C. R. Acad. Sci. Paris, Ser. I 335 (2002) 537–540

Probabilités/*Probability Theory*

Approximation pour la distance de Wasserstein

Nacereddine Belili [a], Henri Heinich [b]

[a] IRSEEM/ESIGELEC, Département TIC, 1, rue Maréchal Juin, 76131 Mont-Saint-Aignan cedex, France

[b] UPRES-A, CNRS 6085, INSA de Rouen, département de génie mathématique, place E. Blondel, 76131 Mont-Saint-Aignan cedex, France

Reçu le 13 juillet 2002 ; accepté le 26 juillet 2002

Note présentée par Jean-Pierre Kahane.

Résumé Pour une classe $\mathcal{C}$ de probabilités et P une probabilité de $\mathbb{R}^d$, nous montrons, sous certaines conditions, l'existence d'une solution au problème de l'approximation de P par $\mathcal{C}$. Il existe une probabilité $Q_0 \in \mathcal{C}$ telle que $l(P, Q_0) \leqslant l(P, Q)$, $\forall Q \in \mathcal{C}$, où l est le carré de la distance de Wasserstein. ***Pour citer cet article : N. Belili, H. Heinich, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 537–540.***

Approximation for the Wasserstein distance

Abstract Let $\mathcal{C}$ be a set of probability-measures and P a probability on $\mathbb{R}^d$. Under some conditions, we show that we have a solution to the approximation problem of P by $\mathcal{C}$. There exists a probability $Q_0 \in \mathcal{C}$, such that $l(P, Q_0) \leqslant l(P, Q)$, $\forall Q \in \mathcal{C}$ where l is the square of the Wasserstein distance. ***To cite this article: N. Belili, H. Heinich, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 537–540.***

1. Problème et résultats

Les variables aléatoires (v.a.) considérées sont définies sur un espace probabilisé suffisament riche, $E[X]$ et $\mathcal{L}(X)$ désignent respectivement l'espérance et la loi de la v.a. X. Pour deux probabilités, P et Q, définies sur $\mathbb{R}^d$, $(P|Q)$ est l'ensemble des probabilités sur $\mathbb{R}^d \times \mathbb{R}^d$ dont les marges sont P et Q et $l(P, Q) = \inf\{E[\|X - Y\|^2],\ \mathcal{L}(X, Y) \in (P|Q)\}$ est le carré de la distance de Wasserstein. Cette expression n'a d'intérêt que si P et Q ont des moments d'ordre deux : $\sigma^2(P) := \int \|x\|^2 \,\mathrm{d}P(x)$ et $\sigma^2(Q)$ finis. Un couple de v.a. (X, Y) est optimal pour (P, Q) si $l(P, Q) = E[\|X - Y\|^2] = \sigma^2(P) + \sigma^2(Q) - 2E[\langle X, Y\rangle]$, $\mathcal{L}(X, Y) \in (P|Q)$. L'existence d'un couple optimal est classique cf. [1] et [2]. Pour une classe $\mathcal{C}$ de probabilités sur $\mathbb{R}^d$ et une probabilité P de $\mathbb{R}^d$, $l(P, \mathcal{C}) := \inf\{l(P, Q),\ Q \in \mathcal{C}\}$ est le carré de la distance de P à la classe $\mathcal{C}$. On s'intéresse au problème $(\mathcal{P})$ de l'existence de $Q \in \mathcal{C}$ telle que $l(P, \mathcal{C}) = l(P, Q)$. Pour la classe $\mathcal{C}_u$ formée par les probabilités uniformes sur les compacts convexes d'intérieur non vide, Cuesta-Albertos, Matràn et Rodríguez-Rodríguez ont montré dans [3], l'existence d'une telle probabilité lorsque P est abolument continue par rapport à la mesure de Lebesgue de $\mathbb{R}^d$.

Adresses e-mail : Nacer.Belili@esigelec.fr (N. Belili); heinich@insa-rouen.fr (H. Heinich).

S1631-073X(02)02522-0/FLA

Nous allons, en simplifiant leur preuve, affaiblir la condition sur P et généraliser la classe $\mathcal{C}_u$. À cette fin, introduisons quelques notations supplémentaires. Π est l'ensemble des projections orthogonales de $\mathbb{R}^d$, et pour $A \subset \mathbb{R}^d$, $ev(A)$ est l'espace vectoriel engendré par A. Une probabilité Q, ou une v.a. $Y = (Y_i)$ de loi Q, est symétrique (relativement à une base), si pour tout $\varepsilon = (\varepsilon_i) \in (\{-1, +1\})^n$, $\mathcal{L}(\varepsilon Y) = Q$, où $\varepsilon Y = (\varepsilon_i Y_i)$.

Nous allons considérer une classe de probabilités de $\mathbb{R}^d$ vérifiant les conditions suivantes.

– $\mathcal{C}$ est formée de probabilités ayant un moment d'ordre 2.

(1) $\mathcal{C}$ est stable par translation : $Q \in \mathcal{C}$ si et seulement si $Q(\cdot - a) \in \mathcal{C}$, $\forall a \in \mathbb{R}^d$.

(2) Pour tout $\pi \in \Pi$, et tout réel $a \neq 0$, $\mathcal{C}$ est stable pour $u = a\pi + \pi^{\perp}$ i.e. si $u(Q)$ est la probabilité définie par $u(Q)(f) = Q(f \circ u)$, alors $u(Q) \in \mathcal{C}$ pour $Q \in \mathcal{C}$.

(3) Pour $\mu \in \overline{\mathcal{C}}$, fermeture pour la convergence en loi, notons S_μ le support de μ, et $\pi^{\perp}$ la projection orthogonale sur $ev(S_\mu)$. Alors il existe $Q \in \mathcal{C}$ telle que $\mu = \pi^{\perp}(Q)$ et $\pi(Q)$ est symétrique, en particulier si $\dim ev(S_\mu) = d$, alors $\mu \in \mathcal{C}$.

(4) Pour tout $\pi \in \Pi$ et $Q \in \mathcal{C}$, $\dim ev(S_{\pi(Q)}) = \dim \pi$.

Le résultat principal est le théorème suivant.

THÉORÈME. – *Soit P une probabilité sur $\mathbb{R}^d$ telle que $\dim ev(S_P) = d$ et $\mathcal{C}$ une classe vérifiant les assertions précédentes, alors il existe $Q \in \mathcal{C}$ telle que $l(P, Q) = l(P, \mathcal{C})$.*

Démonstration. – Réduction du problème. Le problème $(\mathcal{P})$ est équivalent au même problème en supposant les probabilités centrées : $\int x \,\mathrm{d}P(x) = \int x \,\mathrm{d}Q(x) = 0$. Nous adoptons ce cadre, dans toute la suite, en écrivant encore $\mathcal{C}$ pour la classe $\mathcal{C}$ centrée. Il existe une suite $(Q_n) \subset \mathcal{C}$, telle que $Q_n \xrightarrow{\mathcal{L}} \mu$, $l(P, \mu) \leqslant l(P, \mathcal{C})$. Si $\dim ev(S_\mu) = d$, la propriété (3) implique que $\mu \in \mathcal{C}$, et le théorème est prouvé. Supposons $\dim ev(S_\mu) < d$. Le Corollaire 1 assure que $\dim ev(S_\mu) \geqslant 1$. L'assertion (3) donne l'existence d'une probabilité $Q \in \mathcal{C}$ telle que $\mu = \pi(Q)$ et $Q^s = \pi^{\perp}(Q)$ est symétrique et différente de δ_0 par l'assertion (4). Le Corollaire 2, montre que $\pi^{\perp}(P) = \delta_0$, ce qui est contraire à l'hypothèse sur P. $\square$

LEMME. – *Soit P et Q deux probabilités centrées ayant des moments d'ordre deux, Q symétrique et $l(P, Q) = \sigma^2(P) + \sigma^2(Q)$. Alors les supports de P et de Q sont orthogonaux.*

Démonstration. – On utilise l'assertion suivante :

Soit μ une probabilité sur $\mathbb{R}^{n+p}$ dont la marge sur $\mathbb{R}^n$ – respectivement sur $\mathbb{R}^p$ est μ_1 – resp. μ_2 et soit X une v.a. définie sur $(\Omega_1, \mathcal{A}_1, P_1)$ de loi μ_1. Il existe un espace de probabilité $(\Omega = \Omega_1 \times \Omega_2, \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2, P = P_1 \otimes P_2)$, – où $(\Omega_2, \mathcal{A}_2, P_2)$ est un autre espace probabilisé – et un couple de v.a. $(\overline{X}, \overline{Y})$ tels que $\mathcal{L}(\overline{X}, \overline{Y}) = \mu$, $\mathcal{L}(\overline{X}) = \mu_1$, $\mathcal{L}(\overline{Y}) = \mu_2$ et $\overline{X}(\omega_1, \omega_2) = X(\omega_1)$.

Ceci nous assure que pour tout X_0 de loi P, il existe Y_0 de loi Q telle que (X_0, Y_0) soit optimal. Par conséquent $E[\langle X_0, Y_0\rangle] = 0$ et donc $E[\langle X, Y\rangle] \leqslant 0$ si $\mathcal{L}(X, Y) \in (P|Q)$ et, en changeant Y en $-Y$, on a $E[\langle X, Y\rangle] = 0$. Soient X et X' deux v.a. indépendantes de même loi P, la probabilité $P^* = \mathcal{L}(X - X')$ est symétrique. Comme précédemment, pour toute v.a. X_0, $\mathcal{L}(X_0) = P^*$, il existe Y_0, $\mathcal{L}(Y_0) = Q$, telle que $l(P^*, Q) = E[\|X_0 - Y_0\|^2]$. Prenons $X_0 = X - X'$, on en déduit que $E[\langle X_0, Y\rangle] = 0$, $\forall Y$ de loi Q, d'où $l(P^*, Q) = \sigma^2(P^*) + \sigma^2(Q)$. Nous pouvons supposer, sans perte de généralité, que P est symétrique.

Pour une v.a. Z notons $S(Z) = (S(Z_i))$, où $S(Z_i)$ est le signe de Z_i et $|Z| := (|Z_i|)$. Soit (X, Y) un couple de loi $P \otimes Q$, alors $\mathcal{L}(S(Y)|X|) = P$ et $\mathcal{L}(S(Y)|Y|) = Q$, on en déduit $E[\langle |X|, |Y|\rangle] = 0$, donc $\langle X, Y\rangle = 0$ p.s. La preuve s'achève. $\square$

COROLLAIRE 1. – *Soit P une probabilité sur $\mathbb{R}^d$ telle que l'espace vectoriel engendré par son support est de dimension d, alors $l(P, \mathcal{C}) < \sigma^2(P)$.*

Démonstration. – Supposons que $l(P, \mathcal{C}) = \sigma^2(P) = l(P, \delta_0)$. Pour $Q \in \mathcal{C}$ et $a \in \mathbb{R}\backslash\{0\}$, on a $aQ \in \mathcal{C}$. Le minimum de la fonction $a \to l(P, aQ)$ est atteint et vaut $\sigma^2(P) - (E[\langle XY\rangle])^2/\sigma^2(Q)$, où (X, Y) est optimal pour (P, Q). En prenant Q symétrique on en déduit, comme dans le lemme, $\langle X, Y\rangle = 0$ p.s. pour toutes v.a. X de loi P et Y de loi Q symétrique. Par conséquent le support de P est orthogonal au support de Q, ce qui est contradictoire. □

COROLLAIRE 2. – *Soient Q et P deux probabilités de $\mathbb{R}^d$ et $\pi \in \Pi$, $1 \leqslant \dim\pi < d$. Supposons que pour tout réel $a \neq 0$, et $u = \pi^\perp + a\pi$, on a : $l(P, \pi^\perp(Q)) \leqslant l(P, uQ)$. Alors il existe un couple (X, Y) tel que $(\pi^\perp(X), \pi^\perp(Y))$ soit optimal pour $(\pi^\perp(P), \pi^\perp(Q))$. De plus, pour un tel couple, on a* $E[\langle\pi(X), \pi(Y)\rangle] = 0$.

Démonstration. – Notons $Z_1 = \pi(Z)$, $Z_2 = \pi^\perp(Z)$. L'existence d'un couple (X, Y) vérifiant les conditions du corollaire résulte de l'assertion du lemme. Les inégalités $l(P, Q_2) \leqslant l(P, uQ) \leqslant \sigma^2(P) + a^2\sigma^2(Q_1) + \sigma^2(Q_2) - 2E[\langle X_2, Y_2\rangle] - 2aE[\langle X_1, Y_1\rangle]$, donnent, en prenant le minimum en a, $l(P, Q_2) \leqslant l(P, Q_2) - (E[\langle X_1, Y_1\rangle])^2/\sigma^2(Q_2)$. Ce qui permet de conclure. □

2. Exemples

2.1. Lois uniformes sur les convexes compacts d'intérieur non vide

Il est aisé de voir que si $\overset{\circ}{S}$ est l'intérieur du support S de $Q \in \mathcal{C}_u$, alors $0 \in \overset{\circ}{S}$. En fait, cette propriété est valable pour toute probabilité centrée dont l'enveloppe convexe du support est d'intérieur non vide.

Si la suite $(Q_n) \subset \mathcal{C}_u$ converge en loi vers μ, on montre que les supports S_n des Q_n sont contenus dans un compact K et qu'ils convergent p.s. (pour une sous-suite omise) vers S.

PROPOSITION. – *Soient $(Q_n) \subset \mathcal{C}_u$ une suite de probabilités convergeant en loi vers μ et (S_n) une suite de supports de Q_n qui converge p.s. (pour une sous-suite omise) vers S. Alors, on a $S = S_\mu$, où S_μ est le support de μ.*

Preuve succincte. – Montrons que S_μ est contenu dans S. Nous pouvons supposer, sans perte de généralité, que S est fermé. Si B est une boule compacte telle que $B \cap S = \emptyset$, on établit que $B \cap S_n = \emptyset$ dès que n est suffisament grand. Par conséquent $\mu_n(B) = 0$ d'où $\mu(\overset{\circ}{B}) = 0$. Le complémentaire de S est réunion dénombrable de boules ouvertes d'adhérence compacte, par conséquent $\mu(S^c) = 0$, donc $S_\mu \subset S$.

Pour la réciproque, supposons que $\dim ev(S_\mu) = p \geqslant 1$ et notons π la projection orthogonale sur cet espace. Soit a, b deux points de $\mathbb{R}^p$, tels que $a_i < b_i$, $\forall i$ et $[a, b] \subset \pi(\overset{\circ}{S})$, on montre qu'il existe $k > 0$ tel que $\lambda^d(S_n \cap (a \leqslant \pi \leqslant b))/\lambda^d(S_n) = Q_n(a \leqslant \pi \leqslant b) \geqslant k$. On en déduit $\mu(a \leqslant \pi \leqslant b) \geqslant k$, i.e. $S \subset S_\mu$. □

Montrons maintenant le point (3), les autres conditions étant évidentes.

Soit $(Q_n) \subset \mathcal{C}_u$, telle que $Q_n \xrightarrow{\mathcal{L}} \mu$, avec $1 < \dim ev(S) < d$, où S est le support de μ. La projection orthogonale sur $ev(S)$ est notée $\pi^\perp$. Nous avons vu (proposition) que les supports S_n de Q_n tendent vers le support S, pour une sous-suite omise. Choisissons un point $M_n \in S_n$ tel que $|\pi(M_n)| = \sup\pi(S_n)$, on peut supposer $a_n = \pi(M_n) > 0$. Les applications $u_n = (1/a_n)\pi + \pi^\perp$, transforment Q_n en $Q_n^* := u_n(Q_n) \in \mathcal{C}_u$. Pour une sous-suite, omise, $u_n(M_n)$ converge. Montrons la convergence en loi de la suite (Q_n^*). Pour chaque n, S_n est contenu dans le compact convexe $K_n := \{x \mid \pi^\perp(x) \in \pi^\perp(S_n),\ |\pi(x)| \leqslant a_n\}$. L'application u_n transforme K_n en K_n^* qui est contenu dans $\{x \mid \pi^\perp(x) \in \pi^\perp(S_n),\ |\pi(x)| \leqslant 1\}$. Pour $n \geqslant N(\varepsilon)$ on a $K_n^* \subset \{x\ |\pi(x)| \leqslant 1,\ \pi^\perp(S)(\varepsilon)\}$, où $A(\varepsilon)$ est l'ensemble $\{x \mid d(x, A) \leqslant \varepsilon\}$. Par conséquent la suite (Q_n^*) est relativement compacte en loi. On peut supposer que $Q_n^* \xrightarrow{\mathcal{L}} Q$, le support de Q est de mesure strictement positive, donc $Q \in \mathcal{C}_u$. Il est évident que $\pi^\perp(Q_n) = \pi^\perp(Q_n^*) \xrightarrow{\mathcal{L}} \pi^\perp(Q) = \mu$. Enfin, il suffit de symétriser $\pi(Q)$, ce qui conserve l'appartenance à $\mathcal{C}_u$, pour avoir prouvé que $\mathcal{C}_u$ vérifie l'assertion.

2.2. Autres exemples

Soit (Q_n) une suite convergeant en loi vers μ, de support S et $\dim ev(S) \geqslant 1$, notons $\pi^{\perp}$ la projection orthogonale sur cet espace. La méthode précédente montre que $S_n \to S$ p.s. et que la suite $(u_n(Q_n) = Q_n^*)$ est relativement compacte en loi. De plus, si Q est une valeur d'adhérence de cette suite, alors l'e.v. engendré par le support de Q est $\mathbb{R}^d$.

Il en résulte d'autres exemples comme :

A – L'ensemble des probabilités de loi uniforme sur la frontière d'un convexe compact d'intérieur non vide.

B – L'ensemble des probabilités de loi uniforme sur $C_1 \backslash C_2$, C_i convexes compacts d'intérieurs non vide.

C – La classe $\mathcal{C}_o$ formée par les probabilités Q ayant un moment d'ordre deux et telles que

$$Q\big(\theta A + (1-\theta)B\big) \geqslant \inf\big[Q(A), Q(B)\big], \quad \forall A, B, \ \forall \theta \in]0, 1[,$$

et dont le support engendre $\mathbb{R}^d$. Si $Q \in \mathcal{C}_o$, alors Q a une densité g, telle que $\{g \geqslant a\}$ est convexe pout tout $a \in \mathbb{R}$.

Références bibliographiques

[1] N. Belili, H. Heinich, Transport problem and derivation, Appl. Math. 26 (3) (1999) 299–314.
[2] P.J. Bickel, D.A. Freedam, Some asymptotic theory for boostrap, Ann. Statist. 9 (1981) 1196–1217.
[3] J.A. Cuesta-Albertos, C. Matràn, J. Rodríguez-Rodríguez, Approximation to probabilities through uniform laws on convex sets, Preprint, 2002.