Statistics and Probability Letters 84 (2014) 131–134

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A remark on the optimal transport between two probability measures sharing the same copula

A. Alfonsi *, B. Jourdain

Université Paris-Est, CERMICS, Projet MathFi ENPC-INRIA-UMLV, 6 et 8 avenue Blaise Pascal, 77455 Marne La Vallée, Cedex 2, France

ARTICLE INFO

Article history:
Received 16 July 2013
Received in revised form 27 September 2013
Accepted 27 September 2013
Available online 8 October 2013

MSC:
60E05
60E15

Keywords:
Optimal transport
Copula
Wasserstein distance
Inversion of the cumulative distribution function

ABSTRACT

We study the optimal transport between two probability measures on $\mathbb{R}^n$ sharing the same copula C. We investigate the optimality of the image of the probability measure dC by the vectors of pseudo-inverses of marginal distributions.

1. Optimal transport between two probability measures sharing the same copula

Given two probability measures μ and ρ, the optimal transport theory aims at minimizing $\int c(x, y)\nu(dx, dy)$ over all couplings ν with first marginal $\nu \circ ((x, y) \mapsto x)^{-1} = \mu$ and second marginal $\nu \circ ((x, y) \mapsto y)^{-1} = \rho$ for a measurable non-negative cost function c. We use the notation $\nu <^{\mu}_{\rho}$ for such couplings. In the present note, we are interested in the particular case of the so-called Wasserstein distance between two probability measures μ and ρ on $\mathbb{R}^n$:

$$\mathcal{W}_{p,q}(\mu, \rho) = \inf_{\nu <^{\mu}_{\rho}} \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|_q^p \nu(dx, dy) \right)^{1/p} \tag{1.1}$$

obtained for the choice $c(x, y) = \|x - y\|_q^p$. Here $\mathbb{R}^n$ is endowed with the norm $\|(x_1, \dots, x_n)\|_q = \left(\sum_{i=1}^n |x_i|^q\right)^{1/q}$ for $q \in [1, +\infty)$ whereas $p \in [1, +\infty)$ is the power of this norm in the cost function.

In dimension $n = 1$, $\|x\|_q = |x|$ so that the Wasserstein distance does not depend on q and is simply denoted by $\mathcal{W}_p$. Moreover, the optimal transport is given by the inversion of the cumulative distribution functions: whatever $p \in [1, +\infty)$, an optimal coupling is the image of the Lebesgue measure on $(0, 1)$ by $u \mapsto (F_\mu^{-1}(u), F_\rho^{-1}(u))$ where for $u \in (0, 1)$, $F_\mu^{-1}(u) = \inf\{x \in \mathbb{R} : \mu((-\infty, x]) \geq u\}$ and $F_\rho^{-1}(u) = \inf\{x \in \mathbb{R} : \rho((-\infty, x]) \geq u\}$ (see for instance Theorem 3.1.2 in Rachev and Rüschendorf (1998)). This implies that $\mathcal{W}_p^p(\mu, \rho) = \int_{(0,1)} |F_\mu^{-1}(u) - F_\rho^{-1}(u)|^p du$.

* Corresponding author.
E-mail addresses: alfonsi@cermics.enpc.fr (A. Alfonsi), jourdain@cermics.enpc.fr (B. Jourdain).

0167-7152/$ – see front matter
http://dx.doi.org/10.1016/j.spl.2013.09.035

In higher dimensions, according to Sklar's theorem (see for instance Theorem 2.10.11 in Nelsen (2006)),

$$\mu\left(\prod_{i=1}^{n}(-\infty,x_i]\right)=C\left(\mu_1((-\infty,x_1]),\dots,\mu_n((-\infty,x_n])\right)$$

where we denote by $\mu_i=\mu\circ((x_1,\dots,x_n)\mapsto x_i)^{-1}$ the ith marginal of μ and C is a copula function i.e. $C(u_1,\dots,u_n)=m\left(\prod_{i=1}^{n}[0,u_i]\right)$ for some probability measure m on $[0,1]^n$ with all marginals equal to the Lebesgue measure on $[0,1]$. The copula function C is uniquely determined on the product of the ranges of the marginal cumulative distribution functions $x_i\mapsto\mu_i((-\infty,x_i])$. In particular, when the marginals μ_i do not weight points, the copula C is uniquely determined. Sklar's theorem shows that the dependence structure associated with μ is encoded in the copula function C. Last, we give the well-known Fréchet–Hoeffding bounds

$$\forall u_1,\dots,u_n\in[0,1],\quad C_n^-(u_1,\dots,u_n)\le C(u_1,\dots,u_n)\le C_n^+(u_1,\dots,u_n)$$

that hold for any copula function C with $C_n^+(u_1,\dots,u_n)=\min(u_1,\dots,u_n)$ and $C_n^-(u_1,\dots,u_n)=(u_1+\dots+u_n-n+1)^+$ (see Nelsen (2006), Theorem 2.10.12 or Rachev and Rüschendorf (1998), section 3.6). We recall that the copula C_n^+ is the n-dimensional cumulative distribution function of the image of the Lebesgue measure on $[0,1]$ by $\mathbb{R}\ni x\mapsto(x,\dots,x)\in\mathbb{R}^n$. Also the copula C_2^- is the two-dimensional cumulative distribution function of the image of the Lebesgue measure on $[0,1]$ by $\mathbb{R}\ni x\mapsto(x,1-x)\in\mathbb{R}^2$ and, for $n\ge3$, C_n^- is not a copula.

In dimension $n=1$, the unique copula function is $C(u)=u$ and therefore the optimal coupling between μ and ρ, which necessarily share this copula, is the image of the probability measure dC by $u\mapsto(F_\mu^{-1}(u),F_\rho^{-1}(u))$. It is therefore natural to wonder whether, when μ and ρ share the same copula C in higher dimensions, the optimal coupling is still the image of the probability measure dC by $(u_1,\dots,u_n)\mapsto(F_{\mu_1}^{-1}(u_1),\dots,F_{\mu_n}^{-1}(u_n),F_{\rho_1}^{-1}(u_1),\dots,F_{\rho_n}^{-1}(u_n))$. We denote by $\mu\diamond\rho$ this probability law on $\mathbb{R}^{2n}$. It turns out that the picture is more complicated than in dimension one because of the choice of the index q of the norm: optimality is guaranteed only when $p=q$ i.e. when the cost $\|x-y\|^{pq}$ in (1.1) may be decomposed as the sum of coordinatewise costs.

Proposition 1.1. *Let $n\ge2$, μ and ρ be two probability measures on $\mathbb{R}^n$ sharing the same copula C and $\mathcal{W}_{p,q}(\mu,\rho)=\inf_{\nu<^{\mu}_{\rho}}\left(\int_{\mathbb{R}^n\times\mathbb{R}^n}\|x-y\|_q^p\nu(dx,dy)\right)^{1/p}$.*

- *If $p=q$, then an optimal coupling between μ and ρ is given by $\nu=\mu\diamond\rho$ and*

$$\mathcal{W}_{p,p}^p(\mu,\rho)=\int_{[0,1]^n}\sum_{i=1}^{n}|F_{\mu_i}^{-1}(u_i)-F_{\rho_i}^{-1}(u_i)|^p dC(u_1,\dots,u_n)=\int_{[0,1]}\sum_{i=1}^{n}|F_{\mu_i}^{-1}(u)-F_{\rho_i}^{-1}(u)|^p du.$$

- *If $p\ne q$, the coupling $\mu\diamond\rho$ is in general no longer optimal. For $p<q$, if $C\ne C_n^+$, we can construct probability measures μ and ρ on $\mathbb{R}^n$ admitting C as their unique copula such that*

$$\left(\int_{\mathbb{R}^n\times\mathbb{R}^n}\|x-y\|_q^p\mu\diamond\rho(dx,dy)\right)^{1/p}>\mathcal{W}_{p,q}(\mu,\rho).$$

  For $p>q$, the same conclusion holds if $n\ge3$ or $n=2$ and $C\ne C_2^-$.

Remark 1.2. Let μ and ρ be two probability measures on $\mathbb{R}^n$ and $\nu<^{\mu}_{\rho}$. For $n=1$, ν is said to be comonotonic if $\nu((-\infty,x],(-\infty,y])=C_2^+(\mu((-\infty,x]),\rho((-\infty,y]))$. Puccetti and Scarsini (2010) investigate several extensions of this notion for $n\ge2$. In particular, they say that ν is π-comonotonic (resp. c-comonotonic) if μ and ρ have a common copula and $\nu=\mu\diamond\rho$ (resp. ν maximizes $\int_{\mathbb{R}^n\times\mathbb{R}^n}\langle x,y\rangle\tilde\nu(dx,dy)$ over all the coupling measures $\tilde\nu<^{\mu}_{\rho}$). Looking at some connections between their different definitions of comonotonicity, they show in Lemma 4.4 that π-comonotonicity implies c-comonotonicity. Since

$$\int_{\mathbb{R}^n\times\mathbb{R}^n}\|x-y\|_2^2\tilde\nu(dx,dy)=\int_{\mathbb{R}^n}\|x\|_2^2\mu(dx)+\int_{\mathbb{R}^n}\|y\|_2^2\rho(dy)-2\int_{\mathbb{R}^n\times\mathbb{R}^n}\langle x,y\rangle\tilde\nu(dx,dy),$$

this yields our result in the case $p=q=2$.

2. Proof of Proposition 1.1

The optimality in the case $q=p$, follows by choosing $d_1=\dots=d_n=d'_1=\dots=d'_n=d''_1=\dots=d''_n=1$, $c_i(y_i,z_i)=|y_i-z_i|^p$, $\alpha=dC$, and $\varphi_i=F_{\mu_i}^{-1}$, $\psi_i=F_{\rho_i}^{-1}$ in the following Lemma.

Lemma 2.1. *Let α be a probability measure on $\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\times\dots\times\mathbb{R}^{d_n}$ with respective marginals $\alpha_1,\dots,\alpha_n$ on $\mathbb{R}^{d_1},\dots,\mathbb{R}^{d_n}$ and $\varphi_i:\mathbb{R}^{d_i}\to\mathbb{R}^{d'_i}$, $\psi_i:\mathbb{R}^{d_i}\to\mathbb{R}^{d''_i}$ and $c_i:\mathbb{R}^{d'_i}\times\mathbb{R}^{d''_i}\to\mathbb{R}_+$ be measurable functions such that*

$$\forall i\in\{1,\dots,n\},\qquad\inf_{\nu_i<^{\alpha_i\circ\varphi_i^{-1}}_{\alpha_i\circ\psi_i^{-1}}}\int_{\mathbb{R}^{d'_i}\times\mathbb{R}^{d''_i}}c_i(y_i,z_i)\nu_i(dy_i,dz_i)=\int_{\mathbb{R}^{d_i}}c_i(\varphi_i(x_i),\psi_i(x_i))\alpha_i(dx_i).\qquad(2.1)$$