Etudiant : Nicolas BRUYERE
Tuteur : Paul RAYNAUD DE FITTE

Mémoire de maîtrise

LE THEOREME DE KANTOROVICH RUBINSTEIN

Université de ROUEN – Mai 2002

Table des matières

Première partie

Le théorème de Kantorovich-Rubinštein

1 Introduction

Le résultat principal qu'on s'attachera à démontrer dans ce mémoire est le théorème de Kantorovich-Rubinštein (théorème 2.1). Celui-ci voit ses applications en probabilité. Cependant, nul besoin n'est de connaître la théorie des probabilités. Ce ne sont que des résultats de mesure et d'analyse fonctionnelle qui seront utilisés. C'est un résultat récent (1958) qui est démontré (en russe) dans [7], mais il provient d'un vieux problème concret qui est la modélisation du transport d'un tas de charbon de la manière la plus économique étudié vers la fin du XVIIIè siècle par le mathématicien français Gaspard Monge (1781). Ces problèmes de transport font maintenant l'objet d'applications dans d'autres domaines tels celui de l'économie et sont développés dans une branche de la théorie des probabilités. Signalons au passage que Kantorovich n'est pas un mathématicien de formation mais un économiste, d'ailleurs il a obtenu en 1975 le prix Nobel dans cette spécialité. C'est en 1958 que Kantorovich et Rubinštein donnèrent la première démonstration du théorème de Kantorovich-Rubinštein, mais dans le cas où l'espace S était compact. En 1976, Dudley donna dans [4] la première démonstration en anglais du théorème pour S métrique séparable. Son travail est la source essentielle de ce mémoire. Cependant il y avait une faille dans sa démonstration qui a été comblée en 1982 par de Acosta dans [3]. Une autre démonstration est donnée dans [5].

Le mémoire est constitué de deux parties. La première concerne le théorème en lui-même. On y fait quelques commentaires sur son utilité puis on passe à sa démonstration. C'est une longue démonstration car il convient de planter un décor assez chargé. La seconde, intitulée Annexe, regroupe quelques résultats qui servent dans la première partie. Ils sont rejetés en seconde partie car ils ne feraient qu'alourdir la démonstration. Cependant il conviendrait de commencer par la lire, la survoler suffit, car quelques définitions d'espaces fréquemment utilisés y sont donnés. Ensuite elle sera consultée, s'il y a lieu, par les références qui sont faites en première partie.

Avant de débuter, on commence par rappeler l'essentiel. Dans tout le mémoire on travaillera dans un espace métrique séparable (S, d) qui est muni de sa tribu borélienne $\mathcal{B}$. Et on rappelle simplement les définitions des deux distances qui font l'objet du théorème. La distance de Wasserštein (définition 11.3) :

$$\begin{aligned} W(P,Q) &= \inf\{d_1(X,Y) : L(X) = P, L(Y) = Q\} \\ &= \inf_{\pi \in \mathcal{P}(S\times S)} \{\int_{S\times S} d(x,y)\, \pi(dx,dy) : \ \pi \text{ de marges } P \text{ et } Q \ \}. \end{aligned}$$

La distance de Fortet-Mourier (définition 11.2) :

$$F(P,Q) = \sup_{\|f\|_L \leq 1} \left| \int_S f\, d(P-Q) \right| .$$

On rappelle que l'on peut étendre F et P sur des espaces de mesures positives finies non nécessairement probabilisées, d'après les définitions 11.2,11.3 de W et F en annexe.

2 Enoncé et intérêt du théorème

2.1 Enoncé

On commence par énoncer l'objet de toute notre discussion.

Théorème 2.1 (Kantorovich-Rubinštein) *Soient P,$Q \in \mathcal{P}_1(S,d)$, on a alors*

$$W(P,Q) = F(P,Q).$$

Remarques

- Le théorème est vrai dans un cadre plus général où ce n'est plus d qui intervient dans la définition de la distance de Wasserštein mais une fonction de coût c qui vérifie des hypothèses plus faibles que d et qui permet de modéliser plus de problèmes de transport de masse. En effet la fonction à optimiser est justement le coût moyen du transport, ce coût n'étant pas systématiquement une distance.
- Le théorème reste en fait valable pour les extensions de W et F aux espaces de mesures non probabilisées. Mais dans les applications, il est intéressant dans le cas de mesures probabilisées.

Pour le moment on ne se contente que de son énoncé. Sa démonstration, qui fait intervenir de nouveaux concepts, sera faite plus loin. Dans un premier temps on va donc indiquer l'intérêt de ce théorème et le comparer avec un autre théorème qui lui est similaire, le théorème de Strassen.

2.2 Commentaires sur W et F

Les distances W et F, qui servent à modéliser certains problèmes en probabilités, étaient utilisées indépendamment l'une de l'autre, bien avant que le théorème de Kantorovich-Rubinštein ne soit démontré. On voit donc la puissance de ce théorème qui établit que ces deux distances, qui à première vue par leur définition sont différentes, ne font qu'une.

Une des questions intéressantes est de savoir si W est atteinte. On rappelle que W est définie comme un inf. Si $P, Q \in \mathcal{P}_1(S, d)$, le but est de savoir si l'on peut trouver $X, Y \in L^1$ de lois P et Q telles que

$$W(P, Q) = d_1(X, Y).$$

Un des cas que l'on peut traiter est celui où l'espace S est polonais, c'est à dire qu'il existe une distance δ qui engendre la topologie associée à d et telle que (S, δ) soit complet et séparable. On énonce ce résultat dans la proposition qui suit.

Proposition 2.1 *Soient $P, Q \in \mathcal{P}_1(S, d)$. Si S est polonais alors il existe $\pi \in \mathcal{P}(S \times S)$ de marges P et Q telle que*

$$W(P, Q) = \int_{S \times S} d \, d\pi.$$

Démonstration — Comme S est polonais, d'après le corollaire 7.1, toute mesure sur S est tendue. Posons

$$D(P, Q) = \{\pi \in \mathcal{P}(S \times S) \ : \ \pi \text{ de marges } P, Q\}.$$

Alors $D(P, Q)$ est uniformément tendu. En effet soit $\epsilon > 0$, il existe un compact K_ϵ tel que

$$P(K_\epsilon^c) < \frac{\epsilon}{2} \text{ et } Q(K_\epsilon^c) < \frac{\epsilon}{2}.$$

Donc pour tout $\pi \in D(P, Q)$, on a

$$\begin{array}{rcl} \pi((K_\epsilon \times K_\epsilon)^c) & \leq & \pi(K_\epsilon^c \times S) + \pi(S \times K_\epsilon^c) \\ & = & P(K_\epsilon^c) + Q(K_\epsilon^c) \\ & < & \epsilon. \end{array}$$

Comme $D(P, Q)$ est uniformément tendu, alors d'après le critère de Prokhorov (proposition 7.1), $D(P, Q)$ est relativement compact pour la topologie étroite. Or $D(P, Q)$ est fermé, en effet soit $\pi \in \overline{D(P, Q)}$, il existe une suite (π_n) de $D(P, Q)$ convergeant vers π. Montrons que $\pi \in D(P, Q)$. Soit g une fonction continue bornée sur S. On note π^1 et π^2 les première et deuxième marges de π, et p_1 et p_2 les projections de $S \times S$ sur S. On a alors $g \circ p_1$ qui est continue et bornée sur $S \times S$, d'où

$$\int_S g \, d\pi_n^1 = \int_S g \circ p_1 \, d\pi_n \to \int_S g \circ p_1 \, d\pi = \int_S g \, d\pi^1$$

donc

$$P = \pi_n^1 \xrightarrow{\mathcal{L}} \pi^1$$

d'où par unicité de la limite $\pi^1 = P$. De même on montre que $\pi^2 = Q$. D'où $\pi \in D(P,Q)$. Donc finalement, $D(P,Q)$ est compact. Et la fonction

$$\pi \longrightarrow \int_{S\times S} d(x,y)\, \pi(dx,dy)$$

est semi-continue-inférieurement pour la topologie étroite car elle s'écrit comme le sup des fonctions continues

$$\pi \longrightarrow \int_{S\times S} d(x,y) \wedge n\, \pi(dx,dy).$$

L'inf est donc atteint. ♣

Remarques

- En fait par un résultat de Kawabe [8], on peut montrer que dans le cas d'un espace métrique quelconque $D(P,Q)$ est compact et donc la proposition précédente reste valable dans un espace métrique quelconque.
- Dire que $W(P,Q)$ est atteinte pour une mesure π de marges P et Q revient à dire qu'il existe $X, Y \in L^1$ de loi P et Q telles que $W(P,Q) = d_1(X,Y)$.

2.3 Analogie avec le théorème de Strassen

Dans cette partie on ne fera qu'énoncer le théorème de Strassen, on n'en donnera pas la démonstration, qui comme celle de Kantorovich-Rubinštein est longue. On peut la trouver dans [4, théorème 18.3]. Le but est seulement de comparer le contenu : en quoi ces théorèmes sont–ils similaires, et en quoi diffèrent–ils ? Mais commençons d'abord par énoncer le théorème de Strassen.

Théorème 2.2 (Strassen) *Si* $P,Q \in \mathcal{P}(S)$ *alors*

$$\rho(P,Q) = \inf\{\theta(X,Y) : L(X) = P, L(Y) = Q\}.$$

ρ et θ sont définis dans les définitions 11.1 et 10.2 respectivement. L'intérêt de ces théorèmes est que si P et Q sont deux probabilités "proches" pour la distance considérée, c'est-à-dire ρ si $P,Q \in \mathcal{P}(S)$ ou F si elles sont dans $\mathcal{P}_1(S,d)$, alors on peut trouver deux variabes aléatoires dans L^0 dans le premier cas, dans L^1 dans le deuxième cas, ayant pour loi ces mesures de probabilité, "proches" l'une de l'autre, dans le sens de θ dans le premier cas dans celui de d_1 dans le deuxième cas. En effet, par définition de l'inf, si $P_0, Q_0 \in \mathcal{P}(S)$ alors pour tout $\epsilon > 0$ il existe $X_0, Y_0 \in L^0$ de lois P_0 et Q_0 telles que

$$\theta(X_0, Y_0) \quad \leq \quad \rho(P_0, Q_0) + \epsilon. \tag{1}$$

Et si $P_1, Q_1 \in \mathcal{P}_1(S,d)$, il existe $X_1, Y_1 \in L^1$ de lois P_1 et Q_1 telles que

$$d_1(X_1, Y_1) \leq F(P_1, Q_1) + \epsilon.$$

Mais on peut pousser la similarité un peu plus loin. En effet, ici ρ et F ne se ressemblent pas, tout comme θ et d_1. L'idée est donc dans cette comparaison de pouvoir remplacer ρ par β (définition 11.2) et bien sûr θ par $d_0 = \kappa$ (définition 10.1). La similarité serait alors sans équivoque. Ce sont les propositions 10.1 et 11.2 qui nous y aident car alors l'inégalité (1) ci dessus devient

$$\kappa(X_0, Y_0) = d_0(X_0, Y_0) < 4\sqrt{\beta(P_0, Q_0)} + \epsilon.$$

Pour résumer cette comparaison, on dresse le tableau suivant.

	Théorème de Strassen	**Théorème de Kantorovich-Rubinštein**
On prend	$P, Q \in \mathcal{P}(S)$	$P, Q \in \mathcal{P}_1(S)$
Distance	$\beta(P, Q) = \sup\limits_{\|f\|_{L_B} \leq 1} \left\| \int_S f \, d(P - Q) \right\|$	$F(P, Q) = \sup\limits_{\|f\|_L \leq 1} \left\| \int_S f \, d(P - Q) \right\|$
On a alors	Pour tout $\epsilon > 0$, il existe $X, Y \in L^0$ de lois P et Q telles que $d_0(X, Y) \leq 4\sqrt{\beta(P, Q)} + \epsilon$.	Pour tout $\epsilon > 0$,il existe $X, Y \in L^1$ de lois P et Q telles que $d_1(X, Y) \leq F(P, Q) + \epsilon$.

Remarque

– Grossièrement on peut dire que le théorème de Strassen donne un résultat dans l'espace L^0 des variables aléatoires, alors que celui de Kantorovich-Rubinštein en donne un dans l'espace L^1. Il serait alors très intéressant de pouvoir généraliser ces théorèmes aux espaces L^p.

3 L'espace M_{o1}

Le processus de démonstration du théorème va maintenant débuter. Comme annoncé plus haut, il nous faut définir de nouveaux objets.

3.1 Définition et propriétés de M_{o1}

Commençons par donner la définition qui suit.

Définition 3.1 *On note $M_o = M_o(S)$ l'ensemble des mesures signées finies m sur la tribu borélienne de S tel que $m(S) = 0$.*

Si $m \in M_o$ alors par la décomposition de Jordan (théorème 7.1) , il existe m^+ et m^- deux mesures positives finies telles que $m = m^+ - m^-$. Remarquons qu'alors $m^+(S) = m^-(S)$. On pose alors $|m| = m^+ + m^-$. Dans la suite on sera souvent amené à réutiliser m^+ et m^-, mais on ne précisera pas à chaque fois que ces mesures sont issues de la décomposition de Jordan. On donne la définition suivante.

Définition 3.2 *On pose*

$$M_{o1} = \{m \in M_o : \exists x \in S \int d(x,y)\,|m|(dy) < \infty\}.$$

Remarques

– Il est facile de voir que par l'inégalité triangulaire $\exists$ équivaut à $\forall$ dans la définition de M_{o1}.
– L'espace M_{o1} muni de l'addition et de la multiplication par un scalaire est un espace vectoriel.

On a donc défini un nouvel ensemble, il serait donc intéressant de pouvoir le munir d'une topologie. Pour cela on pose d'abord.

Définition 3.3 *Soit $m \in M_{o1}$, on note $\mathcal{X}_m$ l'ensemble des mesures positives finies b sur l'espace produit $S \times S$ muni de la tribu borélienne $\mathcal{B} \otimes \mathcal{B}$ telles que*

$$b(A \times S) - b(S \times A) = m(A) \quad \forall A \in \mathcal{B}.$$

Remarquons que pour tout $m \in M_{o1}$, $\mathcal{X}_m$ est non vide car contient toujours la mesure produit

$$\pi = \frac{m^+ \otimes m^-}{m^+(S)}.$$

En effet pour tout $A \in \mathcal{B}$ on a

$$\begin{aligned} \pi(A \times S) - \pi(S \times A) &= \frac{m^+(A)m^-(S)}{m^+(S)} - \frac{m^+(S)m^-(A)}{m^+(S)} \\ &= m^+(A) - m^-(A) \quad \text{car } m^+(S) = m^-(S) \\ &= m(A). \end{aligned}$$

Voici une proposition qui va nous être utile pour la suite.

Proposition 3.1 *Si P, Q sont deux mesures probabilisées alors $\mathcal{X}_{P-Q}$ contient l'ensemble des mesures probabilisées sur $S \times S$ de marges P et Q.*

Démonstration — Soit π une mesure probabilisée sur $S \times S$ de marges P et Q. Alors pour tout $A \in \mathcal{B}$ on a

$$\pi(A \times S) - \pi(S \times A) = P(A) - Q(A) = (P - Q)(A)$$

d'où $\pi \in \mathcal{X}_{P-Q}$. ♣

3.2 Structure topologique de M_{o1}

On considère sur M_{o1} la fonction

$$\| \,.\, \|_W : m \in M_{o1} \longmapsto \|m\|_W = \inf_{b \in \mathcal{X}_m} \int_{S\times S} d(x,y)\, b(dx,dy).$$

On a le résultat suivant.

Proposition 3.2 $\| \,.\, \|_W$ *est une semi-norme sur* M_{o1}.

Démonstration — Comme $d \geq 0$, $\|m\|_W \geq 0$. Montrons ensuite que $\|m\|_W < \infty$. Si $m \in M_{o1}$ on sait que $\pi = \dfrac{m^+ \otimes m^-}{m^+(S)}$ est dans $\mathcal{X}_m$, il suffit donc de montrer que

$$\int_{S\times S} d(x,y)\, \pi(dx,dy) < \infty.$$

Comme $d \geq 0$, par Fubini, on a

$$\int_{S\times S} d(x,y)\, \pi(dx,dy) = \frac{1}{m^+(S)} \int_S \int_S d(x,y)\, m^+(dx)\, m^-(dy).$$

Soit $z \in S$, par l'inégalité triangulaire et Fubini encore, il vient

$$\begin{aligned}
\int_{S\times S} d(x,y)\, \pi(dx,dy) \quad &\leq \quad \frac{1}{m^+(S)} \int_S \int_S d(x,z)\, m^+(dx)\, m^-(dy) \\
&\qquad + \frac{1}{m^+(S)} \int_S \int_S d(z,y)\, m^-(dy)\, m^+(dx) \\
&= \quad \int_S d(x,z)\, m^+(dx) + \int_S d(z,y)\, m^-(dy) \\
&\leq \quad \int_S d(x,z)\, |m|(dx) + \int_S d(z,y)\, |m|(dy) \\
&< \quad \infty \quad \text{car} \quad m \in M_{o1}.
\end{aligned}$$

Montrons maintenant l'homogénéité. Soit $m \in M_{o1}$, soit $c \in \mathbb{R}$.

– Si $c = 0$ alors $\mathcal{X}_{cm}$ contient la mesure nulle ν sur $S \times S$ et donc $\|0m\|_W$ est atteint avec

$$\|0m\|_W = \int_{S\times S} d(x,y)\, \nu(dx,dy) = 0.$$

D'où $\|0m\| = 0$.

– Si $c > 0$, alors $\mathcal{X}_{cm} = \{cb : b \in \mathcal{X}_m\}$. En effet, soit $b' \in \mathcal{X}_{cm}$, on a alors

$$b'(A \times S) - b'(S \times A) = cm(A).$$

Or on peut écrire $b' = c(b'/c)$ où par l'égalité précédente $b'/c \in \mathcal{X}_m$ et donc $b' \in \{cb : b \in \mathcal{X}_m\}$. Réciproquement, si $b \in \mathcal{X}_m$, on a

$$b(A \times S) - b(S \times A) = m(A)$$

et donc en multipliant des deux cotés par c il vient $cb \in \mathcal{X}_{cm}$. D'où le résultat. Et donc :

$$\begin{aligned}
\|cm\|_W &= \inf\left\{\int_{S\times S} d(x,y)\, b'(dx,dy) : b' \in \{cb : b \in \mathcal{X}_m\}\right\}\\
&= \inf\left\{\int_{S\times S} d(x,y)\, cb(dx,dy) : b \in \mathcal{X}_m\right\}\\
&= c\|m\|_W.
\end{aligned}$$

– Si $c = -1$ alors $\mathcal{X}_{-m} = \{b \circ T^{-1} : b \in \mathcal{X}_m\}$ où $T(x,y) = (y,x)$ (T est bijective). En effet, soit $b' \in \mathcal{X}_{-m}$, on a alors

$$\begin{aligned}
b'(A \times S) - b'(S \times A) &= -m(A)\\
(b' \circ T)(S \times A) - (b' \circ T)(A \times S) &= -m(A)\\
(b' \circ T)(A \times S) - (b' \circ T)(S \times A) &= m(A).
\end{aligned}$$

En posant $b = b' \circ T$ alors $b' = b \circ T^{-1}$ avec $b \in \mathcal{X}_m$. Réciproquement, soit $b \in \mathcal{X}_m$, alors

$$\begin{aligned}
(b \circ T^{-1})(A \times S) - (b \circ T^{-1})(S \times A) &= b(S \times A) - b(A \times S)\\
&= -m(A)
\end{aligned}$$

d'où $b \circ T^{-1} \in \mathcal{X}_{-m}$. D'où le résultat. Et donc :

$$\begin{aligned}
\|-m\|_W &= \inf\left\{\int_{S\times S} d(x,y)\, b'(dx,dy) : b' \in \{b \circ T^{-1} : b \in \mathcal{X}_m\}\right\}\\
&= \inf\left\{\int_{S\times S} d(x,y)\, b \circ T^{-1}(dx,dy) : b \in \mathcal{X}_m\right\}\\
&= \inf\left\{\int_{S\times S} d(T(x,y))\, b(dx,dy) : b \in \mathcal{X}_m\right\}\\
&= \inf\left\{\int_{S\times S} d(x,y)\, b(dx,dy) : b \in \mathcal{X}_m\right\}\\
&= \|m\|_W.
\end{aligned}$$

En conclusion :

$$\|cm\|_W = |c|\|m\|_W \quad \forall c \in \mathbb{R} \quad , \forall m \in M_{o1}.$$

Enfin montrons l'inégalité triangulaire. Soient $m, m' \in M_{o1}$. En utilisant les définitions il est immédiat que si $b \in \mathcal{X}_m$ et $b' \in \mathcal{X}_{m'}$ alors $b + b' \in \mathcal{X}_{m+m'}$. Ainsi :

$$\begin{aligned}
\|m + m'\|_W &\leq \int_{S\times S} d(x,y)\, (b + b')(dx,dy)\\
&= \int_{S\times S} d(x,y)\, b(dx,dy) + \int_{S\times S} d(x,y)\, b'(dx,dy).
\end{aligned}$$

Cette inégalité étant vraie pour tous $b \in \mathcal{X}_m$ et $b' \in \mathcal{X}_{m'}$, on passe à l'inf et donc :

$$\|m + m'\|_W \leq \|m\|_W + \|m'\|_W. \quad \clubsuit$$

L'espace $(M_{o1}, \| \,.\, \|_W)$ est donc un espace vectoriel semi-normé. On pourrait montrer mieux en fait c'est-à-dire que $\| \,.\, \|_W$ est une norme. La démonstration utiliserait un point que nous n'avons pas encore montré. On signale juste cela et on ne cherchera pas à le montrer car ce n'est pas essentiel dans la suite.

3.3 Un résultat de densité

Commençons par définir une certaine classe de mesures sur M_{o1}.

Définition 3.4 *On dit que $m \in M_{o1}$ est* simple *si*

$$m = \sum_{i=1}^{n} a_i(\delta_{x_i} - \delta_{y_i})$$

où $n \in \mathbb{N}$, $x_i, y_i \in S$ et $a_i \in \mathbb{R}$.

Ces mesures, comme leur qualificatif l'indique, sont faciles à manipuler. Et bien, elles font l'objet du très intéressant résultat suivant.

Proposition 3.3 *Les mesures simples sont denses dans M_{o1}.*

Démonstration — Celle-ci se fait en deux étapes.
1ère étape Les mesures à support borné sont denses dans M_{o1}.

Soit $m \in M_{o1}$. Si m est la mesure nulle alors elle est déjà à support (définiton 7.1) borné ($\emptyset$). On prend désormais m non nulle. Fixons $x \in S$ et posons $B_n = B(x, n)$ la boule fermée de centre x et de rayon n. Comme $m^+(B_n) \to m^+(S) \neq 0$ et $m^-(B_n) \to m^-(S) \neq 0$ alors pour n à partir d'un certain rang N, $m^-(B_n), m^+(B_n) > 0$, on peut donc définir, pour $n \geq N$,

$$m_n(A) = m^+(S) \left(\frac{m^+(A \cap B_n)}{m^+(B_n)} - \frac{m^-(A \cap B_n)}{m^-(B_n)} \right) \quad \forall A \in \mathcal{B}.$$

On vérifie facilement que $m_n(B_n^c) = 0$ donc le support de m_n est inclus dans B_n, ainsi le support de m_n est borné pour tout $n \geq N$.

On a, pour tout $A \in \mathcal{B}$,

$$\begin{aligned}
(m - m_n)(A) &= m(A \cap B_n^c) + m(A \cap B_n) - m^+(S)\frac{m^+(A \cap B_n)}{m^+(B_n)} \\
&\quad + m^+(S)\frac{m^-(A \cap B_n)}{m^-(B_n)} \\
&= m(A \setminus B_n) - \left(\frac{m^+(S)}{m^+(B_n)} - 1 \right) m^+(A \cap B_n) \\
&\quad + \left(\frac{m^+(S)}{m^-(B_n)} - 1 \right) m^-(A \cap B_n) \\
&= m(A \setminus B_n) - \epsilon_n m^+(A \cap B_n) + \delta_n m^-(A \cap B_n)
\end{aligned}$$

avec

$$\epsilon_n = \frac{m^+(S)}{m^+(B_n)} - 1 \text{ et } \delta_n = \frac{m^+(S)}{m^-(B_n)} - 1$$

et on remarque que ϵ_n , $\delta_n \to 0$. Pour tout $A \in \mathcal{B}$, on pose

$$\begin{aligned}\mu_n(A) = m^+(A \setminus B_n) + \delta_n m^-(A \cap B_n)\\ \nu_n(A) = m^-(A \setminus B_n) + \epsilon_n m^+(A \cap B_n).\end{aligned}$$

On a μ_n , $\nu_n \geq 0$ et $m - m_n = \mu_n - \nu_n$ et $\mu_n(S) = \nu_n(S)$ donc :

$$b_n = \frac{\mu_n \otimes \nu_n}{\mu_n(S)} \in B_{m-m_n}.$$

Montrons que $\|m - m_n\|_W \to 0$ ce qui achèvera la première étape. Pour cela, il nous suffit de montrer que $\int_{S\times S} d(x,y)\, b_n(dx,dy) \to 0$. On a

$$\begin{aligned}\int_{S\times S} d(x,y)\, b_n(dx,dy) &= \frac{1}{\mu_n(S)} \int_{S\times S} \chi_{B_n^c}(x)\chi_{B_n^c}(y) d(x,y)\, m^+(dx)m^-(dy)\\ &\quad + \frac{1}{\mu_n(S)} \int_{S\times S} \chi_{B_n^c}(x)\chi_{B_n}(y) d(x,y)\epsilon_n\, m^+(dx)m^+(dy)\\ &\quad + \frac{1}{\mu_n(S)} \int_{S\times S} \chi_{B_n}(x)\chi_{B_n^c}(y) d(x,y)\delta_n\, m^-(dx)m^-(dy)\\ &\quad + \frac{1}{\mu_n(S)} \int_{S\times S} \chi_{B_n}(x)\chi_{B_n}(y) d(x,y)\epsilon_n\delta_n\, m^-(dx)m^+(dy)\\ &= I_n^1 + I_n^2 + I_n^3 + I_n^4.\end{aligned}$$

Un calcul simple nous donne : $\mu_n(S) = |m|(B_n^c)$. On va montrer que $I_n^i \to 0$ pour $i = 1,2,3,4$. Dans les calculs d'intégrales que nous allons effectuer les fonctions intégrées sont positives, on pourra donc sans souci utiliser Fubini sans le préciser à chaque fois. Fixons $z \in S$.

– Calcul de I_n^1

$$\begin{aligned}I_n^1 &\leq \frac{1}{\mu_n(S)} \int_{S\times S} \chi_{B_n^c}(x)\chi_{B_n^c}(y) d(x,z)\, m^+(dx)m^-(dy)\\ &\quad + \frac{1}{\mu_n(S)} \int_{S\times S} \chi_{B_n^c}(x)\chi_{B_n^c}(y) d(z,y)\, m^+(dx)m^-(dy)\\ &= \frac{m^-(B_n^c)}{|m|(B_n^c)} \int_S \chi_{B_n^c}(x) d(x,z)\, m^+(dx)\\ &\quad + \frac{m^+(B_n^c)}{|m|(B_n^c)} \int_S \chi_{B_n^c}(y) d(y,z)\, m^-(dy)\\ &\leq \int_S \chi_{B_n^c}(x) d(x,z)\, m^+(dx) + \int_S \chi_{B_n^c}(y) d(y,z)\, m^-(dy).\end{aligned}$$

Comme $\chi_{B_n^c}(x)d(x,z) \to 0$ simplement et $x \mapsto d(x,z)$ est intégrable car $m \in M_{o1}$ alors par le théorème de convergence dominée de Lebesgue la première intégrale converge vers 0. De la même façon, la seconde converge aussi vers 0. D'où $I_n^1 \to 0$.

– Calcul de I_n^2

$$\begin{aligned}
I_n^2 &\leq \frac{1}{\mu_n(S)}\int_{S\times S}\chi_{B_n^c}(x)\chi_{B_n}(y)d(x,z)\epsilon_n\, m^+(dx)m^+(dy)\\
&\quad+\frac{1}{\mu_n(S)}\int_{S\times S}\chi_{B_n^c}(x)\chi_{B_n}(y)d(z,y)\epsilon_n\, m^+(dx)m^+(dy)\\
&= \frac{m^+(B_n)\epsilon_n}{|m|(B_n^c)}\int_S \chi_{B_n^c}(x)d(x,z)\, m^+(dx)\\
&\quad+\frac{m^+(B_n^c)}{|m|(B_n^c)}\int_S \epsilon_n\chi_{B_n}(y)d(y,z)\, m^+(dy)\\
&\leq \int_S \chi_{B_n^c}(x)d(x,z)\, m^+(dx)+\int_S \epsilon_n d(y,z)\, m^+(dy).
\end{aligned}$$

On applique le théorème de convergence dominée de Lebesgue aux deux intégrales, comme dans le premier cas, et on obtient $I_n^2 \to 0$.

– Calcul de I_n^3

Se fait comme I_n^2 et on a $I_n^3 \to 0$.

– Calcul de I_n^4

$$\begin{aligned}
I_n^4 &\leq \frac{1}{\mu_n(S)}\int_{S\times S}\chi_{B_n}(x)\chi_{B_n}(y)d(x,z)\epsilon_n\delta_n\, m^-(dx)m^+(dy)\\
&\quad+\frac{1}{\mu_n(S)}\int_{S\times S}\chi_{B_n}(x)\chi_{B_n}(y)d(z,y)\epsilon_n\delta_n\, m^-(dx)m^+(dy)\\
&\leq \frac{m^+(B_n)\epsilon_n}{|m|(B_n^c)}\int_S \delta_n\chi_{B_n}(x)d(x,z)\, m^-(dx)\\
&\quad+\frac{m^+(B_n^c)\delta_n}{|m|(B_n^c)}\int_S \epsilon_n\chi_{B_n}(y)d(y,z)\, m^+(dy)\\
&\leq \int_S \delta_n d(x,z)\, m^-(dx)+\int_S \epsilon_n d(y,z)\, m^+(dy).
\end{aligned}$$

Toujours par le théorème de Lebesgue, on conclut que $I_n^4 \to 0$.

La première étape est donc terminée.

2ème étape Les mesures simples sont denses dans l'ensemble des mesures à support borné.

Soit $m \in M_{o1}$ à support Σ borné. Quitte à travailler dans l'espace (Σ, d'), où d' est en fait la restriction de d à Σ, on peut supposer directement que l'espace S lui-même est borné. On a $m = m^+ - m^-$. On pose

$$m' = \frac{m}{m^+(S)} = \frac{m^+}{m^+(S)} - \frac{m^-}{m^+(S)} = P - Q$$

où P et Q sont deux mesures de probabilité sur S. Par le théorème de Glivenko-Cantelli-Varadarajan (Théorème 11.1), il existe des suites de probabilités $(P_n), (Q_n)$ sur S tendant en loi respectivement vers P et Q

et telles que, pour tout $A \in \mathcal{B}$,

$$\begin{aligned} P_n(A) &= \frac{1}{n}\sum_{j=1}^{n} \delta_{x_j}(A) \quad \text{où } \ x_j \in S \\ Q_n(A) &= \frac{1}{n}\sum_{j=1}^{n} \delta_{y_j}(A) \quad \text{où } \ y_j \in S \end{aligned}$$

et donc

$$\sigma_n = (P_n - Q_n)(A) = \sum_{j=1}^{n} \frac{1}{n}(\delta_{x_j} - \delta_{y_j})(A)$$

est une mesure simple. Il s'agit maintenant de montrer que σ_n converge vers m' dans M_{o1} on aura alors, par homogénéité de $\| \, . \, \|_W$, $m^+(S)\sigma_n$, qui est simple, converge vers m dans M_{o1}, ce qui achèvera la deuxième étape. Montrons que $\|P - P_n\|_W \to 0$. On pose $k = \operatorname{diam}(S)$. Soit $\epsilon > 0$. Comme la distance de Prohorov ρ (définition 11.1) métrise la convergence en loi (dans un métrique séparable) (proposition 11.3), en prenant δ suffisamment petit, il existe $N \in \mathbb{N}$ tel que, pour tout $n \geq N$,

$$\rho(P, P_n) + \delta < \min(\frac{\epsilon}{2}, \frac{\epsilon}{2k}).$$

Par le théorème 2.2 de Strassen, il existe deux variables aléatoires X et X_n de lois P et P_n respectivement telles que

$$\theta(X, X_n) \leq \rho(P, P_n) + \delta.$$

C'est à dire, par définition de θ,

$$\inf\{\alpha \in \mathbb{R} : Pr(d(X, X_n) > \alpha) \leq \alpha\} < \min(\frac{\epsilon}{2}, \frac{\epsilon}{2k})$$

où Pr désigne la probabilité sur l'univers sur lequel les variables X et X_n sont considérées. D'où, en notant π_n la loi du couple (X, X_n), ça nous donne

$$\inf\{\alpha \in \mathbb{R} : \pi_n(d(x, y) > \alpha) \leq \alpha\} < \min(\frac{\epsilon}{2}, \frac{\epsilon}{2k}).$$

Donc par définition de l'inf, il existe $\alpha \in \mathbb{R}$ tel que :

$$\pi_n(d(x, y) > \alpha) \leq \alpha \ \text{ et } \ \alpha < \min(\frac{\epsilon}{2}, \frac{\epsilon}{2k}).$$

Par conséquent :

$$\pi_n(d(x, y) > \frac{\epsilon}{2}) \leq \pi_n((d(x, y) > \alpha) \leq \frac{\epsilon}{2k}.$$

Finalement :

$$\begin{aligned} \int_S d(x, y)\pi_n(dx, dy) &= \int_S d(x, y)\pi_n(dx, dy) \\ &\leq \int_{\{d(x,y) \leq \frac{\epsilon}{2}\} \cap S} d(x, y)\, \pi_n(dx, dy) \\ &\qquad + \int_{\{d(x,y) > \frac{\epsilon}{2}\} \cap S} d(x, y)\, \pi_n(dx, dy) \\ &\leq \frac{\epsilon}{2}\pi_n(S \times S) + k\frac{\epsilon}{2k} = \epsilon \end{aligned}$$

et ce pour tout $n \geq N$, donc :

$$\int_S d(x,y)\pi_n(dx,dy) \longrightarrow 0.$$

D'autre part π_n, qui est la loi du couple (X, X_n), a pour marges P et P_n, qui sont des mesures probabilisées, d'où $\pi_n \in \mathcal{X}_{P-P_n}$ d'après la proposition 3.1. D'où finalement, on a $\|P - P_n\|_W \to 0$.
De la même manière on montrerait que $\|Q - Q_n\|_W \to 0$, par conséquent, en utilisant l'inégalité triangulaire, on obtient : $\|m' - \sigma_n\|_W \to 0$. Ce qui achève la seconde étape.
Conclusion
Les mesures simples sont denses dans les mesures à support borné pour la topologie de semi-norme $\| \,.\, \|_W$, et les mesures à support borné sont denses dans M_{o1} pour la même topologie. D'où les mesures simples sont denses dans M_{o1}. ♣.

4 Le dual M_{o1}^*

4.1 Définition de M_{o1}^*

L'espace M_{o1}^* designe le dual topologique de M_{o1}, c'est-à-dire l'ensemble des formes linéaires continues sur M_{o1}. L'espace M_{o1}^* sera naturellement munie de la norme duale $\| \,.\, \|_W^*$, qui est bien une norme d'après la proposition 9.1, définie par

$$\|\phi\|_W^* = \sup_{\|m\|_W \leq 1} |\phi(m)|.$$

4.2 Caractérisation de M_{o1}^*

Dans cette partie on va caractériser les éléments de ce dual. Pour cela on va établir un lien entre $L(S)$, l'espace des fonctions lipschtiziennes sur S, et M_{o1}^*

Définition 4.1 *Définissons l'application*

$$\begin{aligned} \Phi \quad : \quad & L(S) \longrightarrow M_{o1}^* \\ & f \longmapsto \Phi(f) = \phi_f \end{aligned}$$

où, pour tout $m \in M_{o1}$,

$$\phi_f(m) = \int_S f(x)\, m(dx).$$

Il convient d'abord de montrer que l'application Φ est bien définie.

Proposition 4.1 *L'application Φ est bien définie.*

Démonstration —Soit $f \in L(S)$. Montrons que $\Phi(f) = \phi_f \in M_{o1}^*$. Soit $m \in M_{o1}$. Pour tout $b \in \mathcal{X}_m$, on a

$$\begin{aligned}
\phi_f(m) &= \int_S f(x)\, m(dx) \\
&= \int_S f(x)\, (b(dx, S) - b(S, dx)) \\
&= \int_S f(x)\, b(dx, S) - \int_S f(y)\, b(S, dy) \\
&= \int_{S\times S} f(x)\, b(dx, dy) - \int_{S\times S} f(y)\, b(dy, dy) \\
&= \int_{S\times S} (f(x) - f(y))\, b(dx, dy)
\end{aligned}$$

d'où, pour tout $b \in \mathcal{X}_m$,

$$|\phi_f(m)| \leq \|f\|_L \int_{S\times S} d(x, y)\, b(dx, dy).$$

En passant à l'inf on obtient alors

$$|\phi_f(m)| \leq \|f\|_L \, \|m\|_W < \infty.$$

Donc ϕ_f est bien à valeurs réelles. D'autre part, l'inégalité ci-dessus montre que ϕ_f est continue avec $\|\phi_f\|_W^* \leq \|f\|_L$. Et par linéarité de l'intégrale il est immédiat que ϕ_f est linéaire. Finalement $\phi_f \in M_{o1}^*$. ♣

On va maintenant étudier cette application Φ de plus près. D'abord un résultat qui fait intervenir la structure de norme et semi-norme des espaces M_{o1}^* et $L(S)$ respectivement. C'est une sorte d'isométrie qui va être montrée maintenant.

Proposition 4.2 *Pour tout $f \in L(S)$ on a :*

$$\|\Phi(f)\|_W^* = \|\phi_f\|_W^* = \|f\|_L.$$

Démonstration — On a déja $\|\phi_f\|_W^* \leq \|f\|_L$, reste à montrer l'inégalité inverse. On vérifie facilement que $\delta_{(x,y)} \in \mathcal{X}_{\delta_x - \delta_y}$. Par conséquent :

$$\begin{aligned}
\|\delta_x - \delta_y\|_W &\leq \int_{S\times S} d(u, v)\, \delta_{(x,y)}(du, dv) \\
&= d(x, y).
\end{aligned}$$

Soit $\epsilon > 0$, par définition de $\| \,.\, \|_L$, il existe $x, y \in S$ distincts tels que

$$\|f\|_L - \epsilon \leq \frac{|f(x) - f(y)|}{d(x, y)} \leq \frac{|\phi_f(\delta_x - \delta_y)|}{\|\delta_x - \delta_y\|_W} \leq \|\phi_f\|_W^*$$

en faisant tendre ϵ vers 0 il vient $\|\phi_f\|_W^* \geq \|f\|_L$. ♣

Enfin, dernière propriété de Φ.

Proposition 4.3 *L'application* Φ *est surjective.*

Démonstration — Soit $\phi \in M_{o1}^*$, montrons qu'il existe $f \in L(S)$ tel que $\phi = \phi_f$. Soit $z \in S$, on pose

$$f(x) = \phi(\delta_x - \delta_z).$$

Alors, pour tout $x, y \in S$ on a,

$$\begin{aligned} |f(x) - f(y)| &= |\phi(\delta_x - \delta_y)| \\ &\leq \|\phi\|_W^* \|\delta_x - \delta_y\|_W \\ &\leq \|\phi\|_W^* \, d(x, y) \end{aligned}$$

d'où $f \in L(S)$ avec $\|f\|_L \leq \|\phi\|_W^*$.
Si $m = \delta_x - \delta_y$, alors

$$\phi_f(m) = f(x) - f(y) = \phi(\delta_x - \delta_y) = \phi(m).$$

Donc par linéarité de ϕ et ϕ_f, pour tout mesure simple m, $\phi_f(m) = \phi(m)$. Soit maintenant $m \in M_{o1}$, comme les mesures simples sont denses dans M_{o1}, il existe (m_n) une suite de mesures simples convergeant vers m dans M_{o1}. Donc par continuité de ϕ et ϕ_f on a

$$\phi(m_n) \to \phi(m) \quad , \quad \phi_f(m_n) = \phi(m_n) \to \phi_f(m)$$

et par unicité de la limite, pour tout $m \in M_{o1}$:

$$\phi(m) = \phi_f(m). \quad \clubsuit$$

Remarques
- L'application Φ n'est pas injective car si $f \in L(S)$ alors $f + cte \in L(S)$ avec $\Phi(f) = \Phi(f + cte)$ car $\Phi(cte) = 0$.
- La proposition 4.3 montre que l'on connait tous les élements de M_{o1}^* via les éléments de $L(S)$, on a en fait

$$M_{o1}^* = \{\phi_f \, : \, f \in L(S)\}.$$

On va établir une conséquence de tout cela.

Proposition 4.4 *Pour tout* $m \in M_{o1}$ *on a :*

$$\|m\|_W = \|m\|_L^*$$

Démonstration — Soit $m \in M_{o1}$, on a

$$\begin{aligned} \|m\|_W &= \sup_{\|\phi\|_W^* \leq 1} |\phi(m)| \qquad \text{(corollaire 9.3)} \\ &= \sup_{\|f\|_L \leq 1} |\phi_f(m)| \qquad \text{(proposition 4.2)} \\ &= \sup_{\|f\|_L \leq 1} \left| \int_S f \, dm \right| \\ &= \|m\|_L^*. \quad \clubsuit. \end{aligned}$$

5 Démonstration du théorème de Kantorovich-Rubinštein

On arrive maintenant au dénouement de notre étude. Le livre de Dudley [4, Lecture 20] ne permettait de démontrer que la première inégalité dans le théorème. C'est grâce à de Acosta [3] qu'a pu être démontrée la seconde inégalité.

Démonstration — Soient $P, Q \in \mathcal{P}_1(S,d)$. On pose $m = P - Q$. On va d'abord montrer que $m \in M_{o1}$. Pour cela on applique la décomposition de Jordan (théorème 7.1), on a alors $m = P' - Q'$. On a de plus $P' \leq P$ et $Q' \leq Q$. Soit $x \in S$, on a :

$$\begin{aligned}
\int_S d(x,y)\,|m|(dy) &= \int_S d(x,y)(P'+Q')(dy) \\
&\leq \int_S d(x,y)\,(P+Q)(dy) \\
&= \int_S d(x,y)\,P(dy) + \int_S d(x,y)\,Q(dy) \\
&< \infty \quad \text{car } P, Q \in \mathcal{P}_1(S,d).
\end{aligned}$$

Ainsi $m \in M_{o1}$. Donc d'après la proposition 4.1 :

$$\|P-Q\|_W = \|P-Q\|_L^*.$$

Reste à montrer que $\|P-Q\|_W = W(P,Q)$. On a vu dans la proposition 3.1 que l'ensemble des mesures sur $S \times S$ de marges P et Q est inclus dans $\mathcal{X}_{P-Q}$. On a donc

$$\|P-Q\|_W \leq W(P,Q).$$

Montrons maintenant l'autre inégalité. Soit $\epsilon > 0$, par définition de la norme de Wasserštein il existe $b \in M^+(S \times S)$ tel que $\pi_1 b - \pi_2 b = P - Q$ et

$$\int_{S\times S} d\,db \leq \|P-Q\|_W + \epsilon.$$

On a en fait

$$\pi_1 b - \pi_2 b = P - Q = P' - Q'.$$

Donc par minimalité de la décomposition de Jordan, il existe $\sigma, \sigma' \in M_1^+(S)$ telles que

$$\begin{aligned}
P = P' + \sigma \qquad & Q = Q' + \sigma \\
\pi_1 b = P' + \sigma' \qquad & \pi_2 b = Q' + \sigma'.
\end{aligned}$$

Et par la proposition 12.4, il vient

$$\begin{aligned}
W(P,Q) &= W(P',Q') \\
&= W(\pi_1 b, \pi_2 b) \\
&\leq \int_{S\times S} d\,db \\
&\leq \|P-Q\|_W + \epsilon.
\end{aligned}$$

Comme ϵ est arbitraire on a l'égalité manquante. ♣

Remarque

– Notons que d'après le corollaire 9.2, le supremum dans

$$\|m\|_W = \sup_{\|\phi\|_W^* \leq 1} |\phi(m)|$$

est atteint. Ce qui, d'après la proposition 4.4, signifie que le supremum dans

$$\|m\|_L^* = \sup_{\|f\|_L \leq 1} \left| \int_S f \, dm \right|$$

est atteint. Donc si $P, Q \in \mathcal{P}_1(S, d)$ alors d'après le théorème 2.1 de Kantorovich-Rubinštein, il existe $f \in L(S)$ avec $\|f\|_L \leq 1$ telle que

$$W(P, Q) = \int_S f \, d(P - Q).$$

Deuxième partie

Annexe

Dans tout le mémoire on travaille dans un espace (S, d) métrique, séparable. On fixe donc une bonne fois pour toute cet espace S muni de sa métrique d lui conférant une structure d'espace séparable. Par ailleurs S est muni de la tribu borélienne $\mathcal{B}$.

6 Un résultat préliminaire

On commence par rappeler le théorème suivant.

Théorème 6.1 (Lindelöf) *Si (S,d) est un espace métrique séparable. Alors de tout recouvrement ouvert on peut extraire un sous recouvrement dénombrable.*

Démonstration — Comme (S, d) est séparable alors la toplogie associée admet une base dénombrable. En effet si A est un sous-ensemble dénombrable dense dans S alors l'ensemble

$$\mathcal{D} = \{B(x, \epsilon) \ : \ x \in A \, , \ \epsilon \in \mathbb{Q}\}$$

convient. Soit maintenant $(U_i)_{i \in I}$ un recouvrement ouvert quelconque de S. On pose

$$\mathcal{C} = \{B \in \mathcal{D} \, / \, \exists i \in I \ : \ B \subset U_i\}$$

qui est dénombrable. Pour tout $B \in \mathcal{C}$ on pose i_B tel que $B \subset U_{i_B}$, alors l'ensemble $J = \{i_B \in I \, / \, B \in \mathcal{C}\}$ est dénombrable. Montrons que

$$\bigcup_{j \in J} U_j = \bigcup_{i \in I} U_i.$$

On a déjà $\subset$ car $J \subset I$. Soit $x \in \bigcup_{i \in I} U_i$, alors il existe $i \in I$ tel que $x \in U_i$. Il existe donc $B \in \mathcal{D}$ tel que $x \in B \subset U_i$ d'où $B \in \mathcal{C}$ et alors

$$x \in B \subset U_{i_B} \subset \bigcup_{j \in J} U_j. \quad \clubsuit$$

On cite maintenant un corollaire qui nous sera utile dans la suite.

Corollaire 6.1 *Si (S, d) est un espace métrique séparable alors on peut le recouvrir de boréliens deux à deux disjoints de diamètre inférieur à ϵ donné.*

Démonstration — Soit $\epsilon > 0$. Comme S est séparable, d'après le théorème 6.1 de Lindelöf, il existe $\left\{B_i = B(x_i, \frac{\epsilon}{2}) \ : \ x_i \in S \ , \ i \in \mathbb{N}\right\}$ un recouvrement ouvert de S. On pose maintenant $A_0 = B_0$ et, pour tout $n \geq 1$,

$$A_n = B_n \setminus \bigcup_{i=0}^{n-1} B_i.$$

Les A_n sont bien des boréliens deux à deux disjoints qui vérifient $\mathrm{diam}(A_n) \leq \epsilon$ et

$$S = \bigcup_{i \in \mathbb{N}} A_i = \bigcup_{i \in \mathbb{N}} B_i. \ \clubsuit$$

Dans notre étude on sera amené à intégrer la distance d. Pour pouvoir donner un sens à cela, il nous faut la proposition suivante.

Proposition 6.1 *La distance d est jointement mesurable sur le produit $S \times S$ muni de la tribu borélienne produit $\mathcal{B} \otimes \mathcal{B}$.*

Démonstration — L'application δ définie sur $S \times S$ par

$$\delta((u, v), (x, y)) = d(u, x) + d(v, y)$$

est une distance sur $S \times S$. Ceci est facile à vérifier car d est elle-même une distance sur S. Et par un résultat connu de topologie, la topologie associée à cette distance est la même que la topologie produit de S par S. Montrons maintenant que $(S \times S, \delta)$ est séparable. Comme S est séparable, il existe $D = \{x_i : i \in \mathbb{N}\}$ dense dans S. On pose alors $E = D \times D$. E est dénombrable, de plus il est dense dans $S \times S$. En effet, soit $(x, y) \in S \times S$ et $\epsilon > 0$, alors il existe $x_i, x_j \in D$ tels que $d(x_i, x) < \epsilon/2$ et $d(x_j, y) < \epsilon/2$ d'où $\delta((x, y), (x_i, x_j)) < \epsilon$. D'où E est séparable. D'autre part, l'inégalité

$$|d(x, y) - d(x', y')| \leq \delta((x, y), (x', y'))$$

obtenue par l'inégalité triangulaire montre que d est continue sur $(S \times S, \delta)$. On a donc par continuité de d, si U est un ouvert de $\mathbb{R}$ alors

$$d^{-1}(U) = \bigcup_{i \in I} U_i \times V_i \quad U_i, V_i \in S.$$

Comme S est séparable, par le théorème de Lindelöf, énoncé ci-dessus, on peut en extraire un sous recouvrement dénombrable et donc $d^{-1}(U) \in \mathcal{B} \otimes \mathcal{B}$ et ceci pour tout ouvert U de $\mathbb{R}$. D'où d est $\mathcal{B} \otimes \mathcal{B}$-mesurable. $\clubsuit$

Fort de cette proposition et du fait que $d \geq 0$, on pourra par la suite définir sans ambiguité, pour une mesure sur $\mathcal{B} \otimes \mathcal{B}$, l'intégrale

$$\int_{S \times S} d(x, y) \, b(dx, dy).$$

7 Quelques résultats sur les mesures

7.1 Décomposition de mesures

Définition 7.1 *Soit m une mesure signée finie sur S. On appelle support de m le plus petit fermé de complémentaire de mesure nulle.*

Voici un résultat de décomposition de mesures qui fait l'objet de la proposition 11.1.1 dans [2].

Théorème 7.1 (Décomposition de Jordan) *Soit m une mesure finie signée sur $\mathcal{B}$ alors il existe un unique couple (m^+, m^-) de mesures sur $\mathcal{B}$ tel que*
- *m^+, m^- soient des mesures positives*
- *m^+, m^- soient mutuellement singulières*
- *$m = m^+ - m^-$.*

De plus cette décomposition est minimale, c'est-à-dire que si $m = m_1 - m_2$ alors $m^+ \leq m_1$ et $m^- \leq m_2$.

7.2 Mesures tendues et propriétés

Définition 7.2 *Soit T un espace topologique. Une mesure de probabilité sur les boréliens de T est dite tendue si pour tout $\epsilon > 0$ il existe un compact K tel que $P(K^c) < \epsilon$. Un ensemble A de lois de probabilité sur T est dit uniformément tendu si pour tout $\epsilon > 0$ il existe un compact K tel que pour tout $P \in A$, $P(K^c) < \epsilon$.*

Théorème 7.2 (Ulam) *Soit (T, d) un espace métrique séparable complet et μ une mesure positive σ-finie sur la tribu borélienne de T. Alors pour tout ensemble borélien A on a*

$$\mu(A) = \sup\{\mu(K) \ : \ K \subset A \, , \ K \text{ compact}\}.$$

Ce théorème ne sera pas démontré, il fait l'objet du théorème 5.1 dans [4]. Ce n'est pas tant ce théorème qui nous sera utile dans l'exposé mais le corollaire qui suit, dont on ne donne pas la démonstration, qui est immédiate.

Corollaire 7.1 *Dans un espace métrique séparable toute mesure probabilisée est tendue.*

On va donner une caractérisation importante des ensembles uniformément tendus dans la propositon qui suit. Celle-ci fait l'objet du théorème 6.1 dans [1].

Proposition 7.1 (Critère de Prokhorov) *Soit T un espace topologique. Soit A un ensemble de probabilités sur T. Alors A est uniformément tendu si et seulement si A est relativement compact pour la topologie de la convergente étroite (définition 11.4).*

8 Les espaces considérés

Enonçons maintenant quelques définitions concernant les objets dont nous aurons besoin.

Définition 8.1 *On pose :*

- $M(S)$ *(resp.* $M(S \times S)$*) l'ensemble des mesures finies signées sur* S *(resp.* $S \times S$*)*
- $M^+(S)$ *(resp.* $M^+(S \times S)$*) l'ensemble des mesures finies positives sur* S *(resp.* $S \times S$*)*
- $M_1^+(S,d) = \{\mu \in M^+(S) : \exists\, x \in S \int_S d(x,y)\,\mu(dy) < \infty\}$
- $M_{1\alpha}^+ = \{\mu \in M_1^+(S,d) \ : \ \mu(S) = \alpha\}$.

La définition qui suit est l'analogue de la précédente mais pour des mesures probabilisées.

Définition 8.2 *On pose*

- $\mathcal{P}(\mathcal{S})$ *(resp.* $\mathcal{P}(S \times S)$*) l'ensemble des mesures probabilisées sur* S *(resp.* $S \times S$*)*
- $\mathcal{P}_1(S,d) = \{P \in \mathcal{P}(S) : \exists\, x \in S \int_S d(x,y)\, P(dy) < \infty\}$.

Remarques

- Si (S,d) est borné alors $M_1^+(S,d) = M^+(S,d)$, et donc $\mathcal{P}_1(S,d) = \mathcal{P}(S)$, car alors pour tout $\mu \in M^+(S)$

$$\int_S d(x,y)\,\mu(dy) \leq MP(S) = M$$

  où M est la constante qui borne d.
- Dans la définition de $M_1^+(S,d)$ et de $\mathcal{P}_1(S,d)$, $\exists\, x \in S$ est équivalent à $\forall\, x \in S$, ceci étant facile à montrer par l'inégalité triangulaire.

Les fonctions lipschitziennes jouent un rôle très important dans la suite, on pose donc la définition suivante.

Définition 8.3 *On appelle* $L(S)$ *l'ensemble des fonctions réelles lipschitziennes définies sur* S*. Pour tout* $f \in L(S)$ *on pose*

$$\|f\|_L = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x,y)}.$$

On note $L_B(S)$ *l'ensemble des fonctions réelles lipschtiziennes bornées définies sur* S*. Pour tout* $f \in L_B(S)$*, on pose*

$$\|f\|_{L_B} = \|f\|_\infty + \|f\|_L.$$

Remarque

- On montre facilement que $L(S)$ muni de $\| \ \|_L$ est un espace vectoriel semi-normé et que $L_B(S)$ muni de $\| \ \|_{L_B}$ est un espace vectoriel normé.

Définition 8.4 *Soit μ une mesure signée finie sur $\mathcal{B}$, on pose :*

$$\|\mu\|_L^* = \sup_{\|f\|_L \le 1} \left| \int_S d\, f \mu \right|.$$

9 Un résultat d'analyse fonctionnelle

Avant de passer à l'objet-même de cette partie qui est le théorème de Hahn-Banach et ses conséquences, on va indiquer le cadre dans lequel on se place. On prend $(E, \| \, . \, \|_E)$ un espace vectoriel semi-normé muni d'une seule semi-norme. On va s'intéresser au dual topologique E', c'est-à-dire l'ensemble des formes linéaires continues sur E. Pour $u \in E'$, on pose

$$\|u\|_{E'} = \sup\{|u(x)| \; : \; x \in E \, , \; \|x\|_E \le 1\}.$$

On a le résultat suivant.

Proposition 9.1 *L'application $\| \, . \, \|_{E'}$ est une norme sur E'.*

Démonstration — Soit $u \in E'$. Comme u est continue, il existe $C > 0$ tel que $|u(x)| \le C\|x\|_E$ pour tout $x \in E$. Si $\|x\|_E = 0$, alors $u(x) = 0$, sinon on a

$$0 \le u \left(\frac{x}{\|x\|_E} \right) \le C$$

donc en passant au sup, il vient $0 \le \|u\|_{E'} \le C < \infty$. L'homogénéité et l'inégalité triangulaire sont immédiates. Reste un dernier point, soit $u \in E'$ tel que $\|u\|_{E'} = 0$ c'est-à-dire que pour tout $x \in E$ tel que $\|x\|_E \le 1$ on a $u(x) = 0$. Soit maintenant $x \in E$. Si $\|x\|_E = 0$ alors par la caractérisation de u continue avec la constante C il vient $u(x) = 0$. Si $\|x\|_E > 0$ alors $\left\| \frac{x}{\|x\|_E} \right\|_E \le 1$ donc $u \left(\frac{x}{\|x\|_E} \right) = 0$ et par linéarité de u il vient $u(x) = 0$. D'où finalement $u(x) = 0$ pour tout $x \in E$. ♣

On rappelle l'énoncé du théorème de Hahn-Banach dans le cadre où il nous intéresse. On n'en donne pas la démonstration qui est faite dans le cours de maîtrise.

Théorème 9.1 (Hahn-Banach) *Soit E un espace vectoriel et p une semi-norme sur E. Soit F un sous-espace vectoriel de E et u une forme linéaire sur F telle que $u \le p$ sur F. Alors il existe $\tilde{u}$ forme linéaire sur E telle que :*

$$\tilde{u}_{|F} = u \quad \text{et} \quad |\tilde{u}| \le p \;\; \text{dans} \;\; E.$$

On commence par citer ce premier corollaire qui n'est autre qu'une généralisation d'un résultat de maîtrise vu dans le cadre de E normé.

Corollaire 9.1 *Soit $(E, \| \,.\, \|_E)$ un espace vectoriel muni de l'unique semi-norme $\| \,.\, \|_E$. Soit F un sous-espace vectoriel de E et $u \in F'$. Alors il existe $\tilde{u} \in E'$ telle que*
$$\tilde{u}_{|F} = u \quad \text{et} \quad \|\tilde{u}\|_{E'} = \|u\|_{E'}.$$

Démonstration — Soit donc $u \in F'$ il est facile de voir que pour tout $x \in F$, on a $|u(x)| \leq \|u\|_{E'}\|x\|_E$. On pose alors pour tout $x \in E$, $p(x) = \|u\|_{E'}\|x\|_E$. Il est facile de vérifier que p est une semi-norme sur E. Donc d'après le théorème 9.1 de Hahn-Banach, il existe $\tilde{u} \in E'$ tel que :

$$\tilde{u}_{|F} = u \;\; \text{et} \;\; |\tilde{u}(x) \leq \|u\|_{E'}\|x\|_E \;\; \forall x \in E.$$

D'où il vient $\|\tilde{u}\|_{E'} \leq \|u\|_{E'}$. De plus comme $\tilde{u}_{|F} = u$ on a l'égalité inverse. ♣

Donnons un deuxième corollaire.

Corollaire 9.2 *Soit $(E, \| \,.\, \|_E)$ l'espace vectoriel semi-normé défini au-dessus. Soit $x \in E$, alors il existe $u \in E'$ tq $\|u\|_{E'} = 1$ et $u(x) = \|x\|_E$.*

Démonstration — Considérons l'application :

$$v : tx \in Vect\{x\} \longmapsto t\|x\|_E.$$

On vérifie facilement que $v \in (Vect\{x\})'$, donc d'après le corollaire 9.1 il existe $u \in E'$ tel que
$$u_{|Vect\{x\}} = v \;\; \text{et} \;\; \|u\|_{E'} = \|v\|_{E'}.$$
On a $u(x) = v(x) = \|x\|_E$ et

$$\|u\|_{E'} = \|v\|_{E'} = \sup_{\|tx\|_E \leq 1} |\, t\|x\|_E \,| = 1. \quad ♣$$

On en arrive maintenant au résultat essentiel de cette qui section. C'est ce résultat qui sera utilisé dans la démonstration du théorème de Kantorovich-Rubinštein.

Corollaire 9.3 *Soit $x \in E$, alors*

$$\|x\|_E = \sup_{\|u\|_{E'} \leq 1} |u(x)|.$$

Démonstration — Soit $x \in E$. Pour tout $u \in E'$ tel que $\|u\|_{E'} \leq 1$, par continuité de u on a $|u(x)| \leq \;\|x\|_E$, d'où en passant au sup il vient

$$\|x\|_E \geq \sup_{\|u\|_{E'} \leq 1} |u(x)|.$$

Et d'après le corollaire 9.2 ce sup est atteint d'où l'égalité. ♣

10 Structures métriques sur les espaces de variables aléatoires

La définition qui suit généralise celle que l'on connait des espaces $L^p(\mathbb{R})$ aux espaces métriques séparables. Ces espaces étant munies d'une distance, non pas d'une norme comme dans $\mathbb{R}$, permettent de définir des modes de convergence de variables aléatoires.

Définition 10.1 *Pour tout $0 < p < \infty$ on définit $L^p(\Omega, S) = L^p(\Omega, \mathcal{B}, P, S, d)$ par l'ensemble des variables aléatoires sur l'espace probabilisé $(\Omega, \mathcal{B}, P)$ à valeurs dans S telles qu'il existe $s \in S$ tel que :*

$$\int_\Omega d(X, s)^p \, dP.$$

Pour $p = 0$, on pose $L^0(\Omega, S)$ l'ensemble des variables aléatoires à valeurs dans S. Pour $X, Y \in L^p$ on définit les distances

$$d_p(X, Y) = \int_\Omega d(X, Y)^p \, dP \quad \text{si } 0 < p < 1$$

$$d_p(X, Y) = (\int_\Omega d(X, Y)^p \, dP)^{\frac{1}{p}} \quad \text{si } 1 \leq p \leq \infty$$

$$d_0(X, Y) = \kappa(X, Y) = \int_\Omega f(d(X, Y)) \, dP \quad \text{où } f(x) = \frac{x}{1 + x}.$$

Remarques

- La distance κ définie ci-dessus est la distance de Ky-Fan, et on peut montrer que cette distance métrise la convergence en probabilité.
- Il existe un espace probabilisé $(\Omega, \mathcal{B}, P)$ tel que $\mathcal{P}_1(S, d) = \{L(X) : X \in L^1(\Omega, \mathcal{B}, P, S, d)\}$ où $L(X)$ est la loi de X définie par $L(X)(A) = P(X^{-1}(A))$ pour tout $A \in \mathcal{B}$.
- Pour ce qui est des modes de convergence, on dit que $X_n \xrightarrow{L^p} X$ si $d_p(X_n - X) \to 0$.

Définition 10.2 *Soient X,Y deux variables aléatoires, on pose*

$$\theta(X, Y) = \inf\{\alpha : Pr(d(X, Y) > \alpha) \leq \alpha\}.$$

Ces distances étant maintenant données, il est souvent intéressant de pouvoir les comparer entre elles.Un cas idyllique étant de pouvoir montrer qu'elles sont équivalentes, ici ce n'est pas le cas. Mais les inégalités qui suivent montrent tout de même grâce à une des remarques précédentes que θ métrise la convergence en probabilité.

Proposition 10.1 *Si X,Y sont deux variables aléatoires alors :*

$$\kappa(X, Y) \leq 2\theta(X, Y) \leq 4\sqrt{\kappa(X, Y)}.$$

Ce résultat est montré dans [4, Lecture 18].

11 Structures métriques sur les espaces de mesures

A l'image de distances construites entre des variables aléatoires, il est aussi souvent utile de considérer des distances entre des mesures probabilisées. Essentiellement, on ne considérera que des métriques sur les espaces de mesures probabilisées car ce sont celles-ci qui nous intéressent dans les résultats. Cependant pour mener nos démonstrations il nous faut étendre certaines d'entre elles à des espaces de mesures non probabilisées. Quand ceci sera nécessaire on l'indiquera en remarque. On munit maintenant les espaces $\mathcal{P}$ et $\mathcal{P}_1$ des distances suivantes.

Définition 11.1 (Distance de Prohorov) *Soient $P, Q \in \mathcal{P}(S)$, on définit la distance de Prohorov par :*

$$\rho(P,Q) = \inf\{\epsilon > 0 : P(A) \leq Q(A^\epsilon) + \epsilon \ \ \forall A \in \mathcal{B}\}$$

où $A^\epsilon = \{y \in S : \exists\, x \in A, d(x,y) < \epsilon\}$.

On cite maintenant deux distances qui font intervenir les fonctions lipschitziennes et leur topologie de norme ou semi-norme définie plus haut.

Définition 11.2
- *Si $P, Q \in \mathcal{P}(S)$ on pose*

$$\beta(P,Q) = \sup_{\|f\|_{L_B} \leq 1} \left| \int_S f\, d(P-Q) \right|.$$

- *Si $P, Q \in \mathcal{P}_1(S)$ on pose*

$$F(P,Q) = \|P-Q\|_L^* = \sup_{\|f\|_L \leq 1} \left| \int_S f\, d(P-Q) \right| \quad \text{(Distance de Fortet-Mourier)}.$$

La distance de Fortet-Mourier est définie dans [6].

Proposition 11.1 *Les applications β et F sont des distances sur $\mathcal{P}(S)$ et $\mathcal{P}_1(S)$ respectivement.*

Démonstration — Si $P, Q \in \mathcal{P}(S)$. On a clairement $\beta(P,Q) \geq 0$, la symétrie et l'inégalité triangulaire. Comme f est bornée par 1 on a aussi, $\beta(P,Q) < \infty$. Supposons maintenant que $\beta(P,Q) = 0$ et montrons que $P = Q$. Soit F un fermé de S, on pose, pour tout $n \in \mathbb{N}$,

$$f_n(x) = \max(0, 1 - nd(x,F)).$$

Alors $\|f_n\|_\infty = 1$ et $\|f_n\|_L \leq n$ d'où $\|f_n\|_{L_B} \leq \infty$, et donc :

$$\int_S f_n\, dP = \int_S f_n\, dQ \ \ \forall n \in \mathbb{N}.$$

Comme $f_n \searrow \chi_F$ le théorème de convergence monotone permet de conclure $P(F) = Q(F)$, et ceci pour tout fermé de S. D'où $P = Q$. Finalement β est bien une distance.

Passons à F, soient $P, Q \in \mathcal{P}_1(S)$. tout se fait exactement comme avant, le seul point délicat est de montrer que $F(P,Q) < \infty$. Soit $f \in L(S)$ telle que $\|f\|_L \leq 1$ et $y \in S$, on a alors

$$\begin{aligned}
\left| \int_S f(x)\, d(P-Q)(x) \right| &= \left| \int_S f(x) - f(y)\, d(P-Q)(x) + \underbrace{\int_S f(y)\, d(P-Q)(x)}_{=0} \right| \\
&\leq \left| \int_S d(x,y)\, d(P-Q)(x) \right| \\
&< \infty \quad \text{car } P, Q \in \mathcal{P}_1.
\end{aligned}$$

F est donc bien une distance. ♣

Remarque

- Pour tout $\alpha > 0$, on peut étendre la distance F à $\tilde{F}$ à l'ensemble $M^+_{1\alpha}$. $\tilde{F}$ est une distance et par commodité on l'appellera encore F.

Définition 11.3 (Distance de Wasserštein) *Si P,$Q \in \mathcal{P}_1(S,d)$, on pose :*

$$\begin{aligned}
W(P,Q) &= \inf\{d_1(X,Y) : L(X) = P, L(Y) = Q\} \\
&= \inf_{\pi \in \mathcal{P}(S\times S)} \{ \int_{S\times S} d(x,y)\, \pi(dx,dy) : \ \pi \text{ de marges } P \text{ et } Q \ \}.
\end{aligned}$$

On ne montrera pas que W est effectivement une distance.

Remarque

- Pour tout $\alpha > 0$, on peut étendre la distance W en $\tilde{W}$ en posant, pour tout $\mu, \nu \in M^+_\alpha$

$$\tilde{W}(\mu,\nu) = \inf_{\pi \in M^+(S\times S)} \{ \int_{S\times S} d(x,y)\, \pi(dx,dy) : \ \pi \text{ de marges } \mu \text{ et } \nu \ \}$$

$\tilde{W}$ est encore une distance et par commodité on l'appellera encore W.

On rappelle maintenant un mode convergence des mesures de probabilités.

Définition 11.4 *On dit qu'une suite P_n de mesures probabilisées sur S converge en loi (ou étroitement) vers une autre mesure probabilisée P, et l'on note $P_n \xrightarrow{\mathcal{L}} P$, si pour toute fonction réelle f définie sur S continue et bornée, on a*

$$\int_S f\, dP_n \longrightarrow \int_S f\, dP.$$

Comme pour les métriques sur les variables aléatoires, on peut aussi comparer ces distances et les modes de convergences sous–jacents.

Proposition 11.2 *Soient P,Q $\in \mathcal{P}(S)$ alors :*

$$\beta(P,Q) < 2\rho(P,Q) < 4\sqrt{\beta(P,Q)}.$$

Proposition 11.3 *Soient P_n, P des mesures probabilisées sur S, alors les conditions suivantes sont équivalentes :*

1. $P_n \xrightarrow{\mathcal{L}} P$
2. $\beta(P_n, P) \longrightarrow 0$
3. $\rho(P_n, P) \longrightarrow 0$

Les deux propositions précédentes sont prouvées dans [4, Lecture 8].

Soit μ une mesure sur $\mathcal{B}$. Il peut être intéressant de trouver des suites de mesures qui convergent vers μ dans un certain sens. C'est ce qui fait l'objet du résultat suivant. Il existe un espace probabilisé (Ω, P) sur lequel on peut trouver une suite de variables aléatoires (X_j) toutes de loi μ. On définit alors les mesures empiriques μ_n par

$$\mu_n(A)(w) = \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j(w)}(A), \quad A \in \mathcal{B}\ ,\ \omega \in \Omega.$$

Théorème 11.1 (Glivenko-Cantelli-Varadarajan) *Si μ une mesure sur $\mathcal{B}$ alors les mesures empiriques (μ_n) convergent presque surement vers μ c'est à dire :*

$$Pr(\{\omega\ :\ \mu_n(.)(\omega) \xrightarrow{\mathcal{L}} \mu\}) = 1.$$

Ce résultat est le théorème 9.1 est démontré dans [4].

Remarquons que pour un ω adéquat (il en existe), on a

$$\mu_n(.)(\omega) \xrightarrow{\mathcal{L}} \mu.$$

C'est sous cette forme que sera utilisé le théorème dans l'exposé.

12 Propriétés de la distance de Wasserštein

On consacre maintenant une partie à la distance W. En effet, cette distance a la propriété très intéressante d'être invariante par translation. Ce que Dudley ne mentionne pas dans [4], mais que de Acosta a signalé dans [3], ce qui lui a permis d'achever la démonstration du théorème de Kantorovich-Rubinštein.

Avant toute chose, on appelle π_1 et π_2 les deux projections de $S \times S$ sur S. Si $b \in M(S \times S)$ les marges de b qui ne sont autres que les mesures images de b par π_1 et π_2 seront notées $\pi_1 b$ et $\pi_2 b$.

La première propriété est la suivante.

Proposition 12.1 *Soient $\mu, \nu \in M^+_{1\alpha}$ et $\lambda \in M^+(S)$, alors*

$$W(\mu + \lambda, \nu + \lambda) \leq W(\mu, \nu).$$

Démonstration — Soit l'application $h : S \to S \times S$ définie par $h(x) = (x,x)$. Soit $\epsilon > 0$, par définition de W on peut choisir $\gamma \in M^+(S \times S)$ telle que $\pi_1\gamma = \mu \quad , \quad \pi_2\gamma = \nu$ et

$$\int d\,d\gamma \leq W(\mu,\nu) + \epsilon.$$

On pose $\tilde{\gamma} = \gamma + \lambda \circ h^{-1}$. On a alors

$$\begin{aligned}
\pi_1\tilde{\gamma}(A) &= \tilde{\gamma}(A \times S)\\
&= \gamma(A \times S) + \lambda(h^{-1}(A \times S))\\
&= \pi_1\gamma(A) + \lambda(A)\\
&= (\mu + \lambda)(A).
\end{aligned}$$

De même on a $\pi_2\tilde{\gamma} = \nu + \lambda$. Par ailleurs, on a

$$\begin{aligned}
\int_{S\times S} d(x,y)\,\tilde{\gamma}(dx,dy) &= \int_{S\times S} d(x,y)\,\gamma(dx,dy) + \int_{S\times S} d(x,y)\,d(\lambda \circ h^{-1})(dx,dy)\\
&= \int_{S\times S} d(x,y)\,\gamma(dx,dy) + \int_{S} (d \circ h)(x)\,\lambda(dx)\\
&= \int_{S\times S} d(x,y)\,\gamma(dx,dy) \quad \text{car } d(x,x) = 0.
\end{aligned}$$

Par conséquent :

$$W(\mu + \lambda, \nu + \lambda) \leq \int_{S\times S} d\,d\tilde{\gamma} \leq W(\mu,\nu) + \epsilon.$$

Or ϵ est arbitraire, la proposition est donc démontrée. ♣

Le but va être maintenant de montrer l'inégalité inverse. C'est un peu plus compliqué, on va démontrer les propositions intermédiaires suivantes.

Proposition 12.2 *Soient $\mu, \nu \in M^+_{1\alpha}$. Soient $c > 0$ et $x \in S$, alors on a*

$$W(\mu,\nu) = W(\mu + c\delta_x, \nu + c\delta_x).$$

Démonstration — Soit $\epsilon > 0$, par définition de W on peut choisir $\gamma \in M^+(S \times S)$ tel que $\pi_1\gamma = \mu + c\delta_x$, $\pi_2\gamma = \nu + c\delta_x$ et

$$\int_{S\times S} d\,d\gamma \leq W(\mu + c\delta_x, \nu + c\delta_x) + \epsilon.$$

Le but est de construire une mesure $\tilde{\gamma} \in M^+(S \times S)$ telle que $\pi_1\tilde{\gamma} = \mu$, $\pi_2\tilde{\gamma} = \nu$ et $\int d\,d\tilde{\gamma} \leq \int d\,d\gamma$. Ce qui achèvera une inégalité car ϵ est arbitraire, l'autre inégalité découlant de la proposition précédente. On pose $p = (x,x)$ et $m = \gamma(\{p\})$.

Cas 1 Si $c \leq m$. On pose $\tilde{\gamma} = \gamma - c\delta_p$. $\tilde{\gamma} \in M^+(S \times S)$ car

– si $p \notin C$, $\tilde{\gamma}(C) = \gamma(C) \geq 0$

– si $p \in C$, $\tilde{\gamma}(C) = \gamma(C \setminus \{p\}) + m - c \geq 0$.

On a :

$$\begin{aligned}\pi_1\tilde{\gamma}(A) &= \mu(A) + c\delta_x(A) - c\delta_p(A\times S)\\ &= \mu(A).\end{aligned}$$

De même $\pi_2\tilde{\gamma} = \nu$. Enfin, on a,

$$\int_{S\times S} d\,d\tilde{\gamma} = \int_{S\times S} d\,d\gamma - cd(x,x) = \int_{S\times S} d\,d\gamma.$$

Cas 2 Si $c > m$. On définit les deux ensembles de $S\times S$ suivants :

$$I_x = \pi_1^{-1}(\{x\}) \setminus \{p\} \text{ et } I^x = \pi_2^{-1}(\{x\}) \setminus \{p\}.$$

On remarque que $\gamma(I_x) = \gamma(\pi_1^{-1}(\{x\})) - m \geq c - m > 0$, de même $\gamma(I^x) > 0$. On peut donc définir les deux mesures probabilisées sur S suivantes :

$$\beta_1(A) = \frac{\gamma(I^x \cap \pi_1^{-1}(A))}{\gamma(I^x)} \text{ et } \beta_2(A) = \frac{\gamma(I_x \cap \pi_2^{-1}(A))}{\gamma(I_x)}.$$

On a

$$\gamma(I_x)(\delta_x \otimes \beta_2)(A\times B) = \delta_x(A)\gamma(I_x \cap \pi_2^{-1}(B)).$$

On veut montrer que

$$\gamma(I_x)(\delta_x \otimes \beta_2)(A\times B) = \gamma(I_x \cap (A\times B)).$$

- Si $x \notin A$ alors $I_x \cap (A\times B) = \emptyset$ d'où l'égalité.
- Si $x \in A$ alors $I_x \cap \pi_2^{-1}(B) = I_x \cap (A\times B)$, d'où l'égalité.

De même on a : $\gamma(I^x)(\beta_1 \otimes \delta_x)(A\times B) = \gamma(I^x \cap (A\times B))$. On pose maintenant $q = c - m$ et :

$$\tilde{\gamma} = \gamma - q\frac{\gamma(I_x \cap\ .\ )}{\gamma(I_x)} - q\frac{\gamma(I^x \cap\ .\ )}{\gamma(I^x)} - m\delta_p + q(\beta_1 \otimes \beta_2).$$

On a $\tilde{\gamma} \in M^+(S\times S)$ et d'après ce qui a été fait juste au-dessus, on a en fait :

$$\tilde{\gamma} = \gamma - q(\delta_x \otimes \beta_2) - q(\beta_1 \otimes \delta_x) - m\delta_p + q(\beta_1 \otimes \beta_2).$$

Il vient donc :

$$\pi_1\tilde{\gamma} = \mu + c\delta_x - q\delta_x - q\beta_1 - m\delta_x + q\beta_1 = \mu.$$

De même $\pi_2\tilde{\gamma} = \nu$. Reste à montrer que $\displaystyle\int_{S\times S} d\,d\tilde{\gamma} \leq \int_{S\times S} d\,d\gamma$. On a

$$\begin{aligned}\int_{S\times S} d(z,y)\,(\beta_1\otimes\beta_2)(dz,dy) &\leq \int_{S\times S} d(z,x)\,(\beta_1\otimes\beta_2)(dz,dy)\\ &\quad + \int_{S\times S} d(x,y)\,(\beta_1\otimes\beta_2)(dz,dy)\\ &= \int_S d(z,x)\,\beta_1(dz) + \int_S d(x,y)\,\beta_2(dy)\\ &= \int_{S\times S} d(z,y)\,(\beta_1\otimes\delta_x)(dz,dy)\\ &\quad + \int_{S\times S} d(z,y)\,(\delta_x\otimes\beta_2)(dz,dy)\end{aligned}$$

Par conséquent :

$$\begin{aligned}\int_{S\times S} d\,d\tilde\gamma &= \int_{S\times S} d\,d\gamma - q\int_{S\times S} d\,d(\delta_x\otimes\beta_2) - q\int_{S\times S} d\,d(\beta_1\otimes\delta_x) + q\int_{S\times S} d\,d(\beta_1\otimes\beta_2)\\ &\leq \int_{S\times S} d\,d\gamma.\end{aligned}$$

Ce qui achève la démonstration. ♣

On dit qu'une mesure $\mu\in\mathcal{P}_1(S,d)$ est simple s'il existe $x_1,\dots,x_n\in S$ et $\alpha_1,\dots,\alpha_n\in\mathbb{R}$ tels que $\alpha_1+\cdots+\alpha_n=1$ et $\mu=\displaystyle\sum_{i=1}^n\alpha_i\delta_{x_i}$. Ces mesures font l'objet de la proposition suivante.

Proposition 12.3 *L'ensembles des mesures simples de $\mathcal{P}_1(S,d)$ est dense dans $\mathcal{P}_1(S,d)$ pour la distance W.*

Démonstration — Soit $y_o\in S$, $\epsilon>0$ et $\mu\in\mathcal{P}_1(S,d)$. $d(\,.\,,y_0)$ est μ-intégrable et d'après le corollaire 6.1 il existe des boréliens A_n deux à deux disjoints de diamètre inférieur à $\dfrac{\epsilon}{2}$ tels que $S=\displaystyle\bigcup_{i\in\mathbb{N}}A_i$. Donc :

$$\int_{\bigcup_{i=1}^n A_i} d(x,y_0)\,\mu(dx)\longrightarrow\int_S d(x,y_0)\,\mu(dx).$$

Pour n assez grand, on a donc

$$\int_D d(x,y_0)\,\mu(dx)<\frac{\epsilon}{2}\quad\text{où}\quad D=\left(\bigcup_{i=1}^n A_i\right)^c.$$

Pour $y\in S$ on définit $h_y(x)=(x,y)$ pour tout $x\in S$. Pour $j=1,\dots,n$ on choisit $y_j\in A_j$ et on définit $\gamma\in\mathcal{P}(S\times S)$ par

$$\gamma(E)=\sum_{j=1}^n\mu(A_j\cap h_{y_j}^{-1}(E))+\mu(D\cap h_{y_o}^{-1}(E))\,,\quad\forall E\in\mathcal{B}\otimes\mathcal{B}.$$

Comme pour tout $y\in S$ on a $\pi_1\circ h_y=Id_S$, on a, pour tout $A\in\mathcal{B}$,

$$\pi_1\gamma(A)=\gamma(\pi_1^{-1}(A))=\sum_{j=1}^n\mu(A_j\cap A)+\mu(D\cap A)=\mu(A).$$

Comme pour tout $x\in S$ on a $(\pi_2\circ h_y)(x)=y$, on a pour tout $A\in\mathcal{B}$:

$$\pi_2\gamma(A)=\gamma(\pi_2^{-1}(A))=\sum_{j=1}^n\mu(A_j)\delta_{y_j}(A)+\mu(D)\delta_{y_0}(A).$$

Ainsi :

$$\begin{aligned}
\int_{S\times S} d\,d\gamma &= \sum_{j=1}^{n}\int_{A_j}(d\circ h_{y_j})(x)\,\mu(dx) + \int_D (d\circ h_{y_0})(x)\,\mu(dx)\\
&= \sum_{j=1}^{n}\int_{A_j} d(x,y_j)\,\mu(dx) + \int_D d(x,y_0)\,\mu(dx)\\
&\leq \frac{\epsilon}{2}\sum_{j=1}^{n}\mu(A_j) + \frac{\epsilon}{2}\\
&\leq \epsilon.
\end{aligned}$$

Finalement

$$W(\mu, \sum_{j=1}^{n}\mu(A_j)\delta_{y_j} + \mu(D)\delta_{y_0}) \leq \epsilon \quad \clubsuit.$$

Remarque

– La proposition précédente reste évidemment valable pour $\mu \in M^+_{1\alpha}(S)$.

On conclut cette partie par cette dernière proposition qui établit ce qui est voulu, à savoir que W est invariante par translation.

Proposition 12.4 *Pour tout $\mu, \nu \in M^+_{1\alpha}(S)$ et pour tout $\lambda \in M^+_{1\beta}(S)$, on a*

$$W(\mu,\nu) = W(\mu+\lambda, \nu+\lambda).$$

Démonstration — Par la proposition 12.1 on a déjà $\geq$. Par la proposition 12.3 il existe une suite (λ_n) de mesures simples de $M^+_{1\beta}(S)$ telles que $W(\lambda_n, \lambda) \to 0$. Par la proposition 12.2 et une récurrence simple, on obtient

$$W(\mu+\lambda_n, \nu+\lambda_n) = W(\mu,\nu).$$

Par conséquent :

$$\begin{aligned}
W(\mu,\nu) &= W(\mu+\lambda_n, \nu+\lambda_n)\\
&\leq W(\mu+\lambda_n, \mu+\lambda) + W(\mu+\lambda, \nu+\lambda) + W(\nu+\lambda, \nu+\lambda_n)\\
&\leq W(\mu+\lambda, \nu+\lambda) + 2W(\lambda, \lambda_n) \quad \text{(prop. 12.1)}
\end{aligned}$$

En faisant $n \to \infty$ il vient l'inégalité manquante. $\clubsuit$

Références

[1] Billingsley (Patrick). – *Convergence of Probability Measures.* – New York, London, J. Wiley, 1968.

[2] Bouziad (Ahmed) et Calbrix (Jean). – *Théorie de la Mesure et de l'Intégration.* – Rouen, France, Publications de l'Université de Rouen, 1993.

[3] de Acosta (Alejandro). – Invariance principles in probability for triangular arrays of B–valued random vectors and some applications. *Ann. Probab.*, vol. 10, n° 2, 1982, pp. 346–373.

[4] Dudley (Richard M.). – *Probabilities and Metrics.* – Aarhus, Denmark, Matematisk Institut, Aarhus Universitet, 1976, *Lecture Notes Series.*

[5] Dudley (Richard M.). – *Real Analysis and Probability.* – Cole, Pacific Grove, CA, Wadsworth and Brooks, 1989.

[6] Fortet (Robert) et Mourier (Édith). – Convergence de la moyenne empirique vers la répartition théorique. *Ann. Sci. École Normale Sup. Ser. 3*, vol. 70, 1953, pp. 266–285.

[7] Kantorovich (Leonid V.) et Rubinštein (Gennadi Sh.). – Sur un espace de fonctions complètement additives. *Vestnik Leningrad Univ.*, vol. 13, n° 7 (Ser. Mat. Astron. 2), 1958, pp. 52–59. – (en russe).

[8] Kawabe (Jun). – A criterion for weak compactness of measures on product spaces with applications. *Yokohama Math. J.*, vol. 42, 1994, pp. 159–169.