JOURNAL OF MULTIVARIATE ANALYSIS **60**, 72–83 (1997)
ARTICLE NO. MV961627

Optimal Transportation Plans and Convergence in Distribution

J. A. Cuesta-Albertos

Universidad de Cantabria, Cantabria, Spain

C. Matrán

Universidad de Valladolid, Valladolid, Spain

and

A. Tuero-Díaz

Universidad de Cantabria, Cantabria, Spain

Explicit expression of mappings optimal transportation plans for the Wasserstein distance in $\Re^p$, $p > 1$, are not generally available. Therefore, it is of great interest to provide results which justify the practical use of simulation techniques to obtain approximate optimal transportation plans. This is done in this paper, where we obtain the consistency of the empirical optimal transportation plans. Our results can also be employed to justify a definition of multidimensional complete dependence.

1. INTRODUCTION

The Monge–Kantorovich mass-transportation problem (MTP) consists in minimizing the cost of transportation of a mass from one location to another. In this kind of problem it can usually be assumed that the mount of the mass does not vary during transportation. Then, without loss of generality, we can assume that the initial and the final distributions of the mass are given, respectively, by two probability measures P and Q. In this

Received December 3, 1993; revised February 1996.

AMS 1991 subject classifications: 60E05, 60B10.

Key words and phrases: Wasserstein distance, Monge–Kantorovich transportation problem, optimal transportation plans, convergence in distribution, consistency, maximal monotone operators, monotone dependence, Skorohod almost sure representation.

0047-259X/97 $25.00

way, it is possible to give a precise mathematical formulation of the MTP as follows.

Let P and Q be two probability measures on the Euclidean space $\Re^p$ equipped with its usual norm $\| \ \|$. Let us assume that $c(x, y) := \|x-y\|^2$ is the cost of transportation of a unit mass from x to y. If $\int \|x\|^2 \, dP$ and $\int \|x\|^2 \, dQ$ are finite, we will denote by the *L^2-Wasserstein distance between P and Q*, $W(P, Q)$, the value defined by means of

$$W^2(P, Q) := \inf\left\{\int \|x-y\|^2 \, \mu(dx, dy), \mu \in M(P, Q)\right\}, \tag{1}$$

where $M(P, Q)$ denotes the set of all probability measures on $\Re^{2p}$ with marginal distributions P and Q. Obviously, $W^2(P, Q)$ is the minimum cost of transportation of the associated MTP.

The infimum in (1) is attained. Thus, to find $W(P, Q)$ it is enough to obtain a pair (X_0, Y_0) of random vector (r.v.'s), with distributions laws $\mathscr{L}(X_0) = P$ and $\mathscr{L}(Y_0) = Q$, verifying

$$E \|X_0 - Y_0\|^2 = \inf\{E \|X - Y\|^2, \mathscr{L}(X) = P, \mathscr{L}(Y) = Q\} \qquad (= W^2(P, Q)).$$

Such a pair is called an *(L^2)-optimal transportation plan* (in short, o.t.p.) between P and Q. (L^2-optimal coupling for (P, Q) is an alternative, sometimes used, terminology).

From the probabilistic point of view, the interest of the Wasserstein distance comes from its relation with the weak convergence of probability measures (see Proposition 2.2). A good reference for the properties and applications of these and related metrics is [9].

In [2] it was proved that, under continuity assumptions on the probability P, if (X, Y) is an o.t.p., then $Y := H(X)$ almost surely (a.s.) for some suitable optimal map H. Moreover, as observed in [4], an interesting consequence of the characterization of optimal transportation plans in [11] is that the optimality of a map H is essentially independent of the distribution of X (see Proposition 2.3). In consequence, with an abuse of notation, we will often refer to such optimal functions as o.t.p.'s.

This kind of result was pioneered by Knott and Smith in [7] by considering the opposite point of view of handling mappings H, possibly multivalued, so that $(X, H(X))$ is an o.t.p. between a pair of probability measures.

The aim of this paper is to study the behavior of o.t.p.'s when the marginal distributions converge (see Theorems 3.2 and 3.4), thus generalizing Theorem 3.1 in [13]. However, the interest of this generalization is shown in the following applications which are not covered by the result in [13].

1.1. *Approximation by Simulations of Optimal Transportation Plans*

Theorems 3.2 and 3.4 permit us to obtain the consistency of o.t.p.'s (Theorem 4.2). This consistency provides the basis for using simulation techniques to obtain approximations to o.t.p.'s as follows. Given two probability measures P and Q, let P_n and Q_n be empirical probability measures obtained through a random sample respectively taken from P and Q. Theorem 4.2 implies that the o.t.p. between P_n and Q_n is a reasonable approximation for the o.t.p. between P and Q.

The interest of this approximation is twofold. On the one hand, it is widely recognized that a major open problem related to MTP is that expressions for o.t.p.'s are not generally available in the multidimensional setting. But there are algorithms which allow computation of the o.t.p. between two discrete probabilities (see, for instance, [1]) which, according to Theorem 4.2, can be used to approximate the o.t.p. we are interested in.

On the other hand, from the statistical point of view, our theorem justifies some reasonable, but previously unjustified, simulation approaches to optimal maps, as in [5], where o.t.p.'s are handled as a multivariate generalization of the quantile–quantile plots.

1.2. *Monotone Dependent Random Vectors*

Theorem 3.2 suggests a reasonable definition for the antithesis of the independence of two r.v.'s (see, for instance, [12] for the state of this question). At first glance, a good definition for this (proposed in [8] in the real case) might be

DEFINITION 1.1. The random vector Y is completely dependent on X if there exists a function H such that $Y=H(X)$.

Just as the limit of a sequence of r.v.'s with independent marginals has independent marginals, it would be desirable for the limit of completely dependent r.v.'s to be completely dependent. However, this does not happen. In fact, some examples are known (see [6]) in which, although the r.v.'s $\{(X_n, Y_n)\}_n$ have the same marginal distributions and $Y_n=H_n(X_n)$, the sequence $\{(X_n, Y_n)\}_n$ converges in distribution to a r.v. (X, Y) with independent marginals. This difficulty was overcome in the real case (in [6]) by assuming in the above definition that the maps H_n are monotone, thus giving way to the definition of *monotone dependent* pairs of real random variables.

Taking into account that, in the one-dimensional case, the o.t.p.'s coincide with the increasing maps, our Theorem 3.2 justifies the following definition of multidimensional monotone dependence.

DEFINITION 1.2. The random vector Y is monotone dependent on X if there exists an optimal transportation plan H such that $Y = H(X)$.

1.3. *Almost Sure Convergent Constructions*

In [10] the following question is studied. Let $\{P_n\}_n$ be a sequence of probability measures on $\mathfrak{R}^2$ which converges weakly. It is well known (by the Skorohod almost sure representation) that there exists a sequence of r.v.'s $\{(X_n, Y_n)\}_n$ which converges almost surely, such that for every n the distribution of (X_n, Y_n) is P_n. Now, let us assume that the first marginal distribution of P_n is constant. The question is: Would it be possible to construct the sequence $\{(X_n, Y_n)\}_n$ in such a way that $X_n = X_1$, for every $n \in \mathcal{N}$? The answer is negative and a counterexample is provided in [10].

However, our Theorem 3.4 shows that the answer is affirmative provided that the representation of the probabilities $\{P_n\}_n$ are monotone dependent.

2. NOTATION AND PRELIMINARY RESULTS

In this section we summarize, for the sake of completeness, some notation, definitions, and results which can be found mainly in [2–4, 13].

The set of probability measures, P, on the Euclidean space $\mathfrak{R}^p$ verifying $\int \|x\|^2 \, dP < \infty$ will be denoted by $\mathscr{P}_2$.

We assume throughout the paper that all the r.v.'s under consideration are $\mathfrak{R}^p$-valued and defined on the same space (Ω, σ, ν). The symbol λ_p will denote the Lebesgue measure in $\mathfrak{R}^p$, and the absolute continuity of a measure μ with respect to λ_p will be denoted by $\mu \ll \lambda_p$. As usual, we say that the set A *is of* μ-*continuity* if the topological boundary of A has μ-probability 0. Weak convergence of probabilities will be denoted by $\xrightarrow{w}$, while $\xrightarrow{\mathscr{L}}$ means convergence in distribution of r.v.'s.

It is well known that in the real case, optimal maps coincide with increasing arrangements. Increasing maps also play an important role in $\mathfrak{R}^p$ but the appropriate concept of increasing multidimensional maps to be handled must be specified. This turns out to be that of a monotone operator in the sense of Zarantarello: Let us denote by $\langle \cdot, \cdot \rangle$ the inner product in $\mathfrak{R}^p$. A mapping $H: D \subset \mathfrak{R}^p \to \mathfrak{R}^p$ is said to be *Z-increasing* when $\langle H(x) - H(x'), x - x' \rangle \geqslant 0$ holds for every x, x' in D. While that condition is not sufficient to characterize optimal maps, it was shown in [2] that it is necessary in $\mathfrak{R}^p$. On the other hand, a characterization of o.t.p.'s has been given in [11] in terms of cyclically monotone maps. These results, together with significant properties obtained in [3], are summarized in the following proposition.

PROPOSITION 2.1. *Consider P and Q in $\mathscr{P}_2$. Let (X, Y) be an o.t.p. between P and Q and assume that $P \ll \lambda_p$. Then there exists a P-probability one set D and a map $H\colon D \to \mathfrak{R}^p$ such that*

(a) *$Y\colon H(X)$, ν-a.s.*

(b) *H is Borel-measurable.*

(c) *H is cyclically monotone (hence Z-increasing) on D.*

(d) *If (X, Y_1) and (X, Y_2) are o.t.p.'s for (P, Q), then $Y_1 = Y_2$ ν-a.s.*

A well known result which relates the Wasserstein distance to the weak convergence of probability measures is the following.

PROPOSITION 2.2. *Let $\{P_n\}_n$ and P be probabilities in $\mathscr{P}_2$. The following are equivalent.*

(a) $\lim_n W(P_n, P) = 0$.

(b) *The sequence $\{P_n\}_n$ converges weakly to P and* $\lim_n \int \|x\|^2 P_n(dx) = \int \|x\|^2 P(dx)$.

The next proposition is a consequence of Theorem 1 in [11].

PROPOSITION 2.3. *Let $(X, H(X))$ be an o.t.p. between P and Q and let P^* be a probability measure in $\mathscr{P}_2$ which is absolutely continuous with respect to P. If the distribution of the random vector X^* is P^*, then $(X^*, H(X^*))$ is an o.t.p.*

The following propositions have been proved in [3]. They are related to the continuity of Z-increasing maps (hence of o.t.p.'s) and to a special kind of regularity of probabilities.

PROPOSITION 2.4. *Let $H\colon D \subset \mathfrak{R}^p \to \mathfrak{R}^p$ be a Z-increasing map and let P be a probability measure such that $P(D) = 1$ and that $P \ll \lambda_p$. Then H is P-a.e. continuous.*

Let us denote by $\text{ang}(x, y)$ the angle defined by the vectors x and y, and, given $x \in \mathfrak{R}^p$ and $\delta > 0$, let $B(x, \delta)$ be the open ball with radius δ centered at x. Also, if $x, z \in \mathfrak{R}^p$ and $\delta, \alpha > 0$, we denote

$$S(x, z, \delta, \alpha) := B(x, \delta) \cap \{y \neq x\colon \text{ang}(y - x, z) < \alpha\}.$$

Given $x \in \mathfrak{R}^p$ and a probability P, we say that *P satisfies property $\mathscr{C}$ at x* if, for every $z \in \mathfrak{R}^p$ and $\delta, \alpha > 0$, we have that $P[S(x, z, \delta, \alpha)] \neq 0$. We will say that *$P$ satisfies property $\mathscr{C}$* if

$$P\{x\colon P[S(x, z, \delta, \alpha)] > 0, \forall z \in \mathfrak{R}^p, \delta \text{ and } \alpha > 0\} = 1.$$

Even when property $\mathscr{C}$ is also satisfied by some probabilities which are not absolutely continuous with respect to λ_p (for instance, consider a probability measure giving positive probability to every point whose coordinates are rational numbers), we have:

PROPOSITION 2.5. *Let P be a probability measure on $\Re^p$ such that $P \ll \lambda_p$. Then P satisfies property $\mathscr{C}$.*

3. JOINT CONVERGENCE IN DISTRIBUTION

We begin with a technical lemma.

LEMMA 3.1. *Let $\delta > 0$, $\varepsilon \in (0, \pi/2)$, $a \in \Re^p$. There exists $h > 0$ such that if $\|u - a\| > \delta$ then*

$$B(a, h) \subset \{ y \colon ang(y - u, u - a) > \pi - \varepsilon \}.$$

Proof. Consider $u \in \Re^p$ such that $\|u - a\| > \delta$ and let $\hat{u} = a + \delta \, \|u - a\|^{-1} (u - a)$. It suffices to prove that

$$\{ y \colon \operatorname{ang}(y - \hat{u}, \hat{u} - a) > \pi - \varepsilon \} \subset \{ y \colon \operatorname{ang}(y - u, u - a) > \pi - \varepsilon \}$$

and then to take $h < \delta \sin \varepsilon$. The computations to prove these two steps are the same as those in Lemma 4.6 in [13]. ∎

As stated in the Introduction, the next theorem contains a basic requirement for a definition of monotone dependence between r.v.'s.

THEOREM 3.2. *Let $\{P_n\}_n$, $\{Q_n\}_n$, P, Q be probability measures in $\mathscr{P}_2$ such that $P \ll \lambda_p$ and $P_n \xrightarrow{w} P$, $Q_n \xrightarrow{w} Q$.*

Let $(X_n, H_n(X_n))$ be an o.t.p. between P_n and Q_n, $n \in \mathscr{N}$, and $(X, H(X))$ an o.t.p. between P and Q. Then

$$(X_n, H_n(X_n)) \xrightarrow{\mathscr{L}} (X, H(X)). \tag{2}$$

Proof. Given $k \in \mathscr{N}$, let A_k be a bounded P and Q-continuity set such that $P(A_k)$ and $Q(A_k) > 1 - k^{-1}$. Given $n \in \mathscr{N}$, let us consider the set $B_n^k \colon = H_n^{-1}(A_k) \cap A_k$ and let P_n^k be the P_n-conditional probability given the set B_n^k, and Q_n^k be the probability distribution generated by H_n from P_n^k; i.e., given $B \in \beta^p$, $Q_n^k(B) := P_n^k[H_n^{-1}(B)]$.

By standard techniques it is possible to show the existence of a subsequence $\{n_k\}_k$ such that

$$P_{n_k}(A_k) \longrightarrow 1 \qquad \text{and} \qquad Q_{n_k}(A_k) \longrightarrow 1,$$

$$P^k_{n_k} \xrightarrow{w} P \qquad \text{and} \qquad Q^k_{n_k} \xrightarrow{w} Q,$$

$$\lim_k \int_{A_k} \|x\|^2 \, P^k_{n_k}(dx) = \int \|x\|^2 \, P(dx)$$

and

$$\lim_k \int_{A_k} \|x\|^2 \, Q^k_{n_k}(dx) = \int \|x\|^2 \, Q(dx).$$

From here, by applying a well known property of weak convergence and taking into account that the support of $P^k_{n_k}$ is contained in A_k, we have that

$$\int \|x\|^2 \, P(dx) \leqslant \liminf_k \int \|x\|^2 \, P^k_{n_k}(dx) \leqslant \limsup_k \int \|x\|^2 \, P^k_{n_k}(dx)$$

$$\leqslant \limsup_k \frac{1}{P_{n_k}(B^k_{n_k})} \int_{A_k} \|x\|^2 \, P_{n_k}(dx) = \int \|x\|^2 \, P(dx).$$

The same relation holds for $\{Q^k_{n_k}\}_k$ and Q. Therefore, as a consequence of Proposition 2.2 and the triangular inequality for the Wasserstein distance, we have that

$$\lim_k W(P^k_{n_k}, Q^k_{n_k}) = W(P, Q). \tag{3}$$

Moreover, given $k \in \mathcal{N}$ let Z_k be any r.v. with distribution $P^k_{n_k}$. From Proposition 2.3 we have that $(Z_k, H_{n_k}(Z_k))$ is an o.t.p. between the probabilities $P^k_{n_k}$ and $Q^k_{n_k}$.

The sequence of probabilities associated with $(Z_k, H_{n_k}(Z_k))$ is tight because both sequences of marginal distributions are tight. Then there exists a new subsequence and a probability measure μ such that $\mathcal{L}(Z_{j_k}, H_{n_{j_k}}(Z_{j_k})) \xrightarrow{w} \mu$, and, from (3), we have that

$$\int \|x - y\|^2 \, \mu(dx, dy) \leqslant \liminf_k \int \|Z_{j_k} - H_{n_{j_k}}(Z_{j_k})\|^2 \, dv$$

$$= \liminf_k W^2(P^{j_k}_{n_{j_k}}, Q^{j_k}_{n_{j_k}}) = W^2(P, Q).$$

Now, since P and Q are the marginals of μ and the optimality of the pair $(X, H(X))$ (essentially unique from Proposition 2.1), we obtain that μ coincides with the law of $(X, H(X))$, and hence the whole sequence $\{(Z_k, H_{n_k}(Z_k))\}_k$ converges in law to $(X, H(X))$.

Let $\{Y_{n_k}\}_k$ be a sequence of r.v.'s which are independent of those in the sequence $\{X_{n_k}\}_k$ and such that the distribution of Y_{n_k} is $P^k_{n_k}$, $k \in \mathcal{N}$. We can assume that

$$Z_{n_k} = \begin{cases} X_{n_k}, & \text{if} \quad X_{n_k} \in B^k_{n_k}, \\ Y_{n_k}, & \text{if} \quad X_{n_k} \notin B^k_{n_k}. \end{cases}$$

Therefore

$$\nu\{\|((X_{n_k}, H_{n_k}(X_{n_k})) - (Z_{n_k}, H_{n_k}(Z_{n_k})))\| \geqslant \varepsilon\} \leqslant \nu\{Z_{n_k} \neq X_{n_k}\} \leqslant \nu\{X_{n_k} \notin B^k_{n_k}\}$$

which converges to zero and the subsequence $\{(X_{n_k}, H_{n_k}(X_{n_k}))\}_k$ converges in law to $(X, H(X))$.

Finally, if we apply the preceding argument not to the original sequence but to any of its subsequences, we have that every subsequence of $\{(X_n, H^n(X_n)\}_n$ contains a new subsequence satisfying (2); hence, this relation is proved. ∎

COROLLARY 3.3. *Under the hypotheses in the previous theorem, if we also assume that the sequence $\{X_n\}_n$ converges a.s. and if C is an open, Q-continuity set, then*

$$\lim_n \nu\{H(X) \in C, H_n(X_n) \notin C\} = 0.$$

Proof. The hypothesis on the sequence $\{X_n\}_n$ and Theorem 3.2 imply that $(X, H_n(X_n)) \xrightarrow{\mathcal{L}} (X, H(X))$. Hence by the P-a.s. continuity of H we have

$$(H(X), H_n(X_n)) \xrightarrow{\mathcal{L}} (H(X), H(X)). \quad ∎$$

THEOREM 3.4. *Let $\{P_n\}_n$, $\{Q_n\}_n$, P, Q be probability measures in $\mathcal{P}_2$ such that $P \ll \lambda_p$ and such that $P_n \xrightarrow{w} P$ and $Q_n \xrightarrow{w} Q$. Let us assume that H_n (resp. H) are o.t.p.'s between P_n and Q_n (resp. P and Q), $n \in \mathcal{N}$.*

Then, if $\{X_n\}_n$ is a sequence of r.v.'s which converges a.s. and $\mathcal{L}(X_n) = P_n$, we have that

$$H_n(X_n) \to H(X) \qquad \nu\text{-a.s.}$$

Proof. Let A be the set in which H is continuous and P satisfies property $\mathscr{C}$. Let

$$\Omega_0 := X^{-1}(A) \cap \{\omega : \lim_n X_n(\omega) = X(\omega)\}.$$

Let $\omega_0 \in \Omega_0$. Let us denote $x_0 = X(\omega_0)$, $x_n = X_n(\omega_0)$, $n \in \mathscr{N}$, and let us assume that the sequence $\{H_n(x_n)\}_n$ does not converge to $H(x_0)$. In this case there exists $\delta > 0$, $z \in \mathfrak{R}^p$, and a subsequence $\{H_{n_k}(x_{n_k})\}_k$ satisfying that $\|H_{n_k}(x_{n_k}) - H(x_0)\| > \delta$, for every k, and that

$$\lim_k \frac{H_{n_k}(x_{n_k}) - H(x_0)}{\|H_{n_k}(x_{n_k}) - H(x_0)\|} = z,$$

and then we can also assume that

$$\operatorname{ang}(H_{n_k}(x_{n_k}) - H(x_0), z) < \eta,$$

for every k, for some prefixed η (to be determined later).

Let $\varepsilon > 0$. According to Lemma 3.1 there exists $h > 0$ such that the open ball $B(H(x_0), h)$ is of Q-continuity and

$$B(H(x_0), h) \subset \{y : \operatorname{ang}(y - H_{n_k}(x_{n_k}), H_{n_k}(x_{n_k}) - H(x_0)) > \pi - \varepsilon\}. \tag{4}$$

Since H is continuous on A, there exists $\gamma > 0$ such that

$$A \cap B(x_0, \gamma) \subset A \cap H^{-1}[B(H(x_0), h)]. \tag{5}$$

Let $\theta > 0$ be such that $\pi/2 + \theta < \pi - \varepsilon$. Since $x_0 \in A$ we have that $P[S(x_0, z, \gamma, \theta)] > 0$ and there exists an open ball B_0 such that $B_0 \subset S(x_0, z, \gamma, \theta)$ and $P(B_0) > 0$. Let $\omega \in \Omega_0$ such that $X(\omega) \in B_0$. Taking into account that $\lim_k x_{n_k} = x_0$ and that $\lim_k X_{n_k}(\omega) = X(\omega)$ we have that there exists k_0 $(= k_0(\omega))$ such that, if $k \geqslant k_0$, then

$$\operatorname{ang}(X_{n_k}(\omega) - x_{n_k}, z) < \theta.$$

Therefore, because of the increasing character of the maps $\{H_n\}_n$, if we fix $k \geqslant k_0$ and $\eta > 0$ such that $\pi/2 + \theta + \eta < \pi - \varepsilon$, we obtain that

$$\begin{aligned} &\operatorname{ang}(H_{n_k}[X_{n_k}(\omega)] - H_{n_k}(x_{n_k}), H_{n_k}(x_{n_k}) - H(x_0)) \\ &\quad \leqslant \operatorname{ang}(H_{n_k}[X_{n_k}(\omega)] - H_{n_k}(x_{n_k}), z) + \operatorname{ang}(z, H_{n_k}(x_{n_k}) - H(x_0)) \\ &\quad < \theta + \frac{\pi}{2} + \eta < \pi - \varepsilon, \end{aligned} \tag{6}$$

thus, from (4), we have that $H_{n_k}[X_{n_k}(\omega)] \notin B(H(x_0), h)$. Moreover, by construction of B_0 and (5), we have that $H(B_0) \subset B(H(x_0), h)$, P-a.s., and then we have that

$$\liminf_k \nu\{H(X) \in B(H(x_0), h); H_{n_k}(X_{n_k}) \notin B(H(x_0), h)\} \geqslant \nu\{X \in B_0\} > 0,$$

and the proof ends because this contradicts Corollary 3.3. ∎

4. CONSISTENCY OF REPRESENTATIONS

Let P, Q be probability measures in $\mathscr{P}_2$ such that $P \ll \lambda_p$ and suppose that H is an o.t.p. between P and Q. Also let P_n^ω (resp. Q_n^ω), $\omega \in \Omega$, be the random sample distribution associated to n independent and identically distributed random variables $U_1, U_2, ..., U_n$ (resp. $V_1, V_2, ..., V_n$), defined on (Ω, σ, ν), with probability law P (resp. Q). Obviously, there exists H_n^ω, which is an o.t.p. between P_n^ω and Q_n^ω.

Note that H_n^ω is defined on the finite support set of P_n^ω, $\mathrm{Supp}(P_n^\omega)$, but for almost every ω the set $S^\omega = \bigcup_{n=1}^\infty \mathrm{Supp}(P_n^\omega)$ is dense on the support of P, $\mathrm{Supp}(P)$. Therefore, our goal is to show that for every ω in a ν-probability one set, if $\{x_k\}_k$ is a sequence in S^ω such that $\lim_k x_k = x \in \mathrm{Supp}(P)$ and $x_k \in \mathrm{Supp}(P_{n_k}^\omega)$, for every k, where $\lim_k n_k = \infty$, then $\lim_k H_{n_k}(x_k) = H(x)$.

By denoting by X_n^ω and X any r.v.'s (defined on some uninteresting probability space (T, α, τ)) such that $\mathscr{L}(X_n^\omega) = P_n^\omega$ and $\mathscr{L}(X) = P$, the Glivenko–Cantelli theorem applied to the sequences P_n^ω and Q_n^ω allows an easy linkage with the results of the preceding section to obtain that

$$(X_n^\omega, H_n^\omega(X_n^\omega)) \xrightarrow{\mathscr{L}} (X, H(X)), \qquad \text{for } (\nu\text{-})\text{a.e. } \omega,$$

and the following specialization of Corollary 3.3. However, note that in both cases direct simpler proofs (by making use of the Strong Law of Large Numbers) are also possible.

LEMMA 4.1. *There exists $\Omega_0 \in \sigma$ such that $\nu(\Omega_0) = 1$ and if $\omega \in \Omega_0$ and C is an open Q-continuity set, then*

$$\lim_n P_n^\omega[H^{-1}(C) \cap H_n^{\omega^{-1}}(C^c)] = 0.$$

On the other hand, since $\mathrm{Supp}(P_n^\omega) \subset \mathrm{Supp}(P)$ a.s., we can improve slightly the proof of Theorem 3.4 to obtain a result in which no reference to the artificial space (Ω, σ, ν) is needed. Its proof is very similar to that of Theorem 3.4 and we only include some hints about it. In its statement, if H is an o.t.p. between P and Q, we denote by $\mathscr{A}$ the set in which H is continuous and P satisfies property $\mathscr{C}$.

THEOREM 4.2. *Let P, Q be probability measures in $\mathscr{P}_2$ such that $P \ll \lambda_p$ and suppose that H is an o.t.p. between P and Q. Let P_n^ω (resp. Q_n^ω), $\omega \in \Omega$, be the sample distribution associated with n independent and identically distributed r.v.'s, with probability law P (resp; Q), and let H_n^ω be an o.t.p. between P_n^ω and Q_n^ω.*

Then there exists $\Omega_0 \in \sigma$ such that $\nu(\Omega_0) = 1$ and, for every $\omega \in \Omega_0$, if $\{x_k\}_k$ is a sequence in $S^\omega = \bigcup_{n=1}^\infty \operatorname{Supp}(P_n^\omega)$ such that $x_k \in \operatorname{Supp}(P_{nk}^\omega)$, for every k, where $\lim_k n_k = \infty$ and $\lim_k x_k = x_0 \in \mathscr{A}$, then

$$\lim_k H_{n_k}(x_k) = H(x).$$

Proof. Let Ω_0 be the ν-probability one set in which $\operatorname{Supp}(P_n^\omega) \subset \mathscr{A}$ and, the Glivenko–Cantelli theorem and Lemma 4.1 are verified. Let $\omega \in \Omega_0$, and assume, on the contrary, that there exists $\delta > 0$ and a sequence $\{n_k\}_{n_k}$, $n_k \to \infty$, such that $x_{n_k} \in \operatorname{Supp}(P_{nk}^\omega)$, $x_{n_k} \to x_0 \in \mathscr{A}$, and $\|H_{nk}^\omega(x_{n_k}) - H(x)\| > \delta$ for every k.

Let $\varepsilon > 0$. As in the proof of Theorem 3.4, we obtain the existence of $h, \gamma, \theta > 0$ and $z \in \Re^p$ such that the ball $B(H(x_0), h)$ is of Q-continuity, (4) and (5) are satisfied, and there exists $B_0 \subset S(x_0, z, \gamma, \theta)$ with $P(B_0) > 0$.

Moreover, B_0 is bounded and $\inf\{\|y - x_0\| : y \in B_0\} > 0$. Therefore, taking into account that $\lim_k x_{n_k} = x_0$, we obtain that there exists k_0 such that if $k \geqslant k_0$ then

$$\sup\{\operatorname{ang}(y - x_{n_k}, z) : y \in B_0\} < \theta.$$

Now, if $y \in B_0$ and $k \geqslant k_0$, in the same way as in (6) we obtain that

$$\operatorname{ang}(H_{n_k}(y) - H_{n_k}(x_{n_k}), H_{n_k}(x_{n_k}) - H(x_0)) < \pi - \varepsilon,$$

and, in consequence, $H_{n_k}(y) \notin B(H(x_0), h)$. Therefore we have that

$$\liminf_k P_{nk}^\infty[H^{-1}(B(H(x), h)) \cap H_{nk}^{\omega^{-1}}(B^c(H(x), h))]$$

$$\geqslant \lim_k P_{nk}^\infty[\mathscr{A} \cap B_0] = \lim_k P_{nk}^\omega[B_0] \geqslant P[B_0] > 0,$$

which contradicts Lemma 4.1. ∎

ACKNOWLEDGMENTS

We thank an editor whose comments have considerably improved the paper.

Note added in Proof. During the proof stage the authors noted the appearance of two papers [14, 15] which treat problems related to the one in this paper. In [14] a result similar to Theorem 3.4 is proved, but one without the integrability hypotheses but keeping probabilities $\{P_n\}$ fixed (i.e., $P_n = P$ for every $n \in \mathcal{N}$); thus this result does not suffice to cover applications 1.1 and 1.2 in the Introduction.

In [15] some results related to the existence and uniqueness of the o.t.p.'s are obtained. As a preliminary step it is also shown that the weak limit of o.t.p.'s is an o.t.p. Nothing related to a.s. convergence appears in that paper.

REFERENCES

[1] Aurenhammer, F., Hoffmann, F., and Aronov, B. (1992). *Minkowski-type Theorems and Least-Squares Partitioning*, Report B-92-09^T. Institute for Computer Science, Freie Universität Berlin, Berlin.

[2] Cuesta, J. A., and Matran, C. (1989). Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.* **17** 1264–1276.

[3] Cuesta, J. A., Matran, C., and Tuero-Diaz, A. (1993). Properties of optimal maps for the L^2-Wasserstein distance, unpublished, manuscript.

[4] Cuesta, J. A., Rüschenforf, L., and Tuero-Diaz, A. (1993). Optimal coupling of multivariate distributions and stochastic processes. *J. Multivariate Anal.* **46** 355–361.

[5] Easton, G. S., and McCullogh, R. E. (1990). A multivariate generalization of quantile-quantile plots. *J. Amer. Stat. Assoc.* **85** (410), 376–386.

[6] Kimeldorf, G., and Sampson, A. (1978). Monotone dependence. *Ann. Statist.* **6** 895–903.

[7] Knott, M., and Smith, C. S. (1984). On the optimal mapping of distributions. *J. Optim. Theory Appl.* **52** 323–329.

[8] Lancaster, H. O. (1963). Correlation and complete dependence of random variables. *Ann. Math. Statist.* **34** 1315–1321.

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

[10] Rachev, S. T., and Rüschendorf, L. (1990). A counterexample to a.s. constructions. *Statist. Probab. Lett.* **9** 307–309.

[11] Rüschendorf, L., and Rachev, S. T. (1990). A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.* **32** 48–54.

[12] Sampson, A. R. (1993). Some applications of monotone transformations in statistics. In *Stochastic Inequalities* (M. Shaked, and Y. L. Tong, Eds.), IMS Lecture Notes, Vol. 22, pp. 359–370. Institute of Mathematical Statistics, Hayward, CA.

[13] Tuero, A. (1993). On the stochastic convergence of representations based on Wasserstein metrics, *Ann. Probab.* **21** 72–85.

[14] Heinich, H., and Lootgieter, J. C. (1996). Convergence of monotone functions. *C.R. Acad. Sci. Paris Ser. I* **322** 869–874.

[15] McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** 309–323.