A SKOROHOD REPRESENTATION THEOREM WITHOUT SEPARABILITY

PATRIZIA BERTI, LUCA PRATELLI, AND PIETRO RIGO

Abstract. Let (S, d) be a metric space, $\mathcal{G}$ a σ-field on S and $(\mu_n : n \geq 0)$ a sequence of probabilities on $\mathcal{G}$. Suppose $\mathcal{G}$ countably generated, the map $(x, y) \mapsto d(x, y)$ measurable with respect to $\mathcal{G} \otimes \mathcal{G}$, and μ_n perfect for $n > 0$. Say that (μ_n) has a Skorohod representation if, on some probability space, there are random variables X_n such that

$$X_n \sim \mu_n \text{ for all } n \geq 0 \quad \text{and} \quad d(X_n, X_0) \xrightarrow{P} 0.$$

It is shown that (μ_n) has a Skorohod representation if and only if

$$\lim_n \sup_f |\mu_n(f) - \mu_0(f)| = 0,$$

where sup is over those $f : S \to [-1, 1]$ which are $\mathcal{G}$-universally measurable and satisfy $|f(x) - f(y)| \leq 1 \wedge d(x, y)$. An useful consequence is that Skorohod representations are preserved under mixtures. The result applies even if μ_0 fails to be d-separable. Some possible applications are given as well.

1. Motivations and results

Throughout, (S, d) is a metric space, $\mathcal{G}$ a σ-field of subsets of S and $(\mu_n : n \geq 0)$ a sequence of probability measures on $\mathcal{G}$. For each probability μ on $\mathcal{G}$, we write $\mu(f) = \int f \, d\mu$ provided $f \in L_1(\mu)$ and we say that μ is *d-separable* if $\mu(B) = 1$ for some d-separable $B \in \mathcal{G}$. Also, we let $\mathcal{B}$ denote the Borel σ-field on S under d.

If

$$\mathcal{G} = \mathcal{B}, \quad \mu_n \to \mu_0 \text{ weakly}, \quad \mu_0 \text{ is } d\text{-separable},$$

there are S-valued random variables X_n, defined on some probability space, such that $X_n \sim \mu_n$ for all $n \geq 0$ and $X_n \to X_0$ almost uniformly. This is Skorohod representation theorem (SRT) as it appears after Skorohod [12], Dudley [5] and Wichura [14]. See page 130 of [6] and page 77 of [13] for some historical notes.

Versions of SRT which allow for $\mathcal{G} \subset \mathcal{B}$ are also available; see Theorem 1.10.3 of [13]. However, separability of μ_0 is still fundamental. Furthermore, unlike μ_n for $n > 0$, the limit law μ_0 must be defined on all of $\mathcal{B}$.

Thus SRT does not apply, neither indirectly, when μ_0 is defined on some $\mathcal{G} \neq \mathcal{B}$ and is not d-separable. This precludes some potentially interesting applications.

For instance, $\mathcal{G}$ could be the Borel σ-field under some distance d^* on S weaker than d, but one aims to realize the μ_n by random variables X_n which converge under the stronger distance d. To fix ideas, S could be some collection of real bounded functions, $\mathcal{G}$ the σ-field generated by the canonical projections and d the uniform distance. Then, in some meaningful situations, $\mathcal{G}$ agrees with the Borel

2010 *Mathematics Subject Classification.* 60B10, 60A05, 60A10.

Key words and phrases. Convergence of probability measures – Perfect probability measure – Separable probability measure – Skorohod representation theorem – Uniform distance.

σ-field under a distance d^* on S weaker than d. Yet, one can try to realize the μ_n by random variables X_n which converge uniformly (and not only under d^*). In such situations, SRT and its versions do not apply unless μ_0 is d-separable.

The following two remarks are also in order.

Suppose first $\mathcal{G} = \mathcal{B}$. Existence of non d-separable laws on $\mathcal{B}$ can not be excluded a priori, unless some assumption beyond ZFC (the usual axioms of set theory) is made; see Section 1 of [2]. And, if non d-separable laws on $\mathcal{B}$ exist, d-separability of μ_0 cannot be dropped from SRT, even if almost uniform convergence is weakened into convergence in probability. Indeed, it may be that $\mu_n \to \mu_0$ weakly but no random variables X_n satisfy $X_n \sim \mu_n$ for all $n \geq 0$ and $X_n \to X_0$ in probability. We refer to Example 4.1 of [2] for details.

More importantly, if $\mathcal{G} \neq \mathcal{B}$, non d-separable laws on $\mathcal{G}$ are quite usual. There are even laws μ on $\mathcal{G}$ such that $\mu(B) = 0$ for all d-separable $B \in \mathcal{B}$. A popular example is

$$S = D[0,1], \quad d = \text{uniform distance}, \quad \mathcal{G} = \text{Borel } \sigma\text{-field under Skorohod topology},$$

where $D[0,1]$ is the set of real cadlag functions on $[0,1]$. To be concise, this particular case is called *the motivating example* in the sequel. In this framework, $\mathcal{G}$ includes all d-separable members of $\mathcal{B}$. Further, the probability distribution μ of a cadlag process with jumps at random time points is typically non d-separable. Suppose in fact that one of the jump times of such process, say τ, has a diffuse distribution. If $B \in \mathcal{B}$ is d-separable, then

$$J_B = \{t \in (0,1] : \Delta x(t) \neq 0 \text{ for some } x \in B\}$$

is countable. Since τ has a diffuse distribution, it follows that

$$\mu(B) \leq \text{Prob}(\tau \in J_B) = 0.$$

This paper provides a version of SRT which applies to $\mathcal{G} \neq \mathcal{B}$ and does not request d-separability of μ_0. We begin with a definition.

The sequence (μ_n) is said to admit a *Skorohod representation* if

> On some probability space $(\Omega, \mathcal{A}, P)$, there are measurable maps $X_n : (\Omega, \mathcal{A}) \to (S, \mathcal{G})$ such that $X_n \sim \mu_n$ for all $n \geq 0$ and
>
> $$P^*\big(d(X_n, X_0) > \epsilon\big) \longrightarrow 0, \quad \text{for all } \epsilon > 0,$$
>
> where P^* denotes the P-outer measure.

Note that almost uniform convergence has been weakened into convergence in (outer) probability. In fact, it may be that (μ_n) admits a Skorohod representation and yet no random variables Y_n satisfy $Y_n \sim \mu_n$ for all $n \geq 0$ and $Y_n \to Y_0$ on a set of probability 1. See Example 7 of [3].

Note also that, if the map $d : S \times S \to \mathbb{R}$ is measurable with respect to $\mathcal{G} \otimes \mathcal{G}$, convergence in outer probability reduces to $d(X_n, X_0) \xrightarrow{P} 0$. In turn, $d(X_n, X_0) \xrightarrow{P} 0$ if and only if

$$(1) \qquad \text{each subsequence } (n_j) \text{ contains a further subsequence } (n_{j_k}) \text{ such that } X_{n_{j_k}} \longrightarrow X_0 \text{ almost uniformly.}$$

Thus, in a sense, Skorohod representations are in the spirit of [8]. Furthermore, as noted in [8], condition (1) is exactly what is needed in most applications.

Let L denote the set of functions $f : S \to \mathbb{R}$ satisfying

$$-1 \leq f \leq 1, \quad \sigma(f) \subset \widehat{\mathcal{G}}, \quad |f(x) - f(y)| \leq 1 \wedge d(x, y) \text{ for all } x, y \in S,$$

where $\widehat{\mathcal{G}}$ is the universal completion of $\mathcal{G}$. If $X_n \sim \mu_n$ for each $n \geq 0$, with the X_n all defined on the probability space $(\Omega, \mathcal{A}, P)$, then

$$\begin{aligned} |\mu_n(f) - \mu_0(f)| &= |E_P f(X_n) - E_P f(X_0)| \leq E_P |f(X_n) - f(X_0)| \\ &\leq \epsilon + 2\, P^*\big(d(X_n, X_0) > \epsilon\big) \quad \text{for all } f \in L \text{ and } \epsilon > 0. \end{aligned}$$

Thus, a necessary condition for (μ_n) to admit a Skorohod representation is

$$\lim_n \sup_{f \in L} |\mu_n(f) - \mu_0(f)| = 0. \tag{2}$$

Furthermore, condition (2) is equivalent to $\mu_n \to \mu_0$ weakly if $\mathcal{G} = \mathcal{B}$ and μ_0 is d-separable. So, when $\mathcal{G} = \mathcal{B}$, it is tempting *to conjecture that*: (μ_n) admits a Skorohod representation if and only if condition (2) holds. If true, this conjecture would be an improvement of SRT, not requesting separability of μ_0. In particular, the conjecture is actually true if d is 0-1 distance; see Proposition 3.1 of [2] and Theorem 2.1 of [11].

We do not know whether such conjecture holds in general, since we were able to prove the equivalence between Skorohod representation and condition (2) only under some conditions on $\mathcal{G}$, d and μ_n. Our main results are in fact the following.

Theorem 1. *Suppose μ_n is perfect for all $n > 0$, $\mathcal{G}$ is countably generated, and $d : S \times S \to \mathbb{R}$ is measurable with respect to $\mathcal{G} \otimes \mathcal{G}$. Then, $(\mu_n : n \geq 0)$ admits a Skorohod representation if and only if condition* (2) *holds.*

Under the assumptions of Theorem 1, $\mathcal{G}$ is the Borel σ-field for some separable distance d^* on S. Condition (2) can be weakened into

$$\lim_n \sup_{f \in M} |\mu_n(f) - \mu_0(f)| = 0, \quad \text{where } M = \{f \in L : \sigma(f) \subset \mathcal{G}\}, \tag{3}$$

provided $d : S \times S \to \mathbb{R}$ is lower semicontinuous in the d^*-topology.

Theorem 2. *Suppose*

(i) *μ_n is perfect for all $n > 0$;*
(ii) *$\mathcal{G}$ is the Borel σ-field under a distance d^* on S such that (S, d^*) is separable;*
(iii) *$d : S \times S \to \mathbb{R}$ is lower semicontinuous when S is given the d^*-topology.*

Then, $(\mu_n : n \geq 0)$ admits a Skorohod representation if and only if condition (3) *holds.*

One consequence of Theorem 2 is that Skorohod representations are preserved under mixtures. Since this fact is useful in real problems, we discuss it in some detail. Let $(\mathcal{X}, \mathcal{E}, Q)$ be a probability space, and for every $n \geq 0$, let

$$\{\alpha_n(x) : x \in \mathcal{X}\}$$

be a measurable collection of probability measures on $\mathcal{G}$. Measurability means that $x \mapsto \alpha_n(x)(A)$ is $\mathcal{E}$-measurable for fixed $A \in \mathcal{G}$.

Corollary 3. *Assume conditions (i)-(ii)-(iii) and*

$$\mu_n(A) = \int \alpha_n(x)(A)\, Q(dx) \quad \text{for all } n \geq 0 \text{ and } A \in \mathcal{G}.$$

Then, $(\mu_n : n \geq 0)$ has a Skorohod representation provided $(\alpha_n(x) : n \geq 0)$ has a Skorohod representation for Q-almost all $x \in \mathcal{X}$. In particular, $(\mu_n : n \geq 0)$ admits a Skorohod representation whenever $\mathcal{G} \subset \mathcal{B}$ and, for Q-almost all $x \in \mathcal{X}$,

$$\alpha_0(x) \text{ is } d\text{-separable} \quad \text{and} \quad \alpha_n(x)(f) \longrightarrow \alpha_0(x)(f) \text{ for each } f \in M.$$

Various examples concerning Theorems 1-2 and Corollary 3 are given in Section 3. Here, we close this section by some remarks.

- (j) Theorems 1-2 unify some known results; see Examples 6 and 7.
- (jj) Theorems 1-2 are proved by joining some ideas on disintegrations and a duality result from optimal transportation theory; see [2] and [10].
- (jjj) Each probability on $\mathcal{G}$ is perfect if $\mathcal{G}$ is the Borel σ-field under some distance d^* such that (S, d^*) is a universally measurable subset of a Polish space. This happens in the motivating example.
- (jv) Even if perfect for $n > 0$, the μ_n may be far from being d-separable. In the motivating example, each probability μ on $\mathcal{G}$ is perfect and yet various interesting μ satisfy $\mu(B) = 0$ for each d-separable $B \in \mathcal{B}$.
- (v) Theorems 1-2 are essentially motivated from the application mentioned at the beginning, where $\mathcal{G}$ is the Borel σ-field under a distance d^* weaker than d. This actually happens in the motivating example and in most examples of Section 3.
- (vj) By Theorem 1, to prove existence of Skorohod representations, one can "argue by subsequences". Precisely, under the conditions of Theorem 1, $(\mu_n : n \geq 0)$ has a Skorohod representation if and only if each subsequence $(\mu_0, \mu_{n_j} : j \geq 1)$ contains a further subsequence $(\mu_0, \mu_{n_{j_k}} : k \geq 1)$ which admits a Skorohod representation.
- (vjj) In real problems, unless μ_0 is d-separable, checking conditions (2)-(3) is usually hard. However, conditions (2)-(3) are necessary for a Skorohod representation (so that they can not be eluded). Furthermore, in some cases, conditions (2)-(3) may be verified with small effort. One such case is Corollary 3. Other cases are exchangeable empirical processes and pure jump processes, as defined in Examples 9-10 of [3]. One more situation, where SRT does not work but conditions (2)-(3) are easily checked, is displayed in forthcoming Example 11.

2. Proofs

2.1. **Preliminaries.** Let $(\mathcal{X}, \mathcal{E})$ and $(\mathcal{Y}, \mathcal{F})$ be measurable spaces.

In the sequel, $\mathcal{P}(\mathcal{E})$ denotes the set of probability measures on $\mathcal{E}$. The universal completion of $\mathcal{E}$ is

$$\widehat{\mathcal{E}} = \bigcap_{\mu \in \mathcal{P}(\mathcal{E})} \overline{\mathcal{E}}^{\mu}$$

where $\overline{\mathcal{E}}^{\mu}$ is the completion of $\mathcal{E}$ with respect to μ.

Let $H \subset \mathcal{X} \times \mathcal{Y}$ and let $\Pi : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ be the canonical projection onto $\mathcal{X}$. By the projection theorem, if $\mathcal{Y}$ is a Borel subset of a Polish space, $\mathcal{F}$ the Borel σ-field and $H \in \mathcal{E} \otimes \mathcal{F}$, then

$$\Pi(H) = \{x \in \mathcal{X} : (x, y) \in H \text{ for some } y \in \mathcal{Y}\} \in \widehat{\mathcal{E}};$$

see e.g. Theorem A1.4, page 562, of [9]. Another useful fact is the following.

Lemma 4. *Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces. If $\mathcal{Y}$ is compact and $H \subset \mathcal{X} \times \mathcal{Y}$ closed, then $\Pi(H)$ is a countable intersection of open sets (i.e., $\Pi(H)$ is a G_δ-set).*

Proof. Let $H_n = \{(x,y) : \rho\big[(x,y),\, H\big] < 1/n\}$, where ρ is any distance on $\mathcal{X} \times \mathcal{Y}$ inducing the product topology. Since H is closed, $H = \cap_n H_n$. Since H_n is open, $\Pi(H_n)$ is still open. Thus, it suffices to prove $\Pi(H) = \cap_n \Pi(H_n)$. Trivially, $\Pi(H) \subset \cap_n \Pi(H_n)$. Fix $x \in \cap_n \Pi(H_n)$. For each n, take $y_n \in \mathcal{Y}$ such that $(x, y_n) \in H_n$. Since $\mathcal{Y}$ is compact, $y_{n_j} \to y$ for some $y \in \mathcal{Y}$ and subsequence (n_j). Hence,

$$\rho\big[(x,y),\, H\big] = \lim_j \rho\big[(x,y_{n_j}),\, H\big] \le \liminf_j \frac{1}{n_j} = 0.$$

Since H is closed, $(x,y) \in H$. Hence, $x \in \Pi(H)$ and $\Pi(H) = \cap_n \Pi(H_n)$. □

A probability $\mu \in \mathcal{P}(\mathcal{E})$ is *perfect* if, for each $\mathcal{E}$-measurable function $f : \mathcal{X} \to \mathbb{R}$, there is a Borel subset B of $\mathbb{R}$ such that $B \subset f(\mathcal{X})$ and $\mu(f \in B) = 1$. If $\mathcal{X}$ is separable metric and $\mathcal{E}$ the Borel σ-field, then μ is perfect if and only if it is tight. In particular, every $\mu \in \mathcal{P}(\mathcal{E})$ is perfect if $\mathcal{X}$ is a universally measurable subset of a Polish space and $\mathcal{E}$ the Borel σ-field.

Finally, in this paper, a disintegration is meant as follows. Let $\gamma \in \mathcal{P}(\mathcal{E} \otimes \mathcal{F})$ and let $\mu(\cdot) = \gamma(\cdot \times \mathcal{Y})$ and $\nu(\cdot) = \gamma(\mathcal{X} \times \cdot)$ be the marginals of γ. Then, γ is said to be *disintegrable* if there is a collection $\{\alpha(x) : x \in \mathcal{X}\}$ such that:

- $\alpha(x) \in \mathcal{P}(\mathcal{F})$ for each $x \in \mathcal{X}$;
- $x \mapsto \alpha(x)(B)$ is $\mathcal{E}$-measurable for each $B \in \mathcal{F}$;
- $\gamma(A \times B) = \int_A \alpha(x)(B)\, \mu(dx)$ for all $A \in \mathcal{E}$ and $B \in \mathcal{F}$.

The collection $\{\alpha(x) : x \in \mathcal{X}\}$ is called a *disintegration* for γ.

A disintegration can fail to exist. However, for γ to admit a disintegration, it suffices that $\mathcal{F}$ is countably generated and ν perfect.

2.2. **Proof of Theorem 1.** The "only if" part has been proved in Section 1. Suppose condition (2) holds. For μ, $\nu \in \mathcal{P}(\mathcal{G})$, define

$$W_0(\mu,\nu) = \inf_{\gamma \in \mathcal{D}(\mu,\nu)} E_\gamma(1 \wedge d) \quad \text{where}$$

$$\mathcal{D}(\mu,\nu) = \{\gamma \in \mathcal{P}(\mathcal{G} \otimes \mathcal{G}) : \gamma \text{ disintegrable}, \ \gamma(\cdot \times S) = \mu(\cdot),\ \gamma(S \times \cdot) = \nu(\cdot)\}.$$

Disintegrations have been defined in Subsection 2.1. Note that $\mathcal{D}(\mu,\nu) \neq \emptyset$ as $\mathcal{D}(\mu,\nu)$ includes at least the product law $\mu \times \nu$.

The proof of the "if" part can be split into two steps.

Step 1. Arguing as in Theorem 4.2 of [2], it suffices to show $W_0(\mu_0,\mu_n) \to 0$. Define in fact $(\Omega, \mathcal{A}) = (S^\infty, \mathcal{G}^\infty)$ and $X_n : S^\infty \to S$ the n-th canonical projection, $n \ge 0$. For each $n > 0$, take $\gamma_n \in \mathcal{D}(\mu_0,\mu_n)$ such that $E_{\gamma_n}(1 \wedge d) < \frac{1}{n} + W_0(\mu_0,\mu_n)$. Fix also a disintegration $\{\alpha_n(x) : x \in S\}$ for γ_n and define

$$\beta_n(x_0, x_1, \ldots, x_{n-1})(B) = \alpha_n(x_0)(B)$$

for all $(x_0, x_1, \ldots, x_{n-1}) \in S^n$ and $B \in \mathcal{G}$. By Ionescu-Tulcea theorem, there is a unique probability P on $\mathcal{A} = \mathcal{G}^\infty$ such that $X_0 \sim \mu_0$ and β_n is a version of the conditional distribution of X_n given $(X_0, X_1, \ldots, X_{n-1})$ for all $n > 0$. Then,

$$P\big(X_0 \in A,\, X_n \in B\big) = \int_A \alpha_n(x)(B)\, \mu_0(dx) = \gamma_n(A \times B)$$

for all $n > 0$ and $A, B \in \mathcal{G}$. In particular, $P(X_n \in \cdot) = \mu_n(\cdot)$ for all $n \geq 0$ and

$$E_P\{1 \wedge d(X_0, X_n)\} = E_{\gamma_n}(1 \wedge d) < \frac{1}{n} + W_0(\mu_0, \mu_n).$$

Step 2. If $\mu, \nu \in \mathcal{P}(\mathcal{G})$ and ν is perfect, then

$$W_0(\mu, \nu) = \sup_{f \in L} |\mu(f) - \nu(f)|. \tag{4}$$

Under (4), $W_0(\mu_0, \mu_n) \to 0$ because of condition (2) and μ_n perfect for $n > 0$. Thus, the proof is concluded by Step 1.

To get condition (4), it is enough to prove $W_0(\mu, \nu) \leq \sup_{f \in L} |\mu(f) - \nu(f)|$. (The opposite inequality is in fact trivial). Define $\Gamma(\mu, \nu)$ to be the collection of those $\gamma \in \mathcal{P}(\mathcal{G} \otimes \mathcal{G})$ satisfying $\gamma(\cdot \times S) = \mu(\cdot)$ and $\gamma(S \times \cdot) = \nu(\cdot)$. By a duality result in [10], since ν is perfect and $1 \wedge d$ bounded and $\mathcal{G} \otimes \mathcal{G}$-measurable, one obtains

$$\inf_{\gamma \in \Gamma(\mu,\nu)} E_\gamma(1 \wedge d) = \sup_{(g,h)} \{\mu(g) + \nu(h)\}$$

where sup is over those pairs (g, h) of real $\mathcal{G}$-measurable functions on S such that

$$g \in L_1(\mu), \quad h \in L_1(\nu), \quad g(x) + h(y) \leq 1 \wedge d(x, y) \text{ for all } x, y \in S. \tag{5}$$

Since $\mathcal{G}$ is countably generated and ν perfect, each $\gamma \in \Gamma(\mu, \nu)$ is disintegrable. Thus, $\Gamma(\mu, \nu) = \mathcal{D}(\mu, \nu)$ and $W_0(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} E_\gamma(1 \wedge d)$. Given $\epsilon > 0$, take a pair (g, h) satisfying condition (5) as well as $W_0(\mu, \nu) < \epsilon + \mu(g) + \nu(h)$.

Since $\{(x, x) : x \in S\} = \{d = 0\} \in \mathcal{G} \otimes \mathcal{G}$, then $\mathcal{G}$ includes the singletons. As $\mathcal{G}$ is also countably generated, $\mathcal{G}$ is the Borel σ-field on S under some distance d^* such that (S, d^*) is separable; see [4]. Then ν is tight, with respect to d^*, for it is perfect. By tightness, $\nu(A) = 1$ for some σ-compact set $A \in \mathcal{G}$. For $(x, a) \in S \times A$, define

$$u(x, a) = 1 \wedge d(x, a) - h(a) \quad \text{and} \quad \phi(x) = \inf_{a \in A} u(x, a).$$

Since A is σ-compact, A is homeomorphic to a Borel subset of a Polish space. (In fact, A is easily seen to be homeomorphic to a σ-compact subset of $[0, 1]^\infty$). Let $b \in \mathbb{R}$ and $\mathcal{G}_A = \{A \cap B : B \in \mathcal{G}\}$. Since $\{u < b\} \in \mathcal{G} \otimes \mathcal{G}_A$, one obtains

$$\{\phi < b\} = \{x \in S : u(x, a) < b \text{ for some } a \in A\} \in \widehat{\mathcal{G}}$$

by the projection theorem applied with $(\mathcal{X}, \mathcal{E}) = (S, \mathcal{G})$, $(\mathcal{Y}, \mathcal{F}) = (A, \mathcal{G}_A)$ and $H = \{u < b\}$. Thus, ϕ is $\widehat{\mathcal{G}}$-measurable. Furthermore,

$$\phi(x) - \phi(y) = \inf_{a \in A} u(x, a) + \sup_{a \in A} \{-u(y, a)\}$$
$$\leq \sup_{a \in A} \{1 \wedge d(x, a) - 1 \wedge d(y, a)\} \leq 1 \wedge d(x, y) \quad \text{for all } x, y \in S.$$

Fix $x_0 \in S$ and define $f = \phi - \phi(x_0)$. Since $|f(x)| = |\phi(x) - \phi(x_0)| \leq 1 \wedge d(x, x_0) \leq 1$ for all $x \in S$, then $f \in L$. On noting that

$$g(x) \leq u(x, a) \text{ for } (x, a) \in S \times A \quad \text{and} \quad \phi(x) + h(x) \leq 1 \wedge d(x, x) = 0 \text{ for } x \in A,$$

one also obtains $g - \phi(x_0) \leq f$ on all of S and $h + \phi(x_0) \leq -f$ on A. Since $\nu(A) = 1$,

$$W_0(\mu, \nu) - \epsilon < \mu(g) + \nu(h) = \mu\{g - \phi(x_0)\} + \nu\{h + \phi(x_0)\}$$
$$\leq \mu(f) - \nu(f) \leq \sup_{\varphi \in L} |\mu(\varphi) - \nu(\varphi)|.$$

This concludes the proof.

2.3. **Proof of Theorem 2.** Assume conditions (i)-(ii)-(iii). Arguing as in Subsection 2.2 (and using the same notation) it suffices to prove that ϕ is $\mathcal{G}$-measurable.

Since A is σ-compact (under d^*),

$$\phi(x) = \inf_n \inf_{a \in A_n} u(x, a)$$

where the A_n are compacts such that $A = \cup_n A_n$. Hence, for proving $\mathcal{G}$-measurability of ϕ, it can be assumed A compact. On noting that

$$\nu(h) = \sup\{\, \nu(k) : k \leq h,\ k \text{ upper semicontinuous}\},$$

the function h can be assumed upper semicontinuous. (Otherwise, just replace h with an upper semicontinuous k such that $k \leq h$ and $\nu(h - k)$ is small). In this case, u is lower semicontinuous, since both $1 \wedge d$ and $-h$ are lower semicontinuous.

Since A is compact and u lower semicontinuous, ϕ can be written as $\phi(x) = \min_{a \in A} u(x, a)$ and this implies

$$\{\phi \leq b\} = \{x \in S : u(x, a) \leq b \text{ for some } a \in A\} \quad \text{for all } b \in \mathbb{R}.$$

Therefore, $\{\phi \leq b\} \in \mathcal{G}$ because of Lemma 4 applied with $\mathcal{X} = S$, $\mathcal{Y} = A$ and $H = \{u \leq b\}$ which is closed for u is lower semicontinuous. This concludes the proof.

2.4. **Proof of Corollary 3.** Fix a countable subset $M^* \subset M$ satisfying

$$\sup_{f \in M^*} |\mu_n(f) - \mu_0(f)| = \sup_{f \in M} |\mu_n(f) - \mu_0(f)| \quad \text{for all } n > 0.$$

The first part of Corollary 3 follows from Theorem 2 and

$$\sup_{f \in M} |\mu_n(f) - \mu_0(f)| \leq \int \sup_{f \in M^*} |\alpha_n(x)(f) - \alpha_0(x)(f)|\, Q(dx) \longrightarrow 0.$$

As to the second part, suppose $\mathcal{G} \subset \mathcal{B}$ and fix a sequence $(\nu_n : n \geq 0)$ of probabilities on $\mathcal{G}$. It suffices to show that (ν_n) has a Skorohod representation whenever

(6) $\qquad \nu_0$ is d-separable and $\nu_n(f) \to \nu_0(f)$ for each $f \in M$.

Let $\mathcal{U}$ be the σ-field on S generated by the d-balls. For all $r > 0$ and $x \in S$, since $\{d < r\} \in \mathcal{G} \otimes \mathcal{G}$ then $\{y : d(x, y) < r\} \in \mathcal{G}$. Thus, $\mathcal{U} \subset \mathcal{G}$. Next, assume condition (6) and take a d-separable set $A \in \mathcal{G}$ with $\nu_0(A) = 1$. Since A is d-separable,

$$A \cap B \in \mathcal{U} \subset \mathcal{G} \quad \text{for all } B \in \mathcal{B}.$$

Define $\lambda_0(B) = \nu_0(A \cap B)$ for all $B \in \mathcal{B}$ and

$$(\Omega_0, \mathcal{A}_0, P_0) = (S, \mathcal{B}, \lambda_0), \quad (\Omega_n, \mathcal{A}_n, P_n) = (S, \mathcal{G}, \nu_n) \text{ for each } n > 0,$$
$$I_n = \text{ identity map on } S \text{ for each } n \geq 0.$$

In view of (6), since $\mathcal{U} \subset \mathcal{G}$ and I_0 has a d-separable law, $I_n \to I_0$ in distribution (under d) according to Hoffmann-Jørgensen's definition; see Theorem 1.7.2, page 45, of [13]. Thus, since $\mathcal{G} \subset \mathcal{B}$, a Skorohod representation for (ν_n) follows from Theorem 1.10.3, page 58, of [13]. This concludes the proof.

Remark 5. Let N be the collection of functions $f : S \to \mathbb{R}$ of the form

$$f(x) = \min_{1 \leq i \leq n} \{1 \wedge d(x, A_i) - b_i\}$$

for all $n \geq 1$, $b_1, \ldots, b_n \in \mathbb{R}$ and $A_1, \ldots, A_n \in \mathcal{G}$. Theorems 1 and 2 are still true if conditions (2) and (3) are replaced by

$$\lim_n \sup_{f \in L \cap N} |\mu_n(f) - \mu_0(f)| = 0 \quad \text{and} \quad \lim_n \sup_{f \in M \cap N} |\mu_n(f) - \mu_0(f)| = 0,$$

respectively. In fact, in the notation of the above proofs, it is not hard to see that h can be taken to be a simple function. In this case, writing down ϕ explicitly, one verifies that $f = \phi - \phi(x_0) \in N$.

3. Examples

As remarked in Section 1, Theorems 1-2 unify some known results and yield new information as well. We illustrate these facts by a few examples.

Example 6. Consider the motivating example, that is, $S = D[0,1]$, d the uniform distance and $\mathcal{G}$ the Borel σ-field under Skorohod distance d^*. Given x, $y \in D[0,1]$, we recall that $d^*(x,y)$ is the infimum of those $\epsilon > 0$ such that

$$\sup_t |x(t) - y \circ \lambda(t)| \leq \epsilon \quad \text{and} \quad \sup_{s \neq t} \Big| \log \frac{\lambda(s) - \lambda(t)}{s - t} \Big| \leq \epsilon$$

for some strictly increasing homeomorphism $\lambda : [0,1] \to [0,1]$. Since $D[0,1]$ is Polish under d^*, conditions (i)-(ii) are trivially true. We now prove that (iii) holds as well. Suppose $d^*(x_n, x) + d^*(y_n, y) \to 0$ where $x_n, x, y_n, y \in D[0,1]$. Define $I = \{t \in [0,1] : x \text{ and } y \text{ are both continuous at } t\}$. Given $\epsilon > 0$, one obtains

$$d(x,y) = \sup_t |x(t) - y(t)| < \epsilon + |x(t_0) - y(t_0)| \quad \text{for some } t_0 \in I \cup \{1\}.$$

Since $x(t_0) = \lim_n x_n(t_0)$ and $y(t_0) = \lim_n y_n(t_0)$, it follows that $d(x,y) \leq \sup_n d(x_n, y_n)$. Hence, if $D[0,1]$ is equipped with the d^*-topology, $\{d \leq b\}$ is a closed subset of $D[0,1] \times D[0,1]$ for all $b \in \mathbb{R}$, that is, d is lower semicontinuous. Thus, conditions (i)-(ii)-(iii) are satisfied, and Theorem 2 implies the main result of [3].

Example 7. Suppose $\mathcal{G}$ countably generated, $\{(x,x) : x \in S\} \in \mathcal{G} \otimes \mathcal{G}$ and μ_n perfect for $n > 0$. By Theorem 1, applied with d the 0-1 distance, $\mu_n \to \mu_0$ in total variation norm if and only if, on some probability space $(\Omega, \mathcal{A}, P)$, there are measurable maps $X_n : (\Omega, \mathcal{A}) \to (S, \mathcal{G})$ satisfying

$$P(X_n \neq X_0) \longrightarrow 0 \;\text{ and }\; X_n \sim \mu_n \text{ for all } n \geq 0.$$

As remarked in Section 1, however, such statement holds without any assumptions on $\mathcal{G}$ or μ_n (possibly, replacing $P(X_n \neq X_0)$ with $P^*(X_n \neq X_0)$). See Proposition 3.1 of [2] and Theorem 2.1 of [11].

Example 8. Suppose $\mathcal{G}$ is the Borel σ-field under a distance d^* such that (S, d^*) is a universally measurable subset of a Polish space. Take a collection F of real functions on S such that

- $\sup_{f \in F} |f(x)| < \infty$ for all $x \in S$;
- If x, $y \in S$ and $x \neq y$, then $f(x) \neq f(y)$ for some $f \in F$.

Then,

$$d(x,y) = \sup_{f \in F} |f(x) - f(y)|$$

is a distance on S. If F is countable and each $f \in F$ is $\mathcal{G}$-measurable, then d is $\mathcal{G} \otimes \mathcal{G}$-measurable. In this case, by Theorem 1, condition (2) is equivalent to

$$\sup_{f \in F} |f(X_n) - f(X_0)| \xrightarrow{P} 0$$

for some random variables X_n such that $X_n \sim \mu_n$ for all $n \geq 0$. In view of Theorem 2, condition (2) can be replaced by condition (3) whenever each $f \in F$ is continuous in the d^*-topology (even if F is uncountable). In this case, in fact, $d : S \times S \to \mathbb{R}$ is lower semicontinuous in the d^*-topology.

Example 9. In Example 8, one starts with a nice σ-field $\mathcal{G}$ and then builds a suitable distance d. Now, instead, we start with a given distance d (similar to that of Example 8) and we define $\mathcal{G}$ basing on d.

Suppose $d(x, y) = \sup_{f \in F} |f(x) - f(y)|$ for some countable class F of real functions on S. Fix an enumeration $F = \{f_1, f_2, \ldots\}$ and define

$$\psi(x) = \big(f_1(x),\, f_2(x),\, \ldots\big) \text{ for } x \in S \quad \text{and} \quad \mathcal{G} = \sigma(\psi).$$

Then, $\psi : S \to \mathbb{R}^\infty$ is injective and d is measurable with respect to $\mathcal{G} \otimes \mathcal{G}$. Also, $(S, \mathcal{G})$ is isomorphic to $(\psi(S), \Psi)$ where Ψ is the Borel σ-field on $\psi(S)$. Thus, Theorem 1 applies whenever $\psi(S)$ is a universally measurable subset of $\mathbb{R}^\infty$.

A remarkable particular case is the following. Let S be a class of real bounded functions on a set T and let d be uniform distance. Suppose that, for some countable subset $T_0 \subset T$, one obtains

$$\text{for each } t \in T, \text{ there is a sequence } (t_n) \subset T_0$$
$$\text{such that } x(t) = \lim_n x(t_n) \text{ for all } x \in S.$$

Then, d can be written as $d(x, y) = \sup_{t \in T_0} |x(t) - y(t)|$. Given an enumeration $T_0 = \{t_1, t_2, \ldots\}$, define $\psi(x) = \big(x(t_1), x(t_2), \ldots\big)$ and $\mathcal{G} = \sigma(\psi)$. It is not hard to check that $\mathcal{G}$ coincides with the σ-field on S generated by the canonical projections $x \mapsto x(t)$, $t \in T$. Thus, Theorem 1 applies to such $\mathcal{G}$ and d whenever $\psi(S)$ is a universally measurable subset of $\mathbb{R}^\infty$.

Example 10. The following conjecture has been stated in Section 1. If $\mathcal{G} = \mathcal{B}$ (and without any assumptions on d and μ_n) condition (2) implies a Skorohod representation. As already noted, we do not know whether this is true. However, suppose that condition (2) holds and d is measurable with respect to $\mathcal{B} \otimes \mathcal{B}$. Then, a Skorohod representation is available on a suitable sub-σ-field $\mathcal{B}_0 \subset \mathcal{B}$ provided the μ_n are perfect on such $\mathcal{B}_0$. In fact, let $\mathcal{I}$ denote the class of intervals with rational endpoints. Since d is $\mathcal{B} \otimes \mathcal{B}$-measurable, for each $I \in \mathcal{I}$ there are A_n^I, $B_n^I \in \mathcal{B}$, $n \geq 1$, such that $\{d \in I\} \in \sigma\big(A_n^I \times B_n^I : n \geq 1\big)$. Define

$$\mathcal{B}_0 = \sigma\big(A_n^I,\, B_n^I : n \geq 1,\, I \in \mathcal{I}\big).$$

Then, d is $\mathcal{B}_0 \otimes \mathcal{B}_0$-measurable, $\mathcal{B}_0$ is countably generated and $\mathcal{B}_0 \subset \mathcal{B}$. By Theorem 1, the sequence $(\mu_n | \mathcal{B}_0)$ admits a Skorohod representation whenever $\mu_n | \mathcal{B}_0$ is perfect for each $n > 0$.

Unless μ_0 is d-separable, checking conditions (2)-(3) looks very hard. This is not always true, however. Our last example exhibits a situation where SRT does not work, and yet conditions (2)-(3) are easily verified. Other examples of this type are exchangeable empirical processes and pure jump processes, as defined in Examples 9-10 of [3].

Example 11. Given $p > 1$, let S be the space of real continuous functions x on $[0,1]$ such that

$$\|x\| := \Big\{|x(0)|^p + \sup\sum_i |x(t_i) - x(t_{i-1})|^p\Big\}^{1/p} < \infty$$

where sup is over all finite partitions $0 = t_0 < t_1 < \ldots < t_m = 1$. Define

$$d(x,y) = \|x - y\|, \quad d^*(x,y) = \sup_t |x(t) - y(t)|,$$

and take $\mathcal{G}$ to be the Borel σ-field on S under d^*. Since S is a Borel subset of the Polish space $(C[0,1], d^*)$, each law on $\mathcal{G}$ is perfect. Further, $d : S \times S \to \mathbb{R}$ is lower semicontinuous when S is given the d^*-topology.

In [1] and [7], some attention is paid to those processes X_n of the type

$$X_n(t) = \sum_k T_{n,k}\, N_k\, x_k(t), \quad n \geq 0,\ t \in [0,1].$$

Here, $x_k \in S$ while $(N_k,\, T_{n,k} : n \geq 0,\ k \geq 1)$ are real random variables, defined on some probability space $(\mathcal{X}, \mathcal{E}, Q)$, satisfying

$$(N_k) \text{ independent of } (T_{n,k}) \quad \text{and} \quad (N_k) \text{ i.i.d. with } N_1 \sim \mathcal{N}(0,1).$$

Usually, X_n has paths in S a.s. but the probability measure

$$\mu_n(A) = Q(X_n \in A), \quad A \in \mathcal{G},$$

is not d-separable. For instance, this happens when

$$0 < \liminf_k |T_{n,k}| \leq \limsup_k |T_{n,k}| < \infty \quad \text{a.s. and}$$

$$x_k(t) = q^{-k/p}\big\{\log(k+1)\big\}^{-1/2} \sin(q^k \pi\, t)$$

where $q = 4^{1+[p/(p-1)]}$. See Theorem 4.1 and Lemma 4.4 of [7].

We aim to a Skorohod representation for $(\mu_n : n \geq 0)$. Since μ_0 fails to be d-separable, SRT and its versions do not apply. Instead, under some conditions, Corollary 3 works. To fix ideas, suppose

$$T_{n,k} = U_n\, \phi_k(V_n, C)$$

where $\phi_k : \mathbb{R}^2 \to \mathbb{R}$ and U_n, V_n, C are real random variables such that

(a) (U_n) and (V_n) are conditionally independent given C;

(b) $E\big\{f(U_n) \mid C\big\} \xrightarrow{Q} E\big\{f(U_0) \mid C\big\}$ for each bounded continuous $f : \mathbb{R} \to \mathbb{R}$;

(c) $Q\big((V_n, C) \in \cdot\big)$ converges to $Q\big((V_0, C) \in \cdot\big)$ in total variation norm.

We next prove the existence of a Skorohod representation for $(\mu_n : n \geq 0)$. To this end, as noted in remark (vj) of Section 1, one can argue by subsequences. Moreover, condition (c) can be shown to be equivalent to

$$\sup_A \Big| Q\big(V_n \in A \mid C\big) - Q\big(V_0 \in A \mid C\big)\Big| \xrightarrow{Q} 0$$

where sup is over all Borel sets $A \subset \mathbb{R}$. Thus (up to selecting a suitable subsequence) conditions (b) and (c) can be strengthened into

(b*) $E\big\{f(U_n) \mid C\big\} \xrightarrow{a.s.} E\big\{f(U_0) \mid C\big\}$ for each bounded continuous $f : \mathbb{R} \to \mathbb{R}$;

(c*) $\sup_A \Big| Q\big(V_n \in A \mid C\big) - Q\big(V_0 \in A \mid C\big)\Big| \xrightarrow{a.s.} 0$.

Let P_c denote a version of the conditional distribution of the array

$$(N_k, U_n, V_n, C : n \geq 0,\, k \geq 1)$$

given $C = c$. Because of Corollary 3, it suffices to prove that $\big(P_c(X_n \in \cdot) : n \geq 0\big)$ has a Skorohod representation for almost all $c \in \mathbb{R}$. Fix $c \in \mathbb{R}$. By (a), the sequences (N_k), (U_n) and (V_n) can be assumed to be independent under P_c. By (b*) and (c*), up to a change of the underlying probability space, (U_n) and (V_n) can be realized in the most convenient way. Indeed, by applying SRT to (U_n) and Theorem 2.1 of [11] to (V_n), it can be assumed that

$$U_n \overset{P_c-a.s.}{\longrightarrow} U_0 \quad \text{and} \quad P_c(V_n \neq V_0) \longrightarrow 0.$$

But in this case, one trivially obtains $X_n \overset{P_c}{\longrightarrow} X_0$, for

$$1 \wedge \|X_n - X_0\| \leq I_{\{V_n \neq V_0\}} + |U_n - U_0| \; \|\sum_k \phi_k(V_0, C)\, N_k\, x_k\|.$$

Thus, $\big(P_c(X_n \in \cdot) : n \geq 0\big)$ admits a Skorohod representation.

The conditions of Example 11 are not so strong as they appear. Actually, they do not imply even $d^*(X_n, X_0) \overset{a.s.}{\longrightarrow} 0$ for the original processes X_n (those defined on $(\mathcal{X}, \mathcal{E}, Q)$). In addition, by slightly modifying Example 11, S could be taken to be the space of α-Holder continuous functions, $\alpha \in (0,1)$, and

$$d(x,y) = |x(0) - y(0)| + \sup_{t \neq s} \frac{|x(t) - y(t) - x(s) + y(s)|}{|t - s|^\alpha}.$$

References

[1] Basse-O'Connor A., Rosinski J. (2012) On the uniform convergence of random series in Skorohod space and representations of cadlag infinitely divisible processes, *Ann. Probab.*, to appear.

[2] Berti P., Pratelli L., Rigo P. (2010) Skorohod representation theorem via disintegrations, *Sankhya*, 72, 208-220.

[3] Berti P., Pratelli L., Rigo P. (2011) A Skorohod representation theorem for uniform distance, *Prob. Theo. Relat. Fields*, 150, 321-335.

[4] Blackwell D. (1955) On a class of probability spaces, *Proc. Third Berkeley Symp. Math. Statist. Probab.*, University of California Press, 1-6.

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

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

[7] Jain N.C., Monrad D. (1983) Gaussian measures in B_p, *Ann. Probab.*, 11, 46-57.

[8] Jakubowski A. (1998) The almost sure Skorokhod representation for subsequences in nonmetric spaces, *Theory Probab. Appl.*, 42, 167-174.

[9] Kallenberg O. (2002) *Foundations of modern probability*, Second edition, Springer.

[10] Ramachandran D., Ruschendorf L. (1995) A general duality theorem for marginal problems, *Prob. Theo. Relat. Fields*, 101, 311-319.

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

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

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

[14] 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: patrizia.berti@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 Matematica "F. Casorati", Universita' di Pavia, via Ferrata 1, 27100 Pavia, Italy
E-mail address: pietro.rigo@unipv.it