**Quaderni di Dipartimento**

# Skorohod Representation Theorem Via Disintegrations

Patrizia Berti
(Università di Modena e Reggio Emilia)

Luca Pratelli
(Accademia Navale di Livorno)

Pietro Rigo
(Università di Pavia)

# 104 (10-09)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Ottobre 2009

# SKOROHOD REPRESENTATION THEOREM VIA DISINTEGRATIONS

PATRIZIA BERTI[1], LUCA PRATELLI[2], AND PIETRO RIGO[3]

Abstract. Let $(\mu_n : n \geq 0)$ be Borel probabilities on a metric space $S$ such that $\mu_n \to \mu_0$ weakly. Say that Skorohod representation holds if, on some probability space, there are $S$-valued random variables $X_n$ satisfying $X_n \sim \mu_n$ for all $n$ and $X_n \to X_0$ in probability. By Skorohod's theorem, Skorohod representation holds (with $X_n \to X_0$ almost uniformly) if $\mu_0$ is separable. Two results are proved in this paper. First, Skorohod representation may fail if $\mu_0$ is not separable (provided, of course, non separable probabilities exist). Second, independently of $\mu_0$ separable or not, Skorohod representation holds if $W(\mu_n, \mu_0) \to 0$ where $W$ is Wasserstein distance (suitably adapted). The converse is essentially true as well. Such a $W$ is a version of Wasserstein distance which can be defined for any metric space $S$ satisfying a mild condition. To prove the quoted results (and to define $W$), disintegrable probability measures are fundamental.

[1]Università di Modena e Reggio Emilia, Italy
[2]Accademia Navale di Livorno, Italy
[3]Università di Pavia, Italy

## 1. Introduction

Throughout, $(S, d)$ is a metric space, $\mathcal{B}$ the Borel $\sigma$-field on $S$, and $(\mu_n : n \geq 0)$ a sequence of probability measures on $\mathcal{B}$.

If $\mu_n \to \mu_0$ weakly and $\mu_0$ is separable, there are $S$-valued random variables $X_n$, defined on some probability space, such that $X_n \sim \mu_n$ for all $n$ and $X_n \to X_0$ almost uniformly. This is Skorohod representation theorem (in its sequential version), as it appears after Skorohod (1956), Dudley (1968) and Wichura (1970). See page 77 of van der Vaart and Wellner (1996) and page 130 of Dudley (1999) for some historical notes.

This paper stems from the following question: Is it possible to drop separability of $\mu_0$ ? Such a question is both natural and subtle. It is natural, since $\mu_n \to \mu_0$ weakly is necessary for the conclusions of Skorohod theorem, and thus separability of $\mu_0$ is the only real assumption. Note also that, if separability of $\mu_0$ would be superfluous, weak convergence of probability measures could be *generally defined* as almost uniform convergence of random variables with given distributions. But the question is subtle as well, since it is consistent with the usual ZFC set theory that non separable probabilities do not exist. So, consistently with ZFC, the question does not arise at all. On the other hand it is currently unknown, and possibly

2000 *Mathematics Subject Classification.* 60B10, 60A05, 60A10.
*Key words and phrases.* Disintegration – Separable probability measure – Skorohod representation theorem – Wasserstein distance – Weak convergence of probability measures.

"unprovable", whether existence of non separable probabilities is consistent with ZFC. Thus, the question makes sense.

And, as shown in Example 3, the answer is no. That is, if non separable probabilities exist, separability of $\mu_0$ cannot be dropped from Skorohod theorem, even if $X_n \to X_0$ almost uniformly is weakened into $X_n \to X_0$ in probability.

So, one cannot dispense with separability of $\mu_0$. On the other hand, when $\mu_0$ is separable, $\mu_n \to \mu_0$ weakly is equivalent to $\rho(\mu_n, \mu_0) \to 0$ where $\rho$ is a suitable distance between probability measures. Well known examples are $\rho$ the Prohorov distance or $\rho$ the bounded Lipschitz metric. Therefore, it is worth investigating versions of Skorohod theorem such as

(SK): *If $\rho(\mu_n, \mu_0) \to 0$, there is a Skorohod representation*

where a Skorohod representation is meant as

*A sequence $(X_n : n \geq 0)$ of $S$-valued random variables, defined on a common probability space, such that $X_n \sim \mu_n$ for all $n$ and $X_n \to X_0$ in (outer) probability.*

Note that almost uniform convergence has been weakened into convergence in probability in a Skorohod representation. Indeed, if non separable probabilities exist, it may be that the sequence $(\mu_n)$ can be realized by random variables $X_n$ such that $X_n \to X_0$ in probability, but not by random variables $Y_n$ such that $Y_n \to Y_0$ on a set of probability 1. See Example 7.

Whether or not (SK) makes some interest depends on the choice of $\rho$. For instance, (SK) is well known if $\rho$ is total variation distance (see Sethuraman (2002) and Proposition 1) but looks intriguing if $\rho$ is Wasserstein distance (suitably adapted) or bounded Lipschitz metric.

Let us suppose $\sigma(d) \subset \mathcal{B} \otimes \mathcal{B}$, where $d$ is the distance on $S$, $\sigma(d)$ the $\sigma$-field on $S \times S$ generated by $(x, y) \mapsto d(x, y)$ and $\mathcal{B} \otimes \mathcal{B} = \sigma\{A \times B : A, B \in \mathcal{B}\}$. This assumption is actually true when $(S, d)$ is separable, as well as for various non separable choices of $(S, d)$. For instance, $\sigma(d) \subset \mathcal{B} \otimes \mathcal{B}$ when $d$ is uniform distance on some space $S$ of cadlag functions, or when $d$ is 0-1 distance and $\text{card}(S) = \text{card}(\mathbb{R})$.

In Theorem 5, (SK) is shown to be true for $\rho = W$ with $W$ defined as follows. Let $X, Y$ be the canonical projections on $S \times S$ and $\mu, \nu$ any probabilities on $\mathcal{B}$. Also, let $\mathcal{D}(\mu, \nu)$ be the class of those probabilities $P$ on $\mathcal{B} \otimes \mathcal{B}$ such that $P \circ X^{-1} = \mu$, $P \circ Y^{-1} = \nu$ and $P$ is *disintegrable* in a suitable sense; see Section 2. Then,

$$W(\mu, \nu) = \frac{W_0(\mu, \nu) + W_0(\nu, \mu)}{2} \quad \text{where} \quad W_0(\mu, \nu) = \inf_{P \in \mathcal{D}(\mu, \nu)} E_P\big\{1 \wedge d(X, Y)\big\}.$$

Such a $W$ is a Wasserstein type distance. If at least one between $\mu$ and $\nu$ is separable, one also obtains

$$W(\mu, \nu) = W_0(\mu, \nu) = \inf_{P \in \mathcal{F}(\mu, \nu)} E_P\big\{1 \wedge d(X, Y)\big\}$$

where $\mathcal{F}(\mu, \nu)$ is the class of laws $P$ on $\mathcal{B} \otimes \mathcal{B}$ such that $P \circ X^{-1} = \mu$ and $P \circ Y^{-1} = \nu$. (That is, members of $\mathcal{F}(\mu, \nu)$ are not requested to be disintegrable).

In checking $W$ is a metric, and even more in proving Theorem 5, restricting to disintegrable probability measures is fundamental. This explains the title of this paper.

Roughly speaking, $W$ is the "right" distance to cope with (SK), as $W(\mu_n, \mu_0) \to 0$ amounts to just a little bit more than a Skorohod representation. In addition, independently of (SK), $W$ looks (to us) a reasonable extension of Wasserstein distance to a large class of metric spaces (those satisfying $\sigma(d) \subset \mathcal{B} \otimes \mathcal{B}$).

Finally, we make a remark on the bounded Lipschitz metric

$$b(\mu, \nu) = \sup_f \Big| \int f \, d\mu - \int f \, d\nu \Big|$$

where sup is over those functions $f : S \to [-1, 1]$ satisfying $|f(x) - f(y)| \leq d(x, y)$ for all $x, y \in S$. It would be nice to have an analogous of Theorem 5 for $b$, that is, to prove that a Skorohod representation is available whenever $b(\mu_n, \mu_0) \to 0$. We do not know whether this is true, but we mention two particular cases. It is trivially true for $d$ the 0-1 distance; see Theorem 2.1 of Sethuraman (2002) and Proposition 1. It is "close to be true" if $S$ is some space of cadlag functions and $d$ the uniform distance. Suppose in fact $(S, d)$ is of this type and $b(\mu_n, \mu_0) \to 0$. Then, by a result in Berti et al. (2009), there are a sub-$\sigma$-field $\mathcal{B}_0 \subset \mathcal{B}$ and $S$-valued random variables $X_n$ such that $X_n \longrightarrow X_0$ in probability and $X_n \sim \mu_n$ on $\mathcal{B}_0$ for all $n$.

## 2. Disintegrations

In this paper, a disintegration is meant as follows. Let $P$ be a probability measure on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, where $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ are arbitrary measurable spaces. Let $X(x, y) = x$ and $Y(x, y) = y$, $(x, y) \in \Omega_1 \times \Omega_2$, denote the canonical projections. Then, $P$ is said to be *disintegrable* if $Y$ admits a regular version of the conditional distribution given $X$ in the space $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P)$. That is, $P$ is disintegrable if there is a collection $\alpha = \{\alpha(x) : x \in \Omega_1\}$ such that:

− $\alpha(x)$ is a probability on $\mathcal{A}_2$ for each $x \in \Omega_1$;

− $x \mapsto \alpha(x)(B)$ is $\mathcal{A}_1$-measurable for each $B \in \mathcal{A}_2$;

− $P(X \in A, Y \in B) = \int_A \alpha(x)(B) \, P \circ X^{-1}(dx)$ for all $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$.

Such an $\alpha$ is called a *disintegration* for $P$.

A disintegration can fail to exist. However, for $P$ to admit a disintegration, it suffices that $P \circ X^{-1}$ is atomic, or that $P(Y \in B) = 1$ for some $B \in \mathcal{A}_2$ which is isomorphic to a Borel set in a Polish space. Furthermore, a countable convex combination of disintegrable laws is disintegrable as well.

The term "disintegration" has usually a broader meaning than in this paper. We refer to Maitra and Ramakrishnan (1988), Berti and Rigo (1999) and references therein for disintegrations in this larger sense.

## 3. Other preliminaries

Let $(\Omega, \mathcal{A}, P)$ be a probability space. The outer and inner measures are

$$P^*(H) = \inf\{P(A) : H \subset A \in \mathcal{A}\}, \quad P_*(H) = 1 - P^*(H^c), \quad H \subset \Omega.$$

Given maps $X_n : \Omega \to S$, $n \geq 0$, say that $X_n$ converges to $X_0$ in (outer) probability, written $X_n \xrightarrow{P} X_0$, in case

$$\lim_n P^*\big(d(X_n, X_0) > \epsilon\big) = 0 \quad \text{for all } \epsilon > 0.$$

Let $(F, \mathcal{F})$ be a measurable space. If $\mu$ and $\nu$ are probabilities on $\mathcal{F}$, their total variation distance is $\|\mu - \nu\| = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$. Recall that

$$\|P \circ X^{-1} - P \circ Y^{-1}\| \leq P_*(X \neq Y)$$

whenever $X, Y : (\Omega, \mathcal{A}) \longrightarrow (F, \mathcal{F})$ are measurable maps.

Next result is essentially well known; see Theorem 2.1 of Sethuraman (2002). We give a proof to make the paper self-contained, and because we need $(F, \mathcal{F})$ to be an arbitrary measurable space (which is not the case in all references known to us).

**Proposition 1.** *Given probabilities $\mu_n$ on $(F, \mathcal{F})$, $n \geq 0$, there are a probability space $(\Omega, \mathcal{A}, P)$ and measurable maps $X_n : (\Omega, \mathcal{A}) \to (F, \mathcal{F})$ such that*

$$P_*(X_n \neq X_0) = P^*(X_n \neq X_0) = \|\mu_n - \mu_0\| \text{ and } X_n \sim \mu_n \text{ for all } n \geq 0.$$

*Proof.* It can be assumed $\mu_n \neq \mu_0$ for all $n \neq 0$. Let $f_n$ be a density of $\mu_n$ with respect to some measure $\lambda$ on $\mathcal{F}$, say $\lambda = \sum_{n=0}^{\infty} (1/2)^{n+1} \mu_n$. Moreover, let $\nu_n(A) = \int_A \frac{(f_n - f_0)^+}{\|\mu_n - \mu_0\|} \, d\lambda$ for $A \in \mathcal{F}$ and $n \geq 1$. On some probability space $(\Omega, \mathcal{A}, P)$, there are independent random variables $X_0$, $U$, $Z$ such that

$$X_0 \sim \mu_0, \quad U = (U_n : n \geq 1) \text{ is an i.i.d. sequence with } U_1 \text{ uniform on } (0,1),$$
$$Z = (Z_n : n \geq 1) \text{ is an independent sequence with } Z_n \sim \nu_n.$$

For every $n \geq 1$, let us define $A_n = \{f_0(X_0) \, U_n > f_n(X_0)\}$ and $X_n = Z_n$ on $A_n$ and $X_n = X_0$ on $A_n^c$. On noting that $P(0 < U_n < 1) = 1$,

$$P(A_n) = \int_{\{f_0 > f_n\}} \frac{f_0 - f_n}{f_0} \, d\mu_0 = \int_{\{f_0 > f_n\}} (f_0 - f_n) \, d\lambda$$
$$= \int (f_0 - f_n)^+ \, d\lambda = \|\mu_n - \mu_0\|.$$

Since $Z_n$ is independent of $(X_0, U_n)$, it follows that

$$P(X_n \in A) = P\big(A_n \cap \{Z_n \in A\}\big) + P\big(A_n^c \cap \{X_0 \in A\}\big)$$
$$= P(A_n) \, P(Z_n \in A) + \mu_0\big(A \cap \{f_0 \leq f_n\}\big) + \int_{A \cap \{f_0 > f_n\}} \frac{f_n}{f_0} \, d\mu_0$$
$$= \int_A (f_n - f_0)^+ \, d\lambda + \int_A (f_n \wedge f_0) \, d\lambda = \int_A f_n \, d\lambda = \mu_n(A)$$

for all $A \in \mathcal{F}$. Thus $X_n \sim \mu_n$, and this in turn implies

$$\|\mu_n - \mu_0\| \leq P_*(X_n \neq X_0) \leq P^*(X_n \neq X_0) \leq P(A_n) = \|\mu_n - \mu_0\|.$$

$\square$

**Remark 2.** In Proposition 1, $P$ can be taken to be *perfect* provided each $\mu_n$ is perfect. We recall that a probability $Q$ on a measurable space $(\Omega_0, \mathcal{A}_0)$ is perfect in case each $\mathcal{A}_0$-measurable function $f : \Omega_0 \to \mathbb{R}$ satisfies $Q(f \in B) = 1$ for some real Borel set $B \subset f(\Omega_0)$; see e.g. Maitra and Ramakrishnan (1988).

## 4. Skorohod representation theorem without separability: an example and a result based on Wasserstein distance

We aim to do three things. First, to show that a Skorohod representation (as defined in Section 1) can fail to exist if $\mu_n \to \mu_0$ weakly but $\mu_0$ is not separable. Second, to introduce a version $W$ of Wasserstein distance for any metric space satisfying a certain (mild) condition. Third to prove that, whether or not $\mu_0$ is separable, a Skorohod representation is available in case $W(\mu_n, \mu_0) \longrightarrow 0$.

4.1. **Separability of $\mu_0$ cannot be dropped in Skorohod theorem.** Given a measurable space $(\Omega, \mathcal{A})$, a map $X : \Omega \to S$ is *measurable*, or a *random variable*, if $X^{-1}(\mathcal{B}) \subset \mathcal{A}$. A probability $\mu$ on $\mathcal{B}$ is *separable* if $\mu(B) = 1$ for some separable $B \in \mathcal{B}$.

**Example 3.** Let $\mathcal{B}_{(0,1)}$ be the Borel $\sigma$-field on $(0,1)$ and $m$ the Lebesgue measure on $\mathcal{B}_{(0,1)}$. Existence of a non separable probability on the Borel $\sigma$-field of a metric space is *equiconsistent* with existence of a countably additive extension of $m$ to the power set of $(0,1)$. See page 403 of Dudley (1999), page 380 of Fuchino et al. (2006) and page 182 of Goldring (1995). This means that, if there is a non separable Borel probability on a metric space, then, possibly in a different model of ZFC, there is a countably additive extension of $m$ to the power set of $(0,1)$.

Suppose that there is a non separable Borel probability on some metric space, or, equiconsistently, that $m$ admits a countably additive extension to the power set.

Let $(f_n : n \geq 1)$ be an i.i.d. sequence of real random variables, on the probability space $((0,1), \mathcal{B}_{(0,1)}, m)$, such that

$$f_n \geq 0, \; E_m(f_n) = 1, \; f_n \text{ has a non degenerate distribution.}$$

Let $S = (0,1)$, equipped with 0-1 distance, and let $\mu_0$ be a countably additive extension of $m$ to the power set of $(0,1)$. Define further

$$\mu_n(B) = E_{\mu_0}(f_n I_B) \quad \text{for all } n \geq 1 \text{ and } B \in \mathcal{B}.$$

(Note that $\mathcal{B}$ is the power set of $(0,1)$). Fix $n > k \geq 1$ and $B \in \sigma(f_1, \ldots, f_k)$. Since $f_n$ and $I_B$ are $\mathcal{B}_{(0,1)}$-measurable and independent under $m$,

$$\mu_n(B) = E_m(f_n I_B) = E_m(f_n) E_m(I_B) = m(B) = \mu_0(B).$$

Hence, $\mu_0(B) = \lim_n \mu_n(B)$ for every $B \in \bigcup_k \sigma(f_1, \ldots, f_k)$. Letting $\mathcal{F} = \sigma(f_1, f_2, \ldots)$, standard arguments imply

$$E_{\mu_0}(g) = \lim_n E_{\mu_n}(g) \quad \text{for all bounded } \mathcal{F}\text{-measurable functions } g.$$

Given $B \in \mathcal{B}$, since $E_{\mu_0}(I_B \mid \mathcal{F})$ is bounded and $\mathcal{F}$-measurable, it follows that

$$\mu_n(B) = E_{\mu_0}\{f_n E_{\mu_0}(I_B \mid \mathcal{F})\} = E_{\mu_n}\{E_{\mu_0}(I_B \mid \mathcal{F})\} \longrightarrow E_{\mu_0}\{E_{\mu_0}(I_B \mid \mathcal{F})\} = \mu_0(B).$$

Thus, $\mu_n \to \mu_0$ weakly. Suppose now that $X_n \sim \mu_n$ for all $n \geq 0$, where the $X_n$ are $S$-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$. As $d$ is 0-1 distance,

$$P^*\big(d(X_n, X_0) > \frac{1}{2}\big) = P^*(X_n \neq X_0) \geq P_*(X_n \neq X_0) \geq \|\mu_n - \mu_0\| = \frac{1}{2} E_m|f_n - 1|.$$

Since $(f_n)$ is an i.i.d. sequence with a nondegenerate distribution, $(f_n)$ fails to converge in $L_1$. Therefore, $X_n$ does not converge to $X_0$ in probability.

Example 3 shows that, unless $\mu_0$ is separable, $\mu_n \to \mu_0$ weakly is not enough for a Skorohod representation. Accordingly, as discussed in Section 1, we focus on results of the type

(SK): If $\rho(\mu_n, \mu_0) \to 0$, then a Skorohod representation is available

where $\rho$ is a suitable distance between probability measures. By Proposition 1, (SK) is true for $\rho$ the total variation distance. Here, we deal with $\rho$ the Wasserstein distance (suitably adapted). In a sense, Wasserstein distance is the "right" distance to cope with Skorohod theorem.

4.2. **A Wasserstein distance.** For any $\sigma$-field $\mathcal{E}$, $\mathcal{M}(\mathcal{E})$ denotes the collection of probability measures on $\mathcal{E}$. Let

$$X(x,y)=x \quad \text{and} \quad Y(x,y)=y, \quad (x,y)\in S\times S,$$

be the canonical projections on $S\times S$ and

$$\mathcal{D}(\mu,\nu)=\{P\in\mathcal{M}(\mathcal{B}\otimes\mathcal{B}):P\circ X^{-1}=\mu,\,P\circ Y^{-1}=\nu,\,P\text{ disintegrable}\}$$

where $\mu$, $\nu\in\mathcal{M}(\mathcal{B})$. Disintegrable probability measures have been defined in Section 2. Note that $\mathcal{D}(\mu,\nu)\neq\emptyset$, as it includes at least the product law $P=\mu\times\nu$.

In the sequel, the distance $d:S\times S\to\mathbb{R}$ is assumed to be measurable, in the sense that

$$\sigma(d)\subset\mathcal{B}\otimes\mathcal{B}.$$

Indeed, $d$ is measurable if $(S,d)$ is separable, as well as in various non separable situations. For instance, $\sigma(d)\subset\mathcal{B}\otimes\mathcal{B}$ if $d$ is uniform distance on some space $S$ of cadlag functions, or if $d$ is 0-1 distance and $\text{card}(S)=\text{card}(\mathbb{R})$. Measurability of $d$ yields $\{(x,y):x=y\}=\{d=0\}\in\mathcal{B}\otimes\mathcal{B}$, which in turn implies $\text{card}(S)\leq\text{card}(\mathbb{R})$. We do not know of any example where $\{(x,y):x=y\}\in\mathcal{B}\otimes\mathcal{B}$ and yet $d$ fails to be measurable. Perhaps, $\{(x,y):x=y\}\in\mathcal{B}\otimes\mathcal{B}$ implies measurability of $d$, or at least measurability of some distance $d^*$ equivalent to $d$.

In any case, if $d$ is measurable, one can define

$$W_0(\mu,\nu)=\inf_{P\in\mathcal{D}(\mu,\nu)}E_P\big\{1\wedge d(X,Y)\big\} \quad \text{and}$$

$$W(\mu,\nu)=\frac{W_0(\mu,\nu)+W_0(\nu,\mu)}{2} \quad \text{for all } \mu,\,\nu\in\mathcal{M}(\mathcal{B}).$$

**Theorem 4.** *Suppose $\sigma(d)\subset\mathcal{B}\otimes\mathcal{B}$ and let $\mu$, $\nu$, $\gamma\in\mathcal{M}(\mathcal{B})$. Then,*

$$W_0(\mu,\nu)=0 \;\Leftrightarrow\; \mu=\nu \quad \textit{and} \quad W_0(\mu,\nu)\leq W_0(\mu,\gamma)+W_0(\gamma,\nu).$$

*In particular, $W$ is a distance on $\mathcal{M}(\mathcal{B})$. In addition, if at least one between $\mu$ and $\nu$ is separable, then*

$$W(\mu,\nu)=W_0(\mu,\nu)=\inf_{P\in\mathcal{F}(\mu,\nu)}E_P\big\{1\wedge d(X,Y)\big\}$$

$$\textit{where } \mathcal{F}(\mu,\nu)=\{P\in\mathcal{M}(\mathcal{B}\otimes\mathcal{B}):P\circ X^{-1}=\mu,\,P\circ Y^{-1}=\nu\}.$$

*(Note: members of $\mathcal{F}(\mu,\nu)$ need not be disintegrable).*

*Proof.* Let $P_\mu(H)=\mu\{x\in S:(x,x)\in H\}$, $H\in\mathcal{B}\otimes\mathcal{B}$. Since $P_\mu\in\mathcal{D}(\mu,\mu)$, then $W_0(\mu,\mu)\leq E_{P_\mu}\big\{1\wedge d(X,Y)\big\}=0$.

Next, if $f:S\to[-1,1]$ satisfies $|f(x)-f(y)|\leq d(x,y)$ for all $x,y\in S$, then

$$\Big|\int f\,d\mu-\int f\,d\nu\Big|=|E_Pf(X)-E_Pf(Y)|\leq E_P|f(X)-f(Y)|$$
$$\leq 2\,E_P\big\{1\wedge d(X,Y)\big\} \quad \text{for all } P\in\mathcal{D}(\mu,\nu).$$

Thus $|\int f\,d\mu-\int f\,d\nu|\leq 2\,W_0(\mu,\nu)$, so that $\mu=\nu$ whenever $W_0(\mu,\nu)=0$.

Next, given $\epsilon>0$, there are $P_1\in\mathcal{D}(\mu,\gamma)$ and $P_2\in\mathcal{D}(\gamma,\nu)$ such that

$$W_0(\mu,\gamma)+W_0(\gamma,\nu)+\epsilon>E_{P_1}\big\{1\wedge d(X,Y)\big\}+E_{P_2}\big\{1\wedge d(X,Y)\big\}.$$

Let $U$, $V$, $Z$ be the canonical projections on $S\times S\times S$ and $\alpha_i$ the disintegration of $P_i$, $i=1,2$. Also, let $Q$ be the probability on $\mathcal{B}\otimes\mathcal{B}\otimes\mathcal{B}$ such that

$$Q\big(U\in A,V\in B,Z\in C\big)=\int\alpha_2(y)(C)\,I_A(x)\,I_B(y)\,P_1(dx,dy), \quad A,B,C\in\mathcal{B}.$$

Then, $(U,V) \sim P_1$ and $(V,Z) \sim P_2$ under $Q$. Let $P_3(\cdot) = Q\big((U,Z) \in \cdot\big)$ denote the distribution of $(U,Z)$ under $Q$. Then,

$$\alpha(x)(\cdot) = \int \alpha_2(y)(\cdot)\,\alpha_1(x)(dy)$$

is a disintegration for $P_3$. Hence, $P_3 \in \mathcal{D}(\mu,\nu)$ so that

$$\begin{aligned} E_{P_1}\big\{1 \wedge d(X,Y)\big\} + E_{P_2}\big\{1 \wedge d(X,Y)\big\} &= E_Q\big\{1 \wedge d(U,V)\big\} + E_Q\big\{1 \wedge d(V,Z)\big\} \\ &\geq E_Q\big\{1 \wedge d(U,Z)\big\} = E_{P_3}\big\{1 \wedge d(X,Y)\big\} \geq W_0(\mu,\nu). \end{aligned}$$

This proves that $W_0(\mu,\nu) \leq W_0(\mu,\gamma) + W_0(\gamma,\nu)$.

Finally, let $W_1(\mu,\nu) = \inf_{P\in\mathcal{F}(\mu,\nu)} E_P\big\{1 \wedge d(X,Y)\big\}$. If $\mu(A) = 1$ for some countable $A$, or if $\nu(A) = 1$ for some $A$ isomorphic to a Borel set in a Polish space, then each $P \in \mathcal{F}(\mu,\nu)$ is disintegrable so that $W_0(\mu,\nu) = W_1(\mu,\nu)$. In particular, $W_0(\mu,\nu) = W_1(\mu,\nu)$ if at least one between $\mu$ and $\nu$ has countable support. Having noted this fact, suppose $\mu$ is separable. Then, given $\epsilon > 0$, there is a Borel partition $A_0, A_1, \ldots, A_k$ of $S$ such that $\mu(A_0) < \epsilon$ and $\mathrm{diam}(A_j) < \epsilon$ for all $j \neq 0$. Fix a point $x_j \in A_j$ and define $T = x_j$ on $\{X \in A_j\}$, $j = 0,1,\ldots,k$. Define also $\lambda = P \circ T^{-1}$, where $P \in \mathcal{F}(\mu,\nu)$ is such that $W_1(\mu,\nu) + \epsilon > E_P\big\{1 \wedge d(X,Y)\big\}$. On noting that $\lambda$ has finite support,

$$\begin{aligned} W_0(\mu,\nu) &\leq W_0(\mu,\lambda) + W_0(\lambda,\nu) = W_1(\mu,\lambda) + W_1(\lambda,\nu) \\ &\leq E_P\big\{1 \wedge d(X,T)\big\} + E_P\big\{1 \wedge d(T,Y)\big\} \\ &\leq 2\,E_P\big\{1 \wedge d(X,T)\big\} + E_P\big\{1 \wedge d(X,Y)\big\} \\ &< 2\sum_{j=0}^{k} E_P\big\{I_{\{X\in A_j\}}\,1 \wedge d(X,x_j)\big\} + W_1(\mu,\nu) + \epsilon \\ &\leq 2\,\big\{P(X \in A_0) + \epsilon\,P(X \notin A_0)\big\} + W_1(\mu,\nu) + \epsilon < W_1(\mu,\nu) + 5\,\epsilon. \end{aligned}$$

Therefore, $W_0(\mu,\nu) \leq W_1(\mu,\nu)$. Since $W_0 \geq W_1$ (by definition), it follows that $W_0(\mu,\nu) = W_1(\mu,\nu)$. Exactly the same proof applies if $\nu$ is separable, so that $W_0(\mu,\nu) = W_1(\mu,\nu)$ even if $\nu$ is separable. Since $W_1(\mu,\nu) = W_1(\nu,\mu)$, one also obtains $W(\mu,\nu) = W_0(\mu,\nu) = W_1(\mu,\nu)$ if $\mu$ or $\nu$ is separable. □

We do not know whether $W_0(\mu,\nu) = W_0(\nu,\mu)$ for *all* $\mu$, $\nu \in \mathcal{M}(\mathcal{B})$.

4.3. **A metric version of Skorohod theorem.** While disintegrability is useful in Theorem 4, it is even crucial in the next version of Skorohod theorem. Indeed, existence of disintegrations makes the proof transparent and simple. We also note that disintegrability underlies the usual proofs of Skorohod theorem. To our knowledge, when $\mu_n \to \mu_0$ weakly and $\mu_0$ is separable, the random variables $X_n$ are constructed such that $\mathcal{L}(X_0,X_n) \in \mathcal{D}(\mu_0,\mu_n)$ where $\mathcal{L}(X_0,X_n)$ is the probability law of $(X_0,X_n)$; see Theorem 1.10.4 of van der Vaart and Wellner (1996) and Theorem 3.5.1 of Dudley (1999).

**Theorem 5.** *Suppose $\sigma(d) \subset \mathcal{B}\otimes\mathcal{B}$. Then, $W_0(\mu_0,\mu_n) \longrightarrow 0$ if and only if there are a probability space $(\Omega,\mathcal{A},P)$ and random variables $X_n : \Omega \to S$ satisfying*

$$X_n \sim \mu_n \text{ and } \mathcal{L}(X_0,X_n) \text{ is disintegrable for each } n \geq 0, \quad X_n \xrightarrow{P} X_0.$$

*In particular, there is a Skorohod representation in case $W(\mu_n,\mu_0) \longrightarrow 0$.*

*Proof.* As to the "if" part, since $\mathcal{L}(X_0, X_n) \in \mathcal{D}(\mu_0, \mu_n)$,

$$W_0(\mu_0, \mu_n) \le E_P\big\{1 \wedge d(X_0, X_n)\big\} \longrightarrow 0.$$

We next turn to the "only if" part. Let $(\Omega, \mathcal{A}) = (S^\infty, \mathcal{B}^\infty)$ and $X_n : S^\infty \to S$ the $n$-th canonical projection, $n \ge 0$. Fix $P_n \in \mathcal{D}(\mu_0, \mu_n)$ such that $E_{P_n}\big\{1 \wedge d(X, Y)\big\} < \frac{1}{n} + W_0(\mu_0, \mu_n)$ and a disintegration $\alpha_n$ for $P_n$. By Ionescu-Tulcea theorem, there is a unique probability $P$ on $\mathcal{B}^\infty$ such that $X_0 \sim \mu_0$ and

$$\beta_n(x_0, x_1, \ldots, x_{n-1})(A) = \alpha_n(x_0)(A), \quad (x_0, x_1, \ldots, x_{n-1}) \in S^n,\ A \in \mathcal{B},$$

is a regular version of the conditional distribution of $X_n$ given $(X_0, X_1, \ldots, X_{n-1})$ for all $n \ge 1$. (Note that $X_n$ is conditionally independent of $(X_1, \ldots, X_{n-1})$ given $X_0$). To conclude the proof, it suffices noting that $\mathcal{L}(X_0, X_n) = P_n$ and

$$\epsilon\, P\big(d(X_0, X_n) > \epsilon\big) \le E_P\big\{1 \wedge d(X_0, X_n)\big\} < \frac{1}{n} + W_0(\mu_0, \mu_n) \longrightarrow 0 \ \text{ for all } \epsilon \in (0, 1).$$

$\square$

**Remark 6.** Let $h : S \times S \to [0, \infty)$ be a function such that $\sigma(h) \subset \mathcal{B} \otimes \mathcal{B}$ and

$$W_h(\mu, \nu) = \inf_{P \in \mathcal{D}(\mu, \nu)} E_P\big\{1 \wedge h(X, Y)\big\} \quad \text{for all } \mu,\ \nu \in \mathcal{M}(\mathcal{B}).$$

For instance, $h$ could be another distance on $S$, stronger than $d$, but measurable with respect to $\mathcal{B} \otimes \mathcal{B}$. Then, $W_h(\mu_0, \mu_n) \longrightarrow 0$ if and only if $h(X_0, X_n) \longrightarrow 0$ in probability for some $S$-valued random variables $X_n$ such that $X_n \sim \mu_n$ and $\mathcal{L}(X_0, X_n)$ is disintegrable for all $n \ge 0$. Up to replacing $d$ with $h$, this can be proved exactly as Theorem 5.

It is not hard to prove that $W(\mu_n, \mu_0) \to 0$ if $\mu_n \to \mu_0$ weakly and $\mu_0$ is separable. Thus, Theorem 5 implies the usual Skorohod theorem provided $\sigma(d) \subset \mathcal{B} \otimes \mathcal{B}$ and almost uniform convergence is weakened into convergence in probability.

A last question is whether convergence in probability can be replaced by almost uniform convergence in a Skorohod representation. More precisely, suppose a Skorohod representation is available, that is, $X_n \sim \mu_n$ for all $n$ and $X_n \to X_0$ in probability for some $S$-valued random variables $X_n$. In this case, are there $S$-valued random variables $Y_n$ such that $Y_n \sim \mu_n$ for all $n$ and $Y_n \to Y_0$ on a set of probability 1 ? By Skorohod theorem, the answer is yes if $\mu_0$ is separable. In particular, the answer is yes if, consistently with ZFC, non separable probability measures fail to exist. As we now prove, however, the answer is no if non separable probabilities exist. Thus, in the spirit of this paper, a Skorohod representation cannot be strengthened by asking almost uniform convergence (or even a.s. convergence) instead of convergence in probability.

**Example 7.** The notation is the same as Example 3. Indeed, we argue essentially as in such example and we use a result from Sethuraman (2002). Recall that existence of a non separable Borel probability on a metric space is equiconsistent with existence of a countably additive extension of $m$ to the power set of $(0, 1)$.

Let $(f_n : n \ge 1)$ be a sequence of real random variables, on the probability space $((0, 1), \mathcal{B}_{(0,1)}, m)$, satisfying

$$f_n \ge 0, \quad E_m(f_n) = 1, \quad \lim_n E_m|f_n - 1| = 0, \quad m\big(\liminf_n f_n < 1\big) > 0.$$

Let $S = (0, 1)$, equipped with 0-1 distance, and let $\mu_0$ be a countably additive extension of $m$ to the power set of $(0, 1)$. Define further $\mu_n(B) = E_{\mu_0}\big(f_n\, I_B\big)$ for all

$n \geq 1$ and $B \in \mathcal{B}$. Since $\|\mu_n - \mu_0\| = \frac{1}{2} E_m |f_n - 1| \longrightarrow 0$, Proposition 1 implies the existence of a Skorohod representation. Suppose now that $Y_n \sim \mu_n$ for all $n \geq 0$, where the $Y_n$ are $S$-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$. Then, since $d$ is 0-1 distance, $\sigma(d) \subset \mathcal{B} \otimes \mathcal{B}$, and $m\big(\liminf_n f_n < 1\big) > 0$, Theorem 3.1 of Sethuraman (2002) implies

$$P(Y_n \longrightarrow Y_0) = P\big(Y_n = Y_0 \text{ ultimately}\big) < 1.$$

## References

Berti P., Rigo P. (1999) Sufficient conditions for the existence of disintegrations, *J. Theoret. Probab.*, 12, 75-86.

Berti P., Pratelli L., Rigo P. (2009) A Skorohod representation theorem for uniform distance, *Submitted*, currently available at: http://economia.unipv.it/pagp/pagine_personali/prigo/skorptrf.pdf

Dudley R.M. (1968) Distances of probability measures and random variables, *Ann. Math. Statist.*, 39, 1563-1572.

Dudley R.M. (1999) *Uniform central limit theorems*, Cambridge University Press.

Fuchino S., Greenberg N., Shelah S. (2006) Models of real-valued measurability, *Ann. Pure Appl. Logic*, 142, 380-397.

Goldring N. (1995) Measures: back and forth between point sets and large sets, *Bull. Symbolic Logic*, vol. 1, n. 2, 170-188.

Maitra A., Ramakrishnan S. (1988) Factorization of measures and normal conditional distributions, *Proc. Amer. Math. Soc.*, 103, 1259-1267.

Sethuraman J. (2002) Some extensions of the Skorohod representation theorem, *Sankhya*, 64, 884-893.

Skorohod A.V. (1956) Limit theorems for stochastic processes, *Theo. Probab. Appl.*, 1, 261-290.

van der Vaart A., Wellner J.A. (1996) *Weak convergence and empirical processes*, Springer.

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.

Patrizia Berti, Dipartimento di Matematica Pura ed Applicata "G. Vitali", Universita' di Modena e Reggio-Emilia, via Campi 213/B, 41100 Modena, Italy
*E-mail address*: berti.patrizia@unimore.it

Luca Pratelli, Accademia Navale, viale Italia 72, 57100 Livorno, Italy
*E-mail address*: pratel@mail.dm.unipi.it

Pietro Rigo (corresponding author), Dipartimento di Economia Politica e Metodi Quantitativi, Universita' di Pavia, via S. Felice 5, 27100 Pavia, Italy
*E-mail address*: prigo@eco.unipv.it