PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 13, Fasc. 2 (1992), pp. 311–320

ON L_p-MINIMAL METRICS

BY

S. T. RACHEV* (SANTA BARBARA, CALIFORNIA) AND A. SCHIEF (MÜNICH)

Abstract. We obtain refine versions of the dual representations for

$$\inf_{(\sup)} \{[\mathrm{E}d^p(X, Y)]^{1/p}: \Pr_X = P, \Pr_Y = Q\}$$

for probabilities P and Q on a separable metric space (U, d).

1. Introduction. Given a separable metric space (s.m.s.) (U, d) let $\mathscr{P}_p(U)$ $(p \geqslant 1)$ be the space of all Borel probability measures (probabilities) P on (U, d) with finite $\int d^p(x, a)P(dx)$. For $P, Q \in \mathscr{P}_p(U)$ let $\mathscr{M}(P, Q)$ be the set of all probabilities on $U \times U$ with fixed marginals P and Q. For $\mu \in \mathscr{M}(P, Q)$ let

$$\mathscr{L}_p(\mu) := \left[\int d^p(x, y)\mu(dx, dy)\right]^{1/p} \tag{1.1}$$

and let

$$l_p(P, Q) := \inf\{\mathscr{L}_p(\mu): \mu \in \mathscr{M}(P, Q)\}, \tag{1.2}$$

$$L_p(P, Q) := \sup\{\mathscr{L}_p(\mu): \mu \in \mathscr{M}(P, Q)\}. \tag{1.3}$$

The problem of dual and explicit solution for the minimal and maximal $\mathscr{L}_p$-metrics, l_p and L_p, has a long history which goes back to the work of G. Monge, C. Gini, M. Fréchet, W. Hoeffding and L. V. Kantorovich (see, e.g., [12] and the survey [15]). The dual forms for l_p and L_p are given by

$$l_p^p(P, Q) = \sup\left\{\int f dP + \int g dQ: (f, g) \in \mathscr{G}_p\right\}, \tag{1.4}$$

$$L_p^p(P, Q) = \inf\left\{\int f dP + \int g dQ: (f, g) \in \mathscr{G}_p^*\right\}, \tag{1.5}$$

where $\mathscr{G}_p$ (resp. $\mathscr{G}_p^*$) is the set of all pairs of bounded continuous functions on U satisfying the dual constraint $f(x)+g(y) \leqslant d^p(x, y)$ (resp. $f(x)+g(y) \geqslant d^p(x, y)$) for all $x, y \in U$ (see [6], [9], [12]). While in the case $p = 1$

* Research supported in part by the Deutsche Forschungsgemeinschaft (DFG) and grant DMS-9103452 from National Science Foundation.

one can replace g with $(-f)$ in (1.4), in general, for $p>1$, there is no dual representation for (1.4) as a $\zeta_{\mathscr{F}}$-metric

$$(1.6)\qquad \zeta_{\mathscr{F}}(P, Q) = \sup_{f\in\mathscr{F}} \left|\int f d(P-Q)\right|,$$

where $\mathscr{F}$ is a class of bounded continuous functions (see [10]).

The aim of this paper is to obtain more informative dual representations than (1.4) and (1.5) by showing that the supremum in (1.4) (resp. the infimum in (1.5)) can be taken over smaller than $\mathscr{G}_p$ (resp $\mathscr{G}_p^*$) set.

Basing on the Kantorovich representation $l_1 = \zeta_{\mathrm{Lip}(1)}$ with

$$\mathrm{Lip}(1) = \{f\colon U\to R, f(x)-f(y) \leqslant d(x, y)\ \forall x, y\in U\},$$

Szulga [18] made the conjecture that for $P, Q\in\mathscr{P}_p(U)$

$$(1.7)\qquad l_p(P, Q) = AS_p(P, Q) := \sup_{f\in\mathrm{Lip}(1)} \left|\left[\int |f|^p dP\right]^{1/p} - \left[\int |f|^p dQ\right]^{1/p}\right|.$$

Despite the fact that l_p and AS_p induce one and the same convergence in $\mathscr{P}_p(U)$ we shall construct an example showing that Szulga's conjecture fails. We shall characterize the optimal solutions μ in (1.3), i.e., those $\mu\in\mathscr{M}(P, Q)$ for which $L_p(P, Q) = \mathscr{L}_p(\mu)$. Finally we shall discuss some open problems.

2. Dual representations for minimal and maximal L_p-metrics. First, we shall show that $l_p^p(P, Q)$ admits a dual form similar to that of $\zeta_{\mathscr{F}}(P, Q)$ (cf. (1.6)) but with $\mathscr{F}$ depending on P and Q. Denote by $\nu_1 = (P-Q)^+$ and $\nu_2 = (P-Q)^-$ the positive and negative part of the Jordan decomposition $P-Q$. Let A_1 be the support of $(P-Q)_+$ and $A_2 = U\backslash A_1$. Define the set $\mathscr{F}_p(P, Q)$ of functions $f = f_1 I_{A_1} + f_2 I_{A_2}$, where f_i are bounded functions on A_i, having finite Lipschitz norms

$$\mathrm{Lip}(f_i; A_i) := \sup\{|f_i(x)-f_i(y)|/d(x, y)\colon x\neq y, x, y\in A_i\} < \infty$$

and satisfying the dual constraint

$$f_1(x)-f_2(y) \leqslant d^p(x, y)\ \forall x\in A_1, y\in A_2.$$

THEOREM 2.1. *For any* $P, Q\in\mathscr{P}_p(U)$,

$$(2.1)\qquad l_p^p(P, Q) = \sup_{f\in\mathscr{F}_p(P,Q)} \int f d(P-Q).$$

Proof. We start with the following dual representation for l_p^p (cf. (1.4)):

$$(2.2)\qquad l_p^p(P, Q) = \sup\left\{\int f dP + \int g dQ\colon (f, g)\in\mathscr{G}_p, \mathrm{Lip}(f; U)+\mathrm{Lip}(g; U) < \infty\right\}$$

(see [12]). Suppose first that

$$(2.3)\qquad P(A_2) = Q(A_1) = 0.$$

By (2.2), and since $f|_{A_1} - g|_{A_2} \in\mathscr{F}_p(P, Q)$ for $(f, g)\in\mathscr{G}_p$, $\mathrm{Lip}(f; A_1) < \infty$, Lip$(g;\, A_2) < \infty$, we have

$$(2.4)\quad l_p^p(P, Q) \leqslant \sup\{\textstyle\int f d(P-Q):\ f\in\mathscr{F}_p(P, Q)\}$$
$$\leqslant \inf_{\mu\in\mathscr{M}(P,Q)} \sup_{f\in\mathscr{F}(P,Q)} \int (f\circ\pi_1 - f\circ\pi_2)\,d\mu$$
$$\leqslant \inf\{\int_{A_1\times A_2} d^p d\mu:\ \mu\in\mathscr{M}(P, Q)\} = \inf\{\textstyle\int d^p d\mu:\ \mu\in\mathscr{M}(P, Q)\}$$
$$= l_p^p(P, Q).$$

To omit the assumption (2.3) set $\hat{P} = (P-Q)^+$, $\hat{Q} = (Q-P)^+$, $\bar{v} = P-\hat{P} = Q-\hat{Q}$ and recall that $\hat{P}(U\backslash A_1) = \hat{Q}(A_1) = 0$. We then get

$$(*)\quad \sup\{\textstyle\int f d(P-Q):\ f\in\mathscr{F}_p(P, Q)\} = \sup\{\textstyle\int f d(\hat{P}-\hat{Q}):\ f\in\mathscr{F}_p(\hat{P}, \hat{Q})\}$$
$$= l_p^p(\hat{P}, \hat{Q})$$
$$= \inf\{\textstyle\int d^p dv:\ \pi_1 v = \hat{P},\ \pi_2 v = \hat{Q}\}$$
$$= \inf\{\textstyle\int d^p d\mu:\ \mu\in\mathscr{M}(P, Q)\}$$
$$= l_p^p(P, Q).$$

The equality $(*)$ can be shown as follows:

($\geqslant$) Given v choose μ by

$$\mu(B) = v(B) + \bar{v}\big(\pi_1^{-1}(B\cap\{(x, x):\ x\in U\})\big).$$

($\leqslant$) Given μ choose v by $v(B_1\times B_2) = \mu(B_1\times B_2) - \bar{v}(B_1\cap B_2)$. ∎

Remark. More interesting would be $\zeta_{\mathscr{F}}$-representation for l_p (not l_p^p) with an $\mathscr{F}$ that depends only on the support of $(P-Q)^+$. The next example shows that this is impossible (cf. [10]).

Example. Suppose $l_p = \zeta_{\mathscr{F}}$. Then for $0 < r < s < 1$ we have

$$l_p\big(r\delta_a + (1-r)\delta_b,\ s\delta_a + (1-s)\delta_b\big)$$
$$= \sup\{\textstyle\int f d\big(r\delta_a + (1-r)\delta_b - s\delta_a - (1-s)\delta_b\big):\ f\in\mathscr{F}\},$$

i.e.,

$$(s-r)^{1/p} d(a, b) = \sup\{(s-r)\big(f(a)-f(b)\big):\ f\in\mathscr{F}\} = (s-r)\,\mathrm{const}.$$

If $d(a, b) > 0$, this yields $(s-r)^{1-1/p} = \mathrm{const}$, and thus $p = 1$. ∎

In the case $p = 1$, the representation (2.1) leads to $l_1(P, Q) = \zeta_{\mathrm{Lip}(1)}$ (see [17], [7], [16]). Taking the dual form for l_1, Szulga's conjecture seems reasonable. First let us show that AS_p and l_p metrize one and the same convergence in $\mathscr{P}_p(U)$. Let π be the Prokhorov metric

$$(2.5)\quad \pi(P, Q) = \inf\{\varepsilon > 0:\ P(C) \leqslant Q(C^\varepsilon) + \varepsilon \text{ for all closed } C\subset U\},$$

where C^ε is an ε-neighborhood of C.

PROPOSITION 2.1. *For any $P, Q \in \mathscr{P}_p(P, Q)$, $p \geqslant 1$, the following inequalities hold*:

$$AS_p(P, Q) \leqslant l_p(P, Q) \tag{2.6}$$

and

$$C_p \pi^2(P, Q) \leqslant AS_p(P, Q), \tag{2.7}$$

where $C_p \geqslant 1/(p \cdot 2^{p-1})$. In particular, for $P_n, P \in \mathscr{P}_p(U)$ the following are equivalent: *as $n \to \infty$,*

(a) $$l_p(P_n, P) \to 0,$$

(b) $$AS_p(P_n, P) \to 0,$$

(c) $$\pi(P_n, P) \to 0 \quad \text{and} \quad \int d^p(x, a)(P_n - P)(dx) \to 0.$$

Proof. The inequality (2.6) is a consequence of the Minkovski inequality. In fact, there exists a rich enough probability space $(\Omega, \mathscr{A}, \Pr)$ such that the space of laws $\Pr_{X,Y}$ coincides with the space of probabilities on $U \times U$, and thus

$$AS_p(P, Q) = AS_p(\Pr_X, \Pr_Y) \leqslant [\mathbb{E} d^p(X, Y)]^{1/p},$$

which implies (2.6). To show (2.7) observe that for any closed $C \subset U$ and

$$f_C(x) = \max\left(0, 1 - \frac{d(x, C)}{\varepsilon}\right), \qquad \varepsilon \in (0, 1),$$

we have

$$P(C)^{1/p} \leqslant \left[\int f_C^p \, dP\right]^{1/p} \leqslant Q(C^\varepsilon)^{1/p} + \varepsilon^{-1} AS_p(P, Q).$$

If $AS_p(P, Q) \leqslant \delta := C_p \varepsilon^2$, then

$$P(C) \leqslant \left(Q(C^\varepsilon)^{1/p} + \varepsilon^{-1} AS_p(P, Q)\right)^p \leqslant \left(Q(C^\varepsilon)^{1/p} + C_p \varepsilon\right)^p \leqslant Q(C^\varepsilon) + \varepsilon.$$

The last inequality follows from $(a^{1/p} + C_p \varepsilon)^p \leqslant a + \varepsilon$ for any $a, \varepsilon \in (0, 1)$. Letting $\delta \to AS_p(P, Q)$ we obtain (2.7). Next, (a)$\Leftrightarrow$(c) (see [12]); (a)$\Rightarrow$(b) (cf. (2.6)); (b)$\Rightarrow$(c) by virtue of (2.7) and

$$AS_p(P, Q) \geqslant \left(\int d^p(x, a) P(dx)\right)^{1/p} - \left(\int d^p(x, a) Q(dx)\right)^{1/p}. \quad \blacksquare$$

Remark. If p is integer, one can get a better estimate for C_p, namely

$$C_2 = \sqrt{2} - 1, \qquad C_n \geqslant 1/2n, \; n \in N.$$

The first indication that Szulga's conjecture is not valid comes from the bound $AS_p \geqslant C_p \pi^2$ and the corresponding bound for l_p, $l_p \geqslant \pi^{1+1/p}$. Note that both estimates have precise order.

The next example shows that $AS_p \neq l_p$. For simplicity we consider the case $p=2$. Let $(U, d) = ([0, 1], |\cdot|)$, $P(\{0\}) = 1-P(\{1\}) = \frac{1}{3}$ and $Q(\{0\}) = 1-Q(\{1\}) = \frac{2}{3}$. Then there exists $\mu \in \mathcal{M}(P, Q)$, $\mathcal{L}_2(\mu) = (\frac{1}{3}d(0, 1))^{1/2} = 1/\sqrt{3}$ and $\mathcal{L}_2(P, Q) = 1/\sqrt{3}$ follow since, for any $\mu = \mathcal{M}(P, Q)$, $\mathcal{L}_2(\mu) \geqslant 1/\sqrt{3}$. For calculating $AS_2(P, Q)$, setting $f(0) = a, f(1) = b$, we have to maximize $|\varphi(a, b)|$,

$$\varphi := (\tfrac{2}{3}a^2 + \tfrac{1}{3}b^2)^{1/2} - (\tfrac{1}{3}a^2 + \tfrac{2}{3}b^2)^{1/2} \quad \text{on } D = \{(a, b)\colon |a-b| \leqslant 1\}.$$

Since $(\partial\varphi/\partial a = 0, \partial\varphi/\partial b = 0) \Leftrightarrow (a = b = 0)$ and the case $a = b = 0$ is trivial, we have to look for the extrema of φ on ∂D. We consider $b = a-1$ (the case $b = a+1$ is similar). Set $g(a) = \varphi(a, a-1)$. Then $g'(a) = 0$ iff

$$(2a-\tfrac{2}{3})^2 (a^2 - \tfrac{4}{3}a + \tfrac{2}{3}) = (2a - \tfrac{4}{3})^2 (a^2 - \tfrac{2}{3}a + \tfrac{1}{3})$$

iff $a = \frac{1}{2}$. Since $g(\frac{1}{2}) = 0$, what is left is to consider the limiting behavior of $\varphi(a, b)$ as $a \to \pm\infty$, $|b-a| \leqslant 1$,

$$\begin{aligned}
\varphi(a, b) &= \big(a^2 + \tfrac{2}{3}a(b-a) + \tfrac{1}{3}(b-a)^2\big)^{1/2} - \big(a^2 + \tfrac{4}{3}a(b-a) + \tfrac{2}{3}(b-a)\big)^{1/2} \\
&= \big((a + \tfrac{1}{3}(b-a))^2 + \tfrac{2}{9}(b-a)^2\big)^{1/2} - \big((a+\tfrac{2}{3}(b-a))^2 + \tfrac{2}{9}(b-a)^2\big)^{1/2} \\
&\underset{|a|\to\infty}{\sim} \left|a + \frac{b-a}{3}\right| - |a + \tfrac{2}{3}(b-a)| = \begin{cases} (b-a)\frac{1}{3}, & a \to +\infty, \\ -(b-a)\frac{1}{3}, & a \to -\infty. \end{cases}
\end{aligned}$$

In both cases, $|\varphi| \leqslant \frac{1}{3}$, and thus $AS_2(P, Q) = \frac{1}{3} \neq l_2(P, Q) = 1/\sqrt{3}$. ∎

Our next theorem is a refinement of the dual representation for L_p (cf. (1.5)) in the case of (U, d) being a separable Banach space, $d(x, y) = \|x-y\|$.

Let f be a function on U. The function f^* on U is *p-conjugate* if

(2.8) $$f^*(y) := \sup_{x \in U} \{\|x-y\|^p - f(x)\}, \quad y \in U.$$

The pair (f, f^*) satisfies the admissibility constraint in (1.5):

(2.9) $$f(x) + f^*(y) \geqslant \|x-y\|^p \ \forall x, y \in U.$$

If $f^{**} = (f^*)^*$ is the second p-conjugate, then

(2.10) $$f \geqslant f^{**}.$$

Moreover, f^{**} is convex and lower semicontinuous (l.s.c.).

THEOREM 2.2. *For any* $P, Q \in \mathcal{P}_p(U)$, $p \geqslant 1$,

(2.11) $$L_p^p(P, Q) = \inf\{\textstyle\int f dP + \int g dQ\colon f, g \text{ convex l.s.c. and for all } x, y \in U, f(x) + g(y) \geqslant \|x-y\|^p\}.$$

Proof. The LHS (left-hand side) of (2.11) is obviously not greater than the RHS (right-hand side). To show LHS $\geqslant$ RHS for any $(f, g) \in \mathcal{G}_p^*$ (cf. (1.5)) consider (f^{**}, f^{***}). Then, by (2.9) and (2.10),

$$g(y) \geqslant \sup\{\|x-y\|^p - f(x)\} = f^*(y) \geqslant f^{***}(y),$$

$f \geqslant f^*$ and $f^{**}(x)+f^{***}(y) \geqslant \|x-y\|^p$. This yields LHS $\geqslant$ RHS. ∎

If $(\Omega, \mathscr{A}, \Pr)$ is a nonatomic space, then

$$L_p(P, Q) = \sup\{(\mathbf{E}\|X-Y\|^p)^{1/p}\colon \Pr_X = P,\ \Pr_Y = Q\} \tag{2.12}$$

and the supremum is attained for an "optimal" pair (X, Y) (cf. [14], Theorem 8.1.1). We shall characterize the set of optimal pairs for (2.12). For any function f on U let us put

$$D_p f(x) := \{y \in U\colon f(x)+f^*(y) = \|x-y\|^p\}. \tag{2.13}$$

The next corollary resembles Theorem 1 of Rüschendorf and Rachev (see [15], p. 334), characterizing the optimal measure for $l_2(P, Q)$.

COROLLARY 2.1. *The pair* (X_0, Y_0) *with* $\Pr_{X_0} = P$, $\Pr_{Y_0} = Q$ *is optimal for* (2.12) *iff*

$$Y_0 \in D_p f(X_0) \ \textit{a.s.} \tag{2.14}$$

for some l.s.c. convex function f.

Proof. Suppose that X_0 and Y_0 — with laws P and Q respectively — satisfy (2.14). Then (X_0, Y_0) is optimal since for any other X and Y with laws P and Q we have

$$\mathbf{E}\|X-Y\|^p \leqslant \mathbf{E}f(X)+\mathbf{E}f^*(Y) = \mathbf{E}f(X_0)+\mathbf{E}f^*(Y_0) = \mathbf{E}\|X_0-Y_0\| \ \text{a.s.}$$

Suppose now that (X_0, Y_0) is an optimal pair. By Theorem 2.21 of [6] there exist f_0, g_0 with $\int|f_0|dP < \infty$, $\int|g_0|dQ < \infty$ satisfying $f_0(x)+g_0(y) \geqslant \|x-y\|^p$ such that

$$\int f_0 dP + \int g_0 dQ = \inf\{\int f dP + \int g dQ\colon \int|f|\,dP < \infty,\ \int|g|dQ < \infty,$$
$$f(x)+g(y) \geqslant \|x-y\|^p \text{ for all } x, y \in U\}.$$

As in Theorem 2.2 we conclude that (f_0^{**}, f_0^{***}) is also optimal, and thus $\|X_0-Y_0\|^p = f_0^{**}(X_0)+f_0^{***}(Y_0)$ a.s., i.e., $Y_0 \in D_p(f_0^{**})$ a.s. ∎

Next we consider the special case $p = 2$ and $U = \boldsymbol{R}^k$ with Euclidean norm $\|\ \|$. Then

$$L_2^2(P, Q) = \sup\{\mathbf{E}\|X-Y\|^2\colon \Pr_X = P,\ \Pr_Y = Q\} \tag{2.15}$$
$$= \mathbf{E}\|X\|^2+\mathbf{E}\|Y\|^2 - 2\inf\{\mathbf{E}\langle X, Y\rangle\colon \Pr_X = P,\ \Pr_Y = Q\}.$$

For any f on $\boldsymbol{R}^k$ define the lower conjugate

$$f_*(y) = \inf_{x \in \boldsymbol{R}^k}\{\langle x, y\rangle - f(x)\}$$

(see [4], p. 172) and let

$$\bar{f}(y) = \sup_{x \in \boldsymbol{R}^k} \{\langle x, y\rangle - f(x)\}.$$

Then $f_* = -\bar{g}_f$, where $g_f(x) = -f(-x)$.

COROLLARY 2.2. *Let $P, Q \in \mathscr{P}_p(\boldsymbol{R}^k)$. Then the random vectors X_0, Y_0 with laws P and Q, respectively, attain the supremum in (2.15) if and only if*

$$f(X_0) + f_*(Y_0) = \langle X_0, Y_0\rangle \ \text{Pr-}a.s. \tag{2.16}$$

for some upper semicontinuous concave function f.

The proof is similar to that of Corollary 2.1, and thus omitted.

Denote the subdifferential of f in x by

$$\partial f(x) = \{y \in \boldsymbol{R}^k \colon f(x) + \bar{f}(y) = \langle x, y\rangle\}.$$

Then (2.16) is equivalent to

$$Y_0 \in \partial g(-X_0) \ \text{Pr-a.s.} \tag{2.17}$$

for some convex l.s.c. function g.

EXAMPLE. Let P and Q be Gaussian measures on $\boldsymbol{R}^k$ with means $\vec{m}_1$ and $\vec{m}_2$ and nonsigular covariance matrices Σ_1 and Σ_2, respectively. Then

$$l_2^2(P, Q) = \|\vec{m}_1 - \vec{m}_2\|^2 + \operatorname{tr}(\Sigma_1) + \operatorname{tr}(\Sigma_2) - 2\operatorname{tr}\left[\left(\sqrt{\Sigma_1}\,\Sigma_2\sqrt{\Sigma_1}\right)^{1/2}\right]$$

(see [11], [3], [2]) and

$$L_2^2(P, Q) = \|\vec{m}_1 - \vec{m}_2\| + \operatorname{tr}(\Sigma_{-1}) + \operatorname{tr}(\Sigma_{-2}) + 2\operatorname{tr}\left[\left(\sqrt{\Sigma_{-1}}\,\Sigma_{-2}\sqrt{\Sigma_{-1}}\right)^{1/2}\right],$$

where Σ_{-1} is the covariance matrix of $P(-dx)$.

OPEN PROBLEM 1. The Kantorovich metric l_1 admits a $\mathscr{G}_{\mathrm{Lip}(1)}$ representation. Is it true that

$$L_1(P, Q) = \inf\{\textstyle\int f d(P_1 + P_2) \colon f \colon U \to \boldsymbol{R},\ \operatorname{Lip}(f; U) < \infty \text{ and}$$
$$f(x) + f(y) \geqslant d(x, y)\ \forall x, y \in U\}? \tag{2.18}$$

On $(U, d) = (\boldsymbol{R}^1, |\cdot|)$ the equality (2.18) holds. In fact, if F and G are the distribution functions of P and Q, then

$$\bar{\mu}(P, Q) := \sup\{\mathbf{E}|X - Y| \colon \Pr_X + \Pr_Y = P + Q\}$$

$$\geqslant L_1(P, Q) = \int_0^1 |F^{-1}(x) - G^{-1}(1-x)|\,dx$$

$$= \int_{-\infty}^{\infty} |x-a|\,(F+G)(dx)$$

$$= \sup\{\mathbf{E}|X-a| + \mathbf{E}|Y-a|\colon \Pr_X + \Pr_Y = P+Q\} \geqslant \bar{\mu}(P_1, P_2),$$

where a is the intersection point of the completed graphs of F and G (see [14], p. 173). The dual representation for $\bar{\mu}(P, Q)$ equals the right-hand side of (2.16) (with $d(x, y) = |x-y|$); see [14], Remark 8.1.1, and Kellerer [7], which completes the proof of (2.16) in this particular case.

OPEN PROBLEM 2. Theorem 2.1 provides the dual form for

$$l_p^p(P, Q) = p\inf\{\int_0^\infty \Pr(d(X, Y) > t)\,t^{p-1}\,dt\colon \Pr_X = P, \Pr_Y = Q\}.$$

What is the dual representation for

$$\lambda_p^p(P, Q) = \inf\{\sup_{t>0}[\Pr(d(X, Y) > t)\,t^{p-1}]\colon \Pr_X = P, \Pr_Y = Q\}?$$

For any $p > 1$, λ_p is a metric. By the Strassen–Dudley theorem (see [1]) we have

$$\lambda_p^p(P, Q) \leqslant \sup_{t>0} t^{p-1} \inf\{\Pr(d(X, Y) > \varepsilon)\colon \Pr_X = P, \Pr_Y = Q\}$$

$$= \sup_{t>0} t^{p-1} \sup\{[P(A) - Q(A^t)]\colon \text{closed } C \subset U\} =\colon Y_p^p(P, Q).$$

The metrics λ_p and Y_p metrize one and the same topology (see [13] and [3]). The difference between λ_p and Y_p was first pointed out by R. Shortt in a private communication. Here we provide one example. Set

$$\Pr(X = 0) = 1 - \Pr(X = 1) = \alpha \quad \text{and} \quad \Pr(Y = 1) = 1 - \Pr(Y = 2) = \beta.$$

The joint distribution of X and Y is then determined by

$$\Pr(X = 0, Y = 1) = \alpha \ (0 < \alpha \leqslant \tfrac{1}{2}), \quad \Pr(X = 0, Y = 2) = \tfrac{1}{2} - \alpha,$$

$$\Pr(X = 1 = Y) = \tfrac{1}{2} - \alpha, \quad \Pr(X = 1, Y = 2) = \alpha.$$

Thus, for $P = \Pr_X$, $Q = \Pr_Y$,

$$\lambda_p^p(P, Q) = \inf_{0<\alpha<1/2} \max\{\max_{0<t<1} \Pr(|X-Y| > t)t^{p-1},$$

$$\max_{1 \leqslant t < 2} \Pr(|X-Y| > t)t^{p-1}\}$$

$$= \inf_{0<\alpha<1/2} \max\{\tfrac{1}{2}+\alpha, 2^{p-1}(\tfrac{1}{2}-\alpha)\} = \tfrac{1}{2}[1 + (2^{p-1}-1)/(2^{p-1}+1)].$$

On the other hand,

$$Y_p^p(P, Q) = \max\Big\{\sup_{0<t<1}\inf_{0<\alpha\leqslant 1/2}\Pr(|X-Y|>t)t^{p-1},$$

$$\sup_{1\leqslant t<2}\inf_{0<\alpha<1/2}\Pr(|X-Y|>t)t^{p-1}\Big\}$$

$$=\max\Big\{\sup_{0<t<1}\inf_{0<\alpha<1/2}(\tfrac{1}{2}+\alpha)t^{p-1},$$

$$\sup_{1\leqslant t<2}\inf_{0<\alpha<1/2}2^{p-1}(\tfrac{1}{2}-\alpha)t^{p-1}\Big\} = \tfrac{1}{2}.$$

REFERENCES

[1] R. M. Dudley, *Real Analysis and Probability*, Wadsworth & Brooks/Cole, Pacific Grove, California, 1989.

[2] M. Gelbricht, *On a formula for the L^2 Wasserstein metric between measures on euclidean and Hilbert spaces*, Math. Nachr. 147 (1990), pp. 185–203.

[3] C. R. Givens and R. M. Shortt, *A class of Wasserstein metrics for probability distributions*, Michigan Math. J. 31 (1984), pp. 231–240.

[4] A. D. Ioffe and V. M. Tihomirov, *Theory of Extremal Problems*, North-Holland, Amsterdam 1979.

[5] A. B. Kakosyan, L. Klebanov and S. T. Rachev, *Quantitative Criteria for Convergence of Probability Measures* (in Russian), Ayastan Press, Erevan 1988.

[6] H. G. Kellerer, *Duality theorems for marginal problems*, Z. Wahrsch. verw. Gebiete 67 (1984), pp. 399–432.

[7] – *Duality theorems and probability metrics*, Proc. 7th Brasov Conv., Bucuresti 1984, pp. 211–220.

[8] – *Measure-theoretic versions of linear programming*, Math. Z. 198 (1988), pp. 367–400.

[9] V. L. Levin, *The mass transfer problem in topological space and probability measures on the product of two spaces with given marginal measures*, Dokl. Akad. Nauk USSR 276 (1984), pp. 1059–1064.

[10] J. Neveu and R. M. Dudley, *On Kantorovich–Rubinstein theorems* (transcript), 1980.

[11] I. Olkin and F. Pukelsheim, *The distances between two random vectors with given dispersion matrices*, Linear Algebra Appl. 48 (1982), pp. 257–263.

[12] S. T. Rachev, *The Monge–Kantorovich mass transfer problem and its stochastic applications*, Theory Probab. Appl. 29 (1984), pp. 647–676.

[13] – *Minimal metrics in the real valued random variables space*, Lecture Notes in Math. 982, Springer-Verlag, Berlin–New York 1985, pp. 172–190.

[14] – *Probability Metrics and Stability of Stochastic Models*, Wiley, London–New York, 1991.

[15] – and L. Rüschendorf, *Recent results in the theory of probability metrics*, Statist. Decisions 9 (1991), pp. 327–372.

[16] S. T. Rachev and R. M. Shortt, *Duality Theorems for Kantorovich–Rubenstein and Wasserstein Functionals*, Dissertationes Math. (Rozprawy Mat.) 299 (1990).

[17] A. Szulga, *On minimal metrics in the space of random variables*, Theory Probab. Appl. 27 (1982), pp. 424–430.
[18] — *Communication at a seminar in Moscow State University*, 1983.

Department of Statistics
and Applied Probability
University of California
Santa Barbara, CA 93106-3110 USA

Mathematisches Institut
Ludwig-Maximilians-Universität
Theresienstr. 39, D-8000
München 2, Germany

Received on 16.4.1991;
revised version on 30.6.1992