



# A Strong Version of the Skorohod Representation Theorem

Luca Pratelli<sup>1</sup> · Pietro Rigo<sup>2</sup>

Received: 30 June 2021 / Revised: 15 January 2022 / Accepted: 18 January 2022 /  
Published online: 3 February 2022  
© The Author(s) 2022

## Abstract

For each  $n \geq 0$ , let  $\mu_n$  be a tight probability measure on the Borel  $\sigma$ -field of a metric space  $S$ . Let  $(T, \mathcal{C})$  be a measurable space such that the diagonal  $\{(t, t) : t \in T\}$  belongs to  $\mathcal{C} \otimes \mathcal{C}$ . Fix a measurable function  $g : S \rightarrow T$  and suppose  $\mu_n = \mu_0$  on  $g^{-1}(\mathcal{C})$  for all  $n \geq 0$ . Necessary and sufficient conditions for the existence of  $S$ -valued random variables  $X_n$ , defined on the same probability space and satisfying

$$X_n \xrightarrow{\text{a.s.}} X_0, \quad X_n \sim \mu_n \quad \text{and} \quad g(X_n) = g(X_0) \quad \text{for all } n \geq 0,$$

are given. Such conditions are then applied to several examples. The tightness condition on  $\mu_0$  can be dropped at the price of some assumptions on  $S$  and  $g$ .

**Keywords** Convergence of probability measures · Coupling · Skorohod representation theorem

**Mathematics Subject Classification 2020** 60B10 · 60A05 · 60A10

## 1 Introduction and Main Results

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

$$\sigma(g) = g^{-1}(\mathcal{C}) = \{g^{-1}(C) : C \in \mathcal{C}\}.$$

---

✉ Pietro Rigo  
pietro.rigo@unibo.it

Luca Pratelli  
pratel@mail.dm.unipi.it

<sup>1</sup> Accademia Navale, viale Italia 72, 57100 Livorno, Italy

<sup>2</sup> Dipartimento di Scienze Statistiche “P. Fortunati”, Università di Bologna, via delle Belle Arti 41, 40126 Bologna, Italy

If  $\mu_n \rightarrow \mu_0$  weakly and  $\mu_0$  is separable (namely,  $\mu_0(A) = 1$  for some separable  $A \in \mathcal{B}$ ), then, on some probability space, there are  $S$ -valued random variables  $X_n$  such that  $X_n \sim \mu_n$  for all  $n \geq 0$  and  $X_n \xrightarrow{\text{a.s.}} X_0$ . This is *Skorohod representation theorem* (SRT) as it appears after Skorohod [21], Dudley [12] and Wichura [23]. See [13, p. 130] and [22, p. 77] for historical notes, and [5] for the case where  $\mu_0$  is not separable. Some other related references are [3,4,8–10,14–16,18,20].

This paper stems from the following question. Suppose  $\mu_0$  is separable,  $\mu_n \rightarrow \mu_0$  in some sense, and

$$\mu_n = \mu_0 \text{ on } \sigma(g) \text{ for all } n \geq 0. \tag{1}$$

Is it possible to take the  $X_n$  in SRT such that  $g(X_n) = g(X_0)$  for all  $n$ ? More precisely, the question is whether, on some probability space, there are  $S$ -valued random variables  $X_n$  satisfying

$$X_n \xrightarrow{\text{a.s.}} X_0, \quad X_n \sim \mu_n \text{ and } g(X_n) = g(X_0) \text{ for all } n \geq 0. \tag{2}$$

Such a question is intriguing, quite natural from the foundational point of view, and also has some practical implications. Examples are in Sect. 3.

Some more notation is needed. If  $\mathcal{X}$  is any topological space,  $\mathcal{B}(\mathcal{X})$  denotes the Borel  $\sigma$ -field on  $\mathcal{X}$  and  $C_b(\mathcal{X})$  the set of real bounded continuous functions on  $\mathcal{X}$ . In case  $\mathcal{X} = S$ , we just write  $\mathcal{B}$  instead of  $\mathcal{B}(S)$ . Moreover, we say that  $(T, \mathcal{C})$  is *diagonal* if

$$\{(t, t) : t \in T\} \in \mathcal{C} \otimes \mathcal{C}.$$

In particular,  $(T, \mathcal{C})$  is diagonal if  $T$  is a separable metric space and  $\mathcal{C} = \mathcal{B}(T)$ .

We are now able to state our main result.

**Theorem 1** *Suppose  $(T, \mathcal{C})$  is diagonal,  $\mu_n$  is tight and  $\mu_n = \mu_0$  on  $\sigma(g)$  for every  $n \geq 0$ . Then, there are  $S$ -valued random variables  $X_n$ , defined on the same probability space, such that*

$$X_n \sim \mu_n \text{ and } g(X_n) = g(X_0) \text{ for all } n \geq 0. \tag{3}$$

Such  $X_n$  can be taken to meet condition (2) (namely, they also satisfy  $X_n \xrightarrow{\text{a.s.}} X_0$ ) if and only if

$$E_{\mu_n}(f | g) \xrightarrow{\mu_0\text{-a.s.}} E_{\mu_0}(f | g) \quad \text{for each } f \in C_b(S). \tag{4}$$

In a nutshell, Theorem 1 states that, under mild conditions on  $\mu_n$  and  $(T, \mathcal{C})$ ,

$$(1) \Leftrightarrow (3) \quad \text{and} \quad (1) \wedge (4) \Leftrightarrow (2).$$

The second equivalence is possibly more meaningful but the first may be useful as well; see, e.g., Example 3. We also note that the notation  $E_{\mu_n}(f \mid g)$  stands for

$$E_{\mu_n}(f \mid g) = E_{\mu_n}[f \mid g^{-1}(\mathcal{C})].$$

Some more remarks are in order.

- (i) Any countably generated sub- $\sigma$ -field  $\mathcal{G} \subset \mathcal{B}$  can be written as  $\mathcal{G} = \sigma(g)$  for some Borel measurable function  $g : S \rightarrow \mathbb{R}$ .
- (ii) Some assumption on  $(T, \mathcal{C})$  is necessary. As an obvious example, take  $T = S$  and  $\mathcal{C}$  the collection of countable and co-countable subsets of  $S$ . If  $g$  is the identity map, conditions (1) and (4) hold true whenever  $\mu_n \rightarrow \mu_0$  weakly and  $\mu_n\{x\} = 0$  for all  $n \geq 0$  and  $x \in S$ . But, since  $g$  is the identity, condition (3) fails unless  $\mu_n = \mu_0$  on all of  $\mathcal{B}$ .
- (iii) Let  $\mu = \sum_{n=0}^{\infty} 2^{-n-1} \mu_n$ . In view of Theorem 1, it is tempting to say that  $\mu_n$  converges to  $\mu_0$  conditionally with respect to  $g$ , written as  $\mu_n \xrightarrow{g} \mu_0$ , whenever

$$E_{\mu_n}(f \mid g) \xrightarrow{\mu\text{-a.s.}} E_{\mu_0}(f \mid g) \quad \text{for each } f \in C_b(S).$$

This notion of convergence allows for a version of SRT and reduces to weak convergence in the special case where  $g$  is constant. Furthermore, if  $S$  is Polish,  $\mu_n \xrightarrow{g} \mu_0$  is equivalent to

$$\gamma_n(x) \rightarrow \gamma_0(x) \text{ weakly for } \mu\text{-almost all } x \in S,$$

where  $\gamma_n = \{\gamma_n(x) : x \in S\}$  is a regular conditional distribution for  $\mu_n$  given  $\sigma(g)$ ; see Sect. 2.

- (iv) In condition (4),  $C_b(S)$  can be replaced by the set of bounded Lipschitz functions on  $S$ . Note also that, if  $g$  is constant, condition (1) is trivially true and condition (4) reduces to  $\mu_n \rightarrow \mu_0$  weakly. Thus, when the  $\mu_n$  are tight, SRT is contained in Theorem 1.
- (v) As an obvious application, suppose the  $\mu_n$  are tight and take a countable Borel partition  $\{H_0, H_1, \dots\}$  of  $S$ . Then, there are random variables  $X_n$  such that

$$X_n \xrightarrow{\text{a.s.}} X_0, \quad X_n \sim \mu_n \text{ and } 1_{H_j}(X_n) = 1_{H_j}(X_0) \text{ for all } n, j \geq 0$$

if and only if

$$\begin{aligned} \mu_n(H_j) &= \mu_0(H_j) \text{ for all } n, j \geq 0 \text{ and} \\ E_{\mu_0}(f 1_{H_j}) &= \lim_n E_{\mu_n}(f 1_{H_j}) \text{ for all } f \in C_b(S) \text{ and } j \geq 0. \end{aligned}$$

This follows from Theorem 1 with  $T = \{0, 1, \dots\}$ ,  $\mathcal{C}$  the power set of  $T$ , and  $g(x) = j$  for all  $x \in H_j$ .

- (vi) Suppose  $S$  is Polish,  $\mathcal{G} \subset \mathcal{B}$  is a countably generated sub- $\sigma$ -field, and some element  $A \in \mathcal{G}$  satisfies  $\{x\} \in \mathcal{G}$  for all  $x \in A$ . Then,  $A \cap \mathcal{G} = A \cap \mathcal{B}$  by a result of Blackwell [7], and one can take  $T = S$  and  $g(x) = x$  for all  $x \in A$ . Therefore, under condition (3), one obtains  $X_n = X_0$  on the set  $\{X_0 \in A\}$ .
- (vii) A weak version of condition (2) is

$$X_n \rightarrow X_0 \text{ in probability, } X_n \sim \mu_n \text{ and } g(X_n) = g(X_0) \text{ for all } n \geq 0.$$

This condition can be characterized by the same argument as Theorem 1. If  $(T, \mathcal{C})$  is diagonal, the  $\mu_n$  are tight and condition (1) holds, the weak version of (2) is actually equivalent to

$$E_{\mu_n}(f | g) \rightarrow E_{\mu_0}(f | g), \text{ in } \mu_0\text{-probability, for each } f \in C_b(S). \quad (5)$$

Indeed, as shown by Example 1, it may be that conditions (1) and (5) hold but condition (4) fails.

To motivate Theorem 1, in addition to the previous remarks, some examples are given in Sect. 3. Here, we close this section with three corollaries.

Given the metric spaces  $S_1$  and  $S_2$ , define

$$S = S_1 \times S_2 \text{ and } g(x, y) = x \text{ for all } (x, y) \in S.$$

If  $S_1$  is separable, it is possible to let  $T = S_1$  and  $\mathcal{C} = \mathcal{B}(S_1)$ , so that

$$\sigma(g) = \{A \times S_2 : A \in \mathcal{B}(S_1)\}.$$

Therefore, Theorem 1 yields:

**Corollary 2** *Let  $v(\cdot) = \mu_0(\cdot \times S_2)$  be the marginal of  $\mu_0$  on  $S_1$ . Suppose  $S_1$  is separable,  $\mu_n$  is tight and*

$$\mu_n(\cdot \times S_2) = v(\cdot) \text{ for every } n \geq 0.$$

*Then, on some probability space, there are random variables  $Y$  and  $(Z_n : n \geq 0)$ , where  $Y$  is  $S_1$ -valued and the  $Z_n$  are  $S_2$ -valued, such that*

$$(Y, Z_n) \sim \mu_n \text{ for all } n \geq 0.$$

*Moreover, the  $Z_n$  can be taken such that  $Z_n \xrightarrow{a.s.} Z_0$  if and only if*

$$E_{\mu_n}(f | g) \xrightarrow{v\text{-a.s.}} E_{\mu_0}(f | g) \text{ for each } f \in C_b(S_2). \quad (6)$$

An obvious application of Corollary 2 is as follows. Let  $(U_n, V_n)$  be a sequence of random variables such that  $U_n \sim U_0$  for all  $n \geq 0$ . For some reason, we would like to replace  $(U_n, V_n)$  with another sequence  $X_n = (Y, Z_n)$ , possibly defined on a different

probability space, such that  $X_n \xrightarrow{\text{a.s.}} X_0$  and  $X_n \sim (U_n, V_n)$  for all  $n$ . Note that all the  $X_n$  have the same first coordinate  $Y$ , and this is basic in various frameworks, including optimal transport and stochastic control. Corollary 2 states that, under mild conditions, to replace  $(U_n, V_n)$  with  $X_n$  is admissible if and only if condition (6) holds with  $\mu_n$  the probability distribution of  $(U_n, V_n)$ . See also [4, Prop. 2] for an analogous result.

The second corollary concerns the tightness condition. The main reason for assuming  $\mu_n$  tight, for every  $n \geq 0$ , is to reduce the proof of Theorem 1 to the special case where  $S$  is a Borel subset of a Polish space. We do not know whether the tightness condition can be weakened for all  $n \geq 0$ . However, arguing as in [3], the tightness condition on  $\mu_0$  can be dropped at the price of requiring something more on  $S$  and  $g$ .

**Corollary 3** *Suppose  $(T, \mathcal{C})$  is diagonal,  $S$  is a subset of a Polish space  $\mathcal{X}$ , and  $g$  can be extended to a function  $\phi : \mathcal{X} \rightarrow T$  such that  $\phi^{-1}(\mathcal{C}) \subset \mathcal{B}(\mathcal{X})$ . Suppose also that  $\mu_n$  is tight and  $\mu_n = \mu_0$  on  $\sigma(g)$  for each  $n \geq 1$ . Then, condition (3) holds for some  $S$ -valued random variables  $X_n$  (defined on the same probability space). Moreover, under condition (4), the  $X_n$  also satisfy condition (2).*

The third corollary deals with the probability space, say  $(\Omega, \mathcal{A}, P)$ , where the  $X_n$  can be defined. In Theorem 1 and Corollary 2, one can take  $\Omega = (0, 1)^2$ ,  $\mathcal{A} = \mathcal{B}(\Omega)$  and  $P$  the Lebesgue measure. In Corollary 3, since  $\mu_0$  is not necessarily tight,  $\Omega$  is a suitable subset of  $(0, 1)^2$  and  $P$  the Lebesgue-outer measure. In our last result, the  $\mu_n$  are the probability distributions of random variables  $U_n$  defined on the same probability space. In this case, a question is whether the  $X_n$  can be defined on the probability space where the  $U_n$  live.

**Corollary 4** *Let  $(U_n : n \geq 0)$  be a sequence of  $S$ -valued random variables on the probability space  $(\Omega, \mathcal{A}, P)$ . Define  $\mu_n(\cdot) = P(U_n \in \cdot)$  and suppose*

- *$(T, \mathcal{C})$  is diagonal and  $(\Omega, \mathcal{A}, P)$  is nonatomic;*
- *Conditions (1) and (4) hold and  $\mu_n$  is tight for each  $n \geq 0$ .*

*Then,  $(\Omega, \mathcal{A}, P)$  supports  $S$ -valued random variables  $X_n$  satisfying condition (2).*

It is worth noting that, for  $(\Omega, \mathcal{A}, P)$  to be nonatomic, a sufficient condition is  $P(U_n = x) = 0$  for some  $n \geq 0$  and all  $x \in S$ ; see, e.g., [3, Lem. 3.3].

## 2 Preliminaries

In the sequel,  $m$  is the Lebesgue measure on  $\mathcal{B}((0, 1))$  and  $\delta_z$  the unit mass at the point  $z$ . We denote by  $\mathcal{P}$  the collection of all probability measures on  $\mathcal{B}$  and we write  $\mu(f) = \int f d\mu$  whenever  $\mu \in \mathcal{P}$  and  $f : S \rightarrow \mathbb{R}$  is a bounded Borel function.

Let  $F \subset C_b(S)$  and  $\mathcal{Q} \subset \mathcal{P}$ . Say that  $F$  is a *convergence determining class* for  $\mathcal{Q}$  if, for any sequence  $(\lambda_n : n \geq 0) \subset \mathcal{Q}$ ,

$$\lambda_n \rightarrow \lambda_0 \text{ weakly} \iff \lambda_0(f) = \lim_n \lambda_n(f) \text{ for each } f \in F.$$

Let  $\mathcal{G} \subset \mathcal{B}$  be a sub- $\sigma$ -field. We recall that a *regular conditional distribution* (r.c.d.) for  $\mu_n$  given  $\mathcal{G}$  is a collection  $\gamma_n = \{\gamma_n(x) : x \in S\}$  such that

- $\gamma_n(x) \in \mathcal{P}$  for each  $x \in S$ ;
- $x \mapsto \gamma_n(x)(B)$  is  $\mathcal{G}$ -measurable for each  $B \in \mathcal{B}$ ;
- $\int_A \gamma_n(x)(B) \mu_n(dx) = \mu_n(A \cap B)$  for all  $A \in \mathcal{G}$  and  $B \in \mathcal{B}$ .

A.r.c.d. for  $\mu_n$  given  $\mathcal{G}$  exists and is  $\mu_n$ -a.s. unique whenever  $\mathcal{B}$  is countably generated and  $\mu_n$  tight.

The following version of SRT is involved in the proof of Theorem 1.

**Theorem 5** (Blackwell and Dubins [8]) *If  $S$  is Polish, there is a Borel map  $\Phi : (0, 1) \times \mathcal{P} \rightarrow S$  such that*

- $m\{\beta \in (0, 1) : \Phi(\beta, \lambda) \in B\} = \lambda(B)$  for all  $\lambda \in \mathcal{P}$  and  $B \in \mathcal{B}$ ;
- $m\{\beta \in (0, 1) : \Phi(\beta, \lambda_n) \rightarrow \Phi(\beta, \lambda_0)\} = 1$  if  $\lambda_n \in \mathcal{P}$  and  $\lambda_n \rightarrow \lambda_0$  weakly.

A clear and detailed proof of Theorem 5 can be found in [16, pp. 52–54] (Blackwell and Dubins actually provide only a sketch of the proof). Moreover, it is straightforward to verify that Theorem 5 is still valid if  $S$  is only a Borel subset of a Polish space.

We finally state two technical lemmas. The first is certainly known, but we give a proof since we are not aware of any explicit reference.

**Lemma 6** *A measurable space  $(T, \mathcal{C})$  is diagonal if and only if there is a countably generated sub- $\sigma$ -field  $\mathcal{C}_0 \subset \mathcal{C}$  such that  $\{t\} \in \mathcal{C}_0$  for all  $t \in T$ . Moreover, if  $(T, \mathcal{C})$  is diagonal and  $Q$  is a probability measure on  $\mathcal{C} \otimes \mathcal{C}$ , then*

$$Q\{(t, t) : t \in T\} = 1 \iff Q(A \times A^c) = 0 \text{ for all } A \in \mathcal{C}. \tag{7}$$

**Proof** Let  $\Delta = \{(t, t) : t \in T\}$ . If  $(T, \mathcal{C})$  is diagonal, there are  $A_n, B_n \in \mathcal{C}$  such that  $\Delta \in \sigma(A_1 \times B_1, A_2 \times B_2, \dots)$ . Let  $\mathcal{C}_0 = \sigma(A_1, B_1, A_2, B_2, \dots)$ . Then,  $\mathcal{C}_0$  is countably generated and  $\Delta \in \mathcal{C}_0 \otimes \mathcal{C}_0$ . Therefore,  $\{t\} = \{u : (t, u) \in \Delta\} \in \mathcal{C}_0$  for all  $t \in T$ . Conversely, if  $\mathcal{C}_0 \subset \mathcal{C}$  is countably generated and includes the singletons, there is a distance  $\rho$  on  $T$  such that  $(T, \rho)$  is a separable metric space and  $\mathcal{C}_0$  is the Borel  $\sigma$ -field on  $T$  under  $\rho$ . Therefore,  $\Delta \in \mathcal{C}_0 \otimes \mathcal{C}_0 \subset \mathcal{C} \otimes \mathcal{C}$ .

Finally, we turn to (7). Fix  $\mathcal{C}_0$  as above, and recall that  $(T, \mathcal{C}_0)$  can be regarded as a separable metric space equipped with the Borel  $\sigma$ -field. Hence,  $Q(\Delta) = 1$  provided  $Q(A \times A^c) = 0$  for all  $A \in \mathcal{C}_0$ . This proves " $\Leftarrow$ " while " $\Rightarrow$ " is trivial.  $\square$

**Lemma 7** *If  $B$  is a  $\sigma$ -compact subset of  $S$ , there is a countable convergence determining class for  $\mathcal{Q} = \{\mu \in \mathcal{P} : \mu(B) = 1\}$ .*

**Proof** Since  $B$  is  $\sigma$ -compact, there is a sequence  $(x_n) \subset B$  such that  $\overline{\{x_1, x_2, \dots\}} = \overline{B}$ . Let  $H = [0, 1]^\infty$  be the Hilbert cube (equipped with the usual metric) and

$$h(x) = (d(x, x_1) \wedge 1, d(x, x_2) \wedge 1, \dots) \quad \text{for all } x \in S,$$

where  $d$  is the distance on  $S$ . Then,  $h : S \rightarrow H$  is continuous, and it is an homeomorphism as a map  $h : B \rightarrow h(B)$ , and  $h(B) \in \mathcal{B}(H)$  (for  $h(B)$  is  $\sigma$ -compact). Take a countable subset  $G \subset C_b(H)$ , dense in  $C_b(H)$  under the sup-norm, and define

$$F = \{g \circ h : g \in G\}.$$

Then,  $F \subset C_b(S)$  is countable. Suppose now that  $\lambda_n \in \mathcal{Q}$  and  $\lambda_n(f) \rightarrow \lambda_0(f)$  for each  $f \in F$ . Then,  $\lambda_n \circ h^{-1} \rightarrow \lambda_0 \circ h^{-1}$  weakly, since  $G$  is dense in  $C_b(H)$ . Hence,  $\lambda_n \rightarrow \lambda_0$  weakly follows from  $h : B \rightarrow h(B)$  is an homeomorphism and  $\lambda_n \circ h^{-1}(h(B)) \geq \lambda_n(B) = 1$  for all  $n \geq 0$ . This concludes the proof.  $\square$

### 3 Examples

It may be that conditions (1) and (5) hold but condition (4) fails. In this case, condition (2) can not be realized (since (4) fails). However, as noted in Remark (vii), some random variables  $X_n$  satisfy a weak version of (2).

**Example 1** Let

$$S = [-1, 1], \quad T = [0, 1], \quad \mathcal{C} = \mathcal{B}(T) \quad \text{and} \quad g(x) = |x|.$$

Take a sequence  $(B_n : n \geq 1) \subset \mathcal{B}[0, 1]$  such that

$$\bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} B_j = [0, 1] \quad \text{and} \quad \lim_n m(B_n) = 0.$$

Using such  $B_n$ , define  $C_n = [-1, 1] \setminus B_n$  and

$$f_n(x) = (1/2) 1_{C_n}(x) + (1/4) \{1_{B_n}(x) + 1_{B_n}(-x)\} \quad \text{for all } x \in [-1, 1].$$

Moreover, let  $f_0 = 1/2$  and

$$\mu_n(dx) = f_n(x) dx \quad \text{for all } n \geq 0.$$

Then,

$$\mu_n \circ g^{-1} = \mu_0 \circ g^{-1} = m \quad \text{for all } n \geq 0.$$

Furthermore, a r.c.d.  $\gamma_n = \{\gamma_n(x) : x \in S\}$  for  $\mu_n$  given  $\sigma(g)$  (see Sect. 2) is

$$\begin{aligned} \gamma_0(x) &= (1/2) \{ \delta_{|x|} + \delta_{-|x|} \} \quad \text{and} \\ \gamma_n(x) &= \gamma_0(x) \text{ if } |x| \notin B_n \quad \text{and} \quad \gamma_n(x) = \frac{\delta_{|x|} + 3 \delta_{-|x|}}{4} \text{ if } |x| \in B_n. \end{aligned}$$

Thus, for each  $f \in C_b[-1, 1]$  and  $\epsilon > 0$ ,

$$\mu_0 \{ x \in [-1, 1] : |\gamma_n(x)(f) - \gamma_0(x)(f)| > \epsilon \} \leq m(B_n) \rightarrow 0.$$

Therefore, conditions (1) and (5) are both satisfied. However, condition (4) fails, since every  $x \in [-1, 1]$  is such that  $|x| \in B_n$  for infinitely many  $n$ .

In a sense, the next example completes the previous one.

**Example 2** Let  $S = S_1 \times S_2$ , where  $S_1$  and  $S_2$  are Polish spaces, and  $(U_n, V_n)$  a sequence of  $S$ -valued random variables such that

$$(U_n, V_n) \rightarrow (U_0, V_0) \text{ in distribution and } U_n \sim U_0 \text{ for all } n \geq 0.$$

Then, by [17, Cor. 2.9], one obtains  $E[f(U_n, V_n)] \rightarrow E[f(U_0, V_0)]$  for each bounded Borel function  $f : S \rightarrow \mathbb{R}$  which is continuous in the second coordinate. Nevertheless, it may be that  $E[f(V_n) | U_n]$  does not converge to  $E[f(V_0) | U_0]$ , even in distribution, for some  $f \in C_b(S_2)$ . As an example, take  $S_1 = S_2 = (0, 1)$  and  $(U_0, V_0)$  uniform on  $S = (0, 1)^2$ . Then,  $(U_n, V_n) \rightarrow (U_0, V_0)$  in distribution for some sequence  $(U_n, V_n)$  such that

$$U_n \sim V_n \sim m \text{ and } V_n \stackrel{\text{a.s.}}{=} \varphi_n(U_n) \text{ for all } n \geq 1$$

where the  $\varphi_n : (0, 1) \rightarrow (0, 1)$  are suitable Borel functions; see, e.g., [1, Prop. 2.7]. Therefore,

$$E(V_0 | U_0) \stackrel{\text{a.s.}}{=} E(V_0) = 1/2 \text{ and } E(V_n | U_n) \stackrel{\text{a.s.}}{=} V_n \sim m \text{ for all } n \geq 1.$$

Another example, similar to the previous one, is [11, Ex. 6.1]. Even in this case,  $E[f(V_n) | U_n]$  does not converge in distribution to  $E[f(V_0) | U_0]$  for some  $f \in C_b(S_2)$ . In addition, one also obtains  $U_n \sim V_n \sim m$  for all  $n \geq 0$  and  $E[h(U_0, V_0)] = \lim_n E[h(U_n, V_n)]$  for each bounded Borel function  $h : S \rightarrow \mathbb{R}$ .

Two remarks are in order. First, to obtain the weak version of (2) involved in Remark (vii), condition (5) cannot be dropped. As noted in [11], however, condition (5) can be weakened into

$$E_{\mu_n}(f | g) \rightarrow E_{\mu_0}(f | g), \text{ in distribution, for each } f \in C_b(S).$$

Second, if  $((U, V_n) : n \geq 0)$  are  $S$ -valued random variables, it may be that  $V_n \rightarrow V_0$  in probability and yet there are not random variables  $(Y, Z_n)$  satisfying  $Z_n \xrightarrow{\text{a.s.}} Z_0$  and  $(Y, Z_n) \sim (U, V_n)$  for all  $n \geq 0$ ; see Example 1.

Let  $d$  be the distance on  $S$ . Sometimes, one aims to realize the  $\mu_n$  by random variables which converge (say in probability) under some distance  $\rho$  stronger than  $d$ ; see [4,5]. This motivates the next example.

**Example 3** Suppose  $S$  is separable and the  $\mu_n$  are tight. Fix  $A \in \mathcal{B}$  and define

$$b(x, y) = 1_A(x) 1 \wedge d(x, y) + 1_{A^c}(x) 1 \wedge \rho(x, y) \text{ for } (x, y) \in S^2,$$

where  $\rho$  is any distance on  $S$  such that the map  $(x, y) \mapsto \rho(x, y)$  is measurable with respect to  $\mathcal{B} \otimes \mathcal{B}$ . For instance,  $\rho$  could be the 0-1 distance. Or else,  $S$  could be the set of real cadlag functions on  $[0, 1]$ ,  $d$  the Skorohod distance, and  $\rho$  the uniform distance.



A question is whether there are  $S$ -valued random variables  $X_n$  such that

$$\lim_n E\{b(X_0, X_n)\} = 0 \quad \text{and} \quad X_n \sim \mu_n \text{ for each } n \geq 0. \tag{8}$$

Equivalently, the question is whether the  $\mu_n$  can be realized by some  $X_n$  such that  $d(X_n, X_0) \rightarrow 0$  in probability on the set  $\{X_0 \in A\}$  and  $\rho(X_n, X_0) \rightarrow 0$  in probability on  $\{X_0 \notin A\}$ .

Corollary 2 allows to answer this question. For any  $\mu, \lambda \in \mathcal{P}$ , let  $\Gamma(\mu, \lambda)$  denote the collection of those probability measures  $\tau$  on  $\mathcal{B} \otimes \mathcal{B}$  such that  $\tau(\cdot \times S) = \mu$  and  $\tau(S \times \cdot) = \lambda$ . By Corollary 2, there are  $X_n$  satisfying condition (8) if and only if

$$\lim_n \tau_n(b) = 0 \quad \text{for some sequence } \tau_n \in \Gamma(\mu_0, \mu_n).$$

In turn, by a duality theorem [19], the above condition can be written as

$$\sup_{(f,h)} \{\mu_0(f) + \mu_n(h)\} \rightarrow 0 \tag{9}$$

where sup is over all pairs  $(f, h)$  of bounded Borel functions on  $S$  such that

$$f(x) + h(y) \leq b(x, y) \quad \text{for all } (x, y) \in S^2.$$

The equivalence between (8) and (9), obtained above, improves [4, Theorem 4]. It also improves [20, Theorem 2.1] in the special case where  $S$  is separable and the  $\mu_n$  are tight.

The next example deals with exchangeable sequences, but the same argument applies to many other types of sequences, including martingale difference and stationary.

**Example 4** Let  $(T_n : n \geq 1)$  be an exchangeable sequence of real random variables on the probability space  $(\Omega, \mathcal{A}, P)$ . Suppose  $E(T_1^2) < \infty$  and define

$$\begin{aligned} \mathcal{T} &= \bigcap_n \sigma(T_n, T_{n+1}, \dots), \quad V = E(T_1^2 | \mathcal{T}) - E(T_1 | \mathcal{T})^2 \\ \text{and} \quad M_n &= \sqrt{n} \left\{ \frac{\sum_{i=1}^n T_i}{n} - E(T_1 | \mathcal{T}) \right\}. \end{aligned}$$

Moreover, let  $N$  be a standard normal random variable independent of  $(T_n)$  and

$$M_0 = \sqrt{V}N.$$

Then,  $M_n \rightarrow M_0$  in distribution (and even *stably*); see, e.g., [2, Theorem 3.1]. In addition, even if the tail  $\sigma$ -field  $\mathcal{T}$  is not countably generated, there is a real random variable  $U$  such that

$$\sigma(\mathcal{T} \cup \mathcal{N}) = \sigma(U \cup \mathcal{N}) \quad \text{where } \mathcal{N} = \{A \in \mathcal{A} : P(A) = 0\}.$$

Apart from trivial cases,  $M_n$  fails to converge in probability. However, thanks to Corollary 2, some real random variables  $Y$  and  $Z_n$  satisfy

$$Z_n \xrightarrow{\text{a.s.}} Z_0 \quad \text{and} \quad (Y, Z_n) \sim (U, M_n) \text{ for each } n \geq 0.$$

Take in fact  $S_1 = S_2 = \mathbb{R}$  and define  $\mu_n$  to be the probability distribution of  $(U, M_n)$ . Conditionally on  $\mathcal{T}$ , the sequence  $(T_n)$  is i.i.d. with mean  $E(T_1 | \mathcal{T})$  and variance  $V$ . Hence, given  $f \in C_b(\mathbb{R})$ , the standard CLT yields

$$E\{f(M_n) | U\} = E\{f(M_n) | \mathcal{T}\} \xrightarrow{\text{a.s.}} \int f(x) N(0, V)(dx),$$

where  $N(0, V)$  denotes the Gaussian law with mean 0 and (random) variance  $V$  with  $N(0, 0) = \delta_0$ . On the other hand,

$$E\{f(M_0) | U\} = E\{f(M_0) | \mathcal{T}\} = \int f(x) N(0, V)(dx).$$

Therefore, condition (6) is satisfied.

Our last example concerns conditionally identically distributed sequences.

**Example 5** Let  $S = \mathbb{R}^\infty$  and  $T$  the set of probability measures on  $\mathcal{B}(\mathbb{R})$  (equipped with the topology of weak convergence). Moreover, let  $\mathcal{C} = \mathcal{B}(T)$  and  $g : S \rightarrow T$  the weak limit of the empirical measures. Precisely, for each  $x = (x_1, x_2, \dots) \in S$ ,

$$g(x) = \lim_n \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \text{ provided the limit exists and } g(x) = \delta_{x_1} \text{ otherwise,}$$

where the limit is meant as a weak limit of probability measures.

A sequence  $Y = (Y_1, Y_2, \dots)$  of real random variables is *conditionally identically distributed* (c.i.d.) if

$$(Y_1, \dots, Y_n, Y_{n+2}) \sim (Y_1, \dots, Y_n, Y_{n+1}) \text{ for all } n \geq 0.$$

An exchangeable sequence is c.i.d. but not conversely; see, e.g., [6] and references therein. However, as in the exchangeable case, if  $Y$  is c.i.d. one obtains

$$\frac{1}{n} \sum_{i=1}^n \delta_{Y_i}(B) \xrightarrow{\text{a.s.}} g(Y)(B) \quad \text{for all } B \in \mathcal{B}(\mathbb{R}).$$

Suppose  $Y$  is c.i.d. and define

$$N = g(Y), \quad \mu_0(\cdot) = E(N^\infty(\cdot)) \quad \text{and} \quad \mu_n(\cdot) = P((Y_n, Y_{n+1}, \dots) \in \cdot).$$

Here,  $N^\infty = N \times N \times \dots$  denotes the random probability measure on  $(S, \mathcal{B})$  which makes the coordinate random variables i.i.d. with common distribution  $N$ . Hence,  $\mu_0$  is exchangeable (for it is a mixture of i.i.d. probability measures). Moreover,  $\mu_n \rightarrow \mu_0$  weakly and  $\mu_n = \mu_0$  on  $\sigma(g)$  because of [2, Theorem 2.6].

Since  $\mu_n = \mu_0$  on  $\sigma(g)$ , Theorem 1 applies. Thus, under condition (4), there are real random sequences

$$X_n = (X_{n,1}, X_{n,2}, \dots), \quad n \geq 0,$$

such that

$$\begin{aligned} X_{0,j} &\stackrel{\text{a.s.}}{=} \lim_n X_{n,j} \quad \text{and} \quad g(X_j) = g(X_0) \quad \text{for each } j \geq 1, \\ X_0 &\sim \mu_0 \quad \text{and} \quad X_n \sim (Y_n, Y_{n+1}, \dots) \quad \text{for each } n \geq 1. \end{aligned}$$

We finally turn to condition (4), namely

$$E\{f(Y_n, Y_{n+1}, \dots) \mid N\} \xrightarrow{\text{a.s.}} \int f(x) N^\infty(dx) \quad \text{for all } f \in C_b(S).$$

We do not know whether this condition holds for any c.i.d. sequence, but it holds in some (meaningful) special cases, including  $N$  a.s. discrete; see also [6, Th. 18].

## 4 Proofs

From now on, to make the notation easier, we let

$$\mathcal{G} = \sigma(g).$$

### 4.1 Proof of Theorem 1

We first prove that (2)  $\Rightarrow$  (4). Suppose condition (2) holds for some  $S$ -valued random variables  $X_n$  defined on the probability space  $(\Omega, \mathcal{A}, P)$ . Fix  $f \in C_b(S)$ . Because of condition (1), for each  $n \geq 0$ , there is a measurable function  $\phi_n : (T, \mathcal{C}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  such that

$$E_{\mu_n}(f \mid g) = \phi_n \circ g, \quad \mu_0\text{-a.s.}$$

It is straightforward to verify that  $E_P\{f(X_n) \mid g(X_n)\} = \phi_n[g(X_n)]$  a.s. Hence,  $g(X_n) = g(X_0)$  implies

$$E_P\{f(X_n) \mid g(X_n)\} = \phi_n[g(X_0)] \quad \text{a.s.}$$

On the other hand, since  $f$  is bounded,  $f(X_n) \xrightarrow{\text{a.s.}} f(X_0)$  and  $g(X_n) = g(X_0)$  for all  $n$ , one also obtains

$$E_P\{f(X_n) \mid g(X_n)\} = E_P\{f(X_n) \mid g(X_0)\} \xrightarrow{\text{a.s.}} E_P\{f(X_0) \mid g(X_0)\}.$$

Therefore,  $\phi_n[g(X_0)] \xrightarrow{\text{a.s.}} \phi_0[g(X_0)]$ , or equivalently

$$\begin{aligned} \mu_0\left(E_{\mu_n}(f \mid g) \rightarrow E_{\mu_0}(f \mid g)\right) &= \mu_0\left(\phi_n \circ g \rightarrow \phi_0 \circ g\right) \\ &= P\left(\phi_n[g(X_0)] \rightarrow \phi_0[g(X_0)]\right) = 1. \end{aligned}$$

Thus, condition (4) holds.

The rest of the proof is split into two steps.

**Step 1** We first suppose that  $S$  is a Borel subset of a Polish space. Then, for every  $n \geq 0$ , we can fix a r.c.d.  $\gamma_n = \{\gamma_n(x) : x \in S\}$  for  $\mu_n$  given  $\mathcal{G}$ . We will write

$$\gamma_n(x)(f) = \int f(y) \gamma_n(x)(dy)$$

for all  $x \in S$  and all bounded Borel functions  $f : S \rightarrow \mathbb{R}$ .

Let  $\Phi : (0, 1) \times \mathcal{P} \rightarrow S$  be the Borel map involved in Theorem 5 and

$$(\Omega, \mathcal{A}, P) = \left((0, 1)^2, \mathcal{B}((0, 1)^2), m^2\right).$$

For each  $n \geq 0$  and  $(\alpha, \beta) \in (0, 1)^2$ , define

$$\phi(\alpha) = \Phi(\alpha, \mu_0) \quad \text{and} \quad X_n(\alpha, \beta) = \Phi\left(\beta, \gamma_n[\phi(\alpha)]\right).$$

The  $X_n$  are  $S$ -valued random variables on  $(\Omega, \mathcal{A}, P)$ . We now prove that they meet condition (3).

Fix  $n \geq 0$  and note that

$$m\left\{\beta : X_n(\alpha, \beta) \in B\right\} = m\left\{\beta : \Phi\left(\beta, \gamma_n[\phi(\alpha)]\right) \in B\right\} = \gamma_n[\phi(\alpha)](B)$$

for all  $\alpha \in (0, 1)$  and  $B \in \mathcal{B}$ . Hence, Fubini’s theorem yields

$$\begin{aligned} P(X_n \in B) &= \int_0^1 m\left\{\beta : X_n(\alpha, \beta) \in B\right\} d\alpha = \int_0^1 \gamma_n[\phi(\alpha)](B) d\alpha \\ &= \int \gamma_n(x)(B) \mu_0(dx) = \int \gamma_n(x)(B) \mu_n(dx) = \mu_n(B), \end{aligned}$$

where the third equality is because  $m \circ \phi^{-1} = \mu_0$  while the fourth depends on  $\mu_n = \mu_0$  on  $\mathcal{G}$  and  $x \mapsto \gamma_n(x)(B)$  is  $\mathcal{G}$ -measurable. This proves  $X_n \sim \mu_n$ .

We next prove  $P(g(X_n) \neq g(X_0)) = 0$ . To this end, since  $(T, \mathcal{C})$  is diagonal, it suffices to show that  $P(g(X_n) \in C, g(X_0) \notin C) = 0$  for all  $C \in \mathcal{C}$ ; see Lemma 6. Fix  $C \in \mathcal{C}$ , define  $A = \{g \in \mathcal{C}\}$ , and note that

$$\begin{aligned} & m\{\beta : X_n(\alpha, \beta) \in A, X_0(\alpha, \beta) \notin A\} \\ & \leq \min\{m\{\beta : X_n(\alpha, \beta) \in A\}, m\{\beta : X_0(\alpha, \beta) \notin A\}\} \\ & = \min\{\gamma_n[\phi(\alpha)](A), \gamma_0[\phi(\alpha)](A^c)\} \quad \text{for all } \alpha \in (0, 1). \end{aligned}$$

Recalling that  $m \circ \phi^{-1} = \mu_0$ , one obtains

$$\begin{aligned} P(g(X_n) \in C, g(X_0) \notin C) &= P(X_n \in A, X_0 \notin A) \\ &= \int_0^1 m\{\beta : X_n(\alpha, \beta) \in A, X_0(\alpha, \beta) \notin A\} \, d\alpha \\ &\leq \int_0^1 \min\{\gamma_n[\phi(\alpha)](A), \gamma_0[\phi(\alpha)](A^c)\} \, d\alpha \\ &= \int \min\{\gamma_n(x)(A), \gamma_0(x)(A^c)\} \, \mu_0(dx). \end{aligned}$$

Let

$$G = \{x \in S : \gamma_n(x)(A) = 1_A(x), \gamma_0(x)(A^c) = 1_{A^c}(x)\}.$$

Since  $\{x \in S : \gamma_n(x)(A) = 1_A(x)\}$  belongs to  $\mathcal{G}$ , condition (1) implies

$$\mu_0(G) = \mu_0\{x \in S : \gamma_n(x)(A) = 1_A(x)\} = \mu_n\{x \in S : \gamma_n(x)(A) = 1_A(x)\} = 1.$$

In turn, this implies

$$P(g(X_n) \in C, g(X_0) \notin C) \leq \int_G \min\{1_A(x), 1_{A^c}(x)\} \, \mu_0(dx) = 0.$$

Hence,  $P(g(X_n) \neq g(X_0)) = 0$ . To get  $g(X_n) = g(X_0)$  everywhere, as prescribed by condition (3), it suffices to modify  $X_n$  on a  $P$ -null set.

This proves condition (3). It remains to show that, under condition (4), one also obtains  $X_n \xrightarrow{\text{a.s.}} X_0$  as  $n \rightarrow \infty$ . Since  $S$  is separable (it is in fact a Borel subset of a Polish space), there is a countable convergence determining class  $F \subset C_b(S)$  for  $\mathcal{P}$ . Let

$$H = \{x \in S : \gamma_0(x)(f) = \lim_n \gamma_n(x)(f) \text{ for each } f \in F\}.$$

Then,  $\gamma_n(x) \rightarrow \gamma_0(x)$  weakly for each  $x \in H$ . Moreover, since  $F$  is countable, condition (4) implies  $\mu_0(H) = 1$ . Therefore, Theorem 5 yields

$$\begin{aligned} P(X_n \rightarrow X_0) &= \int_0^1 m \left\{ \beta : \Phi(\beta, \gamma_n[\phi(\alpha)]) \rightarrow \Phi(\beta, \gamma_0[\phi(\alpha)]) \right\} d\alpha \\ &= \int_H m \left\{ \beta : \Phi(\beta, \gamma_n(x)) \rightarrow \Phi(\beta, \gamma_0(x)) \right\} \mu_0(dx) \\ &= \int_H 1 d\mu_0 = 1. \end{aligned}$$

This concludes the proof when  $S$  is a Borel subset of a Polish space.

**Step 2** Suppose now that  $S$  is an arbitrary metric space. For each  $n \geq 0$ , since  $\mu_n$  is tight, there is a  $\sigma$ -compact set  $B_n \subset S$  such that  $\mu_n(B_n) = 1$ . Let  $B = \cup_n B_n$  and let  $\mathcal{X}$  be the completion of  $B$ . Since  $B$  is  $\sigma$ -compact,  $\mathcal{X}$  is a Polish space and  $B \in \mathcal{B}(\mathcal{X})$  (in fact,  $B$  is still  $\sigma$ -compact as a subset of  $\mathcal{X}$ ). Since  $\mu_n(B) = 1$  for each  $n \geq 0$ , the  $\mu_n$  can be regarded as probability measures on  $\mathcal{B}(B)$ . Furthermore, as shown below, condition (4) implies

$$E_{\mu_n}(f \mid B \cap \mathcal{G}) \xrightarrow{\mu_0\text{-a.s.}} E_{\mu_0}(f \mid B \cap \mathcal{G}) \quad \text{for each } f \in C_b(B). \tag{10}$$

Hence, to conclude the proof, it suffices to replace  $S$  with  $B$ ,  $\mathcal{G}$  with  $B \cap \mathcal{G}$ , and to apply what already proved in Step 1.

We finally prove that (4)  $\Rightarrow$  (10). For each  $n \geq 0$ , as  $\mu_n(B) = 1$  and  $B \in \mathcal{B}(\mathcal{X})$ , there is a r.c.d. for  $\mu_n$  given  $\sigma(\mathcal{G} \cup \{B\})$ , say  $\rho_n = \{\rho_n(x) : x \in S\}$ , such that

$$\rho_n(x)(B) = 1 \quad \text{for all } x \in S. \tag{11}$$

By Lemma 7, there is a countable convergence determining class  $F \subset C_b(S)$  for  $\mathcal{Q} = \{\mu \in \mathcal{P} : \mu(B) = 1\}$ . Moreover, condition (4) implies

$$\rho_n(\cdot)(f) = E_{\mu_n}(f \mid g) \xrightarrow{\mu_0\text{-a.s.}} E_{\mu_0}(f \mid g) = \rho_0(\cdot)(f) \quad \text{for all } f \in F.$$

Hence, because of (11) and  $F$  is countable, there is a set  $H \in \mathcal{G}$  such that

$$\mu_0(H) = 1 \quad \text{and} \quad \rho_n(x) \rightarrow \rho_0(x) \text{ weakly for all } x \in H.$$

On the other hand, thanks to (11),  $\rho_n(x) \rightarrow \rho_0(x)$  weakly if and only if

$$\rho_n(x)(\cdot \cap B) \rightarrow \rho_0(x)(\cdot \cap B) \text{ weakly with respect to the relative topology on } B.$$

Therefore,

$$\rho_0(x)(f \mathbb{1}_B) = \lim_n \rho_n(x)(f \mathbb{1}_B) \quad \text{for all } x \in H \text{ and } f \in C_b(B).$$

Since  $\mu_0(H) = 1$  and

$$E_{\mu_n}(f \mid B \cap \mathcal{G}) = \rho_n(\cdot)(f 1_B), \quad \mu_0\text{-a.s.},$$

this proves condition (10) and concludes the proof of the theorem.

### 4.2 Proof of Remark (vii)

Suppose  $S$  is a Borel subset of a Polish space and define the  $X_n$  as in Step 1 of the proof of Theorem 1. Since condition (3) holds, it suffices to prove  $X_n \rightarrow X_0$  in probability. Equivalently, we have to show that, for each subsequence  $n_j$ , there is a sub-subsequence  $n_{j_k}$  such that  $X_{n_{j_k}} \xrightarrow{\text{a.s.}} X_0$  as  $k \rightarrow \infty$ . Fix a subsequence  $n_j$  and a countable convergence determining class  $F$  for  $\mathcal{P}$ . Since  $F \subset C_b(S)$ , by a diagonalizing argument, there is a sub-subsequence  $n_{j_k}$  satisfying

$$E_{\mu_{n_{j_k}}}(f \mid g) \xrightarrow{\mu_0\text{-a.s.}} E_{\mu_0}(f \mid g), \quad \text{as } k \rightarrow \infty, \text{ for each } f \in F.$$

Since  $F$  is a convergence determining class for  $\mathcal{P}$ , one obtains

$$\gamma_{n_{j_k}}(x) \xrightarrow{\text{weakly}} \gamma_0(x), \quad \text{as } k \rightarrow \infty, \text{ for } \mu_0\text{-almost all } x \in S,$$

where  $\gamma_n$  is a r.c.d. for  $\mu_n$  given  $\mathcal{G}$ . In turn, this implies

$$E_{\mu_{n_{j_k}}}(f \mid g) \xrightarrow{\mu_0\text{-a.s.}} E_{\mu_0}(f \mid g), \quad \text{as } k \rightarrow \infty, \text{ for each } f \in C_b(S).$$

Hence,  $X_{n_{j_k}} \xrightarrow{\text{a.s.}} X_0$ . Finally, the general case (where  $S$  is arbitrary but the  $\mu_n$  are tight) can be handled as in Step 2 of the proof of Theorem 1.

### 4.3 Proof of Corollary 2

Let  $\nu(\cdot) = \mu_0(\cdot \times S_2)$  denote the marginal of  $\mu_0$  on  $S_1$ . In view of Theorem 1, we only have to prove that (6)  $\Rightarrow$  (4).

Let  $S = S_1 \times S_2$  and  $g(x, y) = x$  for all  $(x, y) \in S$ . Suppose condition (6) holds. For any  $f_1 : S_1 \rightarrow \mathbb{R}$  and  $f_2 : S_2 \rightarrow \mathbb{R}$ , define a function  $f_1 \times f_2$  on  $S$  as

$$(f_1 \times f_2)(x, y) = f_1(x) f_2(y) \quad \text{for all } (x, y) \in S.$$

Also, as in the proof of Theorem 1, take a  $\sigma$ -compact set  $B \subset S$  satisfying  $\mu_n(B) = 1$  for each  $n \geq 0$ . Then, by Lemma 7 and separability of  $S_1$ , there are two countable collections  $F_1 \subset C_b(S_1)$  and  $F_2 \subset C_b(S_2)$  such that

$$F = \{f_1 \times f_2 : f_1 \in F_1, f_2 \in F_2\}$$

is a convergence determining class for  $\mathcal{Q} = \{\mu \in \mathcal{P} : \mu(B) = 1\}$ .

Having noted this fact, fix a r.c.d.  $\rho_n = \{\rho_n(x, y) : (x, y) \in S\}$  for  $\mu_n$  given  $\sigma(\mathcal{G} \cup \{B\})$  such that  $\rho_n(x, y)(B) = 1$  for all  $(x, y) \in S$ . Since  $g(x, y) = x$  and  $\mu_n(B) = 1$ , we can write  $\rho_n(x)$  instead of  $\rho_n(x, y)$ . Then, for each  $f = f_1 \times f_2 \in F$ , condition (6) implies

$$\rho_n(x)(f) = f_1(x) \rho_n(x)(f_2) \longrightarrow f_1(x) \rho_0(x)(f_2) = \rho_0(x)(f)$$

for  $\nu$ -almost all  $x \in S_1$ . Since  $F$  is countable, it follows that

$$\nu\{x \in S_1 : \rho_n(x) \rightarrow \rho_0(x) \text{ weakly}\} = 1,$$

which in turn implies condition (4). This concludes the proof.

### 4.4 Proof of Corollary 3

Let

$$\hat{\mathcal{G}} = \phi^{-1}(C) \quad \text{and} \quad \hat{\mu}_n(D) = \mu_n(S \cap D) \quad \text{for all } n \geq 0 \text{ and } D \in \mathcal{B}(\mathcal{X}).$$

The  $\hat{\mu}_n$  are probability measures on  $\mathcal{B}(\mathcal{X})$  and

$$\hat{\mu}_n(\phi^{-1}(C)) = \mu_n(S \cap \phi^{-1}(C)) = \mu_n(g^{-1}(C)) = \mu_0(g^{-1}(C)) = \hat{\mu}_0(\phi^{-1}(C))$$

for all  $n \geq 0$  and  $C \in \mathcal{C}$ . Thus,  $\hat{\mu}_n = \hat{\mu}_0$  on  $\hat{\mathcal{G}}$  for each  $n \geq 0$ .

Let  $L = (0, 1)^2$  and let  $\mathcal{L} = m^2$  be the Lebesgue measure on  $\mathcal{B}(L)$ . Because of Theorem 1 (and its proof), on the probability space  $(L, \mathcal{B}(L), \mathcal{L})$ , there are  $\mathcal{X}$ -valued random variables  $\hat{X}_n$  such that

$$\hat{X}_n \sim \hat{\mu}_n \quad \text{and} \quad \phi(\hat{X}_n) = \phi(\hat{X}_0) \quad \text{for all } n \geq 0.$$

Furthermore,  $\hat{X}_n \xrightarrow{\text{a.s.}} \hat{X}_0$  provided

$$E_{\hat{\mu}_n}(f | \hat{\mathcal{G}}) \xrightarrow{\hat{\mu}_0\text{-a.s.}} E_{\hat{\mu}_0}(f | \hat{\mathcal{G}}) \quad \text{for all } f \in C_b(\mathcal{X}). \tag{12}$$

Let  $\mathcal{L}^*$  denote the  $\mathcal{L}$ -outer measure and

$$\Omega = \{\hat{X}_n \in S \text{ for each } n \geq 0\} \quad \text{and} \quad \mathcal{A} = \mathcal{B}(\Omega) = \Omega \cap \mathcal{B}(L).$$

Suppose  $\mathcal{L}^*(\Omega) = 1$ . Under this assumption, one can define

$$P(B \cap \Omega) = \mathcal{L}(B) \quad \text{for all } B \in \mathcal{B}(L).$$



Such a  $P$  is a probability measure on  $\mathcal{A}$  and condition (3) is satisfied by the  $S$ -valued random variables

$$X_n(\omega) = \hat{X}_n(\omega) \quad \text{for all } n \geq 0 \text{ and } \omega \in \Omega.$$

We next prove  $\mathcal{L}^*(\Omega) = 1$ . Since  $\mu_n$  is tight for  $n \geq 1$ , there is a  $\sigma$ -compact subset  $B \subset S$  such that  $\mu_n(B) = 1$  for each  $n \geq 1$ . On noting that  $B \in \mathcal{B}(\mathcal{X})$  (in fact,  $B$  is a  $\sigma$ -compact subset of  $\mathcal{X}$ ), one obtains

$$\{\hat{X}_n \in B\} \in \mathcal{B}(L) \quad \text{and} \quad \mathcal{L}(\hat{X}_n \in B) = \hat{\mu}_n(B) = \mu_n(B) = 1 \text{ for each } n \geq 1.$$

It follows that

$$\mathcal{L}(\hat{X}_n \in B \text{ for each } n \geq 1) = 1.$$

Moreover, since  $\mathcal{L}$  is tight,

$$\mathcal{L}^*(\hat{X}_0 \in S) = \hat{\mu}_0^*(S) = 1$$

where  $\hat{\mu}_0^*$  is the  $\hat{\mu}_0$ -outer measure; see the proof of [3, Th. 3.1] and [13, Th. 3.4.1]. Thus,  $B \subset S$  implies

$$\mathcal{L}^*(\Omega) \geq \mathcal{L}^*(\hat{X}_0 \in S \text{ and } \hat{X}_n \in B \text{ for each } n \geq 1) = \mathcal{L}^*(\hat{X}_0 \in S) = 1.$$

Hence,  $\mathcal{L}^*(\Omega) = 1$  and this proves condition (3).

Finally, it is not hard to show that condition (4) implies condition (12) (we omit the explicit calculations). Therefore, under (4), one obtains  $\hat{X}_n \xrightarrow{\text{a.s.}} \hat{X}_0$ , which in turn implies  $X_n \xrightarrow{\text{a.s.}} X_0$ . This concludes the proof.

#### 4.5 Proof of Corollary 4

Thanks to the assumptions on  $(T, \mathcal{C})$  and  $(\mu_n : n \geq 0)$ , on the probability space  $((0, 1)^2, \mathcal{B}((0, 1)^2), m^2)$ , there are random variables  $V_n$  such that  $V_n \xrightarrow{\text{a.s.}} V_0$ ,  $V_n \sim \mu_n$  and  $g(V_n) = g(V_0)$  for all  $n \geq 0$ ; see Theorem 1 and its proof. Moreover, being a nonatomic probability space,  $(\Omega, \mathcal{A}, P)$  supports a random variable  $T$  with distribution  $m^2$ , namely,  $T \sim m^2$  for some measurable map  $T : \Omega \rightarrow (0, 1)^2$ ; see, e.g., [3, Th. 3.1]. Therefore, it suffices to let  $X_n = V_n \circ T$  for all  $n \geq 0$ .

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Beiglbock, M., Lacker, D.: Denseness of adapted processes among causal couplings (2020). [arXiv:1805.03185v3](https://arxiv.org/abs/1805.03185v3)
2. Berti, P., Pratelli, L., Rigo, P.: Limit theorems for a class of identically distributed random variables. *Ann. Probab.* **32**, 2029–2052 (2004)
3. Berti, P., Pratelli, L., Rigo, P.: Skorohod representation on a given probability space. *Prob. Theory Relat. Fields* **137**, 277–288 (2007)
4. Berti, P., Pratelli, L., Rigo, P.: A Skorohod representation theorem for uniform distance. *Prob. Theory Relat. Fields* **150**, 321–335 (2011)
5. Berti, P., Pratelli, L., Rigo, P.: A Skorohod representation theorem without separability. *Electron. Commun. Probab.* **18**, 1–12 (2013)
6. Berti, P., Dreassi, E., Pratelli, L., Rigo, P.: A class of models for Bayesian predictive inference. *Bernoulli* **27**, 702–726 (2021)
7. Blackwell, D.: On a class of probability spaces. University of California Press, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, pp. 1–6 (1955)
8. Blackwell, D., Dubins, L.E.: An extension of Skorohod's almost sure representation theorem. *Proc. Am. Math. Soc.* **89**, 691–692 (1983)
9. Chau, H.N., Rasonyi, M.: Skorohod's representation theorem and optimal strategies for markets with frictions. *SIAM J. Control Optim.* **55**, 3592–3608 (2017)
10. Cortissoz, J.: On the Skorohod representation theorem. *Proc. Am. Math. Soc.* **135**, 3995–4007 (2007)
11. Crimaldi, I., Pratelli, L.: Two inequalities for conditional expectations and convergence results for filters. *Stat. Prob. Lett.* **74**, 151–162 (2005)
12. Dudley, R.M.: Distances of probability measures and random variables. *Ann. Math. Stat.* **39**, 1563–1572 (1968)
13. Dudley, R.M.: Uniform Central Limit Theorems. Cambridge University Press, Cambridge (1999)
14. Dumav, M., Stinchombe, M.B.: Skorohod's representation theorem for sets of probabilities. *Proc. Am. Math. Soc.* **144**, 3123–3133 (2016)
15. Fernique, X.: Un modele presque sur pour la convergence en loi. *C.R. Acad. Sci. Paris Ser. I* **306**, 335–338 (1988)
16. Hernandez-Ceron, N: Extensions of Skorohod's almost sure representation theorem Master of Science Thesis, University of Alberta (2010). <https://era.library.ualberta.ca/items/78ddb981-ca1c-4945-a6e9-8298217d6be6>
17. Jacod, J., Memin, J.: Sur un type de convergence intermediaire entre la convergence en loi et la convergence en probabilite. In: *Seminaire de Probabilites de Strasbourg XV*, vol. 850, Lecture Notes in Mathematics, pp. 529–546. Springer, Berlin (1981)
18. Jakubowski, A.: The almost sure Skorohod representation for subsequences in nonmetric spaces. *Theory Probab. Appl.* **42**, 167–174 (1998)
19. Ramachandran, D., Ruschendorf, L.: A general duality theorem for marginal problems. *Prob. Theory Relat. Fields* **101**, 311–319 (1995)
20. Sethuraman, J.: Some extensions of the Skorohod representation theorem. *Sankhya* **64**, 884–893 (2002)
21. Skorohod, A.V.: Limit theorems for stochastic processes. *Theory Probab. Appl.* **1**, 261–290 (1956)
22. van der Vaart, A., Wellner, J.A.: Weak Convergence and Empirical Processes. Springer, New York (1996)
23. Wichura, M.J.: On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *Ann. Math. Stat.* **41**, 284–291 (1970)