Theory of Stochastic Processes
Vol. 20 (36), no. 2, 2015, pp. 1–12

PATRIZIA BERTI, LUCA PRATELLI, AND PIETRO RIGO

A SURVEY ON SKOROKHOD REPRESENTATION THEOREM WITHOUT SEPARABILITY

Let S be a metric space, $\mathcal{G}$ a σ-field of subsets of S and $(\mu_n : n \geq 0)$ a sequence of probability measures on $\mathcal{G}$. Say that (μ_n) admits a Skorokhod representation if, on some probability space, there are random variables X_n with values in $(S, \mathcal{G})$ such that

$$X_n \sim \mu_n \text{ for each } n \geq 0 \quad \text{and} \quad X_n \to X_0 \text{ in probability.}$$

We focus on results of the following type: (μ_n) has a Skorokhod representation if and only if $J(\mu_n, \mu_0) \to 0$, where J is a suitable distance (or discrepancy index) between probabilities on $\mathcal{G}$. One advantage of such results is that, unlike the usual Skorokhod representation theorem, they apply even if the limit law μ_0 is not separable. The index J is taken to be the bounded Lipschitz metric and the Wasserstein distance.

1. Introduction

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 $\mathcal{G}$. A probability μ on $\mathcal{G}$ is said to be *separable* if there is a separable set $A \in \mathcal{G}$ such that $\mu(A) = 1$. We let $\mathcal{R}$ denote the ball σ-field on S and $\mathcal{B}$ the Borel σ-field on S. (Thus, $\mathcal{R}$ and $\mathcal{B}$ are generated by the balls and by the open sets, respectively). Also, M is the class of those functions $f : S \to \mathbb{R}$ satisfying

$$-1 \leq f \leq 1, \quad f \text{ is } \mathcal{G}\text{-measurable}, \quad |f(x) - f(y)| \leq d(x, y) \text{ for all } x, y \in S.$$

Slightly generalizing the usual statement, the Skorokhod representation theorem (SRT) can be stated as follows.

Theorem 1. (SRT). *Suppose $\mathcal{R} \subset \mathcal{G} \subset \mathcal{B}$. If*

$$\mu_0 \text{ is separable} \quad \text{and} \quad \mu_n(f) \to \mu_0(f) \quad \text{for each } f \in M,$$

there are a probability space $(\Omega, \mathcal{A}, P)$ and measurable maps $X_n : (\Omega, \mathcal{A}) \to (S, \mathcal{G})$ such that $X_n \sim \mu_n$ for all $n \geq 0$ and $X_n \to X_0$ almost uniformly.

The usual version of SRT is the special case of Theorem 1 obtained for $\mathcal{G} = \mathcal{B}$. See Skorokhod [16], Dudley [10] and Wichura [20]; see also [11, page 130] and [19, page 77] for historical notes. Theorem 1 is just a slight improvement of the usual version, more suitable for our purposes, which follows trivially from well known facts. For completeness, a proof of Theorem 1 is given in Section 2.

This paper originates from the following questions: Is it possible to drop separability of μ_0 from Theorem 1 ? And if no, are there (reasonable) versions of SRT not requesting separability of μ_0 ? Let us start by the first question.

Example 2. Let d be the uniform distance on

$$S = \{x \in D[0, 1] : x(t) \in \{0, 1\} \text{ for each } t \in [0, 1]\},$$

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

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

where $D[0,1]$ is the set of real cadlag functions on $[0,1]$. Take $\mathcal{G}$ to be the Borel σ-field on S under Skorokhod topology and define

$$X = I_{[U,1]} \quad \text{and} \quad \mu_0(A) = \text{Prob}(X \in A) \text{ for } A \in \mathcal{G},$$

where U is a $(0,1)$-valued random variable with any non-atomic distribution. Such a μ_0 is not separable, as the jump time U of the process X does not have a discrete distribution (see the remarks at the end of this example). Also, since $\mathcal{G}$ is countably generated and $\mu_0\{x\} = 0$ for all $x \in S$, then $(S, \mathcal{G}, \mu_0)$ is a non-atomic probability space. Hence, $(S, \mathcal{G}, \mu_0)$ supports an i.i.d. sequence (f_n) of exponential random variables with mean 1; see e.g. [4, Theorem 3.1]. Let $\mu_n(A) = E_{\mu_0}(I_A\, f_n)$ for $n \geq 1$ and $A \in \mathcal{G}$. If $A \in \sigma(f_1, \ldots, f_k)$, then

$$\mu_n(A) = E_{\mu_0}(I_A\, f_n) = E_{\mu_0}(I_A)\, E_{\mu_0}(f_n) = \mu_0(A) \quad \text{for all } n > k.$$

Thus, $\mu_n(A) \to \mu_0(A)$ for each A in the field $\bigcup_k \sigma(f_1, \ldots, f_k)$. By standard arguments, this implies $\mu_n(A) \to \mu_0(A)$ for all $A \in \mathcal{G}$. Hence, $\mu_n(f) \to \mu_0(f)$ whenever $f \in M$. Finally, take a probability space $(\Omega, \mathcal{A}, P)$ and a sequence of measurable maps $X_n : (\Omega, \mathcal{A}) \to (S, \mathcal{G})$ such that $X_n \sim \mu_n$ for all $n \geq 0$. Then,

$$P\Big(d(X_n, X_0) > 1/2\Big) = P(X_n \neq X_0) \geq \sup_{A \in \mathcal{G}} |\mu_n(A) - \mu_0(A)| = (1/2) \int |f_n - 1|\, d\mu_0.$$

Since (f_n) is i.i.d. with a non-degenerate distribution, $\int |f_n - 1|\, d\mu_0$ does not converge to 0. Thus, X_n fails to converge to X_0 in probability (so that X_n does not converge to X_0 almost uniformly).

Example 2 is less artificial than it appears. Take in fact a real cadlag process X on $[0,1]$ and define

$$S = D[0,1], \quad d = \text{ uniform distance},$$
$$\mathcal{G} = \text{ Borel } \sigma\text{-field on } S \text{ under Skorokhod topology},$$
$$\mu(A) = \text{Prob}(X \in A) \quad \text{for all } A \in \mathcal{G}.$$

It is not hard to see that μ is separable if and only if each jump time of X has a discrete distribution; see [7, page 2]. Thus, a plenty of meaningful probability laws on $\mathcal{G}$ are actually non-separable.

Anyhow, in view of Example 2, separability of μ_0 can not be dropped from Theorem 1. This leads to our second question, namely, to the search of some alternative version of SRT not requesting separability of μ_0.

A first remark is that, if μ_0 needs not be separable, almost uniform convergence should be weakened into convergence in probability. In fact, it may be that $X_n \sim \mu_n$ for all $n \geq 0$ and $X_n \to X_0$ in probability, for some sequence (X_n) of random variables, but no sequence (Y_n) satisfies $Y_n \sim \mu_n$ for all $n \geq 0$ and $Y_n \to Y_0$ almost uniformly or even almost surely. This can be seen by a slight modification of Example 2.

Example 3. Take S, d, $\mathcal{G}$ and μ_0 as in Example 2 but a different sequence (f_n). Precisely, (f_n) is now an independent sequence of random variables on $(S, \mathcal{G}, \mu_0)$ satisfying

$$\mu_0\Big(f_n = (n+1)/n\Big) = n/(n+1) \quad \text{and} \quad \mu_0\big(f_n = 0\big) = 1/(n+1).$$

For $n \geq 1$ and $A \in \mathcal{G}$, define again $\mu_n(A) = E_{\mu_0}(I_A\, f_n)$. Since $E_{\mu_0}(f_n) = 1$, the μ_n are probabilities on $\mathcal{G}$ and

$$\sup_{A \in \mathcal{G}} |\mu_n(A) - \mu_0(A)| = (1/2) \int |f_n - 1|\, d\mu_0 \longrightarrow 0.$$

Hence, by [15, Theorem 2.1], on some probability space $(\Omega, \mathcal{A}, P)$ there is a sequence (X_n) of random variables such that $X_n \sim \mu_n$ for all $n \geq 0$ and

$$P\Big(d(X_n, X_0) > \epsilon\Big) = P(X_n \neq X_0) \longrightarrow 0 \quad \text{for all } \epsilon \in (0,1).$$

Finally, fix any sequence (Y_n) of random variables on a probability space $(\Omega_0, \mathcal{A}_0, P_0)$ such that $Y_n \sim \mu_n$ for all $n \geq 0$. By the Borel-Cantelli lemma, $\liminf_n f_n(x) = 0$ for μ_0-almost all $x \in S$. In particular,

$$\mu_0\Big(\liminf_n f_n < 1\Big) > 0$$

and this implies

$$P_0\Big(d(Y_n, Y_0) \to 0\Big) = P_0(Y_n = Y_0 \text{ eventually}) < 1$$

by results in [15, Theorem 3.1] and [17, Section 5.4]. Therefore, Y_n fails to converge to Y_0 almost uniformly or even almost surely.

In view of Example 3, when μ_0 is not separable we must be content with convergence in probability. Thus, in the sequel, the sequence (μ_n) is said to admit a Skorokhod representation if

> **Skorokhod representation for (μ_n):** On some probability space, there are random variables X_n with values in $(S, \mathcal{G})$ such that $X_n \sim \mu_n$ for all $n \geq 0$ and $X_n \to X_0$ in probability.

We aim to conditions for a Skorokhod representation, possibly necessary and sufficient. To this end, we assume

$$\sigma(d) \subset \mathcal{G} \otimes \mathcal{G},$$

namely, $d : S \times S \to \mathbb{R}$ measurable with respect to $\mathcal{G} \otimes \mathcal{G}$.

For any probabilities μ and ν on $\mathcal{G}$, define

$$B(\mu, \nu) = \sup_{f \in M} |\mu(f) - \nu(f)|.$$

If $f \in M$ and $X_n \sim \mu_n$ for each $n \geq 0$, with the X_n all defined on the probability space $(\Omega, \mathcal{A}, P)$, then

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

Hence, a necessary condition for (μ_n) to have a Skorokhod representation is

$$\lim_n B(\mu_n, \mu_0) = 0.$$

A (natural) question is whether such condition suffices as well. This actually happens if $\mathcal{G} = \mathcal{B}$ and d is 0-1-distance. In this case, in fact, (μ_n) admits a Skorokhod representation if and only if $B(\mu_n, \mu_0) \to 0$; see [15, Theorem 2.1].

More generally, B could be replaced by some other distance (or discrepancy index) between probability measures. This leads to

> **Conjecture:** The sequence (μ_n) admits a Skorokhod representation if and only if $J(\mu_n, \mu_0) \to 0$, where J is some (reasonable) discrepancy index between probabilities on $\mathcal{G}$.

Such a conjecture is the object of this paper. Special attention is paid to $J = B$ and $J = W$, where W is a Wasserstein-type distance. Our main concern is to connect and review some results from [5]-[8], using the conjecture as unifying criterion. In addition to report known facts, some new examples are given as well. The previous Examples 2-3 are actually new. Lemma 4 and Example 11, while implicit in the existing literature, are also of potential interest. Besides [5]-[8], related references are [1], [4], [13], [15], [17], [18].

2. Basic definitions

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

A *kernel* on $(\mathcal{X}, \mathcal{E})$, $(\mathcal{Y}, \mathcal{F})$ is a collection

$$\alpha = \{\alpha(x) : x \in \mathcal{X}\}$$

such that

- $\alpha(x)$ is a probability on $\mathcal{F}$ for each $x \in \mathcal{X}$,
- $x \mapsto \alpha(x)(B)$ is $\mathcal{E}$-measurable for each $B \in \mathcal{F}$.

Let γ be a probability measure on the product space $(\mathcal{X} \times \mathcal{Y},\, \mathcal{E} \otimes \mathcal{F})$. A *disintegration* for γ is a kernel α on $(\mathcal{X}, \mathcal{E})$, $(\mathcal{Y}, \mathcal{F})$ satisfying

$$\gamma(A \times B) = \int_A \alpha(x)(B)\, \mu(dx) \quad \text{for all } A \in \mathcal{E} \text{ and } B \in \mathcal{F}$$

where μ is the marginal of γ on $\mathcal{E}$. A disintegration for γ can fail to exist. However, γ admits a disintegration provided $\mathcal{F}$ is countably generated and the marginal of γ on $\mathcal{F}$ is perfect, where perfect probability measures are defined as follows.

A probability ν on $(\mathcal{Y}, \mathcal{F})$ is *perfect* if, for each measurable function $f : \mathcal{Y} \to \mathbb{R}$, there is a Borel subset A of $\mathbb{R}$ such that $A \subset f(\mathcal{Y})$ and $\nu(f \in A) = 1$. If $\mathcal{Y}$ is separable metric and $\mathcal{F}$ the Borel σ-field, then ν is perfect if and only if it is tight. Therefore, the following result is available. Recall that a metric space T is said to be *universally measurable* if, for any completion of T and any Borel probability λ on such completion, T is λ-measurable.

Lemma 4. *Let $\mathcal{Y}$ be a separable metric space. Each Borel probability on $\mathcal{Y}$ is perfect if and only if $\mathcal{Y}$ is universally measurable.*

Proof. Borel laws on Polish spaces are tight and every completion of $\mathcal{Y}$ is Polish (for $\mathcal{Y}$ is separable). Therefore, if $\mathcal{Y}$ is universally measurable, each Borel probability on $\mathcal{Y}$ is tight (and thus perfect). Conversely, suppose that every Borel law on $\mathcal{Y}$ is perfect. Denote by $\mathcal{F}$ the Borel σ-field on $\mathcal{Y}$ and fix a measurable space $(\mathcal{X},\, \mathcal{E})$ and a law γ on $\mathcal{E} \otimes \mathcal{F}$. Then, γ admits a disintegration since $\mathcal{F}$ is countably generated and the marginal of γ on $\mathcal{F}$ is perfect. Hence, $\mathcal{Y}$ is universally measurable by a result in [9]. □

As already noted, Theorem 1 is a quick consequence of well known results. It follows, for instance, from [19, Part 1].

A proof of Theorem 1. Fix a separable $A \in \mathcal{G}$ such that $\mu_0(A) = 1$. Since $\mathcal{R} \subset \mathcal{G} \subset \mathcal{B}$ and $\mathcal{R}$ includes every separable Borel subset of S, one can define

$$\nu(B) = \mu_0(A \cap B) \quad \text{for each } B \in \mathcal{B}.$$

Let $(\Omega_0, \mathcal{A}_0, P_0) = (S, \mathcal{B}, \nu)$, $(\Omega_n, \mathcal{A}_n, P_n) = (S, \mathcal{G}, \mu_n)$ if $n > 0$, and take I_n to be the identity map on $(\Omega_n, \mathcal{A}_n, P_n)$ for each $n \geq 0$. By [19, Theorem 1.7.2], since ν is separable and $\mu_n(f) \to \mu_0(f) = \nu(f)$ for all $f \in M$, one obtains $I_n \to I_0$ in distribution according to Hoffmann-Jørgensen. Thus, Theorem 1 follows from [19, Theorem 1.10.3].

Finally, to investigate the conjecture of Section 1, we introduce some discrepancy indices between probability measures. One is the index B defined in Section 1. Another is the following version B_0 of B. Let $\mathbb{P}$ be the set of all probabilities on $\mathcal{G}$. Say that a function $f : S \to \mathbb{R}$ is universally measurable with respect to $\mathcal{G}$ if f is μ-measurable for every $\mu \in \mathbb{P}$. Then, B_0 is defined as

$$B_0(\mu,\nu) = \sup_f |\mu(f) - \nu(f)|$$

where $\mu, \nu \in \mathbb{P}$ and sup is over those $f : S \to [-1, 1]$ which are 1-Lipschitz and universally measurable with respect to $\mathcal{G}$.

Let us turn to Wasserstein-type indices. For $\mu, \nu \in \mathbb{P}$, denote by $\mathcal{F}(\mu,\nu)$ the collection of those probabilities γ on $\mathcal{G} \otimes \mathcal{G}$ such that

$$\gamma(A \times S) = \mu(A) \quad \text{and} \quad \gamma(S \times A) = \nu(A) \quad \text{for all } A \in \mathcal{G}.$$

Define also

$$\mathcal{D}(\mu,\nu) = \big\{\gamma \in \mathcal{F}(\mu,\nu) : \gamma \text{ admits a disintegration}\big\}$$

and note that $\mathcal{D}(\mu,\nu) \neq \emptyset$ for it includes at least the product law $\mu \times \nu$.

A basic assumption of this paper is $\sigma(d) \subset \mathcal{G} \otimes \mathcal{G}$. Under this assumption, one can define $E_\gamma(1 \wedge d) = \int 1 \wedge d(x,y)\,\gamma(dx,dy)$ for each $\gamma \in \mathcal{F}(\mu,\nu)$ and

$$W(\mu,\nu) = \inf_{\gamma \in \mathcal{F}(\mu,\nu)} E_\gamma(1 \wedge d),$$
$$W_0(\mu,\nu) = \inf_{\gamma \in \mathcal{D}(\mu,\nu)} E_\gamma(1 \wedge d).$$

Both W and W_0 look reasonable discrepancy indices between elements of $\mathbb{P}$. Also,

$$W(\mu,\nu) = W_0(\mu,\nu) \quad \text{if one between } \mu \text{ and } \nu \text{ is separable.}$$

However, we do not know whether W and W_0 are distances on all of $\mathbb{P}$.

To be precise, we first recall that M is a $\mathcal{G}$-determining class if

$$\mu = \nu \quad \Leftrightarrow \quad \mu(f) = \nu(f) \text{ for each } f \in M,$$

whenever μ, $\nu \in \mathbb{P}$. For instance, M is $\mathcal{G}$-determining if $\mathcal{G} = \mathcal{B}$ or $\mathcal{G} = \mathcal{R}$. Now, if M is $\mathcal{G}$-determining, then

$$W(\mu,\nu) = 0 \;\Leftrightarrow\; W_0(\mu,\nu) = 0 \;\Leftrightarrow\; \mu = \nu.$$

Furthermore, $W(\mu,\nu) = W(\nu,\mu)$ and W_0 meets the triangle inequality. However, we do not know whether $W_0(\mu,\nu) = W_0(\nu,\mu)$ (unless one between μ and ν is separable) and whether W satisfies the triangle inequality; see [5, Theorem 4.1] and [8, Lemma 7]. Note also that $W = W_0$ and W is a distance on the subset {separable laws on $\mathcal{G}$}.

A last remark is that, as easily seen, the previous indices are connected via

$$B \le B_0 \le 2\,W \le 2\,W_0.$$

3. Results

This section collects some (essentially known) results on the conjecture stated in Section 1. The latter is briefly referred to as "the conjecture". Proofs are omitted, except when simple and informative. All the examples are postponed to Section 4.

Our starting point is the following.

Theorem 5. (Theorem 4.2 of [5]**).** *Suppose* $\sigma(d) \subset \mathcal{G} \otimes \mathcal{G}$. *If* $W_0(\mu_0, \mu_n) \to 0$, *then* (μ_n) *admits a Skorokhod representation.*

Proof. Let $(\Omega, \mathcal{A}) = (S^\infty, \mathcal{G}^\infty)$. For each $n \ge 0$, take $X_n : \Omega = S^\infty \to S$ to be the n-th canonical projection. Fix $\gamma_n \in \mathcal{D}(\mu_0, \mu_n)$ such that $E_{\gamma_n}(1 \wedge d) < \frac{1}{n} + W_0(\mu_0, \mu_n)$ and a disintegration α_n for γ_n. By Ionescu-Tulcea theorem, there is a unique probability P on $\mathcal{A} = \mathcal{G}^\infty$ such that $X_0 \sim \mu_0$ and

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

is a regular version of the conditional distribution of X_n given $(X_0, X_1, \ldots, X_{n-1})$. To conclude the proof just note that, under P, one obtains $(X_0, X_n) \sim \gamma_n$ for all $n \geq 1$. Thus, $X_n \sim \mu_n$ for all $n \geq 0$ and

$$E_P\{1 \wedge d(X_0, X_n)\} = E_{\gamma_n}(1 \wedge d) < \frac{1}{n} + W_0(\mu_0, \mu_n) \longrightarrow 0.$$

□

Theorem 5 is surprisingly simple. Its underlying idea is to exploit Ionescu-Tulcea theorem and to take X_n conditionally independent of $(X_1, \ldots, X_{n-1})$ given X_0. Incidentally, this form of conditional independence holds true in every proof (known to us) of SRT. More importantly, Theorem 5 partially addresses the conjecture. The "partially" is because $W_0(\mu_0, \mu_n) \to 0$ is sufficient, but not necessary, for (μ_n) to admit a Skorokhod representation. To get equivalent conditions, some more assumptions are needed.

Theorem 6. (Theorem 1.1 of [7]**).** *Suppose*

(i) μ_n *perfect for all* $n > 0$,
(ii) $\mathcal{G}$ *countably generated and* $\sigma(d) \subset \mathcal{G} \otimes \mathcal{G}$.

Then,

$$W_0(\mu_0, \mu_n) = W(\mu_0, \mu_n) = B_0(\mu_0, \mu_n) \quad \text{for all } n > 0 \text{ and}$$
$$(\mu_n) \text{ has a Skorokhod representation} \quad \Leftrightarrow \quad \lim_n B_0(\mu_0, \mu_n) = 0.$$

Proof. We just give a sketch of the proof. Fix $n > 0$. Since μ_n is perfect and $\mathcal{G}$ countably generated, each $\gamma \in \mathcal{F}(\mu_0, \mu_n)$ admits a disintegration. Hence, $\mathcal{D}(\mu_0, \mu_n) = \mathcal{F}(\mu_0, \mu_n)$ and $W_0(\mu_0, \mu_n) = W(\mu_0, \mu_n)$. In turn, exploiting a duality result in optimal transportation theory [14], it can be shown that $W(\mu_0, \mu_n) = B_0(\mu_0, \mu_n)$. Having proved $W_0(\mu_0, \mu_n) = B_0(\mu_0, \mu_n)$, an application of Theorem 5 concludes the proof of "$\Leftarrow$", while "$\Rightarrow$" is straightforward. □

From the point of view of this paper, the main content of Theorem 6 is that, under (i)-(ii), the conjecture is true with $J = W$ and $J = B_0$.

A stronger (and nicer) result would be the equivalence between $B(\mu_0, \mu_n) \to 0$ and a Skorokhod representation. We do not know whether this is true under (i)-(ii). However, it turns out to be true if (ii) is slightly strengthened. Precisely, under (ii), $\mathcal{G}$ is the Borel σ-field under some separable distance d^* on S. If d is lower semicontinuous (and not only Borel measurable) with respect to such d^*, then B_0 can be replaced with B.

Theorem 7. (Theorem 1.2 of [7]**).** *In addition to (i), suppose*

(jj) $\mathcal{G}$ *is the Borel* σ*-field under a separable distance* d^* *on* S *and* $d : S \times S \to \mathbb{R}$ *is lower semicontinuous when* S *is given the* d^**-topology.*

Then,

$$W_0(\mu_0, \mu_n) = W(\mu_0, \mu_n) = B_0(\mu_0, \mu_n) = B(\mu_0, \mu_n) \quad \text{for all } n > 0 \text{ and}$$
$$(\mu_n) \text{ has a Skorokhod representation} \quad \Leftrightarrow \quad \lim_n B(\mu_0, \mu_n) = 0.$$

A consequence of Theorem 7 is the following.

Corollary 8. (Corollary 1.3 of [7]**).** *For each* $n \geq 0$, *let* $\alpha_n = \{\alpha_n(x) : x \in \mathcal{X}\}$ *be a kernel on* $(\mathcal{X}, \mathcal{E})$, $(S, \mathcal{G})$, *where* $(\mathcal{X}, \mathcal{E})$ *is a measurable space. In addition to (i) and (jj), suppose*

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

where Q is a probability on $\mathcal{E}$. Then, (μ_n) has a Skorokhod representation provided $(\alpha_n(x))$ has a Skorokhod representation for Q-almost all $x \in \mathcal{X}$. In particular, (μ_n) admits a Skorokhod representation whenever $\mathcal{G} \subset \mathcal{B}$ and, for Q-almost all $x \in \mathcal{X}$,

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

Corollary 8 states that Skorokhod representations are preserved under mixtures. This is quite useful in applications, as shown by various examples in Section 4.

Under conditions (i) and (jj), not only the conjecture is true but J can be taken to be W_0, W, B_0 or B indifferently. We next focus on $J = W$ under the only assumption that $\sigma(d) \subset \mathcal{G} \otimes \mathcal{G}$. In a sense, this is the natural choice of J. Contrary to W_0, in fact, $W(\mu_n, \mu_0) \to 0$ is a necessary condition for (μ_n) to admit a Skorokhod representation. Moreover, since $2W \geq B$, the conjecture holds with $J = W$ provided it holds with $J = B$.

Hence, suppose $W(\mu_n, \mu_0) \to 0$. Then, by definition of W, there is a sequence $\gamma_n \in \mathcal{F}(\mu_0, \mu_n)$ such that

$$\lim_n \gamma_n\big\{(x, y) : d(x, y) > \epsilon\big\} = 0 \quad \text{whenever } \epsilon > 0.$$

Thus, the conjecture is automatically true with $J = W$ if one can obtain a sequence (X_n) of random variables, all defined on the same probability space, such that

$$(X_0, X_n) \sim \gamma_n \quad \text{for all } n \geq 1.$$

In turn, one could try to construct such sequence (X_n) by a *gluing* argument. Unfortunately, this line of proof seems to be precluded by [8, Example 1]. However, if $\mathcal{G} = \mathcal{B}$, the gluing argument works in a finitely additive framework.

Theorem 9. (Theorem 8 of [8]**).** *Suppose $\mathcal{G} = \mathcal{B}$ and $\sigma(d) \subset \mathcal{B} \otimes \mathcal{B}$. Then, $\lim_n W(\mu_n, \mu_0) = 0$ if and only if, on a finitely additive probability space $(\Omega, \mathcal{A}, P)$, there are measurable maps $X_n : (\Omega, \mathcal{A}) \to (S, \mathcal{B})$ satisfying*

- $X_n \xrightarrow{P} X_0$,
- $P(X_0 \in A) = \mu_0(A)$ *for all* $A \in \mathcal{B}$,
- *There is a sequence $\gamma_n \in \mathcal{F}(\mu_0, \mu_n)$, $n \geq 1$, such that*

$$P\big[(X_0, X_n) \in C\big] = \gamma_n(C) \quad \text{if } C \in \mathcal{B} \otimes \mathcal{B} \text{ and } \gamma_n^*(\partial C) = 0$$

where γ_n^* is the outer measure.

A drawback of Theorem 9 is that, since P is finitely additive but not necessarily σ-additive, it does not follow that $P\big[(X_0, X_n) \in C\big] = \gamma_n(C)$ for all $C \in \mathcal{B} \otimes \mathcal{B}$. In particular, for $n > 0$, one only obtains $E_P\big\{f(X_n)\big\} = \mu_n(f)$ if f is bounded and continuous, but not necessarily $P(X_n \in A) = \mu_n(A)$ for all $A \in \mathcal{B}$.

4. Examples

Apart from Example 11, the material of this section comes from [6]-[7].

In a sense, most our results are suggested by the following example.

Example 10. (Motivating example). Let $S = D[0, 1]$, $\mathcal{G} = \mathcal{R}$ and d the uniform distance, where $D[0, 1]$ is the set of real cadlag functions on $[0, 1]$. Such a $\mathcal{G}$ agrees with the Borel σ-field on S under Skorokhod topology. Since the latter is a Polish topology, each probability on $\mathcal{G}$ is perfect. Also, d is lower semicontinuous when S is given the Skorokhod topology. Thus, Theorem 7 applies, and (μ_n) has a Skorokhod representation if and only if $\lim_n B(\mu_n, \mu_0) = 0$.

Example 10 has a quite natural extension.

Example 11. (A general version of Example 10). Let $\mathcal{G} = \mathcal{R}$ where (S, d) is any metric space. Suppose $\sigma(d) \subset \mathcal{R} \otimes \mathcal{R}$ and μ_n perfect for all $n > 0$. Since $\sigma(d) \subset \mathcal{R} \otimes \mathcal{R}$, as proved below, $\mathcal{R}$ is countably generated. Hence, by Theorem 6, (μ_n) has a Skorokhod representation if and only if $\lim_n B_0(\mu_n, \mu_0) = 0$. It remains to prove that $\mathcal{R}$ is countably generated. Let $\mathcal{I}$ be the class of intervals with rational endpoints. For each $I \in \mathcal{I}$, since $\{d \in I\} \in \mathcal{R} \otimes \mathcal{R}$, there are A_n^I, $B_n^I \in \mathcal{R}$, $n \geq 1$, such that $\{d \in I\} \in \sigma\big(A_n^I \times B_n^I : n \geq 1\big)$. Define $\mathcal{U} = \sigma\big(A_n^I, B_n^I : n \geq 1, I \in \mathcal{I}\big)$. Then, $\mathcal{U}$ is countably generated, $\mathcal{U} \subset \mathcal{R}$ and $\sigma(d) \subset \mathcal{U} \otimes \mathcal{U}$. Given $x \in S$ and $r > 0$, the ball $\{y : d(x, y) < r\}$ is the x-section of the set $\{d < r\} \in \mathcal{U} \otimes \mathcal{U}$. Thus, $\{y : d(x, y) < r\} \in \mathcal{U}$. It follows that $\mathcal{R} = \mathcal{U}$ is countably generated.

In view of Example 11, it would be useful that $\mathcal{R}$ supports perfect probability measures only. Suppose $\sigma(d) \subset \mathcal{R} \otimes \mathcal{R}$. Then, $\mathcal{R}$ is countably generated (as shown in Example 11) and includes the singletons. Thus, $\mathcal{R}$ is the Borel σ-field under a separable distance d^* on S. By Lemma 4, each probability on $\mathcal{R}$ is perfect if and only if S is universally measurable under such d^*.

The next example comes into play if we are given the sequence (μ_n) but not the distance d.

Example 12. (Uniform convergence over a given class of measurable functions). Suppose $\mathcal{G}$ is the Borel σ-field under a distance d^* on S such that (S, d^*) is separable and universally measurable. Then, $\mathcal{G}$ is countably generated and each μ_n is perfect. Let F be a countable collection of real $\mathcal{G}$-measurable 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,

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

is a distance on S and $\sigma(d_F) \subset \mathcal{G} \otimes \mathcal{G}$. If we are given (μ_n) but not d, it could be reasonable to select a class F as above and to ask for random variables X_n satisfying

$$X_n \sim \mu_n \text{ for all } n \geq 0 \quad \text{and} \quad \sup_{f \in F} |f(X_n) - f(X_0)| \longrightarrow 0 \quad \text{in probability.}$$

This is exactly a Skorokhod representation for (μ_n) with $d = d_F$. Thus, by Theorem 6, such X_n exist if and only if $\lim_n B_0(\mu_n, \mu_0) = 0$. Also, by Theorem 7, B_0 can be replaced with B whenever each $f \in F$ is continuous in the d^*-topology. In this case, in fact, $d_F : S \times S \to \mathbb{R}$ is lower semicontinuous (even if F is uncountable) when S is given the d^*-topology.

Unless μ_0 is separable, checking whether $B(\mu_n, \mu_0) \to 0$ looks very hard. Thus, even if theoretically intriguing, Theorem 7 has a little practical scope. An analogous criticism can be made to Theorems 5 and 6. To address this criticism, we make two remarks. First, even if difficult to check, $B(\mu_n, \mu_0) \to 0$ is a necessary condition for (μ_n) to admit a Skorokhod representation. Hence, it can not be eluded. Second, there are situations where SRT does not work, and yet $B(\mu_n, \mu_0) \to 0$ is easily seen to be true. The remaining examples exhibit some of these situations.

Example 13. (Exchangeable empirical processes). Let (ξ_n) be an exchangeable sequence of $[0, 1]$-valued random variables and

$$F(t) = E\big(I_{\{\xi_1 \leq t\}} \mid \tau\big),$$

where τ is the tail σ-field of (ξ_n). Take F to be regular, i.e., each F-path is a distribution function. The empirical process can be defined as

$$Z_n(t) = \sqrt{n}\,\big\{F_n(t) - F(t)\big\}, \quad 0 \leq t \leq 1,\, n \geq 1,$$

where $F_n(t) = (1/n) \sum_{i=1}^{n} I_{\{\xi_i \leq t\}}$ is the empirical distribution function.

Take (S, d) and $\mathcal{G}$ as in Example 10, that is, $S = D[0, 1]$, d the uniform distance and $\mathcal{G} = \mathcal{R}$. Define also

$$Z_0(t) = W^0_{H(t)},$$

with W^0 a standard Brownian bridge and H an independent copy of F. Since each Z_n is a measurable map with values in $(S, \mathcal{G})$, we can let

$$\mu_n(A) = \text{Prob}\big(Z_n \in A\big) \quad \text{for all } n \geq 0 \text{ and } A \in \mathcal{G}.$$

Then, SRT does not apply for μ_0 may fail to be separable; see [3, Example 11]. However, (μ_n) admits a Skorokhod representation because of Theorem 7. In fact, $B(\mu_n, \mu_0) \to 0$ in the special case where (ξ_n) is i.i.d. (since μ_0 is separable if (ξ_n) is i.i.d.). Therefore, by Corollary 8 and de Finetti's representation theorem, $B(\mu_n, \mu_0) \to 0$ in the general case as well.

Example 14. (Pure jump processes). Take again $S = D[0, 1]$, $\mathcal{G} = \mathcal{R}$, d the uniform distance and $\mu_n(\cdot) = \text{Prob}(Z_n \in \cdot)$. Here, for each $n \geq 0$, Z_n is the process

$$Z_n(t) = \sum_{j=1}^{\infty} C_{n,j}\, I_{\{Y_{n,j} \leq t\}}, \quad 0 \leq t \leq 1,$$

where $C_n = (C_{n,j} : j \geq 1)$ and $Y_n = (Y_{n,j} : j \geq 1)$ are sequences of real random variables satisfying

$$0 \leq Y_{n,j} \leq 1 \quad \text{and} \quad \sum_{j=1}^{\infty} |C_{n,j}| < \infty.$$

Then, by Theorem 7, (μ_n) has a Skorokhod representation under reasonable conditions on C_n and Y_n. For instance, for (μ_n) to admit a Skorokhod representation, it suffices that

$$C_n \text{ independent of } Y_n \text{ for every } n \geq 0,$$
$$\sum_{j=1}^{\infty} |C_{n,j} - C_{0,j}| \longrightarrow 0 \text{ in probability},$$
$$d_{TV}\big(\nu_{n,k},\, \nu_{0,k}\big) \longrightarrow 0 \text{ for all } k \geq 1,$$

where d_{TV} is total variation distance and $\nu_{n,k}$ the probability distribution of $(Y_{n,1}, \ldots, Y_{n,k})$. Note that $\nu_{n,k} = \nu_{0,k}$ (so that the last condition is trivially true) in case $Y_{n,j} = V_{n+j}$ with $V_1, V_2, \ldots$ a stationary sequence. Also, independence between C_n and Y_n can be replaced by

$$\sigma(C_{n,j}) \subset \sigma(Y_{n,1}, \ldots, Y_{n,j}) \quad \text{for all } n \geq 0 \text{ and } j \geq 1.$$

To prove $B(\mu_n, \mu_0) \to 0$, define $Z_{n,k}(t) = \sum_{j=1}^{k} C_{n,j}\, I_{\{Y_{n,j} \leq t\}}$. For each $f \in M$,

$$\begin{aligned}
|\mu_n(f) - \mu_0(f)| &\leq |Ef(Z_n) - Ef(Z_{n,k})| + |Ef(Z_{n,k}) - Ef(Z_{0,k})| + |Ef(Z_{0,k}) - Ef(Z_0)| \\
&\leq E\big\{2 \wedge d(Z_n,\, Z_{n,k})\big\} + |Ef(Z_{n,k}) - Ef(Z_{0,k})| + E\big\{2 \wedge d(Z_0,\, Z_{0,k})\big\} \\
&\leq E\big\{2 \wedge \sum_{j>k} |C_{n,j}|\big\} + |Ef(Z_{n,k}) - Ef(Z_{0,k})| + E\big\{2 \wedge \sum_{j>k} |C_{0,j}|\big\}.
\end{aligned}$$

Given $\epsilon > 0$, take $k \geq 1$ such that $E\big\{2 \wedge \sum_{j>k} |C_{0,j}|\big\} < \epsilon$. Then,

$$\limsup_n B(\mu_n, \mu_0) < 2\,\epsilon + \limsup_n \sup_{f \in M} |Ef(Z_{n,k}) - Ef(Z_{0,k})|.$$

Further, since C_n is independent of Y_n, up to a change of the underlying probability space, it can be assumed

$$\text{Prob}\big(Y_{n,j} \neq Y_{0,j} \text{ for some } j \leq k\big) = d_{TV}\big(\nu_{n,k},\, \nu_{0,k}\big);$$

see [15, Theorem 2.1]. Similarly, if $\sigma(C_{n,j}) \subset \sigma(Y_{n,1}, \ldots, Y_{n,j})$ for all n and j. Then, letting $A_{n,k} = \{Y_{n,j} = Y_{0,j} \text{ for all } j \leq k\}$, one obtains

$$\sup_{f \in M} |Ef(Z_{n,k}) - Ef(Z_{0,k})| \leq E\big\{I_{A_{n,k}}\, 2 \wedge d(Z_{n,k},\, Z_{0,k})\big\} + 2\,\text{Prob}(A^c_{n,k})$$

$$\leq E\Big\{2 \wedge \sum_{j=1}^{\infty} |C_{n,j} - C_{0,j}|\Big\} + 2\, d_{TV}\big(\nu_{n,k},\, \nu_{0,k}\big) \longrightarrow 0 \quad \text{as } n \to \infty.$$

Therefore, $B(\mu_n, \mu_0) \to 0$.

Example 15. (Functions with finite p-variation). 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 [2] and [12], some attention is paid to those processes Z_n of the type

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

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

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

is not 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 [12].

Since μ_0 fails to be separable, SRT does not apply. Instead, under some conditions, Corollary 8 yields a Skorokhod representation for (μ_n). 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 distance.

We finally prove that (μ_n) has a Skorokhod representation. To this end, as noted in point (vj) of [7], it suffices to show that 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 Skorokhod representation. Further, condition (c) can be shown to be equivalent to

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

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

(b*) $E\big\{f(U_n) \mid C\big\} \overset{a.s.}{\longrightarrow} E\big\{f(U_0) \mid C\big\}$ for each bounded continuous $f : \mathbb{R} \to \mathbb{R}$;
(c*) $\sup_A \Big| Q\big(V_n \in A \mid C\big) - Q\big(V_0 \in A \mid C\big) \Big| \overset{a.s.}{\longrightarrow} 0.$

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 8, it suffices to prove that $\big(P_c(Z_n \in \cdot) : n \geq 0\big)$ has a Skorokhod 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 [15, Theorem 2.1] 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 $Z_n \overset{P_c}{\longrightarrow} Z_0$, for

$$1 \wedge \|Z_n - Z_0\| \leq I_{\{V_n \neq V_0\}} + |U_n - U_0| \, \|\sum_k \phi_k(V_0, C)\, N_k\, x_k\|.$$

Thus, $\big(P_c(Z_n \in \cdot) : n \geq 0\big)$ admits a Skorokhod representation, and this concludes the proof.

The conditions of Example 15 are not so strong as they appear. Actually, they do not imply even $d^*(Z_n, Z_0) \overset{a.s.}{\longrightarrow} 0$ for the original processes Z_n (those defined on $(\mathcal{X}, \mathcal{E}, Q)$). In addition, by slightly modifying Example 15, S could be taken to be the space of α-Holder continuous functions, $\alpha \in (0, 1)$, and

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

Acknowledgment: We are grateful to an anonymous referee for reporting an error contained in the first draft of this paper.

References

1. T. O. Banakh, V. I. Bogachev and A. V. Kolesnikov, *Topological spaces with Skorokhod representation property*, Ukrainian Math. J. **57** (2005), 1371–1386.
2. A. Basse-O'Connor and J. Rosinski, *On the uniform convergence of random series in Skorohod space and representations of cadlag infinitely divisible processes*, Ann. Probab. **41** (2013), 4317–4341.
3. P. Berti and P. Rigo, *Convergence in distribution of non measurable random elements*, Ann. Probab. **32** (2004), 365–379.
4. P. Berti, L. Pratelli and P. Rigo, *Skorohod representation on a given probability space*, Prob. Theo. Relat. Fields **137** (2007), 277–288.
5. P. Berti, L. Pratelli and P. Rigo, *Skorohod representation theorem via disintegrations*, Sankhya **72** (2010), 208–220.

6. P. Berti, L. Pratelli and P. Rigo, *A Skorohod representation theorem for uniform distance*, Prob. Theo. Relat. Fields **150** (2011), 321–335.
7. P. Berti, L. Pratelli and P. Rigo, *A Skorohod representation theorem without separability*, Electr. Comm. Probab. **18** (2013), 1–12.
8. P. Berti, L. Pratelli and P. Rigo, *Gluing lemmas and Skorohod representations*, Electr. Comm. Probab. **20** (2015), 1–11.
9. D. Blackwell and A. Maitra, *Factorization of probability measures and absolutely measurable sets*, Proc. Amer. Math. Soc. **92** (1984), 251–254.
10. R. M. Dudley, *Distances of probability measures and random variables*, Ann. Math. Statist. **39** (1968), 1563–1572.
11. R. M. Dudley, *Uniform central limit theorems*, Cambridge University Press, 1999.
12. N. C. Jain and D. Monrad, *Gaussian measures in* B_p, Ann. Probab. **11** (1983), 46–57.
13. A. Jakubowski, *The almost sure Skorokhod representation for subsequences in nonmetric spaces*, Theo. Probab. Appl. **42** (1998), 167–174.
14. D. Ramachandran and L. Ruschendorf, *A general duality theorem for marginal problems*, Prob. Theo. Relat. Fields **101** (1995), 311–319.
15. J. Sethuraman, *Some extensions of the Skorohod representation theorem*, Sankhya **64** (2002), 884–893.
16. A. V. Skorokhod, *Limit theorems for stochastic processes*, Theo. Probab. Appl. **1** (1956), 261–290.
17. H. Thorisson, *Coupling methods in probability theory*, Scand. J. Statist. **22** (1995), 159–182.
18. H. Thorisson, *Convergence in density in finite time windows and the Skorohod representation*, Submitted (2015).
19. A. van der Vaart and J. A. Wellner, *Weak convergence and empirical processes*, Springer, 1996.
20. M. J. Wichura, *On the construction of almost uniformly convergent random variables with given weakly convergent image laws*, Ann. Math. Statist. **41** (1970), 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