Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 338 (2004) 805–808
Comptes Rendus Mathematique

Statistique/Probabilités

# Couplage pour la distance minimale

Jérôme Dedecker [a], Clémentine Prieur [b]

[a] *Laboratoire de statistique théorique et appliquée, Université Paris 6, site Chevaleret, 13, rue Clisson, 75013 Paris, France*
[b] *Laboratoire de statistique et probabilités, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse cedex 4, France*

Reçu le 28 novembre 2003 ; accepté après révision le 15 mars 2004

Présenté par Paul Deheuvels

**Résumé**

Dans cette Note, nous étendons un résultat de couplage pour des variables réelles au cas des variables à valeurs dans un espace polonais. Ce résultat est une conséquence d'une version conditionnelle du théorème de Kantorovitch et Rubinstein. ***Pour citer cet article : J. Dedecker, C. Prieur, C. R. Acad. Sci. Paris, Ser. I 338 (2004).***


**Abstract**

**Coupling for minimal distance.** In this Note, we generalize a coupling result for real variables to the case of variables with values in some Polish space. This result follows from a conditional version of the Kantorovitch and Rubinstein theorem. ***To cite this article: J. Dedecker, C. Prieur, C. R. Acad. Sci. Paris, Ser. I 338 (2004).***


## 1. Une version conditionnelle du théorème de Kantorovitch et Rubinstein

**Définition 1.1.** Soit $\mathcal{A}$ une tribu et $\mathcal{Z}$ un espace polonais. On dira qu'une fonction $P$ de $\mathcal{A} \times \mathcal{Z}$ est une probabilité conditionnelle sur $\mathcal{A} \times \mathcal{Z}$ si

(i) Pour tout $z$ dans $\mathcal{Z}$, $P(\cdot, z)$ est une probabilité sur $\mathcal{A}$.
(ii) Pour tout $A$ de $\mathcal{A}$, $z \to P(A, z)$ est mesurable de $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$ dans $([0,1], \mathcal{B}([0,1]))$.

Si $(\mathcal{X}, d)$ est polonais, $\Lambda_1(\mathcal{X})$ est l'ensemble des fonctions 1-lipschitziennes de $\mathcal{X}$ dans $\mathbb{R}$.

**Proposition 1.2.** *Soient $(\mathcal{X}, d)$ et $\mathcal{Z}$ deux espaces polonais et $\pi$ une probabilité sur $\mathcal{B}(\mathcal{Z})$. Soient $P$ et $Q$ deux probabilités conditionnelles sur $\mathcal{B}(\mathcal{X}) \times \mathcal{Z}$. On note $\mathbf{P}$ et $\mathbf{Q}$ les probabilités sur $\mathcal{B}(\mathcal{X})$ définies par $\mathbf{P}(A) = \int P(A, z)\pi(\mathrm{d}z)$ et $\mathbf{Q}(A) = \int Q(A, z)\pi(\mathrm{d}z)$ respectivement. On suppose que $\int d(x, 0)\,\mathbf{P}(\mathrm{d}x)$ et $\int d(x, 0)\,\mathbf{Q}(\mathrm{d}x)$ sont finies. Il existe $\mu$ de $(\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{X})) \times \mathcal{Z}$ telle que :*

*Adresses e-mail :* dedecker@ccr.jussieu.fr (J. Dedecker), prieur@cict.fr (C. Prieur).

1631-073X/$ – see front matter 
doi:10.1016/j.crma.2004.03.015

(i) *$\mu$ est un probabilité conditionnelle sur $(\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{X})) \times \mathcal{Z}$.*
(ii) *Pour tout $A$ de $\mathcal{B}(\mathcal{X})$, $\mu(A \times \mathcal{X}, \cdot) = P(A, \cdot)$, $\pi$-p.s. et $\mu(\mathcal{X} \times A, \cdot) = Q(A, \cdot)$, $\pi$-p.s.*
(iii) *$\int d(x, y)\mu(\mathrm{d}x, \mathrm{d}y, \cdot) = \sup_{f \in \Lambda_1(\mathcal{X})} |\int f(x)P(\mathrm{d}x, \cdot) - \int f(x)Q(\mathrm{d}x, \cdot)|$, $\pi$-p.s.*

**Remarque 1.** La fonction $M = \sup_{f \in \Lambda_1(\mathcal{X})} |\int f(x)P(\mathrm{d}x, \cdot) - \int f(x)Q(\mathrm{d}x, \cdot)|$ n'est définie que sur l'ensemble $A$ (de $\pi$-mesure 1) des $z$ tels que $\int d(x, 0)P(\mathrm{d}x, z)$ et $\int d(x, 0)Q(\mathrm{d}x, z)$ sont finies. Il découle de la Proposition 1.2 qu'il existe un ensemble $B$ de $\pi$-mesure 1 inclus dans $A$ tel que la fonction $z \to M(z)\mathbf{I}_B(z)$ est $\mathcal{B}(\mathcal{Z})$-mesurable (dans le cas où $\mathcal{X}$ est compact, on peut prendre $B = \mathcal{Z}$). En particulier la définition (5) de la Section 2.1 a bien un sens.

**Preuve de la Proposition 1.2.** On reprend la preuve de Fernique [4].

*Cas $\mathcal{X}$ fini.* Si $\mathcal{X} = \{x_1, \ldots, x_n\}$, alors il existe $\mu_{i,j}(z)$ vérifiant le problème de Monge : $\mu_{i,j}(z) \geqslant 0$, $\sum_{j=1}^n \mu_{i,j}(z) = P(\{x_i\}, z)$, $\sum_{i=1}^n \mu_{i,j}(z) = Q(\{x_j\}, z)$, et la quantité $\sum d(x_i, x_j)\mu_{i,j}(z)$ est minimale pour ces contraintes. C'est un problème d'optimisation convexe. La solution $\mu_{i,j}(z)$ est l'un des sommet du polygone convexe dont les bords sont determinés par les contraintes. Comme les points du sommet sont des fonctions mesurables de $(P(\{x_i\}, z), Q(\{x_j\}, z))$, la fonction $z \to \mu_{i,j}(z)$ est mesurable. Le point (iii) vient du théorème de Kantorovitch et Rubinstein (il est vrai pour tout $z$).

*Cas $\mathcal{X}$ compact.* Pour tout $n$, on choisit des boréliens disjoints $A_1, \ldots, A_{m(n)}$ de diamètre plus petit que $1/n$ et dont la réunion vaut $\mathcal{X}$. Soit $x_i$ un point de $A_i$. On définit $h_n : \mathcal{X} \mapsto \mathcal{X}_n = \{x_1, \ldots, x_{m(n)}\}$ par : $h_n(A_i) = x_i$. On se ramène au cas fini en posant $P_n = P \circ h_n^{-1}$ et $Q_n = Q \circ h_n^{-1}$. Par application du cas précédent, il existe $\mu_n$ supportée par $\mathcal{X}_n$ vérifiant les 3 points de la Proposition 1.2 pour $P_n$ et $Q_n$. Sur l'espace polonais $(\mathcal{X} \times \mathcal{X} \times \mathcal{Z}, \mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Z}))$ on met la probabilité

$$\nu_n(A \times B) = \int_B \mu_n(A, z)\pi(\mathrm{d}z). \tag{1}$$

Puisque $\mathcal{Z}$ est Polonais, $\pi$ est tendue, et comme $\mathcal{X} \times \mathcal{X}$ est compact, la famille $(\nu_n)_{n \geqslant 0}$ est tendue (les marginales le sont). On peut extraire une sous-suite $\nu_{g(n)}$ qui converge étroitement vers une limite $\nu$. La première marginale de $\nu$ est $\pi$. La probabilité $\nu$ peut donc s'écrire $\nu(A \times B) = \int_B \mu(A, z)\pi(\mathrm{d}z)$ pour une fonction $\mu$ vérifiant le point (i) de la Proposition 1.2. Puisque pour tout $z$ et toute $f$ $C$-lipschitzienne, $|P_n(f) - P(f)| \leqslant C/n$ (de même pour $Q_n$ et $Q$), on déduit de l'unicité des probabilités conditionnelles que $\mu$ vérifie le point (ii) de la Proposition 1.2. Par conséquent

$$\int d(x, y)\mu(\mathrm{d}x, \mathrm{d}y, \cdot) \geqslant \sup_{f \in \Lambda_1(\mathcal{X})} \left| \int f(x)P(\mathrm{d}x, \cdot) - \int f(x)Q(\mathrm{d}x, \cdot) \right| \quad \pi\text{-p.s.} \tag{2}$$

Comme $\nu_{g(n)}$ converge étroitement vers $\nu$, on a aussi

$$\int \left( \int d(x, y)\mu(\mathrm{d}x, \mathrm{d}y, z) \right) \pi(\mathrm{d}z) = \int \left( \sup_{f \in \Lambda_1(\mathcal{X})} \left| \int f(x)P(\mathrm{d}x, z) - \int f(x)Q(\mathrm{d}x, z) \right| \right) \pi(\mathrm{d}z),$$

ce qui avec (2) prouve le point (iii) de la Proposition 1.2.

*Cas $\mathcal{X}$ polonais.* Puisque $\mathcal{X}$ est polonais, il existe $K_n$ suite croissante de compacts de $\mathcal{X}$ telle que $\mathbf{P}(K_n)$ et $\mathbf{Q}(K_n)$ soient plus grand que $(n-1)/n$. Soit $a$ un point de $\mathcal{X}$. On se ramène au cas compact en posant $P_n(A, z) = P(A \cap K_n, z) + (1 - P(K_n, z))\delta_a(A)$ (on définit $Q_n$ de la même manière à partir de $Q$ et $K_n$). Il existe $\mu_n$ qui vérifie les 3 points de la Proposition 1.2 pour les contraintes $P_n$ et $Q_n$. Pour toute fonction $f$ $C$-lipschitzienne on a $|P_n(f) - P(f)| \leqslant C \int_{K_n^c} d(x, 0)P(\mathrm{d}x) + Cd(a, 0)P(K_n^c)$ (de même pour $Q_n$ et $Q$). Puisque $\mathbf{P}(K_n)$ tend vers 1, on en déduit que $\pi$-p.s. $P(K_n)$ tend vers 1 (de même pour $Q(K_n)$). Par suite, $\pi$-p.s. $P_n$ et $Q_n$ convergent étroitement vers $P$ et $Q$. On en déduit que $\nu_n$ définie par (1) a des marges étroitement convergentes.

Puisque $\mathcal{Z}$ et $\mathcal{X}$ sont polonais, $(\nu_n)_{n\geqslant 0}$ est tendue et on peut en extraire $\nu_{g(n)}$ qui converge étroitement vers une limite $\nu$. On conclut comme dans le cas compact.

## 2. Application : couplage pour la distance minimale

**Proposition 2.1.** *Soit $(\Omega,\mathcal{A},\mathbb{P})$ un espace probabilisé, $(\mathcal{X},d)$ et $\mathcal{Z}$ deux espaces polonais. Soient $X$ et $Z$ deux variables aléatoires à valeurs dans $\mathcal{X}$ et $\mathcal{Z}$ respectivement. On note $\mathbb{P}_{X|Z=\cdot}$ une distribution conditionnelle de $X$ sachant $Z$, et $\mathbb{P}_Z$ la distribution de $Z$. On suppose que $\Omega$ est assez riche, c'est à dire qu'il existe une variable aléatoire $U$ de $(\Omega,\mathcal{A},\mathbb{P})$ dans $([0,1],\mathcal{B}([0,1]))$, indépendante de $(X,Z)$ et de loi uniforme sur* $[0,1]$. *Soit $Q$ une probabilité conditionnelle sur $\mathcal{B}(\mathcal{X})\times\mathcal{Z}$. Il existe $Y$ de $(\Omega,\mathcal{A},\mathbb{P})$ dans $(\mathcal{X},\mathcal{B}(\mathcal{X}))$, $\sigma(U,X,Z)$-mesurable, telle que $(Y,Z)$ ait pour loi $\mathbb{P}_{Y,Z}(A\times B)=\int_B Q(A,z)\mathbb{P}_Z(\mathrm{d}z)$, et*

$$\mathbb{E}\big(d(X,Y)|Z=\cdot\big)=\sup_{f\in\Lambda_1(\mathcal{X})}\left|\int f(x)\mathbb{P}_{X|Z=\cdot}(\mathrm{d}x)-\int f(x)Q(dx,\cdot)\right|,\quad \mathbb{P}_Z\text{-}p.s. \tag{3}$$

**Démonstration.** Soit $\mu$ la probabilité conditionnelle de la Proposition 1.2, avec $P=\mathbb{P}_{X|Z=\cdot}$ et $\pi=\mathbb{P}_Z$. Sur $(S,\mathcal{B}(S))=(\mathcal{X}\times\mathcal{X}\times\mathcal{Z},\mathcal{B}(\mathcal{X})\otimes\mathcal{B}(\mathcal{X})\otimes\mathcal{B}(\mathcal{Z}))$ on définit

$$\nu(A\times B)=\int_B \mu(A,z)\mathbb{P}_Z(\mathrm{d}z). \tag{4}$$

On note $I=(I_1,I_2,I_3)$ l'identité de $S$ dans $S$. Supposons que l'on puisse construire $Y$ de $(\Omega,\mathcal{A})$ dans $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ telle que $\mathbb{P}_{Y|X=x,\,Z=z}=\mathbb{P}_{I_2|I_1=x,\,I_3=z}$. La distribution de $(X,Y,Z)$ est alors donnée par : $\mathbb{P}_{X,Y,Z}(A\times B\times C)=\int_{A\times C}\mathbb{P}_{I_2|I_1=x,\,I_3=z}(B)\mathbb{P}_{X,Z}(\mathrm{d}x,\mathrm{d}z)$. Grâce au point (ii) de la Proposition 1.2, on a $\mathbb{P}_{X,Z}=\mathbb{P}_{I_1,I_3}$, et donc $\mathbb{P}_{X,Y,Z}=\mathbb{P}_{I_1,I_2,I_3}=\nu$. On déduit de (4) que $\mu$ est une distribution conditionnelle de $(X,Y)$ sachant $Z$, ce qui entraîne (3).

Construisons $Y$. D'après le lemme de Skorohod [8], il existe une application $f$ de $\mathcal{X}$ dans un un borélien de $[0,1]$, bijective et bimesurable pour les tribus boréliennes. Si $F(t,x,z)=\mathbb{P}_{f(I_2)|I_1=x,\,I_3=z}(]-\infty,t])$, alors $F(\cdot,x,z)$ est une fonction de répartition d'inverse càdlàg $F^{-1}(\cdot,x,z)$. On montre facilement que la fonction $(u,x,z)\to F^{-1}(u,x,z)$ est mesurable par rapport à $\mathcal{B}([0,1])\otimes\mathcal{B}(\mathcal{X})\otimes\mathcal{B}(\mathcal{Z})$ (voir [1]). On pose $T(\omega)=F^{-1}(U(\omega),X(\omega),Z(\omega))$, et enfin $Y=f^{-1}(T)$. Il reste à vérifier que $\mathbb{P}_{Y|X=x,\,Z=z}=\mathbb{P}_{I_2|I_1=x,\,I_3=z}$.

$$\mathbb{E}(\mathbf{I}_{X\in A}\mathbf{I}_{Z\in B}\mathbf{I}_{Y\in f^{-1}(]-\infty,t])})=\mathbb{E}(\mathbf{I}_{X\in A}\mathbf{I}_{Z\in B}\mathbf{I}_{T\leqslant t})=\int \mathbf{I}_{X(\omega)\in A}\mathbf{I}_{Z(\omega)\in B}\mathbf{I}_{U(\omega)\leqslant F(t,X(\omega),Z(\omega))}\mathbb{P}(\mathrm{d}\omega).$$

Puisque $U$ est indépendante de $(X,Z)$, on obtient donc

$$\begin{aligned}\mathbb{E}(\mathbf{I}_{X\in A}\mathbf{I}_{Z\in B}\mathbf{I}_{Y\in f^{-1}(]-\infty,t])})&=\int \mathbf{I}_{X(\omega)\in A}\mathbf{I}_{Z(\omega)\in B}F\big(t,X(\omega),Z(\omega)\big)\mathbb{P}(\mathrm{d}\omega)\\&=\int \mathbf{I}_{x\in A}\mathbf{I}_{z\in B}\mathbb{P}_{I_2|I_1=x,I_3=z}\big(f^{-1}\big(]-\infty,t]\big)\big)\mathbb{P}_{I_1,I_3}(\mathrm{d}x,\mathrm{d}z).\end{aligned}$$

Comme $\{f^{-1}(]-\infty,t]),\,t\in[0,1]\}$ est stable par intersection finie et engendre $\mathcal{B}(\mathcal{X})$, le résultat suit.

### *2.1. Couplage et coefficient de dépendance*

Soit $(\Omega,\mathcal{A},\mathbb{P})$ un espace probabilisé, $(\mathcal{X},d)$ et $\mathcal{Z}$ deux espaces polonais. Soient $X$ et $Z$ deux variables aléatoires à valeurs dans $\mathcal{X}$ et $\mathcal{Z}$ respectivement. On pose $\sigma(Z)=\mathcal{M}$, et on définit comme dans [1] et [2]

$$\tau(\mathcal{M},X)=\left\|\sup_{f\in\Lambda_1(\mathcal{X})}\left|\int f(x)\mathbb{P}_{X|Z}(\mathrm{d}x)-\int f(x)\mathbb{P}_X(\mathrm{d}x)\right|\right\|_1. \tag{5}$$

La version uniforme de ce coefficient à été introduite par Rio dans [7]. Ce coefficient est une mesure de dépendance entre $\mathcal{M}$ et $X$. Notons que le coefficient de $\beta$-mélange $\beta(\mathcal{M}, \sigma(X))$ entre $\mathcal{M}$ et $\sigma(X)$ est obtenu à partir de (5) en remplaçant $\Lambda_1(\mathcal{X})$ par l'ensemble des fonctions mesurables bornées par $1/2$. Comme pour $\beta$, on peut définir le coefficient de dépendance d'une suite de variables aléatoires. Dans [1] et [2], nous donnons de nombreux exemples de suites $\tau$-dépendantes qui ne sont pas mélangeantes.

Comme conséquence de la Proposition 2.1 (en prenant pour $Q$ la distribution de $X$), on a le résultat de couplage suivant, qui généralise celui obtenu dans [1] lorsque $\mathcal{X} = \mathbb{R}$.

**Corollaire 2.2.** *Si $\Omega$ est assez riche* (*voir Proposition* 2.1), *il existe $X^*$ mesurable pour la tribu $\sigma(U, X, Z)$, indépendante de $\mathcal{M} = \sigma(Z)$ et de même distribution que $X$, telle que $\tau(\mathcal{M}, X) = \mathbb{E}(d(X, X^*))$.*

**Remarque 2.** Notons $Q_{d(X,x)}$ l'inverse càdlàg de $t \to \mathbb{P}(d(X, x) > t)$ et $T = Q_{d(X,x)}(\beta(\mathcal{M}, \sigma(X)))$. Clairement $d(X, X^*) = d(X, X^*) \wedge T + (d(X, X^*) - T)_+$. En partant de cette égalité et en procédant comme dans [6], p. 174, on obtient que pour tout $x$ dans $\mathcal{X}$,

$$\tau(\mathcal{M}, X) \leqslant 2 \int_0^{\beta(\mathcal{M},\sigma(X))} Q_{d(X,x)}(u)\,\mathrm{d}u. \tag{6}$$

Si $\mathcal{X} = \mathbb{R}$, (6) est valable en remplaçant $\beta(\mathcal{M}, \sigma(X))$ par $\alpha(\mathcal{M}, X) = \sup_{t\in\mathbb{R}} \|\mathbb{E}(\mathbf{I}_{X\leqslant t}|Z) - \mathbb{P}(X \leqslant t)\|_1$ (voir [1]). En revanche, si $\mathcal{X}$ est de dimension infinie, on ne peut remplacer $\beta(\mathcal{M}, \sigma(X))$ par le coefficient plus faible $\alpha(\mathcal{M}, \sigma(X)) = \sup_{A\in\mathcal{B}(\mathbb{R})} \|P_{X|\mathcal{M}}(A) - \mathbb{P}_X(A)\|_1$ (voir [3] pour plus de détails). Notons enfin que le $X^*$ du Corollaire 2.2 minimise $Y \to \mathbb{E}(d(X, Y))$ sur l'ensemble des variables $Y$ de même loi que $X$ et indépendantes de $Z$. Lorsque $\mathcal{X} = \mathbb{R}$, l'existence d'un tel $X^*$ est due à Major [5].

## Remerciements

Nous remercions Emmanuel Rio de nous avoir suggéré que la propriété de couplage de $\tau(\mathcal{M}, X)$ établie dans [1] pouvait s'étendre au cas où $X$ est à valeurs dans un espace polonais.

## Références

[1] J. Dedecker, C. Prieur, Coupling for $\tau$-dependent sequences and applications, Prépublication, 2003. www.ccr.jussieu.fr/lsta/prepublications.html.
[2] J. Dedecker, C. Prieur, New dependence coefficients. Examples and applications to statistics, Prépublication, 2003. www.ccr.jussieu.fr/lsta/prepublications.html.
[3] H. Dehling, A note on a theorem of Berkes and Philipp, Z. Wahrsch. Verw. Gebiete 62 (1983) 39–42.
[4] X. Fernique, Sur le théorème de Kantorovitch–Rubinstein dans les espaces polonais, in: Séminaire de Probabilités XV, in: Lecture Notes in Math., vol. 850, Springer, 1981, pp. 6–10.
[5] P. Major, On the invariance principle for sums of identically distributed random variables, J. Multivariate Anal. 8 (1978) 487–517.
[6] F. Merlevède, M. Peligrad, On the coupling of dependent random variables and applications, in: Empirical Process Techniques for Dependent Data, Birkhäuser, 2002, pp. 171–193.
[7] E. Rio, Sur le théorème de Berry–Esseen pour les suites faiblement dépendantes, Probab. Theory Related Fields 104 (1996) 255–282.
[8] A.V. Skorohod, On a representation of random variables, Theory Probab. Appl. 21 (1976) 628–632.