NACEREDDINE BELILI

Dualité du problème des marges et ses applications

Séminaire de probabilités (Strasbourg), tome 33 (1999), p. 371-387
<http://www.numdam.org/item?id=SPS_1999__33__371_0>

Dualité du problème des marges et ses applications

NACEREDDINE BELILI

Résumé

Cet article présente une synthèse des théorèmes de dualité relatif au problème des marges, ses diverses applications comme le théorème classique de Strassen, la caractérisation de l'ordre stochastique et la représentation des métriques minimales. On y donne une nouvelle preuve du théorème de couplage Goldstein basée sur la représentation de la distance de variation totale.

Abstract

This paper gives a review of duality theorem for marginal problems and its applications like Strassen theorem, representation of stochastic orders and representation of minimal metrics. We give a new proof of Goldstein theorem based on the representation of total variation metric.

1 Introduction et Notations

Soient $(E_i, \mathcal{E}_i, \mu_i)$, pour $i = 1,\ 2$ deux espaces de probabilités. On dit que μ_1 et μ_2 sont les marges d'une mesure de probabilité γ sur $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ si

$$\gamma(A_1 \times E_2) = \mu_1(A_1) \quad \text{pour tout} \quad A_1 \in \mathcal{E}_1$$

et

$$\gamma(E_1 \times A_2) = \mu_2(A_2) \quad \text{pour tout} \quad A_2 \in \mathcal{E}_2.$$

On désigne par $\mathcal{M}(E_i)$ l'ensemble de mesures de probabilités sur E_i, et par $\Gamma(\mu_1, \mu_2)$ l'ensemble de toutes les mesures de probabilités sur $E_1 \times E_2$ ayant μ_1 et μ_2 pour lois marginales. Pour $i = 1,\ 2$, les applications $\pi_i : E_1 \times E_2 \longrightarrow E_i$ désignent les projections canoniques. L'abréviation $\oplus_i g_i$ est utilisée pour $\sum_i g_i \circ \pi_i$.

Si $(\Omega, \mathcal{A})$ est un espace mesurable, alors $f \in \mathcal{A}$ signifie que f est une application réelle $\mathcal{A}$-mesurable. Une mesure de probabilité μ sur $(\Omega, \mathcal{A})$ est dite *parfaite* ou l'espace probabilisé $(\Omega, \mathcal{A}, \mu)$ est *parfait* si pour tout $f \in \mathcal{A}$ on peut trouver un borélien $B_f \subset f(\Omega)$

tel que $\mu(f^{-1}(B_f)) = 1$. L'espace des fonctions μ_i -intégrable $f \in \mathcal{E}_i$ est noté $\mathcal{L}(\mathcal{E}_i, \mu_i)$; en particulier si $\mathcal{E}_i = \mathcal{B}_i$ est la tribu borélienne de E_i, l'ensemble $\mathcal{L}(\mathcal{E}_i, \mu_i)$ sera noté tout simplement $\mathcal{L}(\mu_i)$. On pose $(\mathcal{E}_1 \otimes \mathcal{E}_2)_{\geq} = \{h \in \mathcal{E}_1 \otimes \mathcal{E}_2 : \exists f_i \in \mathcal{L}(\mathcal{E}_i, \mu_i), \ h \geq \oplus_i f_i\}$, et $(\mathcal{E}_1 \otimes \mathcal{E}_2)_{\leq} = \{h \in \mathcal{E}_1 \otimes \mathcal{E}_2 : \exists f_i \in \mathcal{L}(\mathcal{E}_i, \mu_i), \ h \leq \oplus_i f_i\}$.

Définition 1 *Soit h une fonction $\mathcal{E}_1 \otimes \mathcal{E}_2$ -mesurable sur $E_1 \times E_2$. On définit*

$$\begin{aligned} U(h) &:= \sup\left\{ \int_{E_1 \times E_2} h \, d\gamma : \ \gamma \in \Gamma(\mu_1, \mu_2) \right\}, \\ L(h) &:= \inf\left\{ \sum_{i=1}^{2} \int_{E_i} h_i \, d\mu_i : \ h_i \in \mathcal{L}(\mathcal{E}_i, \mu_i) \ \text{et} \ h \leq \oplus_i h_i \right\}. \end{aligned}$$

Si $E_1 = E_2 = E$ est un espace métrique séparable muni d'une distance d, le problème de transport de Monge-Kantorovich (cf. [39], [31] et [32]) consiste à évaluer la fonctionnelle

$$\mathcal{K}_c(\mu_1, \mu_2) := \inf \left\{ \int_{E \times E} c(x, y) \, \gamma(dx, dy) : \ \gamma \in \Gamma(\mu_1, \mu_2) \right\}, \tag{1}$$

où le "coût" $c(x, y)$ est une fonction mesurable ≥ 0 sur $E \times E$. La fonctionnelle $\mathcal{K}_c(\mu_1, \mu_2)$ est appelée *fonctionnelle de Kantorovich* (resp. *métrique de Kantorovich* si $c = d$). En fait, le théorème de dualité du problème des marges (cf. les théorèmes 1 et 2) est une extension en un certain sens du problème de transport de Monge-Kantorovich. Dans la section 3, on prouvera le théorème de dualité, dans le cas des espaces polonais, pour les fonctions continues, au moyen d'une forme géométrique du théorème de Hahn-Banach. On donnera diverses applications de ce théorème : la caractérisation de l'ordre stochastique, un théorème classique de Strassen, ainsi que la représentation des métriques minimales.

La technique de couplage a été introduite par Doeblin (cf. [11]) pour prouver qu'une chaîne de Markov homogène à espace d'états fini est faiblement ergodique. Sa technique a été raffinée pour obtenir de meilleurs résultats et son idée a été explorée pour d'autres applications (cf. [24], [25], [26], [36] et [55]). Dans la section 4.3, on présentera une preuve du théorème de Goldstein (cf. [24], [55] et [36]), basée sur la représentation de la distance de variation totale.

2 Résultats

Énonçons maintenant le théorème de dualité étudié principalement par Kellerer (cf. [33]), Ramachandran, Rüschendorf (cf. [46]).

Théorème 1 *(cf. [33] et [46]). Soient $(E_i, \mathcal{E}_i, \mu_i)$, pour $i = 1, 2$ deux espaces de probabilités et supposons μ_2 parfaite. Alors on a*

$$U(h) = L(h) \quad \text{pour tout } h \in (\mathcal{E}_1 \otimes \mathcal{E}_2)_{\geq}. \tag{2}$$

Le résultat suivant montre que l' inf dans $L(h)$, égal au sup dans $U(h)$, est atteint.

Proposition 1 *(cf. [33]) Pour tout $h \in (\mathcal{E}_1 \otimes \mathcal{E}_2)_{\leq}$ avec $L(h) < \infty$ alors il existe $h_i \in \mathcal{L}(\mathcal{E}_i, \mu_i)$ telle que*

$$h \leq \oplus_i h_i \quad \text{et} \quad L(h) = \sum_{i=1}^{2} \int_{E_i} h_i \, d\mu_i$$

Si on pose

$$\begin{aligned} I(h) &:= \inf\left\{\int_{E_1\times E_2} h\,d\gamma:\ \gamma\in\Gamma(\mu_1,\mu_2)\right\},\\ S(h) &:= \sup\left\{\sum_{i=1}^{2}\int_{E_i} h_i\,d\mu_i:\ h_i\in\mathcal{L}(\mathcal{E}_i,\mu_i)\text{ et } h\geq \oplus_i h_i\right\},\end{aligned}$$

comme $U(h) = -I(-h)$ et $L(h) = -S(-h)$ le théorème 1 est équivalent au théorème suivant.

Théorème 2 *Soient $(E_i, \mathcal{E}_i, \mu_i)$, pour $i = 1, 2$ deux espaces de probabilités et supposons μ_2 parfaite. Alors on a*

$$I(h) = S(h) \quad \textit{pour tout } h \in (\mathcal{E}_1 \otimes \mathcal{E}_2)_{\leq}. \tag{3}$$

On dit qu'un espace probabilisé $(E_2, \mathcal{E}_2, \mu_2)$ est *un espace de dualité* si pour tout espace probabilisé $(E_1, \mathcal{E}_1, \mu_1)$ les propriétés (2) ou (3) sont vérifiées. D'autre part, l'espace $(E_1, \mathcal{E}_1, \mu_1)$ est dit *contenu* dans $(E_2, \mathcal{E}_2, \mu_2)$ si $E_1 \subset E_2$, $\mathcal{E}_1 = \mathcal{E}_2|_{E_1}$ et $\mu_1 = \mu_2^*|_{\mathcal{E}_1}$. Récemment, dans [47] Ramachandran et Rüschendorf donnent une caractérisation des espaces parfaits et ils montrent par le résultat suivant que la réciproque du théorème de dualité reste vraie.

Théorème 3 *(cf.[47]) Soit $(E, \mathcal{E}, \mu)$ un espace probabilisé. Alors les propriétés suivantes sont équivalentes*

(i) L'espace $(E, \mathcal{E}, \mu)$ est parfait;

(ii) Si l'espace $(E, \mathcal{E}, \mu)$ est contenu dans un espace de probabilité $(E_1, \mathcal{E}_1, \mu_1)$, alors $(E_1, \mathcal{E}_1, \mu_1)$ est parfait;

(iii) Si l'espace $(E, \mathcal{E}, \mu)$ est contenu dans un espace de probabilité $(E_1, \mathcal{E}_1, \mu_1)$, alors $(E_1, \mathcal{E}_1, \mu_1)$ est un espace de dualité.

3 Théorème de dualité dans les espaces polonais

On désigne par E un espace polonais muni de sa tribu borélienne $\mathcal{B}$ et par $\mathcal{C}(E)$ l'ensemble des fonctions réelles continues sur E, et on note $\mathcal{C}(E \times E)$ l'ensemble des fonctions réelles continues sur $E\times E$. Par commodité, on désignera désormais par μ et ν deux mesures de $\mathcal{M}(E)$, au lieu de μ_1 et μ_2 utilisés plus haut. Dans cette section, on va donner en détail une preuve du théorème 2 pour les fonctions continues (dans les espaces polonais). Selon Kellerer [33] l'extension aux fonctions boréliennes se fait ensuite en deux temps: d'abord on étend aux fonctions s.c.s. (semicontinue supérieurement) puis, ayant remarqué que la fonctionnelle $I(\cdot)$, convenablement tronquée, est une capacité fonctionnelle de Choquet, on passe des fonctions s.c.s. aux fonctions boréliennes grâce au théorème de capacitabilité (cf. [5] et [9]).

Théorème 4 *Soient μ et ν deux mesures de $\mathcal{M}(E)$. Alors on a*

$$I(h) = S(h) \quad \textit{pour tout} \quad h \in \mathcal{C}(E \times E). \tag{4}$$

On montre d'abord le lemme suivant.

Lemme 1 *Soient μ et ν deux mesures de $\mathcal{M}(E)$. Alors on a*

$$I(h) \geq S(h) \quad \textit{pour tout } h \in \mathcal{C}(E \times E).$$

DÉMONSTRATION DU LEMME 1.
Soient $\gamma \in \Gamma(\mu, \nu)$ et $f,\ g \in \mathcal{C}(E)$ telles que $f(x) + g(y) \leq h(x, y)$ pour tout $x, y \in E$. Alors on a

$$\int_E f(x)\ d\mu(x) + \int_E g(x)\ d\nu(x) = \int (f(x) + g(y))\ d\gamma(x, y) \leq \int_{E \times E} h(x, y)\ d\gamma(x, y)$$

d'où $I(h) \geq S(h)$.

□

DÉMONSTRATION DU THÉORÈME 4.
(a) Supposons E compact. Pour tout $h \in \mathcal{C}(E \times E)$ soit

$$G = \{g \in \mathcal{C}(E \times E)\ :\ g(x, y) < h(x, y),\ \forall x,\ y \in E\};$$

et soit H l'ensemble des fonctions $c(x, y) = f(x) + g(y)$ pour tout f et g dans $\mathcal{C}(E)$. D'autre part, on pose

$$s(c) = \int_E f\ d\mu + \int_E g\ d\nu \quad \text{pour} \quad c \in H.$$

On vérifie aisément que s est une forme linéaire bien définie. L'ensemble G est convexe, ouvert pour la norme sup ; d'autre part, la forme linéaire s sur H est non identiquement nulle, et bornée supérieurement sur l'ensemble convexe non vide $G \cap H$. En effet, comme $f(x) + g(y) < h(x, y)$ pour tout x et y dans E, alors $s(c) \leq \sup(f) + \sup(g) < \sup(h)$. Alors, d'après le théorème de Hahn-Banach (voir appendice théorème 17), il existe une forme linéaire η qui prolonge s sur $\mathcal{C}(E \times E)$ et telle que

$$\sup_G \eta = \sup_{G \cap H} s.$$

D'après le théorème de représentation de Riesz, il existe une unique mesure positive finie γ sur $E \times E$ telle que

$$\eta(c) = \int_{E \times E} c\ d\gamma \quad \text{pour tout} \quad c \in \mathcal{C}(E \times E).$$

Comme $\eta = s$ sur H, on obtient donc

$$\int_{E \times E} f(x)\ d\gamma(x, y) = \int_E f(x)\ d\mu(x) \quad \text{pour tout} \quad f \in \mathcal{C}(E)$$

et

$$\int_{E \times E} g(x)\ d\gamma(x, y) = \int_E g(x)\ d\nu(x) \quad \text{pour tout} \quad g \in \mathcal{C}(E).$$

Par conséquent $\gamma \in \Gamma(\mu, \nu)$. Par ailleurs,

$$S(h) = \sup_{r \in G \cap H} s(r) = \sup_{t \in G} \eta(t) = \int_{E \times E} h\ d\gamma. \tag{5}$$

Ceci entraîne que $S(h) \geq I(h)$ pour tout $h \in \mathcal{C}(E \times E)$. En utilisant le Lemme 1, on en déduit le résultat du théorème 4 dans le cas (a).

(b) Cas général.
Les mesures μ et ν sont tendues, alors pour tout entier $n > 0$, il existe une partie compacte K_n de E telle que

$$\max\left(\mu(E \setminus K_n),\ \nu(E \setminus K_n)\right) \le \frac{1}{n}.$$

On note a un élément de E, définissons alors deux mesures de probabilités μ_n et ν_n à support compact en posant pour tout $A \in \mathcal{B}$,

$$\mu_n(A) := \mu(A \cap K_n) + \mu(E \setminus K_n)\delta_a(A)$$

et

$$\nu_n(A) := \nu(A \cap K_n) + \nu(E \setminus K_n)\delta_a(A).$$

où δ_a est la mesure de Dirac au point a. La preuve précédente montre qu'il existe une mesure de probabilité γ_n sur $E \times E$ ayant pour lois marginales μ_n et ν_n vérifiant, d'après les définitions de μ_n et ν_n

$$\begin{aligned}&\int_{E\times E} h(x,y)\ d\gamma_n(x,y)\\ &\le\ \sup\left\{\int_E f\ d\mu + \int_E g\ d\nu :\ f(x)+g(y) \le h(x,y),\ \forall x,\ y \in E\right\} + \frac{h(a,a)}{n}.\end{aligned}$$

La suite (γ_n) ayant des marges étroitement convergentes est relativement compacte, on peut en extraire une sous-suite étroitement convergente vers une mesure de probabilité γ qui a pour marges μ et ν. D'après le théorème 9 ci-dessous de Rachev-Rüschendorf, il existe un espace de probabilité $(\Omega, \mathcal{A}, P)$ et une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ sur Ω à valeurs dans E telle que la loi de X_n et γ_n pour tout $n = 1, \ldots,$ et la loi de X_0 est γ et de plus X_n converge p.s vers X_0. Par suite en utilisant le lemme de Fatou on obtient,

$$\begin{aligned}\liminf \int h(x,y)\ d\gamma_n(x,y) &= \liminf \int h(X_n)\ dP \ge \int \liminf h(X_n)\ dP\\ &= \int h(X_0)\ dP = \int h(x,y)\ d\gamma(x,y);\end{aligned}$$

et alors

$$\int h(x,y)\ d\gamma(x,y) \le \sup\left\{\int f\ d\mu + \int g\ d\nu :\ f \in \mathcal{L}(\mu),\ g \in \mathcal{L}(\nu),\ f \oplus g \le h\right\}.$$

□

On en déduit le corollaire suivant.

Corollaire 1 *Fixons $h \in \mathcal{C}(E \times E)$ et soient $\mu,\ \nu \in \mathcal{M}(E)$. Alors*

1. *Il existe une mesure "minimale" $\gamma \in \Gamma(\mu,\nu)$ telle que*

$$I(h) = \int h\ d\gamma.$$

2. *Il existe une mesure "maximale" $\gamma \in \Gamma(\mu,\nu)$ telle que*

$$U(h) = \int h\ d\gamma.$$

Dans la partie gauche de l'égalité (4), si $h(x, y)$ est une fonction de coût, on obtient la fonctionnelle de Kantorovich. L'existence d'une mesure minimale appelée aussi *mesure optimale* et la détermination de sa structure ont été récemment explicitées par Gangbo et McCann (cf. [21]), sous des conditions de convexité sur la fonction de coût et dans le cas $E = \mathbb{R}^d$.

4 Applications

Une conséquence très importante du théorème de dualité est le théorème de Strassen (cf. [53], [28], [17] et [51]) qui est en fait la source d'inspiration de toute cette étude. Sa forme dual, en un certain sens, a été établie par Gamboa et Cattiaux dans [4] en utilisant des méthodes de grandes déviations.

Théorème 5 *(cf. [33]) Soient E_1 et E_2 deux espaces polonais, $\mathcal{E}_1$ et $\mathcal{E}_2$ deux σ-algèbres de E_1 et E_2. Et désignons par μ et ν deux mesures de probabilités sur E_1 et E_2. Alors, pour tout $B \in \mathcal{E}_1 \otimes \mathcal{E}_2$ on a*

$$U(B) := U(1_B) = \inf \{\mu(B_1) + \nu(B_2) : \; B \subset \bigcup_{i=1}^{2} {\pi_i}^{-1}(B_i)\} \tag{6}$$

et

$$I(B) := I(1_B) = \sup \{\mu(B_1) + \nu(B_2) - 1 : \; B_1 \times B_2 \subset B\} \tag{7}$$

Maintenant on va donner des applications concrètes du théorème de dualité.

4.1 Ordre stochastique

Soit $(E, \preceq)$ un espace polonais muni d'une relation de préordre, et supposons que

$$G(E) := \{(x, y) \in E \times E : \; x \preceq y\},$$

soit fermé dans $E \times E$. On désigne par $\mathcal{I}(E)$ le cône des fonctions réelles f bornées et croissantes sur E (i.e., $x \preceq y \Longrightarrow f(x) \leq f(y)$), et par $\mathcal{I}^*(E)$ la famille des sous-ensembles A de E croissants (i.e., la fonction indicatrice de A est croissante). On vérifie immédiatement que

$$A \in \mathcal{I}^*(E) \Longleftrightarrow ((x \in A \text{ et } x \preceq y) \Longrightarrow y \in A).$$

On dit qu'une mesure $\nu \in \mathcal{M}(E)$ *domine stochastiquement* $\mu \in \mathcal{M}(E)$ et on note $\mu \leq_{st} \nu$ si et seulement si

$$\int f \, d\mu \leq \int f \, d\nu \quad \text{pour tout } f \in \mathcal{I}(E).$$

Ce qui est équivalent à

$$\mu(A) \leq \nu(A), \quad \forall A \in \mathcal{I}^*(E).$$

(cf. [36], [30] et [38]). D'après le théorème 5, on obtient une caractérisation de l'ordre stochastique.

Théorème 6 *Soient μ et ν deux mesures de $\mathcal{M}(E)$. Alors on a*

1. *(cf. [33]).*

$$U(G(E)) = 1 - \sup\{\mu(F) - \nu(F) : \; F \textit{ fermé, croissant}\} \tag{8}$$

2. (cf. [36] et [33]).

$$\mu \leq_{st} \nu \Longleftrightarrow \exists\, \gamma \in \Gamma(\mu,\nu) \text{ avec } \gamma(G(E)) = 1. \tag{9}$$

4.2 Représentation des métriques minimales

Certaines métriques (pouvant prendre éventuellement la valeur infini) sur l'ensemble des lois de probabilités, très utiles en théorie des probabilités, ont une représentation de "métrique minimale" (cf. [44]).

Soit E un espace polonais muni d'une distance adéquate d et $\mathcal{B}$ sa tribu borélienne. Pour tout $A \subset E$ et $\varepsilon > 0$, le voisinage d'ordre ε de A est défini par

$$A^\varepsilon := \{x \in E : \exists y \in A,\ d(x,y) < \varepsilon\}, \text{ et on pose } A^0 := \overline{A} \text{ (l'adhérence de } A).$$

4.2.1 Métrique de Lévy-Prokhorov

Par le théorème 5 de Strassen, on obtient

Théorème 7 *(cf. Dudley [14], Théorème 18.2)*
Considérons μ, $\nu \in \mathcal{M}(E)$ et soient $\varepsilon \geq 0$ et $\delta > 0$. Alors les assertions suivantes sont équivalentes

1. *Il existe $\gamma \in \Gamma(\mu,\nu)$ vérifiant $\gamma\{d(x,y) > \varepsilon\} < \delta$;*

2. *on a*
$$\mu(A) \leq \nu(A^\varepsilon) + \delta, \text{ pour tout } A \text{ fermé de } E.$$

Le théorème 7 entraîne une représentation de Strassen de la métrique de Lévy-Prokhorov

$$\Pi(\mu,\nu) := \inf\{\varepsilon : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon,\ \forall A \text{ fermé de } E\}. \tag{10}$$

Définissons pour γ mesure de probabilité sur $E \times E$ la métrique de probabilité de Ky-Fan

$$K(\gamma) := \inf\{\epsilon > 0 :\ \gamma(d(x,y) > \varepsilon) < \varepsilon\} \tag{11}$$

et considérons sa métrique correspondante

$$\widehat{K}(\mu,\nu) := \inf\{K(\gamma) :\ \gamma \in \Gamma(\mu,\nu)\}. \tag{12}$$

Théorème 8 *(cf. Strassen [53] et Dudley [13]) Soient μ et ν deux mesures de probabilités sur E. On a*

$$\widehat{K}(\mu,\nu) = \Pi(\mu,\nu). \tag{13}$$

La distance Π induit la topologie de la convergence faible sur l'ensemble de mesures boréliennes tendues (cf. [15] théorème 11.7.1), i.e.

$$\mu_n \longrightarrow \mu \text{ en loi si et seulement si } \Pi(\mu_n,\ \mu) \longrightarrow 0.$$

Un résultat fondamental est le théorème de représentation presque sûre. La preuve de la deuxième partie utilise principalement le théorème 8.

Théorème 9

(i) (théorème de représentation presque sûre) (Skorokhod, Strassen, Dudley, Wichura, cf.[14]). Soit $(\mu_n)_{n\in\mathbb{N}}$ *une suite de mesures de probabilités sur* E*. Alors* $\Pi(\mu_n, \mu_0) \to 0$ *si et seulement s'il existe un espace de probabilité* $(\Omega, \mathcal{A}, P)$ *et une suite de variables aléatoires* $(X_n)_{n\in\mathbb{N}}$ *sur* $(\Omega, \mathcal{A})$ *à valeurs dans* E*, telle que* $P^{X_n} = \mu_n$, $P^{X_0} = \mu_0$ *et* $d(X_n, X_0) \to 0$ *p.s.*

(ii) (Rachev, Rüschendorf et Schief [45], Dudley [15]). Soient $(\mu_n)_{n\in\mathbb{N}}$ *et* $(\nu_n)_{n\in\mathbb{N}}$ *deux suites tendues de mesures de probabilités sur* E*. Alors* $\Pi(\mu_n, \nu_n) \to 0$ *si et seulement s'il existe deux suites de variables aléatoires* $(X_n)_{n\in\mathbb{N}}$ *et* $(Y_n)_{n\in\mathbb{N}}$ *sur* $(\Omega, \mathcal{A}, P)$ *à valeurs dans* E *telles que* $P^{X_n} = \mu_n$, $P^{Y_n} = \nu_n$ *et* $d(X_n, Y_n) \to 0$ *p.s.*

4.2.2 ℓ_p-métriques

Pour γ mesure de probabilité sur $E \times E$, définissons la métrique de probabilité suivante

$$\ell_\infty(\gamma) := \operatorname*{ess\,sup}_{\gamma} d(x,y) = \inf\{\varepsilon > 0 : \ \gamma(d(x,y) > \varepsilon) = 0\},$$

et considérons sa métrique correspondante définie par

$$\widehat{\ell_\infty}(\mu, \nu) := \ \inf\{\ell_\infty(\gamma) : \ \gamma \in \Gamma(\mu, \nu)\},$$

le théorème 7 implique la représentation de la métrique $\widehat{\ell_\infty}$.

Théorème 10 *(cf. [14]. Théorème 18.2)*
Soient μ *et* ν *deux mesures de* $\mathcal{M}(E)$*. Alors on a*

$$\widehat{\ell_\infty}(\mu, \nu) = \inf\{\varepsilon : \ \mu(A) \leq \nu(A^\varepsilon), \ \forall A \text{ fermé de } E\}. \tag{14}$$

Pour $0 < \lambda < \infty$, les métriques de type Prokhorov et de Ky-Fan sont définies respectivement par

$$\Pi_\lambda(\mu, \nu) := \ \inf\{\varepsilon : \mu(A) \leq \nu(A^{\lambda\varepsilon}) + \varepsilon, \ \forall A \text{ fermé de } E\}, \tag{15}$$

et

$$K_\lambda(\gamma) := \ \inf\{\epsilon > 0 : \ \gamma(d(x,y) > \lambda\varepsilon) < \varepsilon\}, \tag{16}$$

remarquons que

$$\lim_{\lambda \longrightarrow \infty} \lambda \Pi_\lambda(\mu, \nu) = \inf\{\varepsilon : \ \mu(A) \leq \nu(A^\varepsilon), \ \forall A \text{ fermé de } E\}.$$

Pour $1 \leq p < \infty$, la ℓ_p-distance entre deux mesures μ et ν de $\mathcal{M}(E)$ est définie par

$$\widehat{\ell_p}(\mu, \nu) := \ \inf\left\{\left(\int d^p(x,y) \ d\gamma(x,y)\right)^{1/p} : \ \gamma \in \Gamma(\mu, \nu)\right\}.$$

Le théorème suivant se déduit immédiatement du théorème 4 de dualité.

Théorème 11 *(cf. [43] et [44])*
Soient μ *et* ν *deux mesures de* $\mathcal{M}(E)$ *telles que* $\int d^p(x, x_0) \ d(\mu + \nu) < \infty$*, pour tout* $p \geq 1$*. Alors on a*

$$\widehat{\ell_p}(\mu, \nu) = \sup\left\{\left(\int f \ d\mu + \int g \ d\nu\right)^{1/p} \ : \ f \in \mathcal{L}(\mu), \ g \in \mathcal{L}(\nu),\right.$$

$$f(x)+g(y)\le d^p(x,y),\ \forall x,y\in E\Big\}.$$

En particulier, pour $p=1$ on a

Théorème 12 *(Kantorovich-Rubinstein, cf. [54], [18], [33] et [44]).*
Si $\int d(x,x_0)d(\mu+\nu)<\infty$, *alors*

$$\widehat{\ell}_1(\mu,\nu)=\sup\left\{\int f\ d(\mu-\nu):\ f\in \mathrm{Lip}_1(E)\right\},$$

où

$$\mathrm{Lip}_1(E):=\{f:E\to I\!R:\ |f(x)-f(y)|\le d(x,y),\ \forall x,\ y\in E\}.$$

Pour $E=I\!R$ les ℓ_p-métriques minimales sont explicitement connues (cf. Dall'Aglio [8], Fréchet [19]-[20] et Vallender [56]) : on a

$$\widehat{\ell}_1(\mu,\nu)=\int\mid M(x)-N(x)\mid\ dx$$
$$\widehat{\ell}_p^{\,p}(\mu,\nu)=\int_0^1|M^{-1}(t)-N^{-1}(t)|^p\ dt,\ p\ge 1,$$

où M et N sont les fonctions de répartitions de μ et ν. Plus généralement, Major [37] montre que si $\int|x|\ d\mu(x)<\infty$ et $\int|x|\ d\nu(x)<\infty$ alors on a

$$\inf\left\{\int\psi(x-y)\ d\gamma(x,y):\ \gamma\in\Gamma(\mu,\nu)\right\}=\int_0^1\psi(M^{-1}(t)-N^{-1}(t))\ dt,$$

avec ψ une fonction convexe.

Maintenant, on va étudier un cas particulier important de la ℓ_p-métrique quand $E=I\!R^d$ et $p=2$. Soit $||\cdot||$ la norme euclidienne dans $I\!R^d$ et $<\cdot,\cdot>$ désigne le produit scalaire dans $I\!R^d$. Pour une fonction réelle f sur $I\!R^d$, semicontinue inférieurement et convexe on désigne par f^* la fonction conjuguée de f définie par

$$f^*(y):=\sup_{x\in I\!R^d}\{<x,y>-f(x)\};$$

et le *sous-différentiel* de f en x est défini par

$$\partial f(x):=\{y\in I\!R^d:\ f(z)-f(x)\ge<z-x,y>,\ z\in I\!R^d\};$$

(cf. Rockafellar [48]). Les éléments de $\partial f(x)$ sont appelés *sous-gradients* de f en x. La ℓ_2-*distance de Lévy-Wasserstein* $W(\mu,\nu)$ entre deux mesures μ et ν de $\mathcal{M}(E)$ est définie par

$$W(\mu,\nu):=\widehat{\ell}_2(\mu,\nu)=\inf\left\{\left(\int||x-y||^2\ \gamma(dx,dy)\right)^{1/2},\ \gamma\in\Gamma(\mu,\nu)\right\}.\qquad(17)$$

Un couple (X,Y) de variables aléatoires sur un espace de probabilité $(\Omega,\mathcal{A},P)$ est dit *W-couple optimal* (en abrégé W-c.o.) pour (μ,ν) si

$$E||X-Y||^2:=\int||X-Y||^2\ dP=W^2(\mu,\nu).$$

Théorème 13

(i) (Knott et Smith [35], Rüschendorf and Rachev [50]). Soient μ et ν deux mesures de $\mathcal{M}(\mathbb{R}^d)$ telles que $\int \|x\|^2\, d\mu(x) < \infty$ et $\int \|x\|^2\, d\nu(x) < \infty$ et X et Y deux variables aléatoires ayant pour lois μ et ν. Alors $E\|X-Y\|^2 = W^2(\mu,\nu)$ si et seulement s'il existe une fonction $f : \mathbb{R}^d \to \mathbb{R}$ semicontinue inférieurement et convexe telle que $Y \in \partial f(X)$.

(ii) (Dowson et Landau [12], Olkin et Pukelsheim [40], Givens et Shortt [23]). Considérons μ et ν deux gaussiennes multidimensionnelles ayant respectivement m_μ, m_ν, Σ_μ et Σ_ν pour moyennes et matrices de covariance. Alors on a

$$W^2(\mu,\nu) = \|m_\mu - m_\nu\|^2 + \operatorname{trace}\left(\Sigma_\mu + \Sigma_\nu - 2\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}\right). \qquad (18)$$

Remarque 1

(i) Une fonction réelle f continue différentiable sur $\mathbb{R}^d$ est convexe si et seulement si

$$\varphi(x) := \nabla f \text{ est monotone, i.e. } < x-y, \varphi(x)-\varphi(y) > \geq 0,\ \forall x,y \in \mathbb{R}^d$$

où ∇f est le gradient de f, (cf. [48], p. 99).
Soient μ et ν deux mesures de $\mathcal{M}(\mathbb{R}^d)$, et X une variable aléatoire de loi μ. Si φ est le gradient d'une fonction différentiable f, et $\varphi(X)$ a pour loi ν, alors $(X,\varphi(X))$ est un W-couple optimale si et seulement si φ est monotone.

Si φ est une fonction continue et différentiable sur $\mathbb{R}^d$, alors $(X,\varphi(X))$ est un W-couple optimal si et seulement si

1. $\dfrac{\partial \varphi_j}{\partial x_i} = \dfrac{\partial \varphi_i}{\partial x_j}$, $\quad \forall i \neq j$, et
2. φ est monotone. (cf. [49]).

Dans le cas d'une fonction linéaire $\varphi(x) = Ax$ (où A est une matrice), le couple $(X,\varphi(X))$ est W-optimal si et seulement si A est symétrique et semidéfinie positive.

En particulier, dans le cas normal multidimensionel avec $\mu = N(0,\Sigma_\mu)$ et $\nu = N(0,\Sigma_\nu)$, si X a pour loi μ et $A = \Sigma_\mu^{-1/2}(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\Sigma_\mu^{-1/2}$, alors (X,AX) est un W-c.o pour (μ,ν), (cf. [49]).

(ii) Le problème de simulation de Monte-Carlo est le suivant :
Pour $\mu_i \in \mathcal{M}(\mathbb{R})$, comment construire des variables aléatoires $X_i \sim \mu_i$ (i.e., X_i a pour loi μ_i) telles que

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) \leq \operatorname{Var}\left(\sum_{i=1}^n Y_i\right) \quad \text{pour tout} \quad Y_i \sim \mu_i.$$

Pour $n=2$, on a une solution bien connue sous le nom de "antithetic variates" (cf. Hammersley et Handscomb [27]). Et pour $n \geq 2$ on connait quelques solutions du problème dans des cas particuliers (cf. [52]).

Le problème correspondant est de déterminer le minimum de

$$\sum_{i<j} E < X_i, X_j > .$$

D'après le théorème 13, pour $n=2$ on obtient une caractérisation de la solution du problème : $E < X_1, X_2 >= \min\{E < Y_1, Y_2 > :\ Y_1 \sim \mu_1, Y_2 \sim \mu_2\}$ si et seulement s'il existe $f : \mathbb{R} \longrightarrow \mathbb{R}$ semicontinue inférieurement et convexe telle que $X_2 \in \partial f(-X_1)$. Ce qui est équivalent à $< -X_1, X_2 >= f(-X_1) + f(X_2)$.

(iii) La preuve de la première partie du théorème 13, est établie grâce au théorème de dualité et d'autres ingrédients d'analyse convexe. On trouve dans [42] d'autres résultats pour la ℓ_2-métrique de Lévy-Wasserstein et ses applications pour les problèmes d'approximations.

(iv) Un cas très intéressant du problème de transport de Monge-Kantorovich consiste à évaluer la fonctionnelle $W^2(\mu,\nu)$. Sous une condition de continuité sur la loi μ, Cuesta et Matrán (cf. [6]) ont montré que les W-c.o. sont de la forme $(X, f(X))$ où X est une variable aléatoire de loi μ et f une application monotone adéquate; récemment dans [7], ils ont étudié les propriétés de la fonction f. D'autre part, une détermination explicite de cette fonction f a été faite par Abdellaoui et Heinich (cf. [1] et [2]).

(v) La deuxième propriété du théorème 13 reste vrai sur un espace de Hilbert séparable si μ et ν sont deux gausssiennes (cf.[22] et [7]).

4.3 Couplage maximal et théorème de Goldstein

Dans cette section, on va présenter une nouvelle démonstration du théorème de Goldstein (cf. [24], [55] et [36]).

Une métrique très intéressante associée à la ℓ_p-métrique est la métrique de variation totale définie par

$$\sigma(\mu,\nu) := \sup_{A\in\mathcal{B}} |\mu(A) - \nu(A)|.$$

En effet, σ peut être obtenue comme cas limite de la métrique $\widehat{\ell_p}$ et celle de Prokhorov Π_λ, c'est-à-dire

$$\sigma(\mu,\nu) = \lim_{p\to\infty} \widehat{\ell_p}(\mu,\nu),$$

et

$$\sigma(\mu,\nu) = \lim_{\lambda\to 0} \Pi_\lambda(\mu,\nu),$$

et admettant donc la représentation suivante

$$\sigma(\mu,\nu) = \inf\left\{\gamma\{(x,y)\in E\times E : x\neq y\} :\ \gamma\in\Gamma(\mu,\nu)\right\}; \tag{19}$$

Maintenant, considérons $A\in\mathcal{B}$ et soit

$$\Delta(A) := \{(x,x)\in E\times E :\ x\in A\}\in\mathcal{B}\otimes\mathcal{B},$$

la diagonale de l'ensemble A. On définit

$$\mu\wedge\nu(A) := \inf\{\mu(A_1)+\nu(A_2) :\ A_1,\ A_2\in\mathcal{B},\ A_1\cup A_2 = A\}.$$

Le résultat suivant nous permet de donner une autre façon d'avoir la représentation (19).

Théorème 14 *Il existe $\gamma\in\Gamma(\mu,\nu)$ telle que*

$$\gamma(\Delta(A)) = U(\Delta(A)) = \mu\wedge\nu(A), \quad \textit{pour tout} \quad A\in\mathcal{B}, \tag{20}$$

où

$$U(\Delta(A)) = \sup\left\{\gamma(\Delta(A)) \ :\ \gamma\in\Gamma(\mu,\nu)\right\}.$$

DÉMONSTRATION DU THÉORÈME 14. L'égalité $U(\Delta(A)) = \mu \wedge \nu(A)$ découle du théorème 5 de Strassen. Par ailleurs, comme $\mu \wedge \nu$ est une mesure sur E, alors U est additive. Alors, on déduit aisément l'existence de $\gamma \in \Gamma(\mu, \nu)$ telle que

$$\gamma(\Delta(A)) = \mu \wedge \nu(A), \quad \text{pour tout} \quad A \in \mathcal{B}.$$

□

Maintenant, si on prend $A = E$, et si on pose $\Delta = \Delta(E)$ pour simplifier les notations, on montre aisément que l'équation (20) implique

$$\sigma(\mu, \nu) = \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)| = 1 - U(\Delta) = I(E \setminus \Delta). \tag{21}$$

Soient $X = (X_n)_0^\infty$ et $Y = (Y_n)_0^\infty$ deux suites de variables aléatoires à valeurs dans $(E^\infty, \mathcal{B}^\infty)$ et de lois μ et ν respectivement, avec

$$E^\infty := E \times E \times E \times \ldots \quad \text{et} \quad \mathcal{B}^\infty := \mathcal{B} \otimes \mathcal{B} \otimes \mathcal{B} \otimes \ldots.$$

Pour $n \in \overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$, définissons le *shift*

$$\theta_n : E^\infty \longrightarrow E^\infty$$

par

$$\theta_n x = \begin{cases} (x_n, x_{n+1}, x_{n+2}, \ldots); & \text{pour } n < \infty, \\ (z, z, z, \ldots); & \text{pour } n = \infty, \end{cases}$$

où $x = (x_0, x_1, x_2, \ldots) \in E^\infty$, et z est un élément fixé dans E. Pour $n \geq 0$, soit $\mathcal{S}_n$ la σ-algèbre $\subset \mathcal{B}^\infty$, définie par

$$\mathcal{S}_n = \theta_n^{-1}(\mathcal{B}^\infty)$$

et

$$\mathcal{S}_\infty = \bigcap_{n=0}^{\infty} \mathcal{S}_n.$$

Désignons par $\mu_{(n)}, \ \mu_{(\infty)}$ la restriction de μ à $\mathcal{S}_n, \ \mathcal{S}_\infty$.

Définition 2 *1. On dit qu'une mesure $\gamma \in \Gamma(\mu, \nu)$ est un* couplage réussi *si*

$$\lim_{n \longrightarrow \infty} \gamma\{(x, y) \in E^\infty \times E^\infty : \ \theta_n x = \theta_n y\} = 1.$$

2. On dit qu'une mesure $\gamma \in \Gamma(\mu, \nu)$ est un couplage maximal *si*

$$\gamma\{(x, y) \in E^\infty \times E^\infty : \ \theta_n x = \theta_n y\} = 1 - \sigma(\mu_{(n)}, \nu_{(n)}), \ \textit{pour tout } n \in \mathbb{N}.$$

Les notions de couplage réussi et maximal sont très utiles pour démontrer des théorèmes d'ergodicité des chaînes de Markov (cf. [3], [25], [26] et [41]).

Théorème 15 *Les propriétés suivantes sont équivalentes:*

1. *Il existe un couplage $\gamma \in \Gamma(\mu, \nu)$ réussi.*
2. *$\mu = \nu$ sur $\mathcal{S}_\infty$.*
3. $\lim_{n \longrightarrow \infty} \sigma(\mu_{(n)}, \nu_{(n)}) = 0.$

Démonstration du théorème 15. La propriété *(3)* entraîne *(2)* provient de l'inégalité suivante

$$\sigma(\mu_{(\infty)}, \nu_{(\infty)}) \leq \sigma(\mu_{(n)}, \nu_{(n)}).$$

Montrons que *(2)* implique *(3)*. Supposons pour tout n, il existe un ensemble $A_n \in \mathcal{S}_n$ tel que

$$|\mu(A_n) - \nu(A_n)| \geq \alpha_\infty,$$

avec

$$\alpha_n = \sigma(\mu_{(n)}, \nu_{(n)}), \text{ et } \alpha_\infty = \lim_{n \longrightarrow \infty} \alpha_n.$$

Sans perdre de généralité, on peut supposer que

$$\mu(A_n) - \nu(A_n) \geq \alpha_\infty \quad \text{pour tout} \quad n.$$

Maintenant, introduisons l'espace de Hilbert

$$\mathcal{H} := L^2(E^\infty, \ \mathcal{E}^\infty, \mu + \nu),$$

et on note $< \cdot, \cdot >_{\mathcal{H}}$ son produit scalaire — rappelons que dans un espace de Hilbert toute suite bornée admet une sous-suite faiblement convergente —.
Soient $f_n = 1_{A_n}$, $g = \dfrac{d\mu}{d(\mu+\nu)}$ et $g' = \dfrac{d\nu}{d(\mu+\nu)}$. La suite $(f_n)_0^\infty$ est bornée, et alors il existe une sous-suite $(f_{k_n})_0^\infty$ qui converge faiblement vers $f_\infty \in [0,1]$. D'autre part, les fonctionnelles linéaires

$$f \longrightarrow \int f \ d\mu =< f, g >_{\mathcal{H}}, \text{ et } f \longrightarrow \int f \ d\nu =< f, g' >_{\mathcal{H}}$$

sont bornées, et donc on obtient

$$\begin{aligned} \int f_\infty \ d\mu - \int f_\infty \ d\nu &= \lim_{n \to \infty} (< f_{k_n}, g >_{\mathcal{H}} - < f_{n_k}, g' >_{\mathcal{H}}) \\ &= \lim_{n \to \infty} \left(\int f_{k_n} \ d\mu - \int f_{k_n} \ d\nu \right) \\ &= \lim_{n \to \infty} (\mu(A_{k_n}) - \mu(A_{k_n})) \geq \alpha_\infty; \end{aligned}$$

et comme l'espace $L^2(E^\infty, \mathcal{S}_n, \mu + \nu)$ est faiblement fermé, on a $f_\infty \in \mathcal{S}_n$, pour tout n, et donc $f_\infty \in \mathcal{S}_\infty$. Et par suite, on en déduit immédiatement le résultat.

Maintenant, prouvons *(3)* est équivalent à *(1)*. D'après le théorème 14 et l'égalité (21), il existe une mesure $\gamma_n \in \Gamma(\mu_{(n)}, \nu_{(n)})$ telle que

$$\begin{aligned} \sigma(\mu_{(n)}, \nu_{(n)}) &= \gamma_n\{(x,y) \in E^\infty \times E^\infty : \ \theta_n x \neq \theta_n y\} \\ &= \gamma\{(x,y) \in E^\infty \times E^\infty : \ \theta_n x \neq \theta_n y\}; \end{aligned}$$

où $\gamma \in \Gamma(\mu, \nu)$ est le prolongement de la mesure γ_n sur $E^\infty \otimes E^\infty$. Et par conséquent, la mesure γ est un couplage maximal, si $\sigma(\mu_{(n)}, \nu_{(n)}) \longrightarrow 0$; ceci est équivalent à γ est un couplage réussi. Ce qui achève la démonstration.

□

5 Appendice

Les théorèmes suivant sont parmi les versions géométriques du théorème de Hahn-Banach (cf. [10], [16], [29] et [34]).

Définition 3 *Soit E un espace vectoriel réel. Un point $x_0 \in A \subset E$ est appelé point* intérieur *de A si pour tout $y \in E$, il existe $\delta > 0$ tel que $x_0 + ty \in A$ pour tout $|t| < \delta$. Ou bien si $A - x_0$ est un ensemble absorbant.*

Théorème 16 (de séparation) *Soit E un espace vectoriel réel, M et N deux sous-ensembles non vides, disjoints et convexes de E. Supposons que M contient au moins un point intérieur, et N ne rencontre aucun point intérieur de M. Alors il existe une forme linéaire f sur E, non triviale, telle que*

$$\inf_{x \in N} f(x) \geq \sup_{y \in M} f(x).$$

Théorème 17 *Soit E un espace vectoriel réel et L un sous espace vectoriel de E. Soit M un sous-ensemble convexe de E. Supposons que M contient un point intérieur appartenant à L. Soit f une forme linéaire non nulle sur L, bornée supérieurement sur $L \cap M$. Alors f se prolonge en une forme linéaire g sur E, vérifiant*

$$\sup_{x \in M} g(x) = \sup_{y \in L \cap M} f(y).$$

DÉMONSTRATION DU THÉORÈME 17. Soit $\alpha := \sup_{x \in L \cap M} f(x)$ et $A := \{x \in L : f(x) > \alpha\}$. Alors L et M sont convexes, disjoints. Par conséquent, d'après le théorème 16 de séparation il existe une forme linéaire h sur E telle que

$$\inf_{x \in A} h(x) \geq \sup_{y \in M} h(y).$$

Soit $x_0 \in L \cap M$ le point intérieur de M. Alors $\sup_{x \in M} h(x) > h(x_0)$, d'où $h(y) > h(x_0)$ pour tout $y \in A$, et par suite h n'est pas une constante sur L. Soit $B := \{x \in L \ : \ h(x) = h(x_0)\}$. Remarquons que f est constante sur B. En effet, supposons que f prend deux valeurs différentes sur deux points quelconques de B. Alors f prend toutes les valeurs sur la droite joignant les deux points, et donc cette droite rencontre A. Mais $h = h(x_0)$ sur cette droite, contradiction. Soit $f = k$ sur B ($k \in \mathbb{R}$). Si $k = 0$, comme f n'est pas constante sur L, on a $0 \in B$ et $h(x_0) = 0$. Considérons $z \in L \setminus B$, $h(z) \neq f(z) \neq 0$ et $h(x) = h(z)f(x)/f(z)$ pour $x = z$ et pour tout $x \in B$, donc pour tout $x \in L$. D'autre part, si $k \neq 0$, alors 0 n'appartient pas à B, $h(x_0) \neq 0$ et $h(x) = h(x_0)f(x)/k$, pour tout $x \in B$ et pour $x = 0$, donc pour tout $x \in L$. Dans les deux cas, $h = \beta f$ sur L pour $\beta \neq 0$. Alors $g := h/\beta$ prolonge f sur E et vérifie

$$\sup_{x \in M} g(x) \leq \inf_{y \in A} h(y)/\beta = \inf_{y \in A} f(y) = \alpha.$$

□

Références

[1] ABDELLAOUI, T. Détermination d'un couple optimal du problème de Monge Kantorovich. *C. R. Acad. Sci. Paris*, 319:981–984, 1994.

[2] ABDELLAOUI, T., ET HEINICH, H. Sur la distance de deux lois dans le cas vectoriel. *C. R. Acad. Sci. Paris*, 319:397–400, 1994.

[3] ALDOUS, D. J. Shift-coupling. *Stoch. Proc. Appl*, 44:1–14, 1993.

[4] CATTIAUX, P., ET F. GAMBOA. Large deviations and variational theorem for marginal problems. *Preprint*, 1996.

[5] CHOQUET, G. Forme abstraite du théorème de capacitabilité. *Ann. Inst. Fourier*, 9:83–89, 1959.

[6] CUESTA-ALBERTOS, J. A., AND MATRÁN, C. Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.*, 17:1264–1276, 1989.

[7] CUESTA-ALBERTOS, J. A., MATRÁN, C, AND TUERO-DIAZ, A . On lower bounds for the l^2-Wasserstein metric in a Hilbert space. *J. of Theoretical Prob.*, 9:263–283, 1996.

[8] DALL'AGLIO. Fréchet classes and compatibility of distribution function. *Sym. Math.*, 9:131–150, 1972.

[9] DELLACHERIE, C., MEYER, P.A. *Probabilités et potentiel.* Herman, Paris, 1983.

[10] DIEUDONNÉ, J. Sur le théorème de Hahn-Banach. *Rev. Sci*, 79:642–643, 1941.

[11] DOEBLIN, W. Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états. *Rev. Math. Union Interbalkanique*, 2:77–105, 1938.

[12] DOWSON, D. C., LANDAU, B. V. The Fréchet distance between multivariate normal distribution. *J. Multivariate Anal.*, 12:450–455, 1982.

[13] DUDLEY, R. M. Distances of probability measures and random variables. *Ann. Math. Stat.*, 39:1563–1572, 1968.

[14] DUDLEY, R. M. *Probability and metrics.* Aarhus Univ., Aarhus, 1976.

[15] DUDLEY, R. M. *Real analysis and probability.* Chapman and Hall, New York London, 1989.

[16] DUNFORD, N., AND SCHWARTZ, J. T. *Linear Operators.* Interscience Publishers, a division of John Wiley and Sons, New York, t. I, 1958.

[17] EDWARDS, D. A. On the existence of probability measures with given marginals. *Ann. Inst. Fourier.*, 28:53–78, 1978.

[18] FERNIQUE, X. *Sur le théorème de Kantorovitch-Rubinstein dans les espaces polonais.* Lecture Notes in Mathematics 850., Springer, 1981.

[19] FRÉCHET, M . Sur les tableaux de corrélation dont les marges sont données. *Annales de l'université de Lyon, Sciences.*, 4:13–84, 1951.

[20] FRÉCHET, M . Sur la distance de deux lois de probabilité. *C. R. Acad. Sci. Paris.*, 244, 1957.

[21] GANGBO, W., AND MCCANN, R. J. The geometry of optimal transportation. *Acta. Math.*, 177:113–161, 1996.

[22] GELBRICH, M . On a formula for the l^2-Wasserstein metric between measures on Euclidean and Hilbert spaces. *Math. Nachr.*, 147:185–203, 1990.

[23] GIVENS, C. R., AND SHORTT, R. M . A class of Wasserstein metrics for probability distributions. *Michigan Math. J.*, 31:231–240, 1984.

[24] GOLDSTEIN, S. Maximal coupling. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 46:193–204, 1979.

[25] GRIFFEATH, D. A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 31:95–106, 1975.

[26] GRIFFEATH, D. Uniform coupling of non-homogenous Markov chains. *J. Appl. Probability*, 12:753–762, 1975.

[27] HAMMERSLEY, I. M., AND HANDSCOMB, D. C. *Monte Carlo methods.* Meth, London, 1964.

[28] HANSEL, G, AND TROALLIC, J. P. Mesures marginales et théorème de Ford-Fulkerson. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 43:245–251, 1978.

[29] HERMES, H., AND LASALLE, J.P. *Functional Analysis and Time optimal control.* Academic Press, New York and London, 1969.

[30] KAMAE, T., KRENGEL, U. AND O'BRIEN. Stochastic inequalities on partially ordered spaces. *Ann. Probab.*, 5:899–912, 1977.

[31] KANTOROVICH, L.V. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[32] KANTOROVICH, L.V. On a problem of Monge (in russian). *Uspekhi Math. Nauk*, 3:225–226, 1948.

[33] KELLERER, H. G. Duality theorems for marginal problems. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 67:399–432, 1984.

[34] KELLEY, J. L., AND NAMIOKA, I. *Linear topological spaces.* D. Van Nostrand Company, Princeton, N. I, 1963.

[35] KNOTT, M., AND SMITH, C. S. On the optimal mapping of distributions. *J. Optim. Th. Appl.*, 43:39–49, 1984.

[36] LINDVALL, T. *Lectures on the coupling method.* Wiley, New York, 1993.

[37] MAJOR, P. On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.*, 8:487–517, 1978.

[38] MARSHALL, A. W., OLKIN, I. *Inequalities theory of majorization and its applications.* Academic Press, New York, 1979.

[39] MONGE, G. Mémoire sur la théorie des déblais et des remblais. *Histoires de l'Académie Royale des Sciences de Paris, avec les mémoires de Mathématiques et de Physique pour la même année*, pages 257–263, 1781.

[40] OLKIN, I., AND PUKELSHEIM, F. The distance between two random vectors with given dispertion matrices. *Linear Algebra Appl.*, 48:257–263, 1982.

[41] PITMAN, J.W. On coupling of Markov chains. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 35:315–322, 1976.

[42] RACHEV, S. T. The Monge Kantorovich mass transference problem and its stochastic applications. *Theory Prob. Appl.*, 29:647–676, 1984.

[43] RACHEV, S. T. On a problem of Dudley. *Soviet. Math. Dokl.*, 29:162–164, 1984.

[44] RACHEV, S. T. *Probability metrics and the stability of the stochastic models.* Wiley, New York, 1991.

[45] RACHEV, S. T., RÜSCHENDORF, L., AND SCHIEF, A. Uniformities for the convergence in law and in probability. *J. of Theoretical Prob.*, 5:33–44, 1992.

[46] RAMACHANDRAN, D., AND RÜSCHENDORF, L. A general duality theorem for marginal problems. *Probab. Theory Relat. Fields*, 101:311–319, 1995.

[47] RAMACHANDRAN, D., AND RÜSCHENDORF, L. Duality and perfect probability spaces. *Proceedings of the American mathematical society*, 124:2223–2228, 1996.

[48] ROCKAFELLAR, R. T. *Convex Analysis.* Princeton, Univ. Press, 1970.

[49] RÜSCHENDORF, L. Fréchet bounds and their applications. In Kotz S Dall'Aglio, G. and Salinetti G, editors, *Advances in probability distributions with given marginals: Beyond the Copulas*, pages 141–176. Dordrecht, Kluwer Academic Publishers, 1991.

[50] RÜSCHENDORF, L., AND RACHEV, S. A characterization of random variables with minimum l^2-distance. *J. of Multivariate Anal.*, 32:48–54, 1990.

[51] SKALA, H. G. The existence of probability measures with given marginals. *Ann. Probab.*, 21:136–142, 1993.

[52] SNIJDERS, T. A. B. Antithetic variates for Monte-Carlo estimation of probabilities. *Statistics Neerlandica*, 38:1–19, 1984.

[53] STRASSEN, V. The existence of measures with given marginals. *Ann. Math. Stat*, 36:423–439, 1965.

[54] SZUGLA, A. On minimal metrics in the space of random variables. *Theory Prob. Appl.*, 27:424–430, 1982.

[55] THORISSON, H. On maximal and distributional coupling. *Ann. Probab.*, 14:873–876, 1986.

[56] VALLENDER, S. S. Calculation of the Wasserstein distance between probability distributions on the line. *Theory. Prob. Appl.*, 18:784–786, 1973.