Parametrized Kantorovich-Rubinštein theorem and application to the coupling of random variables.

Jérôme Dedecker[1], Clémentine Prieur[2], and Paul Raynaud De Fitte[3]

[1] Laboratoire de Statistique Théorique et Appliquée, Université Paris 6, 175 rue du Chevaleret, 75013 Paris, France. dedecker@ccr.jussieu.fr
[2] Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 4, France. prieur@cict.fr
[3] Laboratoire Raphaël Salem, UMR CNRS 6085, UFR Sciences , Université de Rouen, 76821 Mont Saint Aigan Cedex, France. prf@univ-rouen.fr

Summary. We present an approach, going back to Rüschendorf [23], to obtain an optimal coupling for random variables with values in some completely regular topological space $\mathbb{S}$. The main step is to prove a version for random measures of the Kantorovich-Rubinštein duality theorem. This leads to a family of dependence coefficients defined over the class of 1-lipschitz functions with respect to a given metric c on $\mathbb{S}$. In particular the β-mixing coefficient and the well known coupling result of Berbee [1] correspond to the case where c is the discrete metric. To be complete we show that, contrary to Berbee's coupling, the more precise "maximal" coupling result of Goldstein [16] cannot be extended to other metrics than the discrete one.

1 Introduction and notations

Let μ and ν be two probability measures on a Polish space $(\mathbb{S}, d)$. In 1970 Dobrušin [12, page 472] proved that there exists a probability measure λ on $\mathbb{S} \times \mathbb{S}$ with marginals μ and ν, such that

$$\lambda(\{x \neq y, (x,y) \in \mathbb{S} \times \mathbb{S}\}) = \frac{1}{2}\|\mu - \nu\|_v \,, \tag{1}$$

where $\|\cdot\|_v$ is the variation norm. More precisely, Dobrušin gave an explicit solution to (1) defined by

$$\lambda(A \times B) = (\mu - \pi_+)(A \cap B) + \frac{\pi_+(A)\pi_-(B)}{\pi_+(\mathbb{S})} \quad \text{for } A,\, B \text{ in } \mathcal{B}_{\mathbb{S}}\,, \tag{2}$$

where $\mu - \nu = \pi_+ - \pi_-$ is the Hahn decomposition of $\pi = \mu - \nu$.

Starting from (2) (see [1, Proposition 4.2.1]), Berbee obtained the following coupling result ([1, Corollary 4.2.5]): let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, let $\mathcal{M}$ be a σ-algebra of $\mathcal{A}$, and let X be a random variable with values in $\mathbb{S}$. Denote by P_X the distribution of X and by $\mathrm{P}_{X|\mathcal{M}}$ a regular conditional distribution of X given $\mathcal{M}$. If Ω is rich enough, there exists X^* distributed as X and independent of $\mathcal{M}$ such that

$$\mathrm{P}(X \neq X^*) = \frac{1}{2}E(\|\mathrm{P}_{X|\mathcal{M}} - \mathrm{P}_X\|_v)\,. \tag{3}$$

To prove (3), Berbee built a couple (X, X^*) whose conditional distribution given $\mathcal{M}$ is the random probability λ_ω defined by (2), with random marginals $\mu = \mathrm{P}_{X|\mathcal{M}}$ and $\nu = \mathrm{P}_X$.

It is by now well known that Dobrušin's result (1) is a particular case of the Kantorovich-Rubinštein duality theorem (which we recall at the beginning of Section 2) applied to the discrete metric $c(x,y) = \mathbb{1}_{x \neq y}$ (see [20, page 93]). Starting from this simple remark, Berbee's proof can be described as follows: one can find a couple (X, X^*) whose conditional distribution with respect to $\mathcal{M}$ solves the duality problem with cost function $c(x, y) = \mathbb{1}_{x \neq y}$ and random marginals $\mu = \mathrm{P}_{X|\mathcal{M}}$ and $\nu = \mathrm{P}_X$.

A reasonable question is then: for what class of cost functions can we obtain the same kind of coupling than Berbee's? Or, equivalently, given two random probabilities μ_ω and ν_ω on a Polish space $(\mathbb{S}, d)$, for what class of cost functions is there a random probability λ_ω on $\mathbb{S}\times\mathbb{S}$ solution to the duality problem with marginals (μ_ω, ν_ω)? Combining Proposition 4 in [23] and the Kantorovitch-Rubinštein duality Theorem, we shall see in point 1 of Theorem 2.1 that such a λ_ω exists provided the cost function c satisfies

$$c(x, y) = \sup_{u \in \mathrm{Lip}_{\mathbb{S}}^{(c)}} |u(x) - u(y)| , \tag{4}$$

where $\mathrm{Lip}_{\mathbb{S}}^{(c)}$ is the class of continuous bounded functions u on $\mathbb{S}$ such that $|u(x) - u(y)| \leq c(x, y)$. In fact, except for the duality, Rüschendorf proved in [23, Proposition 4] a more general result, which is true for any measurable cost funcion c. In point 2 of Theorem 2.1 we also prove that the parametrized Kantorovich–Rubinštein theorem given in [5, Theorem 3.4.1] still holds for any cost function c satisfying (4).

In Section 3, we give the application of Theorem 2.1 to the coupling of random variables, as done in Section 2 of [23]. In particular, Corollary 1 extends Berbee's coupling in the following way: if $(\Omega, \mathcal{A}, \mathrm{P})$ is rich enough, and if c is a mapping satisfying (4) such that $\int c(X, x_0) d\,\mathrm{P}$ is finite for some x_0 in $\mathbb{S}$, then there exists a random variable X^* distributed as X and independent of $\mathcal{M}$ such that

$$E\left(c(X, X^*)\right) = \left\| \sup_{f \in \mathrm{Lip}_{\mathbb{S}}^{(c)}} \left| \int f(x) \mathrm{P}_{X|\mathcal{M}}(dx) - \int f(x) \mathrm{P}_X(dx) \right| \right\|_1 . \tag{5}$$

If $c(x, y) = \mathbb{1}_{x \neq y}$ is the discrete metric, (5) is exactly Berbee's coupling (3). If $c = d$, (5) has been proved in [23, Proposition 6]. For more details on the coupling property (5) and its applications, see Section 3.2.

In 1979, Goldstein [16] obtained a more precise result than (1) in the case where $\mathbb{S} = \mathbb{S}_1^\infty = \Pi_{k=1}^\infty \mathbb{M}$ is a product space. This result can be written as follows: let μ and ν be two probability measures on $\mathbb{S}_1^\infty$ and let $\mu_{(i)}$ and $\nu_{(i)}$ be the marginals of μ and ν on $\mathbb{S}_i^\infty = \Pi_{k=i}^\infty \mathbb{M}$. There exists a probability measure λ on $\mathbb{S}_1^\infty \times \mathbb{S}_1^\infty$ with marginals $\lambda_{(i)}$ on $\mathbb{S}_i^\infty \times \mathbb{S}_i^\infty$, such that $\lambda(\cdot \times \mathbb{S}_1^\infty) = \mu(\cdot)$, $\lambda(\mathbb{S}_1^\infty \times \cdot) = \nu(\cdot)$, and for any $i \geq 1$,

$$\frac{1}{2} \|\mu_{(i)} - \nu_{(i)}\|_v = \lambda_{(i)}(\{x \neq y, (x, y) \in \mathbb{S}_i^\infty \times \mathbb{S}_i^\infty\}). \tag{6}$$

Starting from (6) (see [1, Theorem 4.3.2]), Berbee obtained the following coupling result ([1, Theorem 4.4.7]): let $X = (X_k)_{k \geq 1}$ be a $\mathbb{S}_1^\infty$-valued random variable and let $X_{(i)} = (X_k)_{k \geq i}$. If Ω is rich enough, there exists X^* distributed as X and independant of $\mathcal{M}$ such that, for any $i \geq 1$,

$$\frac{1}{2} E(\| \mathrm{P}_{X_{(i)}|\mathcal{M}} - \mathrm{P}_{X_{(i)}} \|_v) = \mathrm{P}(X_{(i)} \neq X^*_{(i)}) , \tag{7}$$

where $P_{X_{(i)}}$ is the distribution of $X_{(i)}$ and $P_{X_{(i)}|\mathcal{M}}$ is a regular distribution of $X_{(i)}$ given $\mathcal{M}$. If $(X_k)_{k \in \mathbb{Z}}$ is a strictly stationary sequence of $\mathbb{M}$-valued random variables and $\mathcal{M} = \sigma(X_i, i \leq 0)$, the sequences for which $\mathrm{P}(X_{(i)} \neq X^*_{(i)})$ converges to zero as i tends to infinity are called β-mixing or absolutely regular sequences. The property (7) is very powerful (see [21] and [3] for recent applications).

In Section 4, we shall see that, contrary to (1), the property (6) is characteristic of the discrete metric. Hence, no analogue of (7) is possible if the underlying cost function is not proportional to the discrete metric.

Preliminary notations

For any topological space $\mathfrak{T}$, we denote by $\mathcal{B}_\mathfrak{T}$ the Borel σ–algebra of $\mathfrak{T}$ and by $\mathcal{P}(\mathfrak{T})$ the space of probability laws on $(\mathfrak{T}, \mathcal{B}_\mathfrak{T})$, endowed with the narrow topology, that is, for every mapping $\varphi : \mathfrak{T} \to [0, 1]$, the mapping $\mu \mapsto \int_\mathfrak{T} \varphi\, d\mu$ is l.s.c. if and only if φ is l.s.c.

Throughout, $\mathbb{S}$ is a given completely regular topological space and $(\Omega, \mathcal{A}, \mathrm{P})$ a given probability space. Note that in [23], both Ω and $\mathbb{S}$ were assumed to be Polish. However the results are valid in much more general spaces, without significant changes in the proofs. *The reader who is not interested by this level of generality may assume as well in the sequel that all topological spaces we consider are Polish.* On the other hand, we give in appendix some definitions and references which might be useful for a complete reading.

2 Parametrized Kantorovich–Rubinštein theorem

Most of the ideas of this Section are contained in [23], except for the duality part of point 2 of Theorem 1, which draws inspiration from [5, §3.4].

For any $\mu, \nu \in \mathcal{P}(\mathbb{S})$, let $D(\mu, \nu)$ be the set of probability laws π on $(\mathbb{S} \times \mathbb{S}, \mathcal{B}_{\mathbb{S}\times\mathbb{S}})$ with marginals μ and ν, that is, $\pi(A \times \mathbb{S}) = \mu(A)$ and $\pi(\mathbb{S} \times A) = \nu(A)$ for every $A \in \mathcal{B}_{\mathbb{S}}$. Let us recall the

Kantorovich–Rubinštein duality theorem [18], [20, Theorem 4.6.6] *Assume that $\mathbb{S}$ is a completely regular pre-Radon space*[4]*, that is, every finite τ–additive Borel measure on $\mathbb{S}$ is inner regular with respect to the compact subsets of $\mathbb{S}$. Let $c : \mathbb{S} \times \mathbb{S} \to [0, +\infty]$ be a universally measurable mapping. For every $(\mu, \nu) \in \mathcal{P}(\mathbb{S}) \times \mathcal{P}(\mathbb{S})$, let us denote*

$$\Delta^{(c)}_{\mathrm{KR}}(\mu, \nu) := \inf_{\pi\in D(\mu,\nu)} \int_{\mathbb{S}\times\mathbb{S}} c(x, y)\, d\pi(x, y),$$
$$\Delta^{(c)}_{\mathrm{L}}(\mu, \nu) := \sup_{f\in \mathrm{Lip}^{(c)}_{\mathbb{S}}} (\mu(f) - \nu(f))$$

where $\mathrm{Lip}^{(c)}_{\mathbb{S}} = \{u \in \mathrm{C}_b(\mathbb{S})\,;\, \forall x, y \in \mathbb{S} \quad |u(x) - u(y)| \le c(x, y)\}$. *Then the equality* $\Delta^{(c)}_{\mathrm{KR}}(\mu, \nu) = \Delta^{(c)}_{\mathrm{L}}(\mu, \nu)$ *holds for all* $(\mu, \nu) \in \mathcal{P}(\mathbb{S}) \times \mathcal{P}(\mathbb{S})$ *if and only if (4) holds.*

Note that, if c satifies (4), it is the supremum of a set of continuous functions, thus it is l.s.c. Every continuous metric c on $\mathbb{S}$ satisfies (4) (see [20, Corollary 4.5.7]), and, if $\mathbb{S}$ is compact, every l.s.c. metric c on $\mathbb{S}$ satisfies (4) (see [20, Remark 4.5.6]).

Now, we denote

$$\mathcal{Y}(\Omega, \mathcal{A}, \mathrm{P}; \mathbb{S}) = \{\mu \in \mathcal{P}(\Omega \times \mathbb{S}, \mathcal{A} \otimes \mathcal{B}_{\mathbb{S}});\, \forall A \in \mathcal{A} \quad \mu(A \times \mathbb{S}) = \mathrm{P}(A)\}.$$

When no confusion can arise, we omit some part of the information, and use notations such as $\mathcal{Y}(\mathcal{A})$ or simply $\mathcal{Y}$ (same remark for the set $\mathcal{Y}^{c,1}(\Omega, \mathcal{A}, \mathrm{P}; \mathbb{S})$ defined below). If $\mathbb{S}$ is a Radon space, every $\mu \in \mathcal{Y}$ is *disintegrable*, that is, there exists a (unique, up to P-a.e. equality) $\mathcal{A}^*_\mu$-measurable mapping $\omega \mapsto \mu_\omega$, $\Omega \to \mathcal{P}(\mathbb{S})$, such that

$$\mu(f) = \int_\Omega \int_{\mathbb{S}} f(\omega, x)\, d\mu_\omega(x)\, d\,\mathrm{P}(\omega)$$

for every measurable $f : \Omega \times \mathbb{S} \to [0, +\infty]$ (see [27]). If furthermore the compact subsets of $\mathbb{S}$ are metrizable, the mapping $\omega \mapsto \mu_\omega$ can be chosen $\mathcal{A}$-measurable, see the Appendix.

Let c satisfy (4). We denote

$$\mathcal{Y}^{c,1}(\Omega, \mathcal{A}, \mathrm{P}; \mathbb{S}) = \{\mu \in \mathcal{Y};\, \int_{\Omega\times\mathbb{S}} c(x, x_0)\, d\mu(\omega, x) < +\infty\}$$

[4] In [18] and [20, Theorem 4.6.6], the space $\mathbb{S}$ is assumed to be a universally measurable subset of some compact space. But this amounts to assume that it is completely regular and pre-Radon: see [20, Lemma 4.5.17] and [15, Corollary 11.8].

where x_0 is some fixed element of $\mathbb{S}$ (this definition is independent of the choice of x_0). For any $\mu, \nu \in \mathcal{Y}$, let $\underline{D}(\mu, \nu)$ be the set of probability laws π on $\Omega \times \mathbb{S} \times \mathbb{S}$ such that $\pi(. \times . \times \mathbb{S}) = \mu$ and $\pi(. \times \mathbb{S} \times .) = \nu$. We now define the parametrized versions of $\Delta_{\mathrm{KR}}^{(c)}$ and $\Delta_{\mathrm{L}}^{(c)}$. Set, for $\mu, \nu \in \mathcal{Y}^{c,1}$,

$$\underline{\Delta}_{\mathrm{KR}}^{(c)}(\mu, \nu) = \inf_{\pi \in \underline{D}(\mu,\nu)} \int_{\Omega \times \mathbb{S} \times \mathbb{S}} c(x, y) \, d\pi(\omega, x, y).$$

Let also $\underline{\mathrm{Lip}}^{(c)}$ denote the set of measurable integrands $f : \Omega \times \mathbb{S} \to \mathbb{R}$ such that $f(\omega, .) \in \mathrm{Lip}_{\mathbb{S}}^{(c)}$ for every $\omega \in \Omega$. We denote

$$\underline{\Delta}_{\mathrm{L}}^{(c)}(\mu, \nu) = \sup_{f \in \underline{\mathrm{Lip}}^{(c)}} (\mu(f) - \nu(f)) \, .$$

Theorem 1 (Parametrized Kantorovich–Rubinštein theorem) *Assume that $\mathbb{S}$ is a completely regular Radon space and that the compact subsets of $\mathbb{S}$ are metrizable (e.g. $\mathbb{S}$ is a regular Suslin space). Let $c : \mathbb{S} \times \mathbb{S} \to [0, +\infty[$ satisfy (4). Let $\mu, \nu \in \mathcal{Y}^{c,1}$ and let $\omega \mapsto \mu_\omega$ and $\omega \mapsto \nu_\omega$ be disintegrations of μ and ν respectively.*

1. *Let $G : \omega \mapsto \Delta_{\mathrm{KR}}^{(c)}(\mu_\omega, \nu_\omega) = \Delta_{\mathrm{L}}^{(c)}(\mu_\omega, \nu_\omega)$ and let $\mathcal{A}^*$ be the universal completion of $\mathcal{A}$. There exists an $\mathcal{A}^*$–measurable mapping $\omega \mapsto \lambda_\omega$ from Ω to $\mathcal{P}(\mathbb{S} \times \mathbb{S})$ such that λ_ω belongs to $D(\mu_\omega, \nu_\omega)$ and*

$$G(\omega) = \int_{\mathbb{S} \times \mathbb{S}} c(x, y) \, d\lambda_\omega(x, y).$$

2. *The following equalities hold:*

$$\underline{\Delta}_{\mathrm{KR}}^{(c)}(\mu, \nu) = \int_{\Omega \times \mathbb{S} \times \mathbb{S}} c(x, y) \, d\lambda(\omega, x, y) = \underline{\Delta}_{\mathrm{L}}^{(c)}(\mu, \nu),$$

where λ is the element of $\mathcal{Y}(\Omega, \mathcal{A}, \mathrm{P}; \mathbb{S} \times \mathbb{S})$ defined by $\lambda(A \times B \times C) = \int_A \lambda_\omega(B \times C) \, d\,\mathrm{P}(\omega)$ for any A in $\mathcal{A}$, B and C in $\mathcal{B}_{\mathbb{S}}$. In particular, λ belongs to $\underline{D}(\mu, \nu)$, and the infimum in the definition of $\underline{\Delta}_{\mathrm{KR}}^{(c)}(\mu, \nu)$ is attained for this λ.

Remark 1. In the case where both Ω and $\mathbb{S}$ are Polish spaces, point 1 and the first equality in point 2 of Theorem 1 are contained in Proposition 4 of Rüschendorf [23]. The proof we give below follows that of Proposition 4 in [23] and of Theorem 3.4.1 in [5]. As in [23], the main argument is a measurable selection lemma given in [6].

The set of compact subsets of a topological space $\mathfrak{T}$ is denoted by $\mathcal{K}(\mathfrak{T})$.

Lemma 1 (A measurable selection lemma) *Assume that $\mathbb{S}$ is a Suslin space. Let $c : \mathbb{S} \times \mathbb{S} \to [0, +\infty]$ be an l.s.c. mapping. Let $\mathcal{B}^*$ be the universal completion of the σ–algebra $\mathcal{B}_{\mathcal{P}(\mathbb{S}) \times \mathcal{P}(\mathbb{S})}$. For any $\mu, \nu \in \mathcal{P}(\mathbb{S})$, let*

$$r(\mu, \nu) = \inf_{\pi \in D(\mu,\nu)} \int c(x, y) \, d\pi(x, y) \in [0, +\infty].$$

The function r is $\mathcal{B}^*$–measurable. Furthermore, the multifunction

$$K : \begin{cases} \mathcal{P}(\mathbb{S}) \times \mathcal{P}(\mathbb{S}) \to \mathcal{K}\left(\mathcal{P}(\mathbb{S} \times \mathbb{S})\right) \\ (\mu, \nu) \qquad\qquad\;\; \mapsto \left\{\pi \in D(\mu, \nu); \; \int c(x, y) \, d\pi(x, y) = r(\mu, \nu)\right\} \end{cases}$$

has a $\mathcal{B}^*$–measurable selection, that is, there exists a $\mathcal{B}^*$–measurable mapping $\lambda : (\mu, \nu) \mapsto \lambda_{\mu,\nu}$ defined on $\mathcal{P}(\mathbb{S}) \times \mathcal{P}(\mathbb{S})$ with values in $\mathcal{K}\left(\mathcal{P}(\mathbb{S} \times \mathbb{S})\right)$, such that $\lambda_{\mu,\nu} \in K(\mu, \nu)$ for all $\mu, \nu \in \mathcal{P}(\mathbb{S})$.

Proof. Observe first that the mapping r can be defined as

$$r:\ (\mu,\nu)\mapsto \inf\left\{\psi(\pi);\ \pi\in D(\mu,\nu)\right\},$$

with

$$\psi:\begin{cases}\mathcal{P}(\mathbb{S}\times\mathbb{S})\to[0,+\infty]\\ \pi\qquad\qquad\ \mapsto \int_{\mathbb{S}\times\mathbb{S}} c(x,y)\,d\pi(x,y).\end{cases}$$

The mapping ψ is l.s.c. because it is the supremum of the l.s.c. mappings $\pi\mapsto\pi(c\wedge n)$, $n\in\mathbb{N}$ (if c is bounded and continuous, ψ is continuous). Furthermore, we have $D=\Phi^{-1}$, where Φ is the continuous mapping

$$\Phi:\begin{cases}\mathcal{P}(\mathbb{S}\times\mathbb{S})\to\mathcal{P}(\mathbb{S})\times\mathcal{P}(\mathbb{S})\\ \lambda\qquad\qquad\ \mapsto(\lambda(.\times\mathbb{S}),\lambda(\mathbb{S}\times .))\end{cases}$$

(recall that $D(\mu,\nu)$ is the set of probability laws π on $\mathbb{S}\times\mathbb{S}$ with marginals μ and ν). Therefore, the graph $\operatorname{gph}(D)$ of D is a closed subset of the Suslin space $\mathbb{X}=\big(\mathcal{P}(\mathbb{S})\times\mathcal{P}(\mathbb{S})\big)\times\mathcal{P}(\mathbb{S}\times\mathbb{S})$. Applying Lemma III.39 of [6] as done in [23], we infer that r is $\mathcal{B}^*$–measurable. Now the fact that K has a $\mathcal{B}^*$–measurable selection follows from the application of Lemma III.39 given in paragraph 39 of [6]. □

Proof of Theorem 1. By the Radon property, the probability measures $\mu(\Omega\times .)$ and $\nu(\Omega\times .)$ are tight, that is, for every integer $n\geq 1$, there exists a compact subset K_n of $\mathbb{S}$ such that $\mu(\Omega\times(\mathbb{S}\setminus K_n))\leq 1/n$ and $\nu(\Omega\times(\mathbb{S}\setminus K_n))\leq 1/n$. Now, we can clearly replace $\mathbb{S}$ in the statements of Theorem 1 by the smaller space $\cup_{n\geq 1}K_n$. But $\cup_{n\geq 1}K_n$ is Suslin (and even Lusin), so we can assume without loss of generality that $\mathbb{S}$ is a regular Suslin space.

We easily have

$$\begin{aligned}\underline{\Delta}_{\mathrm{L}}^{(c)}(\mu,\nu)&=\sup_{f\in\underline{\mathrm{Lip}}^{(c)}}\int_\Omega\int_{\mathbb{S}}\int_{\mathbb{S}}\left(f(\omega,x)-f(\omega,y)\right)\,d\mu_\omega(x)\,d\nu_\omega(y)\,d\,\mathrm{P}(\omega)\\&\leq\int_\Omega\int_{\mathbb{S}}\int_{\mathbb{S}}c(x,y)\,d\mu_\omega(x)\,d\nu_\omega(y)\,d\,\mathrm{P}(\omega)\\&\leq\underline{\Delta}_{\mathrm{KR}}^{(c)}(\mu,\nu).\end{aligned}\tag{8}$$

So, to prove Theorem 1, we only need to prove that $\underline{\Delta}_{\mathrm{KR}}^{(c)}(\mu,\nu)\leq\underline{\Delta}_{\mathrm{L}}^{(c)}(\mu,\nu)$ and that the minimum in the definition of $\underline{\Delta}_{\mathrm{KR}}^{(c)}(\mu,\nu)$ is attained.

Using the notations of Lemma 1, we have $G(\omega)=r(\mu_\omega,\nu_\omega)$, thus G is $\mathcal{A}^*$–measurable (indeed, the mapping $\omega\mapsto(\mu_\omega,\nu_\omega)$ is measurable for $\mathcal{A}^*$ and $\mathcal{B}^*$ because it is measurable for $\mathcal{A}$ and $\mathcal{B}_{\mathcal{P}(\mathbb{S})\times\mathcal{P}(\mathbb{S})}$). From Lemma 1, the multifunction $\omega\mapsto D(\mu_\omega,\nu_\omega)$ has an $\mathcal{A}^*$–measurable selection $\omega\mapsto\lambda_\omega$ such that, for every $\omega\in\Omega$, $G(\omega)=\int_{\mathbb{S}\times\mathbb{S}}c(x,y)\,d\lambda_\omega(x,y)$. We thus have

$$\underline{\Delta}_{\mathrm{KR}}^{(c)}(\mu,\nu)\leq\int_{\Omega\times\mathbb{S}\times\mathbb{S}}c(x,y)\,d\lambda(\omega,x,y)=\int_\Omega G(\omega)\,d\,\mathrm{P}(\omega).\tag{9}$$

Furthermore, since $\mu,\nu\in\mathcal{Y}^{c,1}$, we have $G(\omega)<+\infty$ a.e. Let Ω_0 be the almost sure set on which $G(\omega)<+\infty$. Fix an element x_0 in $\mathbb{S}$. We have, for every $\omega\in\Omega_0$,

$$G(\omega)=\sup_{g\in\mathrm{Lip}_{\mathbb{S}}^{(c)}}\left(\mu_\omega(g)-\nu_\omega(g)\right)=\sup_{g\in\mathrm{Lip}_{\mathbb{S}}^{(c)},\ g(x_0)=0}\left(\mu_\omega(g)-\nu_\omega(g)\right).$$

Let $\epsilon>0$. Let $\widetilde{\mu}$ and $\widetilde{\nu}$ be the finite measures on $\mathbb{S}$ defined by

$$\widetilde{\mu}(B)=\int_{\Omega\times B}c(x_0,x)\,d\mu(\omega,x)\quad\text{ and }\quad\widetilde{\nu}(B)=\int_{\Omega\times B}c(x_0,x)\,d\nu(\omega,x)$$

for any $B \in \mathcal{B}_{\mathbb{S}}$. Let $\mathbb{S}_0$ be a compact subset of $\mathbb{S}$ containing x_0 such that $\widetilde{\mu}(\mathbb{S} \setminus \mathbb{S}_0) \leq \epsilon$ and $\widetilde{\nu}(\mathbb{S} \setminus \mathbb{S}_0) \leq \epsilon$. For any $f \in \underline{\mathrm{Lip}}^{(c)}$, we have

$$\left| \int_\Omega (\mu_\omega - \nu_\omega)(f(\omega, .)) \, d\,\mathrm{P}(\omega) - \int_\Omega (\mu_\omega - \nu_\omega)(f(\omega, .)\, \mathbb{1}_{\mathbb{S}_0}) \, d\,\mathrm{P}(\omega) \right| \\ = \left| \int_\Omega (\mu_\omega - \nu_\omega)(f(\omega, .)\, \mathbb{1}_{\mathbb{S} \setminus \mathbb{S}_0}) \, d\,\mathrm{P}(\omega) \right| \leq 2\epsilon. \quad (10)$$

Set, for all $\omega \in \Omega_0$,

$$G'(\omega) = \sup_{g \in \mathrm{Lip}_{\mathbb{S}}^{(c)},\ g(x_0)=0} (\mu_\omega - \nu_\omega)(g\, \mathbb{1}_{\mathbb{S}_0}).$$

We thus have

$$\left| \int_{\Omega_0} G \, d\,\mathrm{P} - \int_{\Omega_0} G' \, d\,\mathrm{P} \right| \leq 2\epsilon. \quad (11)$$

Let $\mathrm{Lip}_{\mathbb{S}}^{(c)}\big|_{\mathbb{S}_0}$ denote the set of restrictions to $\mathbb{S}_0$ of elements of $\mathrm{Lip}_{\mathbb{S}}^{(c)}$. The set $\mathbb{S}_0$ is metrizable, thus $\mathrm{C}_b\,(\mathbb{S}_0)$ (endowed with the topology of uniform convergence) is metrizable separable, thus its subspace $\mathrm{Lip}_{\mathbb{S}}^{(c)}\big|_{\mathbb{S}_0}$ is also metrizable separable. We can thus find a dense countable subset $D = \{u_n;\, n \in \mathbb{N}\}$ of $\mathrm{Lip}_{\mathbb{S}}^{(c)}$ for the seminorm $\|u\|_{\mathrm{C}_b(\mathbb{S}_0)} := \sup_{x \in \mathbb{S}_0} |u(x)|$. Set, for all $(\omega, x) \in \Omega_0 \times \mathbb{S}$,

$$N(\omega) = \min\Big\{ n \in \mathbb{N};\ \int_{\mathbb{S}} u_n(x) \, d(\mu_\omega - \nu_\omega)(x) \geq \Delta_{\mathrm{L}}^{(c)}(\mu_\omega, \nu_\omega) - \epsilon G'(\omega) - \epsilon \Big\}, \quad \text{and}$$

$$f(\omega, x) = u_{N(\omega)}(x).$$

We then have, using (10) and (11),

$$\underline{\Delta}_{\mathrm{L}}^{(c)}(\mu, \nu) \geq \int_{\Omega_0 \times \mathbb{S}} f \, d(\mu - \nu) \geq \int_{\Omega_0 \times \mathbb{S}_0} f \, d(\mu - \nu) - 2\epsilon \\ \geq \int_{\Omega_0} G' \, d\,\mathrm{P} - 3\epsilon \geq \int_{\Omega_0} G \, d\,\mathrm{P} - 5\epsilon.$$

Thus, in view of (8) and(9),

$$\underline{\Delta}_{\mathrm{KR}}^{(c)}(\mu, \nu) = \int_{\Omega \times \mathbb{S} \times \mathbb{S}} c(x, y) \, d\lambda(\omega, x, y) = \underline{\Delta}_{\mathrm{L}}^{(c)}(\mu, \nu).$$

□

3 Application: coupling for the minimal distance

In this section $\mathbb{S}$ is a completely regular Radon space with metrizable compact subsets, $c : \mathbb{S} \times \mathbb{S} \to [0, +\infty]$ is a mapping satisfying (4) and $\mathcal{M}$ is a sub-σ-algebra of $\mathcal{A}$. Let X be a random variable with values in $\mathbb{S}$, let P_X be the distribution of X, and let $\mathrm{P}_{X|\mathcal{M}}$ be a regular conditional distribution of X given $\mathcal{M}$ (see Section 5 for the existence). We assume that $\int c(x, x_0)\, \mathrm{P}_X(dx)$ is finite for some (and therefore any) x_0 in $\mathbb{S}$ (which means exactly that the unique measure of $\mathcal{Y}(\mathcal{M})$ with disintegration $\mathrm{P}_{X|\mathcal{M}}(\cdot, \omega)$ belongs to $\mathcal{Y}^{c,1}(\mathcal{M})$). The proof of the following result is comparable to that of Corollary 4.2.5 in [1] and of Proposition 5 in [23].

Theorem 2 (A general coupling theorem) *Assume that Ω is rich enough, that is, there exists a random variable U from $(\Omega, \mathcal{A})$ to $([0,1], \mathcal{B}([0,1]))$, independent of $\sigma(X) \vee \mathcal{M}$ and uniformly distributed over $[0,1]$. Let Q be any element of $\mathcal{Y}^{c,1}(\mathcal{M})$. There exists a $\sigma(U) \vee \sigma(X) \vee \mathcal{M}$-measurable random variable Y, such that $Q_.$ is a regular conditional probability of Y given $\mathcal{M}$, and*

$$E\left(c(X,Y)|\mathcal{M}\right) = \sup_{f \in \mathrm{Lip}^{(c)}_{\mathbb{S}}} \left| \int f(x) \mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x) Q_.(dx) \right| \quad \text{P-}a.s.. \tag{12}$$

Proof. We apply Theorem 1 to the probability space $(\Omega, \mathcal{M}, \mathrm{P})$ and to the disintegrated measures $\mu_\omega(\cdot) = \mathrm{P}_{X|\mathcal{M}}(\cdot, \omega)$ and $\nu_\omega = Q_\omega$. As in the proof of Theorem 1, we assume without loss of generality that $\mathbb{S}$ is Lusin regular. From point 1 of Theorem 1 we infer that there exists a mapping $\omega \mapsto \lambda_\omega$ from Ω to $\mathcal{P}(\mathbb{S} \times \mathbb{S})$, measurable for $\mathcal{M}^*$ and $\mathcal{B}_{\mathcal{P}(\mathbb{S}\times\mathbb{S})}$, such that λ_ω belongs to $D(\mathrm{P}_{X|\mathcal{M}}(\cdot,\omega), Q_\omega)$ and $G(\omega) = \int_{\mathbb{S}\times\mathbb{S}} c(x,y)\lambda_\omega(dx,dy)$.

On the measurable space $(\mathbb{M}, \mathcal{T}) = (\Omega \times \mathbb{S} \times \mathbb{S}, \mathcal{M}^* \otimes \mathcal{B}_{\mathbb{S}} \otimes \mathcal{B}_{\mathbb{S}})$ we put the probability

$$\pi(A \times B \times C) = \int_A \lambda_\omega(B \times C)\, \mathrm{P}(d\omega)\,.$$

If $I = (I_1, I_2, I_3)$ is the identity on $\mathbb{M}$, we see that a regular conditional distribution of (I_2, I_3) given I_1 is given by $\mathrm{P}_{(I_2,I_3)|I_1=\omega} = \lambda_\omega$. Since $\mathrm{P}_{X|\mathcal{M}}(\cdot,\omega)$ is the first marginal of λ_ω, a regular conditional probability of I_2 given I_1 is given by $\mathrm{P}_{I_2|I_1=\omega}(\cdot) = \mathrm{P}_{X|\mathcal{M}}(\cdot,\omega)$. Let $\lambda_{\omega,x} = \mathrm{P}_{I_3|I_1=\omega, I_2=x}$ be a regular conditional distribution of I_3 given (I_1, I_2), so that $(\omega, x) \mapsto \lambda_{\omega,x}$ is measurable for $\mathcal{M}^* \otimes \mathcal{B}_{\mathbb{S}}$ and $\mathcal{B}_{\mathcal{P}(\mathbb{S})}$. From the uniqueness (up to P-a.s. equality) of regular conditional probabilities, it follows that

$$\lambda_\omega(B \times C) = \int_B \lambda_{\omega,x}(C) \mathrm{P}_{X|\mathcal{M}}(dx, \omega) \quad \text{P-a.s.}. \tag{13}$$

Assume that we can find a random variable $\tilde{Y}$ from Ω to $\mathbb{S}$, measurable for $\sigma(U) \vee \sigma(X) \vee \mathcal{M}^*$ and $\mathcal{B}_{\mathbb{S}}$, such that $\mathrm{P}_{\tilde{Y}|\sigma(X)\vee\mathcal{M}^*}(\cdot,\omega) = \lambda_{\omega,X(\omega)}(\cdot)$. Since $\omega \mapsto \mathrm{P}_{X|\mathcal{M}}(\cdot,\omega)$ is measurable for $\mathcal{M}^*$ and $\mathcal{B}_{\mathcal{P}(\mathbb{S})}$, one can check that $\mathrm{P}_{X|\mathcal{M}}$ is a regular conditional probability of X given $\mathcal{M}^*$. For A in $\mathcal{M}^*$, B and C in $\mathcal{B}_{\mathbb{S}}$, we thus have

$$\begin{aligned}
E\left(\mathbb{1}_A \mathbb{1}_{X\in B} \mathbb{1}_{\tilde{Y}\in C}\right) &= E\left(\mathbb{1}_A E\left(\mathbb{1}_{X\in B} E\left(\mathbb{1}_{\tilde{Y}\in C}|\sigma(X)\vee\mathcal{M}^*\right)|\mathcal{M}^*\right)\right) \\
&= \int_A \Big(\int_B \lambda_{\omega,x}(C)\mathrm{P}_{X|\mathcal{M}}(dx,\omega)\Big)\mathrm{P}(d\omega) \\
&= \int_A \lambda_\omega(B\times C)\mathrm{P}(d\omega)\,.
\end{aligned}$$

We infer that λ_ω is a regular conditional probability of $(X, \tilde{Y})$ given $\mathcal{M}^*$. By definition of λ_ω, we obtain that

$$E\left(c(X,\tilde{Y})|\mathcal{M}^*\right) = \sup_{f \in \mathrm{Lip}^{(c)}_{\mathbb{S}}} \left| \int f(x) \mathrm{P}_{X|\mathcal{M}}(dx) - \int f(x) Q_.(dx) \right| \quad \text{P-a.s.}. \tag{14}$$

Since $\mathbb{S}$ is Lusin, it is standard Borel (see Section 5). Applying Lemma 2, there exists a $\sigma(U)\vee\sigma(X)\vee\mathcal{M}$-measurable modification Y of $\tilde{Y}$, so that (14) still holds for $\mathbb{E}(c(X,Y)|\mathcal{M}^*)$. We obtain (12) by noting that $E\left(c(X,Y)|\mathcal{M}^*\right) = E\left(c(X,Y)|\mathcal{M}\right)$ P-a.s.

It remains to build $\tilde{Y}$. Since $\mathbb{S}$ is standard Borel, there exists a one to one map f from $\mathbb{S}$ to a Borel subset of $[0,1]$, such that f and f^{-1} are measurable for $\mathcal{B}([0,1])$ and $\mathcal{B}_{\mathbb{S}}$. Define $F(t,\omega) = \lambda_{\omega,X(\omega)}(f^{-1}(]-\infty,t]))$. The map $F(\cdot,\omega)$ is a distribution function with càdlàg inverse $F^{-1}(\cdot,\omega)$. One can see that the map $(u,\omega) \to F^{-1}(u,\omega)$ is $\mathcal{B}([0,1]) \otimes \mathcal{M}^* \vee \sigma(X)$-measurable. We now use the fact that Ω is rich enough:

the existence of the random variable U uniformly distributed over $[0,1]$ and independent of $\sigma(X)\vee\mathcal{M}$ allows some independent randomization. Let $T(\omega)=F^{-1}(U(\omega),\omega)$ and $\tilde{Y}=f^{-1}(T)$. It remains to see that $\mathrm{P}_{\tilde{Y}|\sigma(X)\vee\mathcal{M}^*}(\cdot,\omega)=\lambda_{\omega,X(\omega)}(\cdot)$. For any A in $\mathcal{M}^*$, B in $\mathcal{B}_{\mathbb{S}}$ and t in $\mathbb{R}$, we have

$$E\Big(\mathbb{1}_A\,\mathbb{1}_{X\in B}\,\mathbb{1}_{\tilde{Y}\in f^{-1}(]-\infty,t])}\Big)=\int_A \mathbb{1}_{X(\omega)\in B}\,\mathbb{1}_{U(\omega)\le F(t,\omega)}\mathrm{P}(d\omega).$$

Since U is independent of $\sigma(X)\vee\mathcal{M}$, it is also independent of $\sigma(X)\vee\mathcal{M}^*$. Hence

$$\begin{aligned}E\Big(\mathbb{1}_A\,\mathbb{1}_{X\in B}\,\mathbb{1}_{\tilde{Y}\in f^{-1}(]-\infty,t])}\Big)&=\int_A \mathbb{1}_{X(\omega)\in B}F(t,\omega)\,\mathrm{P}(d\omega)\\&=\int_A \mathbb{1}_{X(\omega)\in B}\lambda_{\omega,X(\omega)}(f^{-1}(]-\infty,t]))\,\mathrm{P}(d\omega).\end{aligned}$$

Since $\{f^{-1}(]-\infty,t]),t\in[0,1]\}$ is a separating class, the result follows. $\square$

Coupling and dependence coefficients

Define the coefficient

$$\tau_c(\mathcal{M},X)=\Big\|\sup_{f\in\mathrm{Lip}_{\mathbb{S}}^{(c)}}\Big|\int f(x)\mathrm{P}_{X|\mathcal{M}}(dx)-\int f(x)\mathrm{P}_X(dx)\Big|\Big\|_1\,. \tag{15}$$

If $\mathrm{Lip}_{\mathbb{S}}^{(c)}$ is a separating class, this coefficient measures the dependence between $\mathcal{M}$ and X ($\tau_c(\mathcal{M},X)=0$ if and only if X is independent of $\mathcal{M}$). From point 2 of Theorem 1, we see that an equivalent definition is

$$\tau_c(\mathcal{M},X)=\sup_{f\in\mathrm{Lip}_{\mathbb{S},\mathcal{M}}^{(c)}}\int f(\omega,X(\omega))\mathrm{P}(d\omega)-\int\Big(\int f(\omega,x)\mathrm{P}_X(dx)\Big)\mathrm{P}(d\omega)\,.$$

where $\mathrm{Lip}_{\mathbb{S},\mathcal{M}}^{(c)}$ is the set of integrands f from $\Omega\times\mathbb{S}\to\mathbb{R}$, measurable for $\mathcal{M}\otimes\mathcal{B}_{\mathbb{S}}$, such that $f(\omega,.)$ belongs to $\mathrm{Lip}_{\mathbb{S}}^{(c)}$ for any $\omega\in\Omega$.

Let $c(x,y)=\mathbb{1}_{x\neq y}$ be the discrete metric and let $\|\cdot\|_v$ be the variation norm. From the Riesz-Alexandroff representation theorem (see [29, Theorem 5.1]), we infer that for any (μ,ν) in $\mathcal{P}(\mathbb{S})\times\mathcal{P}(\mathbb{S})$,

$$\sup_{f\in\mathrm{Lip}_{\mathbb{S}}^{(c)}}|\mu(f)-\nu(f)|=\frac{1}{2}\|\mu-\nu\|_v\,.$$

Hence, for the discrete metric $\tau_c(\mathcal{M},X)=\beta(\mathcal{M},\sigma(X))$ is the β-mixing coefficient between $\mathcal{M}$ and $\sigma(X)$ introduced in [24]. If c is a distance for which $\mathbb{S}$ is Polish, $\tau_c(\mathcal{M},X)$ has been introduced in [23, Inequality (10)] in its"dual" form, and in [8], [10] in its present from (obviously the reference to [23] is missing in these two papers).

Applying Theorem 2 with $Q=\mathrm{P}\otimes\mathrm{P}_X$, we see that this coefficient has a characteristic property which is often called the *coupling* or *reconstruction* property.

Corollary 1 (reconstruction property) *If Ω is rich enough (see Theorem 2), there exists a $\sigma(U)\vee\sigma(X)\vee\mathcal{M}$-measurable random variable X^*, independent of $\mathcal{M}$ and distributed as X, such that*

$$\tau_c(\mathcal{M},X)=E\left(c(X,X^*)\right)\,. \tag{16}$$

If $c(x,y) = \mathbb{1}_{x \neq y}$, (16) is given in [1, Corollary 4.2.5] (note that in Berbee's corollary, $\mathbb{S}$ is assumed to be standard Borel. For other proofs of Berbee's coupling, see [4], [23, Proposition 5 and Remark 2 page 123] and [22, Section 5.3]). If c is a distance for which $\mathbb{S}$ is a Polish space, (16) has been proved in [23, Proposition 6] (in [23] a more general result for sequences is given, in the spirit of [2]. For an other proof of (16) when $(\mathbb{S}, c)$ is Polish, see [8]).

Coupling is a very useful property in the area of limit theorems and statistics. Many authors have used Berbee's coupling to prove various limit theorems (see for instance the review paper [19] and the references therein) as well as exponential inequalities (see for instance the paper [13] for Bernstein-type inequalities and applications to empirical central limit theorems). Unfortunately, these results apply only to β-mixing sequences, but this property is very hard to check and many simple processes (such as iterates of maps or many non-irreducible Markov chains) are not β-mixing. In many cases however, this difficulty may be overcome by considering another distance c, more adapted to the problem than the discrete metric (typically c is a norm for which $\mathbb{S}$ is a separable Banach space). The case $\mathbb{S} = \mathbb{R}$ and $c(x,y) = |x-y|$, is studied in the paper [9], where many non β-mixing examples are given. In this paper the authors used the coefficients τ_c to prove Bernstein-type inequalities and a strong invariance principle for partial sums. In the paper [10, Section 4.4] the same authors show that if T is an uniformly expanding map preserving a probability μ on $[0,1]$, then $\tau_c(\sigma(T^n), T) = O(a^n)$ for $c(x,y) = |x-y|$ and some a in $[0,1[$.

The following inequality (which can be deduced from [19, page 174]) shows clearly that $\beta(\mathcal{M}, \sigma(X))$ is in some sense the more restrictive coefficient among all the $\tau_c(\mathcal{M}, X)$: for any x in $\mathbb{S}$, we have that

$$\tau_c(\mathcal{M}, X) \leq 2 \int_0^{\beta(\mathcal{M}, \sigma(X))} Q_{c(X,x)}(u) du \,, \tag{17}$$

where $Q_{c(X,x)}$ is the generalized inverse of the function $t \mapsto \mathrm{P}(c(X,x) > t)$. In particular, if c is bounded by M, $\tau_c(\mathcal{M}, X) \leq 2M\beta(\mathcal{M}, \sigma(X))$.

A simple example

Let $(X_i)_{i \geq 0}$ be a stationary Markov chain with values in a Polish space $\mathbb{S}$, satisfying the equation $X_{n+1} = F(X_n, \xi_{n+1})$, where $(\xi_i)_{i>0}$ is a sequence of independent and identically distributed random variables with values in some measurable space $\mathbb{M}$ and independent of X_0, and F is a measurable function from $\mathbb{S} \times \mathbb{M}$ to $\mathbb{S}$. Let X_0^* be a random variable distributed as X_0 and independent of $(X_0, (\xi_i)_{i>0})$, and let $X_{n+1}^* = F(X_n^*, \xi_{n+1})$. The sequence $(X_i^*)_{i \geq 0}$ is independent of X_0 and distributed as $(X_i)_{i \geq 0}$. From the definition (15) of τ_c, we easily infer that

$$\tau_c(\sigma(X_0), X_k) \leq E(c(X_k, X_k^*)) \,.$$

Let μ be the distribution of X_0 and $(X_n^{(x)})_{n \geq 0}$ the chain starting from $X_0^{(x)} = x$. With these notations, we have that

$$E(c(X_k, X_k^*)) = \iint E(c(X_k^{(x)}, X_k^{(y)}))\mu(dx)\mu(dy) \,.$$

If there exists a sequence $(\delta_i)_{i \geq 0}$ of nonnegative numbers such that $E(c(X_k^{(x)}, X_k^{(y)})) \leq \delta_k c(x,y)$, then

$$\tau_c(\sigma(X_0), X_k) \leq \delta_k E(c(X_0, X_0^*)) \,.$$

For instance, in the case where $E(c(F(x,\xi_0), F(y,\xi_0))) \leq \kappa c(x,y)$ for some $\kappa < 1$, we can take $\delta_k = \kappa^k$. An important example is the case where $\mathbb{S} = \mathbb{M}$ is a separable Banach space and $X_{n+1} = f(X_n) + \xi_{n+1}$ for some κ lipschitz function f with respect to c.

Let us consider the well known example $2X_{n+1} = X_n + \xi_{n+1}$, where X_0 has uniform distribution λ over $[0,1]$ and ξ_1 is Bernoulli distributed with parameter $1/2$. If $c(x,y) = |x-y|$, it follows from our preceding remarks that $\tau_c(\sigma(X_0), X_k) \le 2^{-k}$. However, it is well known that this chain is not β mixing. Indeed, it is a stationary Markov chain with invariant distribution λ and transition kernel

$$K(x, \cdot) = \frac{1}{2}(\delta_{x/2} + \delta_{(x+1)/2}),$$

so that $\|K^k(x, .) - \lambda\|_v = 2$. Consequently $\beta(\sigma(X_0), \sigma(X_k)) = 1$ for any $k \ge 0$.

A simple application

Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of real-valued random variables with common distribution function F. Let $\mathcal{M}_0 = \sigma(X_k, k \le 0)$, and let $F_{X_k|\mathcal{M}_0}$ be a conditional distribution function of X_k given $\mathcal{M}_0$. Let $F_n = n^{-1}\sum_{i=1}^n \mathbb{1}_{X_i \le t}$ be the empirical distribution function. Let μ be a finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. In [7, Example 2, Section 2.2], it is proved that the process $\{t \mapsto \sqrt{n}(F_n(t) - F(t))\}$ converges weakly in $\mathbb{L}^2(\mu)$ to a mixture of $\mathbb{L}^2(\mu)$-valued Gaussian random variables as soon as

$$\sum_{k>0} E\Big(\int |F_{X_k|\mathcal{M}_0}(t) - F(t)|^2 \mu(dt)\Big)^{1/2} < \infty. \tag{18}$$

Let X_k^* be a random variable distributed as X_k and independent of $\mathcal{M}_0$, and let $F_\mu(x) = \mu(]-\infty, x[)$. Since $F = F_{X_k^*|\mathcal{M}_0}$, it follows that

$$E\Big(\int |F_{X_k|\mathcal{M}_0}(t) - F(t)|^2 \mu(dt)\Big)^{1/2} \le E\Big(\sqrt{|F_\mu(X_k) - F_\mu(X_k^*)|}\Big).$$

Let $d_\mu(x, y) = \sqrt{|F_\mu(x) - F_\mu(y)|}$. From (16) it follows that one can choose X_k^* such that

$$E\Big(\sqrt{|F_\mu(X_k) - F_\mu(X_k^*)|}\Big) = \tau_{d_\mu}(\mathcal{M}_0, X_k).$$

Consequently (18) holds as soon as $\sum_{k>0} \tau_{d_\mu}(\mathcal{M}_0, X_k) < \infty$. This is an example where the natural cost function d_μ is not the discrete metric $c(x, y) = \mathbb{1}_{x\ne y}$ nor the usual norm $c(x, y) = |x - y|$.

4 A counter example to maximal coupling

In this section we prove that no analogue of Goldstein's maximal coupling (see [16]) is possible if the cost function is not proportional to the discrete metric.

More generally, we consider the following problem. Let $\mathbb{M}$ be a Polish space and $\mathbb{S} = \mathbb{M} \times \mathbb{M}$. Let c be any symmetric measurable function from $\mathbb{M} \times \mathbb{M}$ to $\mathbb{R}^+$, such that $c(x, y) = 0$ if and only if $x = y$. Let $\mathcal{F}$ be the class of symmetric measurable functions φ from $\mathbb{R}^+ \times \mathbb{R}^+$ to $\mathbb{R}^+$, such that $x \mapsto \varphi(0, x)$ is increasing. For $\varphi \in \mathcal{F}$, we define the cost function $c_\varphi((x_1, x_2), (y_1, y_2)) = \varphi(c(x_1, y_1), c(x_2, y_2))$ on $\mathbb{S} \times \mathbb{S}$.

The question $\mathbf{Q}$ is the following. For which couples (φ, c) do we have the property: for any probability measures μ, ν on $\mathbb{S}$ with marginals $\mu_{(2)}(A) = \mu(\mathbb{M} \times A)$ and $\nu_{(2)}(A) = \nu(\mathbb{M} \times A)$, there exists a probability measure λ in $D(\mu, \nu)$ with marginal $\lambda_{(2)}(A \times B) = \lambda(\mathbb{M} \times A \times \mathbb{M} \times B)$, such that

$$\Delta_{\mathrm{KR}}^{(c_\varphi)}(\mu, \nu) = \int \varphi(c(x_1, y_1), c(x_2, y_2))\lambda(dx_1, dx_2, dy_1, dy_2) \tag{19}$$

$$\Delta_{\mathrm{KR}}^{(c)}(\mu_{(2)}, \nu_{(2)}) = \int c(x_2, y_2)\lambda_{(2)}(dx_2, dy_2)\ ? \tag{20}$$

From Goldstein's result we know that the couple $(\varphi(x, y) = x \vee y, c(x, y) = \mathbb{1}_{x\ne y})$ is a solution to $\mathbf{Q}$. The following proposition shows that, if c is not proportional to the discrete metric, no couple (φ, c) can be a solution to $\mathbf{Q}$.

Proposition 1 *Suppose that c is not proportional to the discrete metric. There exist a_1, b_1, a_2, b_2 in $\mathbb{M}$ such that $a_1 \neq b_1$ and $a_2 \neq b_2$ and two probabilities μ and ν on $\{(a_1, a_2), (a_1, b_2), (b_1, a_2), (b_1, b_2)\}$ for which, for any $\varphi \in \mathcal{F}$, there is no λ in $D(\mu, \nu)$ satisfying (19) and (20) simultaneously.*

Proof. Since c is not proportional to the discrete metric, there exist at least two points (a_1, b_1) and (a_2, b_2) in $\mathbb{M} \times \mathbb{M}$ such that $a_1 \neq b_1$, $a_2 \neq b_2$ and $c(a_1, b_1) > c(a_2, b_2) > 0$. Define the probabilities μ and ν by

$$\begin{array}{ll} \mu(a_1,a_2) = \frac{1}{2} & \nu(a_1,a_2) = 0 \\ \mu(a_1,b_2) = 0 & \nu(a_1,b_2) = \frac{1}{2} \\ \mu(b_1,a_2) = 0 & \nu(b_1,a_2) = \frac{1}{2} \\ \mu(b_1,b_2) = \frac{1}{2} & \nu(b_1,b_2) = 0 \end{array}$$

The set $D(\mu, \nu)$ is the set of probabilities λ_α such that $\lambda_\alpha(a_1, a_2, a_1, b_2) = \lambda_\alpha(b_1, b_2, b_1, a_2) = \alpha$, $\lambda_\alpha(a_1, a_2, b_1, a_2) = \lambda_\alpha(b_1, b_2, a_1, b_2) = 1/2 - \alpha$, for α in $[0, 1/2]$. Consequently, for any φ in $\mathcal{F}$,

$$\int \varphi(c(x_1, y_1), c(x_2, y_2))\lambda_\alpha(dx_1, dx_2, dy_1, dy_2) = 2\alpha\, \varphi(0, c(a_2, b_2)) + (1 - 2\alpha)\, \varphi(c(a_1, b_1), 0)\,. \tag{21}$$

Since $c(a_1, b_1) > c(a_2, b_2)$, since φ is symmetric, and since $x \mapsto \varphi(0, x)$ is increasing, $\varphi(c(a_1, b_1), 0) > \varphi(0, c(a_2, b_2))$. Therefore, the unique solution to (19) is $\lambda_{1/2}$. Now

$$\int c(x_2, y_2)\lambda_{1/2}(dx_1, dx_2, dy_1, dy_2) = c(a_2, b_2) > 0\,.$$

Since $\mu_{(2)} = \nu_{(2)}$, $\Delta^{(c)}_{\mathrm{KR}}(\mu_{(2)}, \nu_{(2)}) = 0$. Hence $\lambda_{1/2}$ does not satisfy (20). $\square$

Remark 2. If now c is the discrete metric $c(x, y) = \mathbb{1}_{x \neq y}$, the right hand term in equality (21) is $\varphi(c(a_1, b_1), 0)$. Consequently, any λ_α is solution to (19) and λ_0 is solution to both (19) and (20). We conjecture that if c is the discrete metric, then any couple (φ, c), $\varphi \in \mathcal{F}$, is a solution to **Q**.

5 Appendix: topological and measure-theoretical complements

Topological spaces

Let us recall some definitions (see [25, 15] for complements on Radon and Suslin spaces). A topological space $\mathbb{S}$ is said to be

- *regular* if, for any $x \in \mathbb{S}$ and any closed subset F of $\mathbb{S}$ which does not contain x, there exist two disjoint open subsets U and V such that $x \in U$ and $F \subset V$,
- *completely regular* if, for any $x \in \mathbb{S}$ and any closed subset F of $\mathbb{S}$ which does not contain x, there exists a continuous function $f : \mathbb{S} \to [0, 1]$ such that $f(x) = 0$ and $f = 1$ on F (equivalently, $\mathbb{S}$ is *uniformizable*, that is, the topology of $\mathbb{S}$ can be defined by a set of semidistances),
- *pre-Radon* if every finite τ–additive Borel measure on $\mathbb{S}$ is inner regular with respect to the compact subsets of $\mathbb{S}$ (a Borel measure μ on $\mathbb{S}$ is *τ–additive* if, for any family $(F_\alpha)_{\alpha \in A}$ of closed subsets of $\mathbb{S}$ such that $\forall \alpha, \beta \in A \quad \exists \gamma \in A \quad F_\gamma \subset F_\alpha \cap F_\beta$, we have $\mu(\cap_{\alpha \in A} F_\alpha) = \inf_{\alpha \in A} \mu(F_\alpha)$),
- *Radon* if every finite Borel measure on $\mathbb{S}$ is inner regular with respect to the compact subsets of $\mathbb{S}$,
- *Suslin*, or *analytic*, if there exists a continuous mapping from some Polish space onto $\mathbb{S}$,
- *Lusin* if there exists a continuous injective mapping from some Polish space onto $\mathbb{S}$. Equivalently, $\mathbb{S}$ is Lusin if there exists a Polish topology on $\mathbb{S}$ which is finer than the given topology of $\mathbb{S}$.

Obviously, every Lusin space is Suslin and every Radon space is pre-Radon. Much less obviously, every Suslin space is Radon. Every regular Suslin space is completely regular.

Many usual spaces of Analysis are Lusin: besides all separable Banach spaces (e.g. L^p $(1 \le p < +\infty)$, or the Sobolev spaces $\mathrm{W}^{s,p}(\Omega)$ $(0 < s < 1$ and $1 \le p < +\infty))$, the spaces of distributions $\mathcal{E}'$, $\mathcal{S}'$, $\mathcal{D}'$, the space $\mathcal{H}(\mathbb{C})$ of holomorphic functions, or the topological dual of a Banach space, endowed with its weak*–topology are Lusin. See [25, pages 112–117] for many more examples.

Standard Borel spaces

A measurable space $(\mathbb{M}, \mathcal{M})$ is said to be *standard Borel* if it is Borel-isomorphic with some Polish space $\mathbb{T}$, that is, there exists a mapping $f : \mathbb{T} \to \mathbb{M}$ which is one-one and onto, such that f and f^{-1} are measurable for $\mathcal{B}_{\mathbb{T}}$ and $\mathcal{M}$. We say that a topological space $\mathbb{S}$ is *standard Borel* if $(\mathbb{S}, \mathcal{B}_{\mathbb{S}})$ is standard Borel.

If τ_1 and τ_2 are two comparable Suslin topologies on $\mathbb{S}$, they share the same Borel sets. In particular, every Lusin space is standard Borel.

A useful property of standard Borel spaces is that every standard space $\mathbb{S}$ is Borel-isomorphic with a Borel subset of $[0, 1]$. This a consequence of e.g. [17, Theorem 15.6 and Corollary 6.5], see also [26] or [11, Théorème III.20]. (Actually, we have more: every standard Borel space is countable or Borel-isomorphic with $[0, 1]$. Thus, for standard Borel spaces, the Continuum Hypothesis holds true!)

Another useful property of standard Borel spaces is that, if $\mathbb{S}$ is a standard Borel space, if $X : \Omega \mapsto \mathbb{S}$ is a measurable mapping, and if $\mathcal{M}$ is a sub-σ-algebra of $\mathcal{A}$, there exists a regular conditional distribution $\mathrm{P}_{X|\mathcal{M}}$ (see e.g. [14, Theorem 10.2.2] for the Polish case, which immediately extends to standard Borel spaces from their definition). Note that, if $\mathbb{S}$ is radon, then the distribution P_X of X is tight, that is, for every integer $n \ge 1$, there exists a compact subset K_n of $\mathbb{S}$ such that $\mathrm{P}_X(\mathbb{S} \setminus K_n) \ge 1/n$. Hence one can assume without loss of generality that X takes its values in $\cup_{n \ge 1} K_n$. If moreover $\mathbb{S}$ has metrizable compact subsets, then $\cup_{n \ge 1} K_n$ is Lusin (and hence standard Borel), and there exists a regular conditional distribution $\mathrm{P}_{X|\mathcal{M}}$. Thus, if $\mathbb{S}$ is Radon with metrizable compact subsets, every element μ of $\mathcal{Y}$ has an $\mathcal{A}$-measurable disintegration. Indeed, denoting $\mathcal{A}' = \mathcal{A} \otimes \{\emptyset, \mathbb{S}\}$, one only needs to consider the conditional distribution $\mathrm{P}_{X|\mathcal{A}'}$ of the random variable $X : (\omega, x) \mapsto x$ defined on the probability space $(\Omega \times \mathbb{S}, \mathcal{A} \otimes \mathcal{B}_{\mathbb{S}}, \mu)$.

For any σ–algebra $\mathcal{M}$ on a set $\mathbb{M}$, the *universal completion* of $\mathcal{M}$ is the σ-algebra $\mathcal{M}^* = \cap_\mu \mathcal{M}^*_\mu$, where μ runs over all finite nonegative measures on $\mathcal{M}$ and $\mathcal{M}^*_\mu$ is the μ–completion of $\mathcal{M}$. A subset of a topological space $\mathbb{S}$ is said to be *universally measurable* if it belongs to $\mathcal{B}^*_{\mathbb{S}}$. The following lemma can be deduced from e.g. [28, Exercise 10 page 14] and the Borel-isomorphism theorem.

Lemma 2 *Assume that $\mathbb{S}$ is a standard Borel space. Let $X : \Omega \to \mathbb{S}$ be $\mathcal{A}^*$–measurable. Then there exists an $\mathcal{A}$–measurable modification $Y : \Omega \to \mathbb{S}$ of X, that is, Y is $\mathcal{A}$–measurable and satisfies $Y = X$ a.e.*

References

1. Henry C. P. Berbee, *Random walks with stationary increments and renewal theory*, Mathematical Centre Tracts, vol. 112, Mathematisch Centrum, Amsterdam, 1979.
2. I. Berkes and W. Philipp, *Approximation theorems for independent and weakly dependent random vectors*, Ann. Probab. **7** (1979), 29-54.
3. S. Borovkova, R. Burton and H. Dehling, *Limit theorems for functionals of mixing processes with application to U-statistics and dimension estimation*, Trans. Amer. Math. Soc. **353** (2001) 4261-4318.
4. W. Bryc, *On the approximation theorem of I. Berkes and W. Philipp*, Demonstratio Mathematica **15** (1982), no. 3, 807–816.
5. Charles Castaing, Paul Raynaud de Fitte, and Michel Valadier, *Young measures on topological spaces. With applications in control theory and probability theory*, Kluwer Academic Publishers, Dordrecht, 2004.

6. Charles Castaing and Michel Valadier, *Convex analysis and measurable multifunctions*, Lecture Notes in Math., no. 580, Springer Verlag, Berlin, 1977.
7. Jérôme Dedecker and Florence Merlevède, *The conditional central limit theorem in Hilbert spaces*, Stochastic Process. Appl. **108** (2003), 229-262.
8. Jérôme Dedecker and Clémentine Prieur, *Couplage pour la distance minimale*, C. R. Math. Acad. Sci. Paris **338** (2004), no. 10, 805–808.
9. Jérôme Dedecker and Clémentine Prieur, *Coupling for τ-dependent sequences and applications*, J. Theor. Probab **17** (2004), no. 4, 861-885.
10. Jérôme Dedecker and Clémentine Prieur, *New dependence coefficients. Examples and applications to statistics*, Probab. Theory Relat. Fields **132** (2005), 203-236.
11. Claude Dellacherie and Paul André Meyer, *Probabilités et potentiel. Chapitres I à IV*, Hermann, Paris, 1975.
12. R. L. Dobrušin, *Prescribing a system of random variables by conditional distributions*, Theor. Probability Appl. **15** (1970), 458-486.
13. Paul Doukhan, Pascal Massart, and Emmanuel Rio, *Invariance principles for absolutely regular empirical processes*, Annales Inst. H. Poincaré Probab. Statist. **31** (1995), 393–427.
14. R. M. Dudley, *Real analysis and probability*, Cambridge University Press, Cambridge, 2002.
15. R. J. Gardner and W. F. Pfeffer, *Borel measures*, Handbook of set-theoretic topology, North-Holland, Amsterdam, 1984, pp. 961–1043.
16. S. Goldstein, *Maximal coupling*, Z. Wahrsch. Verw. Gebiete **46** (1979), 193-204.
17. Alexander S. Kechris, *Classical descriptive set theory*, Graduate Texts in Mathematics, no. 156, Springer Verlag, New York, 1995.
18. V. L. Levin, *The problem of mass transfer in a topological space and probability measures with given marginal measures on the product of two spaces*, Dokl. Akad. Nauk SSSR **276** (1984), no. 5, 1059–1064, English translation: Soviet Math. Dokl. 29 (1984), no. 3, 638–643.
19. Florence Merlevède and Magda Peligrad, *On the coupling of dependent random variables and applications*, Empirical process techniques for dependent data, Birkhäuser, 2002, pp. 171–193.
20. S. T. Rachev and L. Rüschendorf, *Mass transportation problems. Volume I: Theory*, Probability and its Applications, Springer Verlag, New York, Berlin, 1998.
21. Emmanuel Rio, *Processus empiriques absolument réguliers et entropie universelle*, Probab. Theory Relat. Fields **111** (1998), no. 4, 585–608.
22. Emmanuel Rio, *Théorie asymptotique des processus aléatoires faiblement dépendants*, Mathématiques et Applications, no. 31, Springer, Berlin, Heidelberg, 2000.
23. L. Rüschendorf, *The Wasserstein Distance and Approximation Theorems*, Z. Wahrsch. Verw. Gebiete **70** (1985), 117-129.
24. Y. A. Rozanov and V. A. Volkonskii, *Some limit theorems for random functions I*, Teor. Verojatnost. i Primenen. **4** (1959), 186–207.
25. Laurent Schwartz, *Radon measures on arbitrary topological spaces and cylindrical measures*, Tata Institute of Fundamental Research Studies in Mathematics, Oxford University Press, London, 1973.
26. A. V. Skorohod, *On a representation of random variables*, Teor. Verojatnost. i Primenen. **21** (1976), no. 3, 645–648, English translation: Theor. Probability Appl. 21 (1976), no. 3, 628–632 (1977).
27. Michel Valadier, *Désintégration d'une mesure sur un produit*, C. R. Acad. Sci. Paris Sér. I **276** (1973), A33–A35.
28. Aad W. van der Vaart and Jon A. Wellner, *Weak convergence and empirical processes. With applications to statistics*, Springer Series in Statistics, Springer Verlag, Berlin, 1996.
29. Robert F. Wheeler, *A survey of Baire measures and strict topologies*, Exposition. Math. **1** (1983), no 2, 97-190.

Mathematics Subject Classifications (2000): 60A10, 60G05.
Key words: parametrized Kantorovich–Rubinštein theorem, random probabilities, coupling, dependence coefficients, maximal coupling