The Annals of Probability
1993, Vol. 21, No. 1, 72–85

ON THE STOCHASTIC CONVERGENCE OF REPRESENTATIONS BASED ON WASSERSTEIN METRICS[1]

BY ARACELI TUERO

Universidad de Cantabria, Spain

Suppose that P and P_n, $n \in \mathscr{N}$, are probabilities on a real, separable Hilbert space, V. It is known that if P satisfies some regularity conditions and X is such that $P_X = P$, then there exist mappings $H_n: V \to V$, such that $P_{H_n(X)} = P_n$ and the Wasserstein distance between P_n and P coincides with $(\int \|x - H_n(x)\|^2 \, dP)^{1/2}$, $n \in \mathscr{N}$. In this paper we prove that the weak convergence of $\{P_n\}$ to P is enough to ensure that $\{H_n(X)\}$ converges to X in measure, and that, if $V = \Re^p$, then the convergence is also a.e. This property seems to be characteristic of finite-dimensional spaces, because we include an example, with V infinite-dimensional and P Gaussian, where a.e. convergence does not hold.

1. Introduction. Wasserstein metrics can be defined in the following way:

Let V be a Polish space and β its Borel σ-algebra. If $\sigma: V \times V \to \Re$ is a measurable, nonnegative map and P and Q are probabilities defined on β, then the Wasserstein distance between P and Q with respect to σ is related, through a suitable increasing function, to

$$\sigma(P,Q) = \inf\left\{\int \sigma \, d\lambda; \lambda \in M(P,Q)\right\}, \tag{1}$$

where $M(P,Q)$ is the set of all probability measures defined on $\beta \times \beta$ whose marginal distributions are, respectively, P and Q.

These metrics are related to the so-called Kantorovich–Rubinstein problem and were first developed by Kantorovich [9]. We use the terminology Wasserstein metrics as do most articles on the subject. Good surveys are [14] and [15].

Given two random variables (r.v.) X and Y defined on the probability space (Ω, α, μ), we say that (X,Y) is an optimal coupling (o.c.) between P and Q with respect to σ, if the marginal distributions of (X,Y) are P and Q, respectively, and

$$\sigma(P,Q) = \int \sigma(X,Y) \, d\mu.$$

Received November 1990; revised July 1991.

[1]Research partially supported by the Spanish DGICYT under Grants PB87-0905-C02-00, 01 and 02.

AMS 1991 *subject classifications*. Primary 60E05, 60B10.

Key words and phrases. Wasserstein distance, stochastic convergence of representations, Skorohod's representation theorem, Hilbert spaces, weak convergence, increasing functions.

We often use the simpler expressions "(X, Y) is an o.c. with respect to σ" or "(X, Y) is an o.c."

If V is a separable Banach space with norm $\| - \|$, $\sigma(x, y) = \|x - y\|^r$, $1 \leq r < \infty$, and P and Q are two probabilities on β that verify $\int \|x\|^r \, dP < \infty$ and $\int \|x\|^r \, dQ < \infty$, then the o.c. between P and Q always exists; see [1]. However, explicit representations of the o.c. are only known when $V = \mathfrak{R}$; see, for instance, [10, 20, 12]. If $V = \mathfrak{R}^p$, $p \geq 2$, only some partial answers are known even if we restrict our attention to $\sigma(x, y) = \|x - y\|^2$.

Rüschendorf and Rachev give in [17] a necessary and sufficient condition for a given pair be an o.c. This condition allows them to find an o.c. in some situations. Knott and Smith [11] show that if ϕ is a regular invertible function, then $(X, \phi(X))$ is an o.c. between P and $P\phi^{-1}$. This includes the case in which P and Q are normal. In this situation, o.c.'s have also been proposed in [5, 13, 8]. Rüschendorf [16] solves a problem quite similar to (1). Finally, when V is a Hilbert space and $\sigma(x, y) = \|x - y\|^2$, it is proved in [3] that if P satisfies a continuity property, then there exists an o.c. between P and Q of the form $(X, H(X))$, where H is increasing in the Zarantarello sense (see definition and Proposition 2.3 below).

On the other hand, if $V = \mathfrak{R}$, the result in [12] mentioned above is equivalent to the following proposition:

Let P and Q be probabilities on β, with P continuous. Let H be an increasing function and let X be a r.v. such that the distribution of X is P and that of $H(X)$ is Q. Then $(X, H(X))$ is an o.c. between P and Q.

The required increasing function always exists (take $H = G^{-1} \circ F$, where F and G^{-1} are the distribution function of P and the quantile function of Q, respectively).

This construction is related to the following Skorohod representation theorem:

"If $V = \mathfrak{R}$ and $\{P_n\}$ are probabilities on β that converge weakly to the probability P ($P_n \to_w P$), P is continuous, X is a r.v. with distribution P and H_n, $n \in \mathcal{N}$, are increasing functions such that the distribution of $H_n(X)$ is P_n, then $H_n(X)$ converges almost everywhere (a.e.) to X ($H_n(X) \to_{\text{a.s.}} X$)."

Note that the key to both results is the increasing character of H. This and the result in [3] suggest the following question:

Let V be a Hilbert space and $(X, H_n(X))$ an o.c. between P and P_n, $n \in \mathcal{N}$, with respect to $\sigma(x, y) = \|x - y\|^2$. Let us suppose that $P_n \to_w P$. Is it then true that $H_n(X) \to_{\text{a.s.}} X$?

Surprisingly the answer is yes, without any additional hypothesis if $V = \mathfrak{R}^p$ (Theorem 3.1), but Theorem 2.6 and Example 3.4 show that the convergence is only in probability when V is an infinite-dimensional Hilbert space, even if P is a Gaussian distribution.

Finally, let us mention that some generalizations of Skorohod's representation theorem are available: The theorem was proved by Skorohod in [19] for Polish spaces; more general versions have been given in [6, 21, 18]. The existence of simultaneous representations has been obtained in [2] and independently in [7].

2. Convergence in measure. First we present the notation to be employed. From now on we denote by V a real, separable Hilbert space, by $\langle -, - \rangle$, $\| - \|$ and β its inner product, norm and Borel σ-algebra, respectively.

I_d is the identity and $X, Y, \ldots$ (with or without subscripts or superscripts) represent V-valued r.v.'s defined on the same probability space, (Ω, α, μ), and such that $\int \|X\|^2 \, d\mu$, $\int \|Y\|^2 \, d\mu, \ldots$ are finite. $P_X, P_Y, \ldots$ denote their probability distributions.

$\mathscr{L}[v, w, \ldots]$ is the linear subspace spanned by $v, w, \ldots$ and $\mathscr{L}[v, w, \ldots]^\perp$ is the corresponding orthogonal subspace. The angle between the vectors v and w is denoted by $\widehat{v, w}$ or by $\operatorname{ang}(v, w)$.

Let $w \in V$, $w \neq 0$, $v \in \mathscr{L}[w]^\perp$. Let P be a probability on β. We represent by P_w and $P_w(-/v)$ the marginal probability on $\mathscr{L}[w]$ and a regular conditional probability on $\mathscr{L}[w]$ given v, respectively.

Given $t \in V$ and $r > 0$; $B(t, r)$, $\overline{B}(t, r)$ and $S(t, r)$ are, respectively, the open ball, the closed ball and the sphere of center x and radius r. If A is a set in V, then A^0, $\overline{A}$ and $\nu(A)$ denote its topological interior, closure and boundary, respectively.

We only use (1) in the case $\sigma(x, y) = \|x - y\|^2$. Therefore, given two probability measures P, Q defined on β such that $\int \|x\|^2 \, dP < \infty$ and $\int \|x\|^2 \, dQ < \infty$, the Wasserstein distance between them is $(W(P, Q))^{1/2}$, where

$$W(P, Q) = \inf\left\{ \int \|X - Y\|^2 \, d\mu, P_X = P \text{ and } P_Y = Q \right\}$$

and we say that the pair (X, Y) is an o.c. between P and Q if $P_X = P$, $P_Y = Q$ and $W(P, Q) = \int \|X - Y\|^2 \, d\mu$.

Finally, given a function H: $D \subset V \to V$, we say it is *increasing* if it satisfies $\langle H(x) - H(y), x - y \rangle \geq 0$ for every x, y in D.

We use the following condition to solve some measurability problems (see Proposition 2.2).

Let P be a probability measure on β. We say that P satisfies $\mathscr{C}$ if there exists a dense sequence $\{w_n\}$ in $S(0, 1)$ that contains a complete orthonormal system and for every n and almost everywhere w in $\mathscr{L}[w_n]^\perp$, $P_{w_n}(-/w)$ is atomless.

$\mathscr{C}$ is related to the continuity of P and holds in many common situations. For instance, it is easy to prove that P satisfies $\mathscr{C}$ if $V = \mathfrak{R}^p$ and P is absolutely continuous with respect to the Lebesgue measure. We will see next that if P is a nondegenerate Gaussian measure in an infinite-dimensional Hilbert space, then it satisfies $\mathscr{C}$.

PROPOSITION 2.1. *Let P be a Gaussian distribution on β and let S be its covariance operator. If $V = \mathscr{L}[v \in V\colon Sv = \lambda_v v; \lambda_v \neq 0]$, then P satisfies $\mathscr{C}$.*

PROOF. Let $\{e_n\}$ be a complete orthonormal system in V of eigenvectors of S. We will show that it suffices to take as $\{w_n\}$ the set of the vectors in $S(0, 1)$ which are finite linear combinations with rational coefficients of vectors in $\{e_n\}$.

Let $w = \sum_{i=1}^{n} \alpha_i e_i$ with α_i rational, $i = 1, 2, \dots n$, and let $v_1 = w, v_2, \dots, v_n$ be an orthonormal basis of $\mathscr{L}[e_1, \dots, e_n]$. Let π_i be the projection from V on $\mathscr{L}[v_i]$ for $1 \leq i \leq n$ and from V on $\mathscr{L}[e_i]$ for $i > n$, respectively. Let A be a set in the Borel σ-algebra in $\Re$. Taking into account that

$$E\left[I_{\{\pi_1 \in A\}}/\sigma(\pi_2, \dots, \pi_n, \pi_{n+1}, \dots)\right] = E\left[I_{\{\pi_1 \in A\}}/\sigma(\pi_2, \dots, \pi_n)\right],$$

we obtain that a version of the regular conditional probability of A, given $u \in \mathscr{L}[v_1]^{\perp}$, coincides with a version of the regular conditional probability of A given π^*, where π^* is the projection on $\mathscr{L}[v_2, \dots, v_n]$ which is a nondegenerate normal. □

With the next proposition, we solve some measurability problems. Its proof is given in the Appendix.

Proposition 2.2. *Let P be a probability measure satisfying $\mathscr{C}$. If $D \in \beta$ and H: $D \to V$ is an increasing mapping, then H is measurable in the P-completed σ-algebra.*

$\mathscr{C}$ implies the hypotheses of Theorem 2.8 in [3], which is stated here for further reference.

Proposition 2.3. *Let P, Q be two probabilities on β such that $\int \|x\|^2 \, dP$ and $\int \|x\|^2 \, dQ$ are finite and suppose that P satisfies $\mathscr{C}$. Let X be a r.v. such that $P_X = P$. Then there exists a mapping H: $D \subset V \to V$ such that:*

(i) *For every x, x' in D: $\langle H(x) - H(x'), x - x' \rangle \geq 0$.*
(ii) $P_{H(X)} = Q$.
(iii) $\int \|X - H(X)\|^2 \, d\mu = W(P, Q)$.

Remark. Note that D in the previous proposition is a P-probability 1 set, so often D will not appear in the notation.

On the other hand, after Propositions 2.2 and 2.3, whenever we have a probability P which satisfies $\mathscr{C}$ and we need to choose an o.c. between P and another probability measure, we will always take it as $(X, H(X))$, where X is a r.v. such that $P_X = P$ and H is increasing in a P-probability 1 set.

We will show next that the o.c.'s give a Skorohod representation for the convergence in distribution if we just want to get convergence in measure. We start with the following proposition, which is a trivial consequence of Theorem 2 in [17], and continue with Lemma 2.5, which is a trivial consequence of Lemma 8.3 in [1].

Proposition 2.4. *Let P, Q be two probability measures on β such that P satisfies $\mathscr{C}$. Suppose that H: $D \subset V \to V$ is such that $(X, H(X))$ is an o.c. between P and Q. Let $A \in \beta$ such that $P(A) > 0$. Let P_1 be the P-conditional probability measure given A. Let Y be a r.v. such that $P_Y = P_1$ and let $P_2 = P_{H(Y)}$.*

Then $(Y, H(Y))$ *is an o.c. between* P_1 *and* P_2 *and*

$$W(P_1, P_2) = \frac{1}{P(A)} \int_A \|x - H(x)\|^2 \, dP.$$

LEMMA 2.5. *Let* P *be a probability measure defined on* β *such that* $\int \|x\|^2 \, dP$ *is finite. Let* $\{B_n\}$ *be a sequence of measurable sets such that* $\{P(B_n)\}$ *converges to* 1. *Let* Q_n *be the* P*-conditional probability measure given* B_n, $n \in \mathcal{N}$. *Then*

$$\lim_n W(Q_n, P) = 0.$$

The next theorem is proved by using a truncation technique for the construction of two sequences of probability measures such that their Wasserstein distances converge to 0.

THEOREM 2.6. *Let* $\{P_n\}$, P *be probability measures on* β *such that* P *satisfies* $\mathscr{C}$, $\int \|x\|^2 \, dP < \infty$, $\int \|x\|^2 \, dP_n < \infty$, $n \in \mathcal{N}$, *and* $P_n \to_w P$. *Let* X *be a r.v. with* $P_X = P$ *and* H_n: $D_n \subset V \to V$, $n \in \mathcal{N}$, *such that* $(X, H_n(X))$ *is an o.c. between* P *and* P_n, $n \in \mathcal{N}$.
Then $H_n \to_{c.p.} I_d$.

PROOF. Let $\{A_r\}_r$ be a sequence of bounded, measurable, P-continuity sets, such that $P(A_r) > 1 - (1/r)$, $r \in \mathcal{N}$. Let $r, n \in \mathcal{N}$ and consider the probability measures defined by

$$P^1_{n,r}(B) = \frac{1}{P[H_n^{-1}(A_r)]} P[H_n^{-1}(A_r) \cap B],$$

$$P^2_{n,r}(B) = P^1_{n,r}[H_n^{-1}(B)],$$

$$Q_r(B) = \frac{1}{P(A_r)} P[A_r \cap B].$$

First we prove the existence of a subsequence, $\{W(P^1_{n_r,r}, P^2_{n_r,r})\}_r$, which converges to 0.

By the triangular inequality, we have

$$\begin{aligned} (2) \qquad & \left(W(P^1_{n,r}, P^2_{n,r})\right)^{1/2} \\ & \leq \left(W(P^1_{n,r}, P)\right)^{1/2} + \left(W(P, Q_r)\right)^{1/2} + \left(W(Q_r, P^2_{n,r})\right)^{1/2}. \end{aligned}$$

Let $r \in \mathcal{N}$ be fixed and let $B \in \beta$ such that $Q_r(\nu(B)) = 0$. Then, $A_r \cap B$ is a P-continuity set and Lemma 8.3 in [1] implies that, for every $r \in \mathcal{N}$, $W(Q_r, P^2_{n,r}) \to_{n \to \infty} 0$.

On the other hand, varying r, we obtain the existence of a subsequence $\{n_r\}_r$ such that

$$(3) \qquad P\left(H_{n_r}^{-1}(A_r)\right) \to_{r \to \infty} 1.$$

Then, by Lemma 2.5 and (2),

$$W\left(P^1_{n_r,r}, P^2_{n_r,r}\right) \to_{r\to\infty} 0. \tag{4}$$

Now note that if X_{n_r} is such that $P_{X_{n_r}} = P^1_{n_r,r}$, $r \in \mathscr{N}$, then $P_{H_n(X_{n_r})} = P^2_{n_r,r}$, so by Proposition 2.4,

$$W\left(P^1_{n_r,r}, P^2_{n_r,r}\right) = \frac{1}{P\left(H^{-1}_{n_r}(A_r)\right)} \int_{H^{-1}_{n_r,r}(A_r)} \left\| x - H_{n_r,r}(x) \right\|^2 dP,$$

and from (4):

$$I_{H^{-1}_{n_r}(A_r)}\left(I_d - H_{n_r}\right) \underset{n\to\infty}{\overset{\text{c.p.}}{\to}} 0,$$

where I_C denotes the indicator function of the set C, and then

$$H_{n_r} \underset{n\to\infty}{\overset{\text{c.p.}}{\to}} I_d.$$

Therefore we have proved that every subsequence of $\{H_n\}$ possesses a new subsequence, $\{H_{n_k}\}$, such that $H_{n_k} \to^{\text{c.p.}}_{n\to\infty} I_d$, and, consequently, $H_n \to^{\text{c.p.}} I_d$. □

3. Almost everywhere convergence. In the preceding section we have needed the fact that $\{H_n\}$ are increasing only to prove their measurability. The increasing character of $\{H_n\}$ is essential in this section because the main result (Theorem 3.2) relies on Theorem 3.1 which is, essentially, a property of increasing functions.

THEOREM 3.1. *Let $V = \Re^p$, $p \in \mathscr{N}$. Let $\{P_n\}$ and P be probability measures on β such that P is atomless, P satisfies $\mathscr{E}$, $P(S(P)^0) = 1$, where S denotes support, $\int \|x\|^2 \, dP < \infty$, $\int \|x\|^2 \, dP_n < \infty$, $n \in \mathscr{N}$, and $P_n \to_w P$. Let X be a r.v. with $P_X = P$ and H_n: $D_n \subset V \to V$, $n \in \mathscr{N}$, be increasing functions such that $(X, H_n(X))$ is an o.c. between P_n and P, $n \in \mathscr{N}$.*
Then

$$H_n \to_{a.e.} I_d.$$

PROOF. Note that it suffices to prove that $\{H_n(x)\}$ converges to x for every x in $S(P)^0 \cap \liminf D_n$.

Therefore, let $x \in S(P)^0 \cap \liminf D_n$ and $n_x \in \mathscr{N}$ such that $x \in D_n$, $\forall$ $n \geq n_x$. Let $n \geq n_x$. We use the following notation:

$$\varepsilon_n = \| H_n(x) - x \|,$$

$$B_n = \begin{cases} B\left(x, \left(\varepsilon_n/2\sqrt{2}\right)\right) \cap \left\{y \neq x: \operatorname{ang}\left(H_n(x) - x, y - x\right) \leq \pi/4\right\} \cap D, \\ \qquad \text{if } \varepsilon_n > 0, \\ \varnothing, \quad \text{if } \varepsilon_n = 0. \end{cases}$$

Let $y \in B_n$; then $y \neq x$, and, taking into account that H_n is increasing,

$$\| H_n(y) - x \| \geq \frac{\langle H_n(y) - x, y - x \rangle}{\|y - x\|} \geq \frac{\varepsilon_n \sqrt{2}}{2}.$$

Hence

$$\frac{\varepsilon_n}{\sqrt{2}} \leq \| H_n(y) - x \| \leq \| H_n(y) - y \| + \|y - x\| \leq \| H_n(y) - y \| + \frac{\varepsilon_n}{2\sqrt{2}}.$$

Thus

$$\| H_n(y) - y \| \geq \frac{\varepsilon_n}{2\sqrt{2}} \geq \|y - x\|,$$

which shows that $B_n \subset V_n := \{y \in D_n : \|H_n(y) - y\| \geq \|y - x\|\}$.

Let $\varepsilon > 0$. Then

$$\begin{aligned} 0 &\leq \limsup P(V_n) \\ &\leq \limsup\big(P\{y : \| H_n(y) - y \| \geq \varepsilon\} + P[B(x, \varepsilon)]\big) = P[B(x, \varepsilon)] \end{aligned}$$

since $H_n \to_{\text{c.p.}} I_d$ by Theorem 2.6. Now letting ε tend to zero, we have that $\lim P(B_n) = 0$, because P has no atoms.

Let us next suppose that $(\varepsilon_n\}$ does not converge to zero. In this case there exists $\varepsilon > 0$ and a subsequence $\{\varepsilon_{n_k}\}$ such that $\varepsilon_{n_k} > \varepsilon$, $\forall\, k$.

Let y_{n_k} be the projections of $H_{n_k}(x)$ on $S(x, (\varepsilon/2\sqrt{2}))$. Then there exists a subsequence (which we still denote $\{y_{n_k}\}$) which converges to x_0 in $S(x, (\varepsilon/2\sqrt{2}))$.

Let $A = B(x, (\varepsilon/2\sqrt{2})) \cap \{y \neq x : \operatorname{ang}(x_0 - x, y - x) < \pi/4\}$; then $P(A) > 0$ because $x \in S(P)^0$ and

$$A \subset \liminf\left[B\left(x, \frac{\varepsilon}{2\sqrt{2}}\right) \cap \left\{y \neq x : \operatorname{ang}\big(H_{n_k}(x) - x, y - x\big) < \frac{\pi}{4}\right\}\right],$$

which contradicts that $\{P(B_n)\}$ converges to zero.

So, $\{\varepsilon_n\}$ converges to zero and the theorem is proved. □

We now summarize Proposition 2.3 and the preceding theorem in the following result which shows that the o.c.'s, besides their importance from the Wasserstein metric viewpoint, can be used to get a.e. representations for the convergence in distribution in $\mathfrak{R}^p$.

THEOREM 3.2. *Let $V = \mathfrak{R}^p$, $p \in \mathscr{N}$. Let $\{P_n\}$, P be probability measures on β such that P is atomless, P satisfies $\mathscr{C}$, $P(S(P)^0) = 1$, $\int \|x\|^2\, dP < \infty$, $\int \|x\|^2\, dP_n < \infty$, $n \in \mathscr{N}$, and $P_n \to_w P$. Then, there exists a sequence of r.v. $\{(X_n, X)\}$ such that:*

(i) *(X_n, X) is an o.c. between P_n and P, $n \in \mathscr{N}$.*
(ii) *$X_n \to_{a.e.} X$.*
(iii) *$X_n = H_n(X)$ with H_n increasing, $n \in \mathscr{N}$.*

Next, as stated in the Introduction, we are going to construct a counterexample to Theorem 3.1 in infinite-dimensional Hilbert spaces. The key is that in these spaces, $\overline{B}(x, 1)$ is not compact but weakly compact and this precludes the use of the arguments leading to a.e. convergence. Moreover, the distributions $\{P_n\}$ and P in our counterexample are such that P is Gaussian and, from equality (5) below,

$$\sup_{A \in \beta} |P_n(A) - P(A)| \to 0.$$

In the construction we use the following lemma:

LEMMA 3.3. *There exist two sequences of real numbers, $\{\alpha_n\}$ and $\{\sigma_n\}$, such that $\alpha_n < 0$, $\sigma_n > 0$, $\Sigma \sigma_n^2 < \infty$ and $\mu[Z_n \geq 0] = 1/(n+1)$, where Z_n is a real r.v. with normal law $N(\alpha_n, \sigma_n^2)$, $n \in \mathscr{N}$.*

PROOF. Let Y be a real $r.v.$ with normal distribution with mean 0 and variance 1 and let γ_n and δ_n (>0) be such that Y satisfies $\mu[Y \geq \gamma_n] = 1/(n+1)$ and $\mu[-\delta_n \leq Y \leq \delta_n] = 1 - 2^{-n}$. It suffices to take

$$\sigma_n = \frac{1}{n(\gamma_n + \delta_n)} \quad \text{and} \quad \alpha_n = \frac{-\gamma_n}{n(\gamma_n + \delta_n)}. \qquad \square$$

The example is as follows.

EXAMPLE 3.4. Let $\{e_n\}$ be a complete orthonormal system in a real, separable, infinite-dimensional Hilbert space V. Let $\{\alpha_n\}$ and $\{\sigma_n\}$ be the two sequences in the preceding lemma and let S be the operator given by $S(\Sigma x_n e_n) = \Sigma \sigma_n^2 x_n e_n$. Then there exists a Gaussian measure P on β with mean vector $m = \Sigma \alpha_n e_n$ and covariance operator S. The measure P satisfies $\mathscr{C}$ by Proposition 2.1. Let $n \in \mathscr{N}$ and $E_n = \{x \in V: \langle x, e_n \rangle \geq 0\}$. We define P_n as follows:

Let $A \in \beta$:

$$P_n(A \cap E_n^c) = P(A \cap E_n^c),$$

$$P_n(A \cap (E_n + e_n)) = P((A - e_n) \cap E_n),$$

$$P_n\big(A \cap E_n \cap (E_n + e_n)^c\big) = 0.$$

Then $P_n \to_w P$ because for every $A \in \beta$, $n \in \mathscr{N}$:

$$P_n(A) = P(A \cap E_n^c) + P((A - e_n) \cap E_n) \tag{5}$$

and $\{P(E_n^c)\}$ converges to 1.

Let X be a r.v. such that $P_X = P$ and

$$H_n(x) = \begin{cases} x, & \text{if } x \in E_n^c, \\ x + e_n, & \text{if } x \in E_n. \end{cases}$$

It takes an elementary computation to see that H_n is increasing and $P_{H_n(X)} = P_n$. Since P satisfies $\mathscr{C}$, by Proposition 2.3, we can assume that there exists $(X, G_n(X))$, that is an o.c. between P and P_n. Then, we have

$$W(P, P_n) \geq W(P^n, P_n^n) = P^n[0, \infty) = \int \|x - H_n(x)\|^2 \, dP,$$

where P^n, P_n^n are the nth marginal distributions of P and P_n and then $(X, H_n(X))$ is an o.c. between P and P_n.

To end, note that $\{x\colon H_n(x) \nrightarrow x\} = \limsup E_n$ and that $P(E_n) = 1/(n+1)$. Then, by the Borel–Cantelli lemma, the conclusion holds if we prove that $\{E_n\}$ are independent events. For this, let $i_1, i_2, \ldots, i_r$ be natural numbers. Now

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) = P^{i_1, i_2, \ldots, i_r}[0, \infty) \times \cdots \times [0, \infty),$$

where $P^{i_1, i_2, \ldots, i_r}$ is the rth-dimensional marginal distribution of P on the subspace $\mathscr{L}[e_{i_1}, e_{i_2}, \ldots, e_{i_r}]$, which in turn, is a normal distribution with independent marginals.

Then

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) = P(E_{i_1}) \times \cdots \times P(E_{i_r}).$$

APPENDIX

To end the paper we give a proof of the measurability of increasing functions. We need some auxiliary results, most of which are of a geometrical nature.

LEMMA 4.1. *Let $t \in V$ and $\delta > 0$. Let $\{w_n\}$ be a sequence of elements from $S(t, \delta)$, dense in it. We define $R_n = \{x \in V\colon \langle x - w_n, t - w_n \rangle < 0\}$, $n \in \mathscr{N}$. Then*

$$\overline{B}(t, \delta) = \bigcap R_n^c.$$

PROOF. Let $x \in \overline{B}(t, \delta)$. We have that

$$\langle x - t, t - w_n \rangle \geq -\delta^2 = -\langle t - w_n, t - w_n \rangle$$

and, therefore, $\langle x - w_n, t - w_n \rangle \geq 0$. On the other hand, let $x \notin \overline{B}(t, \delta)$. It suffices to consider $y = t + (x - t)(\delta/\|x - t\|)$ and a subsequence of $\{w_n\}$, $\{w_{n_m}\}$, such that $w_{n_m} \to_{m \to \infty} y$ in order to get the existence of $n \in \mathscr{N}$ with $x \in R_n$. □

LEMMA 4.2. *Let $\delta > 0$ and $x, y \in V$, $y \neq x$. Then*

$$\operatorname{ang}(z - y, x - y) \leq \arccos \frac{\|x - y\|}{\left(\|x - y\|^2 + \delta^2\right)^{1/2}}$$

for every $z \in \overline{B}(x, \delta)$ such that $\langle z - x, y - x \rangle \leq 0$.

PROOF. If $\langle z - x, y - x \rangle = 0$,

$$\cos\left(\widehat{z - y, x - y}\right) = \frac{\|x - y\|}{\left(\|z - x\|^2 + \|x - y\|^2\right)^{1/2}} \geq \frac{\|x - y\|}{\left(\|x - y\|^2 + \delta^2\right)^{1/2}}.$$

Suppose that $\langle z - x, y - x \rangle < 0$. There exists $\alpha_0 \in (0,1)$ such that if $z^* = \alpha_0 y + (1 - \alpha_0)z$, then $\langle z^* - x, y - x \rangle = 0$ and

$$\cos\left(\widehat{z^* - y, x - y}\right) = \cos\left(\widehat{z - y, x - y}\right).$$

Therefore the lemma holds. □

PROPOSITION 4.3. *Let $t \in V$ and $r > 0$. Let $p \in V$ be such that $\|p - t\| = r$ and define $R_{(p)} = \{y \in V\colon \langle y - p, p - t \rangle > 0\}$. Let $y \in R_{(p)}$. Then, there exists $\varepsilon = (\varepsilon(y, p, t)) \in (0, \pi/2)$ that satisfies*

$$\{z \neq y\colon \operatorname{ang}(z - y, p - t) < \pi/2 + \varepsilon\} \cap \overline{B}(t, r) = \varnothing.$$

PROOF. To simplify the expressions set $w = p - t$. Let $\alpha \in \Re$ be such that $x = y + \alpha w$ is the unique point in the set $[y + \mathscr{L}[w]] \cap [p + \mathscr{L}[w]^{\perp}]$. It follows from $\langle x - p, w \rangle = 0$ that $\alpha < 0$ because $y \in R_{(p)}$.

Let $\delta = \|t - x\| + r$. It is evident that $\overline{B}(x, \delta) \supset \overline{B}(t, r)$.

Let $\varepsilon^* = \arccos \|x - y\|/(\|x - y\|^2 + \delta^2)^{1/2}$. Note that $\varepsilon = \pi/2 - \varepsilon^* \in (0, \pi/2)$. Let us show that ε satisfies the proposition.

First, note that if z satisfies $\langle z - p, w \rangle \leq 0$, we have

$$\langle z - x, y - x \rangle = \langle z - p, -\alpha w \rangle + \langle p - x, -\alpha w \rangle \leq 0.$$

Then, if $z \in \overline{B}(x, \delta)$ and $\langle z - p, w \rangle \leq 0$, by Lemma 4.2, we may conclude that $\operatorname{ang}(z - y, x - y) \leq \varepsilon^*$.

By the same argument as in the proof of the Lemma 4.1, if $\langle z - p, w \rangle > 0$, then $z \notin \overline{B}(t, r)$. So, $\overline{B}(t, r) \subset \{z\colon \operatorname{ang}(z - y, \alpha w) \leq \varepsilon^*\}$.

Finally, taking into account that $\alpha < 0$, we have

$$\{z\colon \operatorname{ang}(z - y, \alpha w) > \varepsilon^*\} = \{z\colon \operatorname{ang}(z - y, w) < \pi/2 + \varepsilon\}. \qquad \square$$

PROPOSITION 4.4. *Let $w \in V$ and $H\colon D \subset V \to V$ be an increasing function. Then for every $\varepsilon \in (0, \pi/2)$, there exists $\delta_\varepsilon \in (0, \pi/2)$ such that if* $\operatorname{ang}(y - x, w) < \delta_\varepsilon$; $y, x \in D$, *then* $\operatorname{ang}(H(y) - H(x), w) < \pi/2 + \varepsilon$.

PROOF. To prove this proposition we use the following function (which has been suggested as a definition of the angle between two vectors in [4], page 74):

$$\phi(u, v) = \left\| \frac{u}{\|u\|} - \frac{v}{\|v\|} \right\|; \qquad u, v \in V.$$

ϕ is related to the usual definition of angle through the function

$$\varphi\colon [0, \Pi] \to [0, 2],$$

$$\theta \to 2\left(1 - \cos^2(\theta/2)\right)^{1/2}.$$

In fact, after some computations, we obtain that

$$\varphi(\widehat{u,v}) = \phi(u,v).$$

Therefore, by the triangular inequality,

$$\begin{aligned}(6)\qquad &\varphi(\operatorname{ang}(H(y) - H(x), w)) \\ &\leq \varphi(\operatorname{ang}(H(y) - H(x), y - x)) + \varphi(\operatorname{ang}(y - x, w)).\end{aligned}$$

Let $\rho \in (0,2)$ be such that $\varphi(\pi/2) + \rho < \varphi((\pi/2) + \varepsilon)$. Then there exists $\delta_\varepsilon \in (0, \pi)$ such that $\varphi(\delta_\varepsilon) = \rho$. Thus, if $\operatorname{ang}(y - x, w) < \delta_\varepsilon$, $y, x \in D$, by (6) we obtain that $\varphi[\operatorname{ang}(H(y) - H(x), w)] < \varphi(\pi/2 + \varepsilon)$ and therefore

$$\operatorname{ang}(H(y) - H(x), w) < \pi/2 + \varepsilon. \qquad \square$$

REMARK. Note that in the preceding proposition, δ_ε does not depend on x or w. It only depends on ε.

The following result is a corollary of Propositions 4.3 and 4.4.

COROLLARY 4.5. *Let H: $D \subset V \to V$ be an increasing function. Let $t \in V$ and $r > 0$, and let $p \in V$ such that $\|p - t\| = r$. We define the set $R_{(p)} = \{y \in V: \langle y - p, p - t\rangle > 0\}$. Let $x \in D$ such that $H(x) \in R_{(p)}$. Then, there exists $\delta(= \delta(x, p, t)) \in (0, \pi/2)$ such that*

$$y \in D, \qquad \operatorname{ang}(y - x, p - t) < \delta \Rightarrow H(y) \notin \overline{B}(t, r).$$

LEMMA 4.6. *Let $x, w \in V$ such that $\|w\| = 1$ and let $\theta \in (0, \pi/2)$. Then the set $C_{x,w,\theta} = \{y \neq x: \operatorname{ang}(y - x, w) < \theta\}$ is open.*

PROOF. Let $y \in C_{x,w,\theta}$ and $y_1 = y - sw$, where $s > 0$ is such that $y_1 \in C_{x,w,\theta}$. Let $z \in C_{y_1,w,\theta}$. Then $z \neq x$ and

$$\frac{\langle z - x, w\rangle}{\|z - x\|\,\|w\|} = \frac{\langle z - y_1, w\rangle}{\|z - y_1\|\,\|w\|}\frac{\|z - y_1\|}{\|z - x\|} + \frac{\langle y_1 - x, w\rangle}{\|y_1 - x\|\,\|w\|}\frac{\|y_1 - x\|}{\|z - x\|}$$

$$> \cos\theta\left(\frac{\|z - y_1\| + \|y_1 - x\|}{\|z - x\|}\right) \geq \cos\theta.$$

Therefore $C_{y_1,w,\theta} \subset C_{x,w,\theta}$.

Now let v be such that $\|v\| = 1$ and $\widehat{v,w} = \theta$. Let $r = \inf\{\|y - (y_1 + hv)\|, h \in \mathfrak{R}\}$. Some computations give that $r = s(1 - \cos^2\theta)^{1/2}$ independently of v.

Note that $y_1 \notin B(y, r)$. Let $z \neq y_1$ be such that $\operatorname{ang}(z - y_1, w) \geq \theta$. Then there exists $\rho_0 \in (0, 1]$ such that $\operatorname{ang}(y + \rho_0(z - y) - y_1, w) = \theta$ and therefore

$$r \leq \|y - y_1 - (y + \rho_0(z - y) - y_1)\| \leq \|z - y\|,$$

so that the lemma is proved. $\square$

Before starting with the next proposition, let us present some notation. Given $t \in V$, $\delta > 0$ and $\{w_i\}$ a denumerable set in $S(t, \delta)$ with the same properties as those in Lemma 4.1, let us denote $\overline{B} = \overline{B}(t, \delta)$ and $R_i = \{x \in V:$

$\langle x - w_i, t - w_i \rangle < 0\}$. Now, if H: $D \subset V \to V$ is increasing, we set $\mathscr{N}_0 = \{i \in \mathscr{N}: H^{-1}(R_i) \neq \varnothing\}$ and, for every $i \in \mathscr{N}_0$:

(i) Let $x \in \mathscr{L}(w_i - t)^{\perp}$ and ρ be such that $y = x + \rho(w_i - t) \in H^{-1}(R_i)$. We set $C_{x,\rho,i} = C_{y, w_i - t, \theta}$, where θ is a positive real number chosen in such a way that $D \cap C_{x,\rho,i} \subset H^{-1}(\overline{B}^c)$. Such a θ exists by Corollary 4.5.

(ii)
$$\Gamma_i = \bigcup_{x+\rho(w_i - t) \in H^{-1}(R_i)} C_{x,\rho,i}.$$

(iii)
$$f_i: \mathscr{L}[w_i - t]^{\perp} \to \Re,$$
$$x \to \inf\{\alpha: x + \alpha(w_i - t) \in \Gamma_i\}.$$

(iv)
$$A_i = \{x: x \in \mathscr{L}(w_i - t)^{\perp} \text{ and } f_i(x) \neq \infty\}.$$

(v)
$$N_i = \{x + f_i(x)(w_i - t), x \in A_i\} \cap H^{-1}(\overline{B}^c).$$

Note that for every $x \in \mathscr{L}(w_i - t)^{\perp}$, $\{\alpha: x + \alpha(w_i - t) \in \Gamma_i\}$ is a nonempty interval without upper bound.

PROPOSITION 4.7. *Let $t \in V$, $r > 0$ and assume that H: $D \subset V \to V$ is increasing. With the notation just presented we have that $H^{-1}(\overline{B}^c) = \bigcup_{i \in \mathscr{N}_0}[(\Gamma_i \cap D) \cup N_i]$.*

PROOF. $H^{-1}(\overline{B}^c) \supset \bigcup_{i \in \mathscr{N}_0}((\Gamma_i \cap D) \cup N_i)$ by the definitions of Γ_i and of N_i.

Now, let $x \in H^{-1}(\overline{B}^c)$. Lemma 4.1 implies that there exists i_0 such that $H(x) \in R_{i_0}$. Then $i_0 \in \mathscr{N}_0$ and, if we write $x = x_1 + \alpha(w_{i_0} - t)$, with $x_1 \in \mathscr{L}(w_{i_0} - t)^{\perp}$, then $f_{i_0}(x_1) \leq \alpha$. Thus, if $\alpha = f_{i_0}(x_1)$, then $x_1 \in A_{i_0}$ and $x \in N_{i_0}$, by definition of N_{i_0}.

Otherwise, $\alpha > f_{i_0}(x_1)$ and there exists $\varepsilon > 0$ such that $x_1 + (\alpha - \varepsilon)(w_{i_0} - t) \in \Gamma_{i_0}$ and $x \in \Gamma_{i_0} \cap D$ and the proof is complete. □

THEOREM 4.8. *Let P be a probability measure on β satisfying $\mathscr{C}$ and $\overline{\beta}$ be the P-completion of β. If $D \in \overline{\beta}$ and H: $D \subset V \to V$ increasing, then H is $\overline{\beta}$-measurable.*

PROOF. We use the same notation as in the preceding proposition. Let $t \in V$, $r > 0$, and we denote $\overline{B} = \overline{B}(t, r)$.

We have

$$H^{-1}(\overline{B}^c) = (\Gamma \cap D) \cup N, \quad \text{where } \Gamma = \bigcup_{i \in \mathscr{N}_0} \Gamma_i \quad \text{and} \quad N = \bigcup_{i \in \mathscr{N}_0} N_i.$$

Γ is an open set by Lemma 4.6. Thus, the proof will be complete if we prove that N_i, $i \in \mathscr{N}_0$, is a P-null set. □

Let $i \in \mathscr{N}_0$. We define $g_i: V \to \mathfrak{R}$ by $g_i = f_i \pi_i^* - \pi_i$, where π_i and π_i^* are the projections from V on $\mathscr{L}[w_i - t]$ and on $\mathscr{L}[w_i - t]^\perp$, respectively. g_i is a measurable mapping and $g_i^{-1}(0) = \{x + f_i(x)(w_i - t), x \in A_i\}$. Then this set is measurable.

On the other hand, $V = \mathscr{L}[w_i - t] \oplus \mathscr{L}[w_i - t]^\perp$. Then

$$P\{x + f_i(x)(w_i - t), x \in A_i\} = \int_{A_i} P_{w_i - t}(\{f_i(x)\}|w)\, d\mu_i(w),$$

where $P_{w_i - t}(-|w)$ is an atomless version of the regular conditional probability on $\mathscr{L}[w_i - t]$ given $w \in \mathscr{L}[w_i - t]^\perp$ and μ_i is the marginal of P on $\mathscr{L}[w_i - t]^\perp$. Then $P\{x + f_i(x)(w_i - t), x \in A_i\} = 0$ and so N_i is a P-null set. □

Acknowledgment. I would like to express my sincere thanks to Professor J. A. Cuesta, under whose guidance this work was completed.

REFERENCES

[1] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

[2] Blackwell, D. and Dubins, L. E. (1983). An extension of Skorohod's almost sure representation theorem. *Proc. Amer. Math. Soc.* **89** 691–692.

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

[4] Diestel, J. (1975). *Geometry of Banach Spaces. Lecture Notes in Math.* **485**. Springer, Berlin.

[5] Dowson, D. C. and Landau, B. V. (1982). The Frechet distance between multivariate normal distributions. *J. Multivariate Anal.* **12** 450–455.

[6] Dudley, R. M. (1968). Distance of probability measures and random variables. *Ann. Math. Statist.* **39** 1563–1572.

[7] Fernique, X. (1988). Un modèle presque sûre pour la convergence in loi. *C. R. Acad. Sci. Paris* **306** 335–338.

[8] Givens, C. R. and Shortt, R. M. (1984). A class of Wasserstein metrics for probability distributions. *Michigan Math. J.* **31** 231–242.

[9] Kantorovich, L. V. (1942). On the transfer of masses. *Dokl. Akad. Nauk SSSR* **37** 7–8.

[10] Kantorovich, L. V. and Rubinstein, G. SH. (1958). On a space of completely additive functions. *Vestnik Leningrad Univ. Math.* **13** 52–59.

[11] Knott, M. and Smith, C. S. (1985). On the optimal mapping of distributions. *J. Optim. Theory Appl.* **43** 39–49.

[12] Major, P. (1978). On the invariance principle for sums of independent, identically distributed random variables. *J. Multivariate Anal.* **48** 487–517.

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

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

[15] Rachev, S. T. and Rüschendorf, L. (1990). Recent results in the theory of probability metrics. Unpublished manuscript.

[16] Rüschendorf, L. (1985). The Wasserstein distance and approximation theorems. *Z. Wahrsch. Verw. Gebiete* **70** 117–129.

[17] RÜSCHENDORF, L. and RACHEV, S. T. (1990). A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.* **32** 48–54.
[18] SCHIEF, A. (1989). Almost surely convergent random variables with given laws. *Probab. Theory Related Fields* **81** 559–567.
[19] SKOROHOD, A. V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl.* **3** 261–290.
[20] VALLENDER, S. S. (1973). Calculation of the Wasserstein distance between distributions of the line. *Theory Probab. Appl.* **18** 784–786.
[21] WICHURA, M. J. (1970). On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *Ann. Math. Statist* **41** 284–291.

DEPARTAMENTO DE MATEMÁTICAS
ESTADÍSTICA Y COMPUTACIÓN
FACULTAD DE CIENCIAS
UNIVERSIDAD DE CANTABRIA
39005 SANTANDER
SPAIN