leqslant k \leqslant d$, en notant $(Y_T)_k$ la k-ième coordonnée de Y_T et $[m_k(i,j)]_{i,j}$ la matrice des $k-ième$ composantes de m, $((Y_T)_k)_{T\in\mathbb{N}}$ est un accroissement de martingale par rapport à $(\mathcal{F}_T)_{T\geqslant 1}$ (c'est-à-dire $M_T := \sum_{t=1}^T Y_t$ est une martingale par rapport à $(\mathcal{F}_T)$). En effet, on a : $(Y_T)_k$ est $\mathcal{F}_T$-mesurable, et

$$\begin{aligned} \mathbb{E}[(Y_{T+1})_k|\mathcal{F}_T] &= \mathbb{E}\left[m_k(I_{T+1}, J_{T+1}) - m_k(p_{T+1}, J_{T+1})|\mathcal{F}_T\right] \\ &= 0 \end{aligned}$$

d'après (2) (en utilisant l'égalité pour chaque coordonnée).

Le théorème d'Hoeffding-Azuma donne des informations sur la vitesse de convergence : la suite de variables aléatoires $((Y_T)_k)_{T\in\mathbb{N}}$ vérifie les hypothèses nécessaires à l'application du théorème :

$$\begin{aligned} \forall T \in \mathbb{N}, \qquad |m_k(I_T, J_T) - m_k(p_T, q_T)| &\leqslant \|m(I_T, J_T) - m(p_T, J_T)\|_\infty \\ &\leqslant \|m(I_T, J_T) - m(p_T, J_T)\|_2 \\ &\leqslant 2M \ , \end{aligned}$$

d'où

$$\forall T \in \mathbb{N}, \qquad |(Y_T)_k| \leqslant 2M \qquad \text{ps.}$$

On obtient donc directement :

$$\mathbb{P}\left(\left|\sum_{t=1}^T (Y_t)_k\right| \leqslant \sqrt{8TM^2 \ln\left(\frac{2}{\delta}\right)}\right) \geqslant 1 - \delta.$$

D'où, comme ce qui précède est vrai pour toutes les coordonnées, et en divisant par T :

$$\mathbb{P}\left(\left\|\frac{1}{T}\sum_{\tau=1}^T m(I_\tau, J_\tau) - \frac{1}{T}\sum_{\tau=1}^T m(p_\tau, J_\tau)\right\|_\infty \leqslant 2M\sqrt{\frac{2\ln(\frac{2}{\delta})}{T}}\right) \geqslant 1 - d\delta \ ,$$

c'est-à-dire, en remplaçant δ par $\frac{\delta}{d}$:

$$\mathbb{P}\left(\left\|\frac{1}{T}\sum_{\tau=1}^T m(I_\tau, J_\tau) - \frac{1}{T}\sum_{\tau=1}^T m(p_\tau, J_\tau)\right\|_\infty \leqslant 2M\sqrt{\frac{2\ln(\frac{2d}{\delta})}{T}}\right) \geqslant 1 - \delta.$$

Or $\|\cdot\|_2 \leqslant \sqrt{d}\|\cdot\|_\infty$, donc avec probabilité au moins $1-\delta$:

$$\left\|\frac{1}{T}\sum_{\tau=1}^T m(I_\tau, J_\tau) - \frac{1}{T}\sum_{\tau=1}^T m(p_\tau, J_\tau)\right\|_2 \leqslant 2M\sqrt{\frac{2d\ln(\frac{2d}{\delta})}{T}}.$$

Or le choix de p_T par le joueur 1 ne dépend que des J_t pour $1 \leqslant t \leqslant T-1$ et on obtiendrait donc le même p_T si le joueur 2 jouait $\forall t, \ q_t = \delta_{J_t}$. On est alors ramené au cas des stratégies déterministes (car $m(p_t, q_t) = m(p_t, J_t)$ dans ce nouveau cas) pour lesquelles on a l'inégalité :

$$\inf_{c\in C}\left\|c - \frac{1}{T}\sum_{t=1}^T m(p_t, J_t)\right\|_2 \leqslant \frac{2M}{\sqrt{T}}.$$

En combinant les deux inégalités, on a qu'avec probabilité au moins $1-\delta$:

$$D_T = \inf_{c\in C} \left\| c - \frac{1}{T}\sum_{t=1}^{T} m(I_t, J_t) \right\|_2 \leqslant \frac{2M}{\sqrt{T}} \left(1 + \sqrt{2d \ln\left(\frac{2d}{\delta}\right)} \right).$$

On a donc prouvé que D_T tendait en probabilité vers 0, et donné une majoration de la vitesse de convergence. Il reste à prouver que D_T tend vers 0 presque sûrement. On va même prouver que $D_T \leqslant K\sqrt{\frac{\ln T}{T}}$ avec K une constante.

On utilise pour cela le lemme de Borel-Cantelli (Lemme 9.3.1 [3]) : on vient de voir qu'avec probabilité au moins $1-\frac{1}{T^2}$, pour tout $T \geqslant 2$:

$$\begin{aligned}
D_T &\leqslant \frac{2M}{\sqrt{T}}\left(1+\sqrt{2d\ln(2dT^2)}\right) \\
&\leqslant \frac{2M}{\sqrt{T}}\left(1+\sqrt{2d(2\ln(T)+\ln(2d))}\right) \\
&\leqslant \frac{2M}{\sqrt{T}}\left(1+\sqrt{4d\ln(T)}+\sqrt{2d\ln(2d)}\right) \qquad \text{car} \qquad \sqrt{a+b} \leqslant \sqrt{a}+\sqrt{b} \\
&\leqslant \frac{2M}{\sqrt{T}}\left(\left(1+\sqrt{2d\ln(2d)}\right)\sqrt{\frac{\ln T}{\ln(2)}}+\sqrt{4d\ln(T)}\right) \\
&\leqslant K\sqrt{\frac{\ln T}{T}} \, ,
\end{aligned}$$

avec $K = 2M\left(\left(1+\sqrt{2d\ln(2d)}\right)\frac{1}{\sqrt{\ln(2)}}+\sqrt{4d\ln(2dT)}\right)$.

Donc si on considère $A_T = \left\{ D_T \leqslant K\sqrt{\frac{\ln T}{T}} \right\}$ on a : $\mathbb{P}(A_T) \geqslant 1-\frac{1}{T^2}$,

c'est-à-dire

$$\mathbb{P}({}^cA_T) \leqslant \frac{1}{T^2} \, ,$$

donc par Borel-Cantelli,

$$\mathbb{P}(\limsup \, {}^cA_T) = 0.$$

Ainsi ps il existe un nombre fini de T tels que $D_T > K\sqrt{\frac{\ln T}{T}}$.

Finalement ps, $\limsup_{T\to\infty} D_T \sqrt{\frac{T}{\ln T}} \leqslant K$.

□

3 Le cas bandits

Jusqu'à présent, dans les cas que nous avons vus, l'action q_T ou J_T était révélée au joueur 1 à la fin de chaque tour. Dans le cas déterministe, il suffisait de connaître $m(p_T, q_T)$ pour construire l'algorithme d'approchabilité. On aimerait avoir un résultat similaire dans le cas où l'action J_T n'est pas révélée. C'est ce qui se passe dans le cas "bandits" : on se place de nouveau avec des stratégies aléatoires, mais cette fois-ci, seul $m(I_T, J_T)$ est révélé au joueur 1 à la fin de chaque tour.

Un théorème de Blackwell existe toujours, mais la démonstration est légèrement différente.

Définition 5. *On appelle stratégie (pour le joueur 1), dans le cas bandits, une suite de fonctions* $f_T : (\mathbb{R}^d)^{T-1} \mapsto \Delta(\mathcal{A})$*, qui à partir des gains passés* $m(I_t, J_t)$ $(t < T)$ *donne le coup suivant,* p_T *pour le joueur 1 (ou* $q_T \in \Delta(\mathcal{B})$ *pour le joueur 2).*

La définition d'un ensemble approchable est la même que dans le cas des stratégies aléatoires.

Théorème 4 (Blackwell, cas bandit)**.**
Soit C *un convexe fermé. Alors* C *est approchable si et seulement si*

$$\forall q \in \Delta(\mathcal{B}), \exists p \in \Delta(\mathcal{A}), \qquad m(p,q) \in C.$$

Si on note $\overline{m}_T := \frac{1}{T}\sum_{\tau=1}^{T} m(I_\tau, J_\tau)$ *et* π_T *sa projection sur* C*, la stratégie utilisée pour approcher* C *est la suivante :* p_{T+1} *résout*

$$\min_{p\in\Delta(\mathcal{A})} \max_{q\in\Delta(\mathcal{B})} < \overline{m}_T - \pi_T, m(p,q) > ,$$

De plus, si l'on note $M = \sup_{(i,j)\in\mathcal{A}\times\mathcal{B}} \|m(i,j)\|_2$*, alors on a :*

$$\mathbb{E}[D_T^2] \leqslant \frac{4M}{T} \text{ et avec probabilité au moins } 1-\delta, \sup_{t\geqslant T} D_t \leqslant \sqrt{\frac{8M^2}{\delta T}}.$$

Remarque : dans le cas précédent, on obtenait la majoration "avec probabilité au moins $1-\delta$, $D_T \leqslant \frac{2M}{\sqrt{T}}\left(1 + 2\sqrt{d \ln\left(\frac{2d}{\delta}\right)}\right)$".

Le résultat est donc complètement différent : même si la vitesse de convergence vis-à-vis de T est la même, la majoration porte cette fois sur le $\sup_{t\geqslant T}$ et on a un résultat plus faible pour ce qui est de la précision en δ : si la norme de la matrice vaut 1 en dimension 2, et que la condition nécessaire et suffisante est vérifiée, le théorème donne par exemple qu'avec au moins 99% de chance, on aura pour $T = 3200$ coups, que $\sup_{t\geqslant T} D_t \leqslant \frac{1}{2}$ et pour savoir qu'avec au moins $99,99\%$ de chance on a $\sup_{t\geqslant T} D_t \leqslant \frac{1}{2}$, il faut attendre 320 000 coups environ. Une très bonne certitude est cette fois très couteuse. On pourrait obtenir un résultat de majoration uniforme dans le cas aléatoire avec une inégalité de Doob (c'est-à-dire avec une méthode comparable à celle de cette partie). Dans cas, puisque le joueur dispose de plus d'information, on obtiendrait peut-être une meilleure majoration.

Démonstration. On reprend ici le schéma de preuve donné dans [5]. Cette preuve est très semblable à celle donnée dans la partie consacrée aux stratégies déterministes, mais cette fois-ci, on travaille avec des variables aléatoires.

On introduit la tribu $\mathcal{F}_T$ définie par $\mathcal{F}_T := \sigma(I_t, m(I_t, J_t), 1 \leqslant t \leqslant T)$. Puisque $\overline{m}_T$ et π_T sont $\mathcal{F}_T$-mesurables, en utilisant de même que pour les stratégies déterministes le théorème du minimax et les propriétés de projection sur les convexes,

$$\mathbb{E}\left[< \overline{m}_T - \pi_T, m(I_{T+1}, J_{T+1}) - \pi_T > |\mathcal{F}_T\right] =< \overline{m}_T - \pi_T, m(p_{T+1}, q_{T+1}) - \pi_T >\leqslant 0,$$

c'est à dire :

$$\mathbb{E}\left[< \overline{m}_T - \pi_T, m(I_{T+1}, J_{T+1}) > |\mathcal{F}_T\right] \leqslant < \overline{m}_T - \pi_T, \pi_T >, \tag{3}$$

$$\begin{aligned} D_{T+1}^2 &\leqslant \|\overline{m}_{T+1} - \pi_T\|^2 \\ &= \|\overline{m}_T - \pi_T\|^2 + 2 < \overline{m}_T - \pi_T, \overline{m}_{T+1} - \overline{m}_T > + \|\overline{m}_{T+1} - \overline{m}_T\|^2. \end{aligned}$$

Or, tout comme en page 5 :

$$\begin{aligned} \overline{m}_{T+1} - \overline{m}_T &= \frac{1}{T+1}\left[m(I_{T+1}, J_{T+1}) - \overline{m}_T\right] \\ &= \frac{1}{T+1}\left[(m(I_{T+1}, J_{T+1}) - \pi_T) - (\overline{m}_T - \pi_T)\right]. \end{aligned}$$

D'où :

$$D_{T+1}^2 \leqslant \|\overline{m}_T - \pi_T\|^2 + \|\overline{m}_{T+1} - \overline{m}_T\|^2 + \\ + \frac{2}{T+1}\left(< m(I_{T+1}, J_{T+1}) - \pi_T, \overline{m}_T - \pi_T > - \|\overline{m}_T - \pi_T\|^2\right)$$

$$D_{T+1}^2 \leqslant \left(1 - \frac{2}{T+1}\right)\|\overline{m}_T - \pi_T\|^2 + \|\overline{m}_{T+1} - \overline{m}_T\|^2 + \\ + \frac{2}{T+1} < m(I_{T+1}, J_{T+1}) - \pi_T, \overline{m}_T - \pi_T > . \tag{4}$$

On a donc :

1. D_T est $\mathcal{F}_T$-mesurable,
2. $\mathbb{E}\left[< m(I_{T+1}, J_{T+1}) - \pi_T, \overline{m}_T - \pi_T > |\mathcal{F}_T\right] \leqslant 0$ par l'inégalité (3),
3. $\overline{m}_{T+1} - \overline{m}_T = \frac{1}{T+1}\left[m(I_{T+1}, J_{T+1}) - \overline{m}_T\right]$.

Donc en prenant l'espérance conditionnelle selon $\mathcal{F}_T$ (qui passe sans problème aux inégalités ps) dans (4) on obtient :

$$\mathbb{E}\left[D_{T+1}^2|\mathcal{F}_T\right] \leqslant \left(1 - \frac{2}{T+1}\right) D_T^2 + \frac{1}{(T+1)^2}\mathbb{E}\left[\|m(I_{T+1}, J_{T+1}) - \overline{m}_T\|^2|\mathcal{F}_T\right].$$

On remarque que par inégalité triangulaire,

$$\mathbb{E}\left(\|m(I_{T+1}, J_{T+1}) - \overline{m}_T\|^2\right) \leqslant 4M^2.$$

On va prouver que pour toute suite de variables aléatoires (D_T) $\mathcal{F}_T$-mesurables vérifiant :

$$\mathbb{E}\left[D_{T+1}^2|\mathcal{F}_T\right] \leqslant \left(1 - \frac{2}{T+1}\right) D_T^2 + \frac{1}{(T+1)^2}\mathbb{E}\left[u_T|\mathcal{F}_T\right] , \tag{5}$$

avec (u_T) une suite de variables aléatoires telle que $\forall t, \quad \mathbb{E}[u_t] \leqslant 4M^2$, on a la conclusion du théorème :

$$\mathbb{E}[D_T^2] \leqslant \frac{4M^2}{T} \text{ et avec probabilité au moins } 1 - \delta, \sup_{t \geqslant T} D_t \leqslant \sqrt{\frac{8M^2}{\delta T}}.$$

La première conclusion s'obtient, tout comme à la partie I, par récurrence sur $\mathbb{E}[D_T^2]$, en passant aux espérances dans (5) et en utilisant le fait que pour toute variable aléatoire X et pour toute tribu $\mathcal{F}$, $\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{F}\right]\right] = \mathbb{E}[X]$. En effet on obtient :

$$\mathbb{E}[D_{T+1}^2] \leqslant \left(1 - \frac{2}{T+1}\right)\mathbb{E}[D_T^2] + \frac{1}{(T+1)^2}4M^2 \ ,$$

la même relation de récurrence que dans la partie I. Donc $\mathbb{E}[D_T^2] \leqslant \frac{4M}{T}$, d'où $\mathbb{E}(D_T^2) \to 0$ et par l'inégalité de Markov, comme $D_T^2 \geqslant 0$, $D_T \to 0$ en probabilité.
Pour montrer la convergence presque sûre et déterminer un majorant de la vitesse de convergence, on va utiliser un argument de surmartingales. En effet, si on considère :

$$Z_T := D_T^2 + \mathbb{E}\left[\sum_{i=T}^{\infty} \frac{u_i}{(i+1)^2}\middle|\mathcal{F}_T\right],$$

on vérifie bien que :

$$\begin{aligned}
\mathbb{E}[Z_{T+1}|\mathcal{F}_T] &= \mathbb{E}\left[D_{T+1}^2 + \mathbb{E}\left[\sum_{i=T+1}^{\infty} \frac{u_i}{(i+1)^2}\middle|\mathcal{F}_{T+1}\right]\middle|\mathcal{F}_T\right] \\
&= \mathbb{E}\left[D_{T+1}^2|\mathcal{F}_T\right] + \mathbb{E}\left[\mathbb{E}\left[\sum_{i=T+1}^{\infty} \frac{u_i}{(i+1)^2}\middle|\mathcal{F}_{T+1}\right]\middle|\mathcal{F}_T\right] \\
&= \mathbb{E}\left[D_{T+1}^2|\mathcal{F}_T\right] + \mathbb{E}\left[\sum_{i=T+1}^{\infty} \frac{u_i}{(i+1)^2}\middle|\mathcal{F}_T\right] \text{ car } \mathcal{F}_T \subseteq \mathcal{F}_{T+1}, \\
&\leqslant \left(1 - \frac{2}{T+1}\right) D_T^2 + \mathbb{E}\left[\frac{u_T}{(T+1)^2}\middle|\mathcal{F}_T\right] + \mathbb{E}\left[\sum_{i=T+1}^{\infty} \frac{u_i}{(i+1)^2}\middle|\mathcal{F}_T\right] \text{ par (5)}, \\
&\leqslant Z_T,
\end{aligned}$$

et Z_T est bien $\mathcal{F}_T$-mesurable, car par définition d'une espérance conditionnelle, $\mathbb{E}[X|\mathcal{F}]$ est $\mathcal{F}$-mesurable. Donc $(Z_T)_{T\in\mathbb{N}}$ est bien une surmartingale positive. Donc $(Z_T)_{T\in\mathbb{N}}$ converge presque-sûrement vers une variable aléatoire Z. En outre,

$$\begin{aligned}
& \mathbb{E}[Z_T] &=& \ \mathbb{E}[D_T^2] + \mathbb{E}\left[\sum_{i=T}^{\infty} \tfrac{u_i}{(i+1)^2}\right] \\
\text{donc par Fubini-Tonnelli,} \quad & \mathbb{E}[Z_T] &\leqslant& \ \mathbb{E}[D_T^2] + 4M^2 \sum_{i=T}^{\infty} \tfrac{1}{(i+1)^2} \\
\text{car } \mathbb{E}[u_i] \leqslant 4M^2\text{, et finalement} \quad & \mathbb{E}[Z_T] &\leqslant& \ \tfrac{4M^2}{T} + \tfrac{4M^2}{T} = \tfrac{8M^2}{T}.
\end{aligned}$$

Donc la limite est nulle ps (car par exemple, par le lemme de Fatou, Z vérifie $\mathbb{E}(Z) \leqslant \liminf \mathbb{E}(Z_t) = 0$ avec $Z \geqslant 0$), et comme $0 \leqslant D_T^2 \leqslant Z_T$, on a bien $D_T^2 \to 0$ ps.

Pour trouver la vitesse de convergence, on va utiliser l'inégalité maximale de Doob suivante, dont on donnera en annexe la démonstration et les liens avec l'inégalité maximale de Doob du cours d'intégration :

Théorème 5 (Inégalité maximale de Doob)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une surmartingale positive, alors :*

$$\forall \lambda > 0, \qquad \lambda\, \mathbb{P}\left(\sup_{n\leqslant k} X_k \geqslant \lambda\right) \leqslant \mathbb{E}(X_n).$$

Appliqué à la surmartingale Z_T, cela donne :

$$\mathbb{P}\left(\sup_{T\leqslant t} Z_t \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon}\mathbb{E}(Z_T) \leqslant \frac{8M^2}{\varepsilon T},$$

c'est-à-dire avec probabilité au moins $1-\delta$:

$$\begin{aligned}
\sup_{T \leqslant t} Z_t &\leqslant \frac{8M^2}{\delta T}, \text{ d'où} \\
\sup_{T \leqslant t} D_t &\leqslant \sqrt{\frac{8M^2}{\delta T}} \text{ ce qui est le résultat désiré.}
\end{aligned}$$

□

4 Annexe

4.1 Le théorème du minimax

En théorie des jeux, le théorème du minimax de von Neumann joue un rôle fondamental. Nous nous proposons ici d'en donner une démonstration, issue de [6] (p.15). Pour cela, nous avons besoin du lemme suivant :

Lemme 1. *Soit $A = (a_{i,j})$ une matrice $m \times n$, alors l'une des deux propriétés suivantes est vérifiée :*

(i) le point $\underline{0}$ (de $\mathbb{R}^m$) est dans l'enveloppe convexe des n colonnes de A et des m vecteurs de la base canonique, c'est-à-dire :

$$a_1 = \begin{bmatrix} a_{1,1} \\ \vdots \\ a_{m,1} \end{bmatrix}, \quad \ldots, \quad a_n = \begin{bmatrix} a_{1,n} \\ \vdots \\ a_{m,n} \end{bmatrix}$$

et

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad e_m = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

(ii) il existe $x_1, \ldots, x_m \in \mathbb{R}$ tels que :
$\forall j, \quad x_j > 0$,
$\sum_{i=1}^{m} x_i = 1$,
$\sum_{i=1}^{m} a_{i,j} x_i > 0 \quad$ *pour tout j de 1 à n.*

Démonstration. On notera $x = (x_1, ..., x_m)$ les points de $\mathbb{R}^m$. Supposons que (i) soit faux, par le théorème de Hahn Banach ($2^{\text{ème}}$ forme géométrique), il existe un hyperplan H séparant strictement $\underline{0}$ de l'enveloppe convexe S des $m + n$ points précédents, c'est à dire :

$$\exists p \in \mathbb{R}^m, \exists a, \exists \varepsilon > 0, \quad \sum_{j=1}^{m} p_j \times 0 \leqslant a \quad \text{et} \quad \forall y \in S, \quad \sum_{j=1}^{m} p_j \times y_j \geqslant a + \varepsilon$$

(note : si les inégalités sont dans l'autre sens, c'est à dire

$$\exists p \in \mathbb{R}^m, \exists a, \exists \varepsilon > 0, \quad \sum_{j=1}^{m} p_j \times 0 \geqslant a + \varepsilon \quad \text{et} \quad \forall y \in S, \quad \sum_{j=1}^{m} p_j \times y_j \leqslant a,$$

on peut se ramener à ce cas en changeant $a + \varepsilon$ en $-a$ et p en $-p$).

On a donc $a \geqslant 0$ et

$$\forall y \in S, \qquad \sum_{j=1}^{m} p_j y_j > 0.$$

En particulier, quand y est l'un des $m + n$ vecteurs, on obtient :
$\forall j, \quad \sum_{i=1}^{n} a_{i,j} p_i > 0$,
$\forall i, \quad p_i > 0$.

On peut alors poser

$$x_i := p_i \Big/ \sum_{j=1}^{n} p_j.$$

Et on obtient donc

$x_i > 0$,

$\sum a_{i,j} x_i > 0$,

$\sum x_i = 1$.

Ce qui est le résultat attendu. $\square$

Nous allons maintenant montrer le théorème suivant (dû à von Neumann).

Théorème 6 (Théorème du minimax)**.**
Soit $A = (a_{i,j})$ une matrice $m \times n$, $X = \{x \in \mathbb{R}^m \ / \ \sum x_i = 1 \ \text{et} \ x_i \geqslant 0\}$ et $Y = \{y \in \mathbb{R}^n \ / \ \sum y_i = 1 \ \text{et} \ y_i \geqslant 0\}$. Alors si

$$m_1 := \max_{x \in X} \min_{y \in Y} {}^t x A y,$$

$$m_2 := \min_{y \in Y} \max_{x \in X} {}^t x A y,$$

on a :

$$m_1 = m_2.$$

Démonstration. Tout d'abord, pour m_1, remarquons que ${}^t x A y = \sum_j y_j \sum_i x_i m_{i,j}$ et donc pour le minimiser, il suffit de choisir y_j valant 0 pour tout j, sauf là où $\sum_i x_i m_{i,j}$ est minimale. D'où $\min_{y \in Y} {}^t x A y = \min_j \sum_i x_i m_{i,j}$ et le *min* d'un nombre fini de fonctions continues est continu. Donc $x \mapsto \min_{y \in Y} {}^t x A y$ est continue et atteint son *max* sur le compact X, ce qui justifie le fait que les inf et les *sup* soient des *min* et des *max* (même raisonnement pour m_2).

On va raisonner par double inégalité :

1. L'inégalité $m_1 \leqslant m_2$ est évidente : il suffit de remarquer que pour $x \in X$ et $y \in Y$ on a

$$\min_{y' \in Y} {}^t x A y' \leqslant {}^t x A y \leqslant \max_{x' \in X} {}^t x' A y$$

   et donc en passant au *max* sur x à gauche, on obtient :

$$\max_{x \in X} \min_{y' \in Y} {}^t x A y' \leqslant \max_{x' \in X} {}^t x' A y,$$

   puis en passant au *min* sur y à droite, on a bien :

$$\max_{x \in X} \min_{y' \in Y} {}^t x A y' \leqslant \min_{y \in Y} \max_{x' \in X} {}^t x' A y.$$

2. Pour l'autre inégalité, nous allons utiliser le lemme précédent pour montrer que selon que l'on a (i) ou (ii), on a soit $m_2 \leqslant 0$, soit $m_1 > 0$, et on en déduira que $m_1 \leqslant 0 < m_2$ est impossible puis, par translation, que $m_1 < k < m_2$ n'est vrai pour aucun k.

   Si (i) est vrai, $\underline{0}$ est combinaison convexe des $m + n$ vecteurs précédents, donc il existe $s_1, \dots, s_{m+n}$ tels que

$$\sum_{j=1}^{n} s_j a_j + \sum_{i=1}^{m} s_{n+i} e_i = \underline{0} \ ,$$

   soit

$$\sum_{j=1}^{n} s_j a_{i,j} + s_{n+i} = 0 \qquad \forall i = 1, \dots, m$$

$$s_j \geqslant 0 \qquad \forall j = 1, \dots, m+n$$

$$\sum_{j=1}^{m+n} s_j = 1.$$

$\underline{0}$ n'étant pas combinaison linéaire non triviale des $e_1, \dots, e_m$, l'un des s_j, $1 \leqslant j \leqslant n$ est non nul, et donc $\sum_{j=1}^{n} s_j > 0$. Si l'on pose alors $y_j := s_j / \sum_{j=1}^{n} s_j$, on a :

$$y_j \geqslant 0$$
$$\sum_{j=1}^{n} y_j = 1$$
$$\sum_{j=1}^{n} a_{i,j} y_j = -s_{n+i} / \sum_{j=1}^{n} s_j \leqslant 0.$$

Donc chaque composante de Ay est négative, d'où :

$$m_2 := \min_{y' \in Y} \max_{x \in X} {}^t x A y' \leqslant \max_{x \in X} {}^t x A y \leqslant 0.$$

Si au contraire (ii) est vrai, en considérant $x \in \mathbb{R}$ donné par le lemme, on a alors :

$$m_1 := \max_{x' \in X} \min_{y \in Y} {}^t x' A y \geqslant \min_{y \in Y} {}^t x A y > 0,$$

car toutes les composantes de ${}^t x A$ sont strictement positives.

Donc il n'est pas possible d'avoir $m_1 \leqslant 0 < m_2$. Or en remplaçant la matrice A par la matrice $B := (a_{i,j} - k)$, $k \in \mathbb{R}$, on obtient pour $(x, y) \in X \times Y$

$${}^t x B y = {}^t x A y - k.$$

Donc

$$m_1(B) = m_1(A) - k,$$
$$m_2(B) = m_2(A) - k.$$

Puisque $m_1(B) < 0 < m_2(B)$ est impossible, on en déduit qu'on ne peut pas avoir

$$m_1(A) < k < m_2(A)$$

et ceci pour aucun $k \in \mathbb{R}$. Donc on a $m_1 \geqslant m_2$, ce qui conclut la preuve.

□

4.2 Inégalité de Hoeffding-Azuma

Les démonstrations du lemme et du théorème qui suivent peuvent être trouvées dans le livre [2].

Lemme 2 (Lemme de Hoeffding)**.**
Soit Y une variable aléatoire réelle bornée. Soient $\mathcal{F}$ une tribu, c une variable aléatoire $\mathcal{F}-$mesurable et $\ell > 0$ une constante. On suppose $\mathbb{E}[Y|\mathcal{F}] = 0$ et $\mathbb{P}(c \leqslant Y \leqslant c + \ell) = 1$. Alors pour tout réel s on a :

$$\mathbb{E}[e^{sY}|\mathcal{F}] \leqslant \exp\left(\frac{s^2 \ell^2}{8}\right) \quad ps.$$

Démonstration. On utilise tout d'abord la convexité de e^{st} sur $[c; c+\ell]$: en écrivant

$$t = \frac{c + \ell - t}{\ell} c + \frac{t - c}{\ell}(c + \ell),$$

on obtient

$$\forall t \in [c; c+\ell], \quad e^{st} \leqslant \frac{c + \ell - t}{\ell} e^{sc} + \frac{t - c}{\ell} e^{s(c+\ell)}.$$

En appliquant cette inégalité à Y (qui est tel que $c \leqslant Y \leqslant c+\ell$ p.s) et en passant à l'espérance conditionnelle, on a :

$$\mathbb{E}[e^{sY}|\mathcal{F}] \leqslant \frac{c+\ell}{\ell}e^{sc} - \frac{c}{\ell}e^{s(c+\ell)} =: f(s),$$

car $\mathbb{E}[Y|\mathcal{F}] = 0$. On pose alors :

$$\begin{array}{rcl} u & = & \ell s, \\ \psi(u) & = & \ln\left(f(s)\right), \\ p & = & \dfrac{-c}{\ell}, \\ 1-p & = & \dfrac{c+\ell}{\ell}. \end{array}$$

Ainsi

$$\begin{array}{rcl} \psi(u) & = & \ln\left(f\left(\dfrac{u}{\ell}\right)\right) \\ & = & \ln\left(\dfrac{c+\ell}{\ell}e^{\frac{uc}{\ell}} + \dfrac{-c}{\ell}e^{\frac{u(c+\ell)}{\ell}}\right) \\ & = & \ln\left(e^{\frac{uc}{\ell}}\left(\dfrac{c+\ell}{\ell} + \dfrac{-c}{\ell}e^{u}\right)\right) \\ & = & \dfrac{uc}{\ell} + \ln\left(1+\dfrac{c}{\ell} - \dfrac{c}{\ell}e^{u}\right). \end{array}$$

En remplaçant :

$$\psi(u) = -pu + \ln(1-p+pe^u).$$

On constate que : $\psi(0) = 0$ et

$$\forall u, \quad \psi'(u) = -p + \frac{pe^u}{1-p+pe^u}.$$

Donc en particulier, $\psi'(0) = -p + \frac{pe^0}{1-p+pe^0} = -p+p = 0$ et de plus,

$$\begin{array}{rcl} \psi''(u) & = & \dfrac{pe^u(1-p+\cancel{pe^u}) - \cancel{pe^u pe^u}}{(1-p+pe^u)^2} \\ & = & \dfrac{(1-p)e^u}{(1-p+pe^u)^2} \\ & = & \dfrac{\alpha\beta}{(\alpha+\beta)^2} \text{ (avec des notations évidentes)} \\ & \leqslant & \dfrac{1}{4} \text{ car} (\alpha+\beta)^2 - 4\alpha\beta = (\alpha-\beta)^2 \geqslant 0. \end{array}$$

Finalement, par l'inégalité de Taylor Lagrange, pour tout u il existe $\theta \in [0;1]$ tel que :

$$\psi(u) = \psi(0) + u\psi'(0) + \psi''(\theta u)\frac{u^2}{2} \leqslant \frac{u^2}{8},$$

ce qui se réécrit :

$$\mathbb{E}[e^{sY}|\mathcal{F}] \leqslant f(s) \leqslant \exp\left(\frac{s^2\ell^2}{8}\right).$$

$\square$

Théorème 7 (Inégalité d'Hoeffding-Azuma)**.**
Soit $(\mathcal{F}_n)_{n\geqslant 0}$ *une filtration,* (c_n) *et* (ℓ_n) *deux suites réelles et* $(Y_n)_{n\geqslant 1}$ *une suite de variables aléatoires telles que* $\mathbb{E}[Y_n|\mathcal{F}_{n-1}] = 0$ *et* $c_n \leqslant Y_n \leqslant c_n + \ell_n$ *ps. Alors pour tout* $\lambda \geqslant 0$*,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n Y_i\right| \geqslant \lambda\right) \leqslant 2\exp\left(-\frac{2\lambda^2}{\sum_{i=1}^n \ell_i^2}\right).$$

En particulier, si $|Y_n| \leqslant c$ *ps, alors si* $0 < \delta < 1$*, on obtient qu'avec probabilité au moins* $1-\delta$*,*

$$\left|\sum_{i=1}^n Y_i\right| \leqslant c\sqrt{2n\ln\left(\frac{2}{\delta}\right)}.$$

Remarque : si on suppose de plus Y_n $\mathcal{F}_n$-mesurable et si on pose $M_0 := 0$ et $M_n := Y_1 + \ldots + Y_n$ pour $n \geqslant 1$, on remarque que $\mathbb{E}[Y_n|\mathcal{F}_{n-1}] = 0$ est équivalent à dire que (Y_n) est une accroissement de martingale (ie : $Y_n = M_n - M_{n-1}$ avec (M_n) une martingale).

Démonstration. Pour tout $s > 0$ l'inégalité de Markov donne :

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^n Y_i\right| \geqslant \lambda\right) &= \mathbb{P}\left(e^{s\sum_{i=1}^n Y_i} \geqslant e^{s\lambda}\right)\\ &\leqslant \mathbb{E}\left[e^{s\sum_{i=1}^n Y_i}\right] e^{-\lambda s}\\ &= \mathbb{E}\left[e^{s\sum_{i=1}^n Y_i} e^{-\lambda s}\right]\end{aligned}$$

On va chercher à majorer $\mathbb{E}[e^{sY_m}|F_{m-1}]$. On a

$$\mathbb{E}[Y_n|\mathcal{F}_{n-1}] = 0,$$

$$\mathbb{P}(c_n \leqslant Y_n \leqslant c_n + \ell_n|\mathcal{F}_{n-1}) = 1.$$

Or la variable c_n est $\mathcal{F}_{n-1}$-mesurable, donc on peut appliquer le Lemme 2 et obtenir la majoration désirée :

$$\mathbb{E}\left[e^{sY_m}\middle|F_{m-1}\right] \leqslant \exp\left(\frac{s^2\ell_m^2}{8}\right).$$

Par suite :

$$\begin{aligned}\mathbb{E}\left[e^{s(Y_1+Y_2+\ldots+Y_n)}\right] &= \mathbb{E}\left[\mathbb{E}[e^{s(Y_1+Y_2+\ldots+Y_n)}\middle|\mathcal{F}_{n-1}]\right]\\ &= \mathbb{E}\left[e^{s(Y_1+Y_2+\ldots+Y_{n-1})}\mathbb{E}[e^{s(Y_n)}\middle|\mathcal{F}_{n-1}]\right]\\ &\leqslant \mathbb{E}\left[e^{s(Y_1+Y_2+\ldots+Y_{n-1})}\right] e^{\frac{s^2\ell_m^2}{8}}.\end{aligned}$$

Puis on obtient immédiatement par récurrence :

$$\mathbb{E}\left[e^{s(Y_1+Y_2+\ldots+Y_n)}\right] \leqslant \exp\left(\frac{1}{8}\sum_{i=1}^n \ell_i^2 s^2\right).$$

On a alors prouvé :

$$\mathbb{P}\left(\left|\sum_{i=1}^n Y_i\right| \geqslant \lambda\right) \leqslant \exp\left(\frac{1}{8}\sum_{i=1}^n \ell_i^2 s^2\right) e^{-\lambda s},$$

inégalité vraie pour tout $s > 0$. On va donc chercher un s_0 pour lequel la majoration est optimale. La fonction exponentielle étant croissante il suffit de trouver le point s_0 où le polynôme en s : $\frac{1}{8}\sum_{i=1}^{n} \ell_i^2 s^2 - \lambda s$ atteint son minimum. On obtient immédiatement par dérivation :

$$s_0 = \frac{4\lambda}{\sum\limits_{i=1}^{n} \ell_i^2},$$

d'où on tire la majoration :

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geqslant \lambda\right) \leqslant \exp\left(-\frac{2\lambda^2}{\sum_{i=1}^{n} \ell_i^2}\right),$$

qui par symétrie donne en passant à des valeurs absolues :

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i\right| \geqslant \lambda\right) \leqslant 2\exp\left(-\frac{2\lambda^2}{\sum_{i=1}^{n} \ell_i^2}\right),$$

ce qui est l'inégalité d'Hoeffding-Azuma.

Pour la deuxième partie du théorème, il s'agit du cas particulier où $\ell_i = 2c$. On a alors

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i\right| \geqslant \lambda\right) \leqslant 2\exp\left(-\frac{\lambda^2}{2nc^2}\right),$$

il suffit alors de poser

$$\lambda := c\sqrt{2n\ln\left(\frac{2}{\delta}\right)},$$

et on a bien avec probabilité au moins $1 - \delta$,

$$\left|\sum_{i=1}^{n} Y_i\right| \leqslant c\sqrt{2n\ln\left(\frac{2}{\delta}\right)}.$$

□

4.3 Inégalités maximales de Doob

Dans le cas Bandits, une inégalité maximale de Doob est nécessaire. Ces inégalités permettent de décrire le comportement du $\sup_{m\leqslant t\leqslant M}$ d'une martingale par rapport au comportement de la martingale en m ou M.

L'inégalité la plus classique, dont on trouvera une preuve dans [3] (Théorème 12.4.2), concerne les sous-martingales. Elle décrit la déviation (probabilité d'un écart à 0 plus grand qu'un λ) du *sup* entre 0 et n par rapport à l'état en n.

Théorème 8 (Inégalité maximale de Doob, sous-martingales). *Soit $(X_n)_{n\in\mathbb{N}}$ une sous-martingale, alors :*

$$\forall\lambda > 0,\ \forall n \in \mathbb{N}, \qquad \lambda\,\mathbb{P}\left(\sup_{0\leqslant k\leqslant n} X_k \geqslant \lambda\right) \leqslant \mathbb{E}\left[X_n \mathbb{1}_{\{\sup_{0\leqslant k\leqslant n} X_k \geqslant \lambda\}}\right] \leqslant \mathbb{E}(X_n^+).$$

L'inégalité dont nous avons besoin concerne les surmartingales et est beaucoup moins connue. La démonstration donnée est donc inspirée de celle du théorème ci-dessus. Cette inégalité décrit la déviation du *sup* après n par rapport à l'état en n.

Théorème 9 (Inégalité maximale de Doob, surmartingales)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une surmartingale positive, alors :*

$$\forall\lambda > 0,\ \forall n \in \mathbb{N}, \qquad \lambda\ \mathbb{P}\left(\sup_{n\leqslant k} X_k \geqslant \lambda\right) \leqslant \mathbb{E}(X_n).$$

Démonstration.
Soit (X_n) une surmartingale (ie : $\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] \leqslant 0$), et $\lambda > 0$.

On pose $T := \inf\{n \geqslant 0/X_n \geqslant \lambda\}$ et $A := \left\{\max\limits_{0\leqslant k\leqslant n} X_k \geqslant \lambda\right\} = \{T \leqslant n\}$. T est un temps d'arrêt, et donc $T \wedge n$ est un temps d'arrêt borné. Par théorème d'arrêt, on sait alors que $(X_{T\wedge k})_{k\leqslant n}$ est une surmartingale, et en particulier :

$$\mathbb{E}[X_{T\wedge n}] \leqslant \mathbb{E}[X_0]$$

Or

$$X_{T\wedge n} \geqslant \lambda \mathbb{1}_A + X_n \mathbb{1}_{A^c}.$$

D'où :

$$\mathbb{E}[X_0] \geqslant \mathbb{E}[X_{T\wedge n}] \geqslant \mathbb{E}[\lambda\mathbb{1}_A + X_n\mathbb{1}_{A^c}] \geqslant \lambda\mathbb{P}(A)$$

car $X_n \geqslant 0$. On a donc :

$$\lambda\ \mathbb{P}\left(\max_{0\leqslant k\leqslant n} X_k \geqslant \lambda\right) \leqslant \mathbb{E}[X_0],$$

puis par suite (en appliquant le résultat à $(X_{p+n})_{p\geqslant 0}$), pour $n \leqslant N$:

$$\lambda\ \mathbb{P}\left(\max_{n\leqslant k\leqslant N} X_k \geqslant \lambda\right) \leqslant \mathbb{E}[X_n].$$

Cette inégalité étant vraie pour tout N, on a :

$$\begin{aligned}
\mathbb{P}\left(\max_{n\leqslant k} X_k \geqslant \lambda\right) &= \mathbb{P}\left(\bigcup_{N\in\mathbb{N}, N\geqslant n} \left\{\max_{n\leqslant k\leqslant N} X_k \geq \lambda\right\}\right) \\
&= \lim \uparrow \mathbb{P}\left(\max_{n\leqslant k\leqslant N} X_k \geqslant \lambda\right) \text{ car il s'agit d'une union croissante,} \\
&\leqslant \frac{1}{\lambda}\,\mathbb{E}[X_n], \text{ l'inégalité est vraie pour chaque terme de la suite.}
\end{aligned}$$

Finalement on a bien :

$$\mathbb{P}\left(\sup_{n\leqslant k} X_k \geqslant \lambda\right) \leqslant \frac{1}{\lambda}\,\mathbb{E}[X_n].$$

□

Références

[1] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1) :1–8, 1956.

[2] N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games.* Cambridge Univ Press, 2006.

[3] J.F. Le Gall. Intégration, probabilités et processus aléatoires. *Ecole Normale Supérieure de Paris*, 2006.

[4] S. Mannor, V. Perchet, and G. Stoltz. Robust approachability and regret minimization in games with partial monitoring. *Arxiv preprint arXiv :1105.4995*, 2011.

[5] J.F. Mertens, S. Sorin, and S. Zamir. Repeated games. Part A : Background material. *CORE Discussion Papers*, 1994.

[6] Guillermo Owen. *Game theorie second edition.* Academic press, 1982.

[7] F. Paulin. Topologie, analyse et calcul différentiel. *Notes de cours, École Normale Supérieure http ://www. fimfa. ens. fr/fimfa/IMG/File/cours/polyanalyse12009. pdf.*