On c-optimal Random Variables

Ludger Rüschendorf, University of Freiburg

Abstract

A characterization is proved for random variables which are optimal couplings w.r.t. a general function c. It turns out that on very general probability spaces optimal couplings can be characterized by subgradients of c-convex functions. An interesting application of optimal couplings are minimal l^p-metrics.

AMS subject classification: 62H20

Key words and phrases: optimal couplings, c-cyclically monotone functions, marginal problem

1 Introduction

Let P_i be probability measures on $(\Omega_i, \mathcal{A}_i), i = 1,2$ and let $c : \Omega_1 \times \Omega_2 \longrightarrow \mathbb{R}^1$ be measurable w.r.t. the product σ-algebra. Call a pair of random variables $X_1 \stackrel{d}{=} P_1, Y_2 \stackrel{d}{=} P_2$ c-optimal if

$$Ec(X_1, X_2) = \sup\{Ec(U,V); U \stackrel{d}{=} P_1, V \stackrel{d}{=} P_2\}. \tag{1.1}$$

The underlying probability space is assumed to support sufficiently many rv's. (1.1) is the basis of the optimal coupling problem and optimal solutions have been characterized in several cases (cf. [1], [6], [7], [8], [9]). An interesting special case of problem (1.1) is given when $\Omega_1 = \Omega_2$ is a metric space and $c(x,y) = -d^p(x,y), p \geq 1$, is the p-th power of the underlying metric. Then (1.1) leads to the problem to determine the minimal l_p-metric (w.r.t. distance d), i.e.

$$l_p(P_1, P_2) = \inf\left\{(Ed^p(Y_1,Y_2))^{1/p}; Y_i \stackrel{d}{=} P_i\right\}. \tag{1.2}$$

For the relevance and wide field of applications of this metric cf. [3].

The characterization of optimal solutions of (1.1) is closely related to the investigation of inequalities from conjugate duality theory. Define a subset $\Gamma \subset \Omega_1 \times \Omega_2$ to be c-cyclically monotone, if for all $(x_1,y_1), \ldots, (x_n,y_n) \in \Gamma, x_{n+1} := x_1$:

$$\sum_{i=1}^{n} (c(x_{i+1}, y_i) - c(x_i, y_i)) \leq 0 \tag{1.3}$$

and for functions f on Ω_1, g on Ω_2 define the c-subgradient in $x \in \Omega_1$ resp. $y \in \Omega_2$

$$\begin{aligned} \partial_c f(x) &= \{y \in \Omega_2; f(z) - f(x) \geq c(z,y) - c(x,y), \forall z \in \mathrm{dom}(f)\} \\ \partial_c g(y) &= \{x \in \Omega_1; g(z) - g(y) \geq c(x,z) - c(x,y), \forall z \in \mathrm{dom}(g)\} \end{aligned} \tag{1.4}$$

(cf.[2], [7]).

f is called c-convex if

$$f(x) = \sup_{i \in I}(c(x,y_i) + a_i) \tag{1.5}$$

for some y_i, a_i and index set I. The c-conjugate of f is defined by

$$f^*(y) = \sup_{x \in \mathrm{dom} f}(c(x,y) - f(x)), \qquad y \in \Omega_2 \tag{1.6}$$

and the doubly c-conjugate

$$f^{**}(x) = \sup_{y \in \mathrm{dom} f^*}(c(x,y) - f^*(y)). \tag{1.7}$$

Then f is c-convex if and only if $f = f^{**}$(cf. [2]). The aim of this note is to relate problems (1.1) to (1.3) in a general situation.

2 Optimal c-couplings

We first establish a relation between c-cyclically monotone sets and c-subgradients.

Lemma 2.1 *$\Gamma \subset \Omega_1 \times \Omega_2$ is c-cyclically monotone if and only if there ex. a c-convex function f on Ω_1 such that $\Gamma \subset \partial_c f$ (i.e. $\Gamma_x \subset \partial_c f(x)$ for all $x \in \Omega_1$).*

Proof: If $\Gamma \subset \partial_c f$ and $(x_i, y_i) \in \Gamma$, $1 \leq i \leq n$, then by definition of $\partial_c f(x_i)$, $\sum_{i=1}^n (c(x_{i+1}, y_i) - c(x_i, y_i)) \leq \sum_{i=1}^n (f(x_{i+1}) - f(x_i)) = 0$, i.e. Γ is c-cyclically monotone.

If, conversely, Γ is c-cyclically monotone, and $(x_0, y_0) \in \Gamma$, then define $f : \Omega_1 \longrightarrow \overline{\mathbb{R}}$

$$f(x) = \sup_{(x_i,y_i)\in\Gamma, 1\leq i\leq n} (c(x, y_n) - c(x_n, y_n) + \ldots + c(x_1, y_0) - c(x_0, y_0)). \tag{2.1}$$

Then f is c-convex and $f(x_0) = 0$ as Γ is c-cyclically monotone. We establish that $\Gamma \subset \partial_c f$. Let $(x', y') \in \Gamma$ and $\lambda < f(x')$, then there exist $(x_i, y_i) \in \Gamma, 1 \leq i \leq m$, with $\lambda < c(x', y_m) - c(x_m, y_m) + \ldots + c(x_1, y_0) - c(x_0, y_0)$. Define $x_{m+1} = x', y_{m+1} = y'$, then for $x \in \Omega_1$

$$\begin{aligned} f(x) &\geq c(x, y_{m+1}) - c(x_{m+1}, y_{m+1}) + c(x_{m+1}, y_m) \\ &\quad - c(x_m, y_m) + \ldots + c(x_1, y_0) - c(x_0, y_0) \\ &\geq c(x, y_{m+1}) - c(x_{m+1}, y_{m+1}) + \lambda. \end{aligned}$$

This implies

$$f(x) - f(x') \geq c(x, y') - c(x', y'), \forall x \in \Omega_1$$

and since $f(x_0) = 0, f(x') < \infty$. Therefore, $y' \in \partial_c f(x')$ and so $\Gamma \subset \partial_c f$. □

Let $(\mathcal{A}_1 \otimes \mathcal{A}_2)_m$ denote the set of all lower majorized $\mathcal{A}_1 \otimes \mathcal{A}_2$ measurable functions c on $\Omega_1 \times \Omega_2$, i.e. $c(x, y) \geq f_1(x) + f_2(y)$ for some $f_i \in \mathcal{L}^1(P_i)$. Recall that P_i is called perfect if for every measurable function $f_i : \Omega_i \longrightarrow \mathbb{R}^1$ one can find a Borel set $B_i \subset f_i(\Omega_i)$ such that $P_i(f_i^{-1}(B_i)) = 1$. Perfectness is a weak regularity condition on P_i. For properties of this notion we refer to [4]. Define for $f_i \in \mathcal{L}^1(P_i), f_1 \oplus f_2(x, y) = f_1(x) + f_2(y)$. The following theorem gives a very general characterization of c-optimal random variables. Special cases of this result are in [6], [7], [8], [9].

Theorem 2.2 *Let P_1 or P_2 be perfect, $c \in (\mathcal{A}_1 \otimes \mathcal{A}_2)_m$ and*

$$I(c) = \inf \left\{ \sum_{i=1}^{2} \int f_i dP_i; f_i \in \mathcal{L}^1(\mathcal{A}_i, P_i), c \leq f_1 \oplus f_2 \right\} < \infty. \tag{2.2}$$

Then $X_i \stackrel{d}{=} P_i, i = 1, 2$, are c-optimal if and only if $X_2 \in \partial_c f(X_1)$ a.s. for some c-convex function f or if and only if the support Γ of the distribution of (X_1, X_2) is c-cyclically monotone.

Proof: If $X_i \stackrel{d}{=} P_i$ and $X_2 \in \partial_c f(X_1)$ a.s. for some c-convex function f, then for any rv's $Y_i \stackrel{d}{=} P_i$ we have the following chain of inequalities. If f^* denotes the c-conjugate of f, then $f(x) + f^*(y) \geq c(x,y)$ for all x, y and

$$Ec(Y_1, Y_2) \leq E(f(Y_1) + f^*(Y_2)) = E(f(X_1) + f^*(X_2)) = Ec(X_1, X_2), \tag{2.3}$$

i.e. the pair (X_1, X_2) is c-optimal.

For the converse note that by Theorem 1 in [5] the following duality theorem holds:

$$\sup\{Ec(Y_1, Y_2); Y_i \stackrel{d}{=} P_i\} = I(c) = \inf\left\{\sum_{i=1}^{2} \int f_i dP_i; f_i \in \mathcal{L}^1(P_i), c \leq f_1 \oplus f_2\right\}. \tag{2.4}$$

Let (f_1, f_2) be a solution of the dual problem which exists by Proposition 3 in [5]. Then with

$$\begin{aligned} f^*(y) &= \sup_x (c(x,y) - f_1(x)) \quad \text{and} \\ f^{**}(x) &= \sup_y (c(x,y) - f^*(y)) \end{aligned}$$

the pair (f^{**}, f^*) is admissible, i.e. $f^{**}(x) + f^*(y) \geq c(x,y), f^*, f^{**}$ are c-convex and f^{**} is the largest c-convex function majorized by f, and $f_1 \oplus f_2 \geq f^{**} \oplus f^*$. Therefore, also (f^{**}, f^*) is a solution of the dual problem.

From the equality $Ec(X_1, X_2) = E(f^{**}(X_1) + f^*(X_2))$ we conclude that $c(X_1, X_2) = f^{**}(X_1) + f^*(X_2)$ a.s. and so $X_2 \in \partial_c f^{**}(X_1)$ a.s. (equivalently, also $X_1 \in \partial_c f^*(X_2)$ a.s.) □

Examples and Remark:

a) Let $\Omega_i = \mathbb{R}^k$ and $c(x,y) = -|x-y|^p, p > 1, |\quad|$ the euclidean metric, i.e. we consider the problem to determine the minimal l_p-metric as in the introduction. $\Phi : \mathbb{R}^k \longrightarrow \mathbb{R}^k$ is called cyclically monotone if $\sum_{i=1}^n \Phi(x_i)(x_{i+1} - x_i) \leq 0$ for all $x_1, \ldots, x_n \in \mathbb{R}^k, x_{n+1} := x_1$. Cyclically monotone functions are well studied in convex analysis. They arise essentially as gradients of convex functions. From [6] cyclically monotone functions lead to optimal couplings w.r.t. $-|\quad|^2$. For a cyclically monotone function Φ define

$$\Psi(x) = |\Phi(x)|^{-\frac{p-2}{p-1}} \Phi(x) + x, \tag{2.5}$$

then Ψ is c-cyclically monotone, and for any r.v. X_1 in the domain of Ψ, the pair $(X_1, \Psi(X_1))$ is an optimal c-coupling.

For the proof note that by concavity of $c(x,y) = -|x-y|^p$ we have

$$\begin{aligned} &\sum_{i=1}^n (c(x_{i+1}, \Psi(x_i)) - c(x_i, \Psi(x_i)) \\ \leq\ &\sum_{i=1}^n c_1(x_i, \Psi(x_i))(x_{i+1} - x_i) \\ =\ &p \sum_{i=1}^n \Phi(x_i)(x_{i+1} - x_i) \leq 0. \end{aligned}$$

The case $p = 2$ leads to the optimality of $\Phi(x) + x$ for the squared euclidean distance (cf. [6]), the case $1 < p < 2$ of this result has been dealt with in [9]. From the result for $p = 2$ one can see that the sufficient condition for optimality in (2.5) is not too far from being necessary.

The case $p = 1$, i.e. the Kantorovich l_1-metric has been studied in [8]. If Ψ satisfies the normalized angle monotonicity condition

$$(x - y)\left(\frac{\Psi(x) - x}{|\Psi(x) - x|} - \frac{\Psi(y) - x}{|\Psi(y) - x|}\right) \geq 0, \tag{2.6}$$

then $(X_1, \Psi(X_1))$ is an optimal l_1-coupling for the l_1-metric w.r.t. the euclidean distance for any r.v. X_1 in the domain of Ψ.

b) There remain two central open problems with the application of Theorem 2.2. The first one is to find characterizations of c-convex functions and c-subgradients. Only in few cases as $c(x, y) = -|x - y|^2$ this problem has been dealt with satisfactorily. A second problem is to find to given P, Q an optimal coupling function Φ. If P, Q are on $\mathbb{R}^k$ with densities f, g and if a regular invertible solution Φ exists, then by the transformation formula the problem to be solved is a Monge type nonlinear partial differential equation. Find Φ regular, c-cyclically monotone such that in the support of Q

$$g(x) = f(\Phi^{-1}(x))|\det D_{\Phi^{-1}}(x)|. \tag{2.7}$$

The usual boundary conditions of PDE's are replaced by the condition of c-cyclical monotonicity. □

References

[1] Cuesta-Albertos, J.A., Rüschendorf, L. and Tuero-Diaz, A. (1993) Optimal coupling of multivariate distributions and stochastic processes. J. Mult. Anal. 46, 335-361

[2] Dietrich, H. (1988) Zur c-Konvexität und c-Subdifferenzierbarkeit von Funktionalen. Optimization 19, 355-371

[3] Rachev, S.T. (1991) Probability Metrics and the Stability of Stochastic Models. Wiley New York

[4] Ramachandran, D. (1979) Perfect Measures, I and II. ISI Lecture Notes Ser. 5 and 7, New Delhi, McMillan

[5] Ramachandran, D. and Rüschendorf, L. (1994) A general duality theorem for marginal problems. To appear in: Probability Theory Rel. Fields

[6] Rüschendorf, L. and Rachev, S.T. (1990) A characterization of random variables with minimum L^2-distance. J. Mult. Analysis 32, 48-54

[7] Rüschendorf, L. (1991) Fréchet bounds and their applications. In: Advances in Probability Measures with given Marginals. Eds. Dall'Aglio, Katz and Salinetti, Kluver, 151-188

[8] Rüschendorf, L. (1993) Optimal solutions of multivariate coupling problems. Preprint. To appear in: Applicationes Mathematicae

[9] Smith, C. and Knott, M. (1992) On Hoeffding Fréchet bounds and cyclic monotone relations. J. Mult Anal. 40, 328-334

L. Rüschendorf
Institut für Mathematische Stochastik
Hebelstr. 27
79104 Freiburg
Germany