Probab. Theory Relat. Fields (2007) 137:277–288
DOI 10.1007/s00440-006-0018-1

Skorohod representation on a given probability space

Patrizia Berti · Luca Pratelli · Pietro Rigo

Received: 6 February 2006 / Revised: 27 May 2006 /
Published online: 26 July 2006

Abstract Let $(\Omega, \mathcal{A}, P)$ be a probability space, S a metric space, μ a probability measure on the Borel σ-field of S, and $X_n : \Omega \to S$ an arbitrary map, $n = 1, 2, \ldots$. If μ is tight and X_n converges in distribution to μ (in Hoffmann–Jørgensen's sense), then $X \sim \mu$ for some S-valued random variable X on $(\Omega, \mathcal{A}, P)$. If, in addition, the X_n are measurable and tight, there are S-valued random variables $\tilde{X}_n$ and X, defined on $(\Omega, \mathcal{A}, P)$, such that $\tilde{X}_n \sim X_n$, $X \sim \mu$, and $\tilde{X}_{n_k} \to X$ a.s. for some subsequence (n_k). Further, $\tilde{X}_n \to X$ a.s. (without need of taking subsequences) if $\mu\{x\} = 0$ for all x, or if $P(X_n = x) = 0$ for some n and all x. When P is perfect, the tightness assumption can be weakened into separability up to extending P to $\sigma(\mathcal{A} \cup \{H\})$ for some $H \subset \Omega$ with $P^*(H) = 1$. As a consequence, in applying Skorohod representation theorem with separable probability measures, the Skorohod space can be taken $((0,1), \sigma(\mathcal{U} \cup \{H\}), m_H)$, for some $H \subset (0,1)$ with outer Lebesgue measure 1, where $\mathcal{U}$ is the Borel σ-field on $(0,1)$ and m_H the only extension of Lebesgue measure such that $m_H(H) = 1$.

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

L. Pratelli
Accademia Navale, viale Italia 72, 57100 Livorno, Italy
e-mail: pratel@mail.dm.unipi.it

P. Rigo (✉)
Dipartimento di Economia Politica e Metodi Quantitativi, Universita' di Pavia,
via S. Felice 5, 27100 Pavia, Italy
e-mail: prigo@eco.unipv.it

In order to prove the previous results, it is also shown that, if X_n converges in distribution to a separable limit, then X_{n_k} converges stably for some subsequence (n_k).

Keywords Empirical process · Non measurable random element · Skorohod representation theorem · Stable convergence · Weak convergence of probability measures

Mathematics Subject Classification (2000) 60B10 · 60A05 · 60A10

1 Introduction

Let S be a metric space, μ a probability measure on the Borel subsets of S, and X_n an S-valued random variable on some probability space $(\Omega_n, \mathcal{A}_n, P_n)$, $n = 1, 2, \ldots$. According to Skorohod representation theorem and its subsequent generalizations by Dudley and Wichura, if μ is separable and $P_n \circ X_n^{-1} \to \mu$ weakly then, on a *suitable probability space*, there are S-valued random variables Z_n and Z such that $Z_n \sim X_n$, $Z \sim \mu$, and $Z_n \to Z$ a.s.. See Theorem 3.5.1 of [4] and Theorem 1.10.4 of [8]; see also p. 77 of [8] for historical notes. Let us call *Skorohod space* the probability space where Z_n and Z are defined.

In a number of real problems, the X_n are all defined on the same probability space, that is,

$$(\Omega_n, \mathcal{A}_n, P_n) = (\Omega, \mathcal{A}, P) \quad \text{for all } n.$$

In this case, provided $P \circ X_n^{-1} \to \mu$ weakly, a first question is:

(a) Is there an S-valued random variable X, defined on $(\Omega, \mathcal{A}, P)$, such that $X \sim \mu$?

One more question is:

(b) Is it possible to take $(\Omega, \mathcal{A}, P)$ as the Skorohod space ? In other terms, are there S-valued random variables $\widetilde{X}_n$ and X, defined on $(\Omega, \mathcal{A}, P)$, such that $\widetilde{X}_n \sim X_n$, $X \sim \mu$, and $\widetilde{X}_n \to X$ a.s. ?

Answering questions (a) and (b), the main purpose of this paper, can be useful at least from the foundational point of view.

As to (a), unlike Skorohod theorem, separability of μ is not enough for X to exist. However, a sufficient condition for X to exist is that μ is tight. Under this assumption, moreover, the $X_n : \Omega \to S$ can be taken to be arbitrary functions (not necessarily measurable) converging in distribution to μ in Hoffmann–Jørgensen's sense. Thus, for example, the result applies to convergence in distribution of empirical processes under uniform distance. See Corollary 5.4 and Examples 5.1 and 5.6.

As to (b), in addition to μ tight, suppose the X_n are (measurable and) tight. This happens, in particular, whenever S is Polish (and the X_n measurable). In spite of these assumptions, (b) can have a negative answer all the same. However, there are S-valued random variables $\widetilde{X}_n$ and X on $(\Omega, \mathcal{A}, P)$, with the given marginal distributions, such that $\widetilde{X}_{n_k} \to X$ a.s. for some subsequence (n_k). Furthermore, $\widetilde{X}_n \to X$ a.s. (without need of taking subsequences) in case $\mu\{x\} = 0$ for all $x \in S$, or in case $P(X_n = x) = 0$ for some $n \geq 1$ and all $x \in S$. See Examples 5.2 and 5.7, Theorem 5.3, and Corollary 5.5.

So far, one basic assumption is tightness. If P is perfect, tightness can be weakened into separability. In this case, however, $\widetilde{X}_n$ and X are to be defined on the enlarged probability space

$$(\Omega, \sigma(\mathcal{A} \cup \{H\}), P_H),$$

where $H \subset \Omega$ is a suitable subset with $P^*(H) = 1$ and P_H is the only extension of P to $\sigma(\mathcal{A} \cup \{H\})$ such that $P_H(H) = 1$.

The latter fact has, among others, the following consequence; cf. Theorem 3.2. Let m be Lebesgue measure on the Borel σ-field $\mathcal{U}$ on $(0,1)$. Suppose $\mu_n \to \mu$ weakly, where μ and μ_n are separable probabilities on the Borel subsets of S. Then, the corresponding Skorohod space can be taken to be $((0,1), \sigma(\mathcal{U} \cup \{H\}), m_H)$, for some $H \subset (0,1)$ with $m^*(H) = 1$, where m_H is the only extension of m such that $m_H(H) = 1$. Roughly speaking, provided all probabilities are separable, the Skorohod space can be obtained by just extending m to one more set, without need of taking some involved product space.

As a main tool for proving the previous results, we also get a proposition, of possible independent interest, on *stable convergence*. If X_n converges in distribution to a separable limit (the X_n being possibly non measurable), then X_{n_k} converges stably for some subsequence (n_k); see Theorem 4.1.

This paper is organized as follows. Section 2 includes notation and Sect. 3 provides answers to questions (a) and (b) in case P is nonatomic. The nonatomicity condition is removed in Sect. 5, after dealing with stable convergence in Sect. 4.

2 Notation

Throughout, S is a metric space, $\mathcal{B}$ the Borel σ-field on S, μ a probability on $\mathcal{B}$, $(\Omega, \mathcal{A}, P)$ a probability space, and $X_n : \Omega \to S$ an *arbitrary function*, $n = 1, 2, \ldots$. We let d denote the distance on S. A probability ν on $\mathcal{B}$ is separable in case $\nu(S_0) = 1$ for some separable set $S_0 \in \mathcal{B}$. In particular, ν is separable whenever it is tight. A map $Z : \Omega \to S$ is called measurable, or a random variable, in case $Z^{-1}(\mathcal{B}) \subset \mathcal{A}$. If Z is measurable, we write $Z \sim \nu$ to mean that $\nu = P \circ Z^{-1}$ and Z is said to be separable or tight in case $P \circ Z^{-1}$ is separable or tight. Similarly, $Z \sim Z'$ means that Z and Z' are identically distributed. Moreover, $\mathcal{U}$ is the Borel σ-field on $(0,1)$ and m the Lebesgue measure on $\mathcal{U}$.

A set $A \in \mathcal{A}$ is a *P-atom* in case $P(A) > 0$ and $P(A \cap H) \in \{0, P(A)\}$ for all $H \in \mathcal{A}$, and P is said to be *nonatomic* in case there are not P-atoms. If P is not nonatomic, there are countably many pairwise disjoint P-atoms, $A_1, A_2, \ldots$, such that either $\sum_{j \geq 1} P(A_j) = 1$ or $P\big(\cdot \mid (\cup_{j \geq 1} A_j)^c\big)$ is nonatomic.

The probability P is *perfect* in case, for each measurable $f : \Omega \to \mathbb{R}$, there is a real Borel set $B \subset f(\Omega)$ such that $P(f \in B) = 1$. For instance, P is perfect if Ω is a universally measurable subset of a Polish space and $\mathcal{A}$ the Borel σ-field on Ω.

Given any probability space $(\mathcal{X}, \mathcal{F}, Q)$, we let Q^* and Q_* denote outer and inner probabilities, i.e., for all $H \subset \mathcal{X}$ we let

$$Q^*(H) = \inf\{Q(A) : A \in \mathcal{F}, A \supset H\}, \quad Q_*(H) = 1 - Q^*(H^c).$$

If $Q^*(H) = 1$, Q admits an unique extension Q_H to $\sigma(\mathcal{F} \cup \{H\})$ such that $Q_H(H) = 1$, that is, $Q_H\Big((A_1 \cap H) \cup (A_2 \cap H^c)\Big) = Q(A_1)$ for all $A_1, A_2 \in \mathcal{F}$.

Finally, if Z_n and Z are S-valued maps on some probability space $(\mathcal{X}, \mathcal{F}, Q)$, $Z_n \to Z$ *almost surely* (a.s.) means that $Q_*\big(Z_n \to Z\big) = 1$. If the Z_n are measurable and Z is measurable and separable, this is equivalent to $Z_n \to Z$ *almost uniformly*, i.e., for each $\epsilon > 0$ there is $A \in \mathcal{F}$ with $Q(A^c) < \epsilon$ and $Z_n \to Z$ uniformly on A; see Lemma 1.9.2 and Theorem 1.9.6 of [8].

3 Existence of random variables with given distribution on a nonatomic probability space

We start by giving conditions for $(\Omega, \mathcal{A}, P)$ to support a random variable with given distribution ν, where ν is a (separable) probability on $\mathcal{B}$. To this end, if ν is not tight, nonatomicity and perfectness of P are not enough; see Example 5.1. However, a random variable with distribution ν is available up to extending P to one more subset of Ω. In the sequel, *given $H \subset \Omega$ with $P^*(H) = 1$, P_H denotes the only extension of P to $\sigma(\mathcal{A} \cup \{H\})$ such that $P_H(H) = 1$.* We also recall that $((0,1), \mathcal{U}, m)$ supports a random variable with distribution ν provided S is Polish.

Theorem 3.1 *Let P be nonatomic and ν a separable probability on $\mathcal{B}$. Then:*

(i) *If ν is tight, $X \sim \nu$ for some S-valued random variable X on $(\Omega, \mathcal{A}, P)$.*
(ii) *If P is perfect, there are $H \subset \Omega$ with $P^*(H) = 1$ and $X : \Omega \to S$ such that*

$$X^{-1}(\mathcal{B}) \subset \sigma(\mathcal{A} \cup \{H\}) \quad \textit{and} \quad X \sim \nu \textit{ under } P_H.$$

Proof Since P is nonatomic, there is a measurable map $U : \Omega \to (0,1)$ such that $U \sim m$; see, e.g., the proof of Lemma 2 of [2]. Take a separable set $S_0 \in \mathcal{B}$ with $\nu(S_0) = 1$, and fix a countable subset $\{x_1, x_2, \ldots\} \subset S_0$, dense in S_0. Define

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

Letting $C = [0,1]^\infty$ be the Hilbert cube, $h : S \to C$ is continuous and it is an homeomorphism as a map $h : S_0 \to h(S_0)$. Since C is Polish and $\nu \circ h^{-1}$ is a probability on the Borel subsets of C, there is a C-valued random variable Z on $((0,1),\mathcal{U},m)$ such that $Z \sim \nu \circ h^{-1}$. Fix $x_0 \in S$ and define

$$H = \{Z \circ U \in h(S_0)\}, \quad X = h^{-1}(Z \circ U) \text{ on } H, \quad X = x_0 \text{ on } H^c.$$

Given $B \in \mathcal{B}$, since $h : S_0 \to h(S_0)$ is an homeomorphism, $h(B \cap S_0) = h(S_0) \cap D$ for some Borel set $D \subset C$. Hence,

$$\{X \in B\} \cap H = \{Z \circ U \in h(B \cap S_0)\} = \{Z \circ U \in D\} \cap H \in \sigma(\mathcal{A} \cup \{H\}).$$

If ν is tight, S_0 can be taken σ-compact, and thus $h(S_0)$ is Borel in C (it is in fact σ-compact). It follows that $H \in \mathcal{A}$ and $X^{-1}(\mathcal{B}) \subset \mathcal{A}$. On noting that $P(H) = \nu \circ h^{-1}(h(S_0)) = \nu(S_0) = 1$, one easily obtains $X \sim \nu$.

If P is perfect, Lemma 1 of [2] (see also Theorem 3.4.1 of [4]) implies

$$P^*(H) = (\nu \circ h^{-1})^*(h(S_0)) \geq \nu(S_0) = 1.$$

If $B \in \mathcal{B}$ and $h(B \cap S_0) = h(S_0) \cap D$ for some Borel set $D \subset C$, then

$$\begin{aligned} P_H(X \in B) &= P_H(Z \circ U \in D) = P(Z \circ U \in D) \\ &= \nu \circ h^{-1}(D) \geq \nu(B \cap S_0) = \nu(B). \end{aligned}$$

Taking complements yields $P_H \circ X^{-1} = \nu$ and concludes the proof. □

Our next result is a consequence of Theorem 3.1. Let μ_n be probabilities on $\mathcal{B}$ such that $\mu_n \to \mu$ weakly, where μ is separable. Then, Skorohod theorem applies, and a question is whether $((0,1),\mathcal{U},m)$ can be taken as Skorohod space. As shown in [6], this is possible in case μ and the μ_n are tight. *Up to extending m to one more subset of* $(0,1)$, this is still possible in case the μ_n are only separable. Indeed, it suffices to let $(\Omega,\mathcal{A},P) = ((0,1),\mathcal{U},m)$ in the following Theorem 3.2.

Theorem 3.2 *Suppose P is nonatomic, μ and each μ_n are separable probabilities on $\mathcal{B}$, and $\mu_n \to \mu$ weakly. Then:*

(i) *If μ and each μ_n are tight, there are S-valued random variables $\widetilde{X}_n$ and X on $(\Omega,\mathcal{A},P)$ such that $\widetilde{X}_n \sim \mu_n$, $X \sim \mu$, and $\widetilde{X}_n \to X$ a.s.*

(ii) *If P is perfect, there are $H \subset \Omega$ with $P^*(H) = 1$ and S-valued random variables $\widetilde{X}_n$ and X on $(\Omega,\sigma(\mathcal{A} \cup \{H\}),P_H)$ satisfying $\widetilde{X}_n \sim \mu_n$, $X \sim \mu$, and $\widetilde{X}_n \to X$ a.s.*

Proof By Skorohod theorem, on some probability space $(\mathcal{X},\mathcal{F},Q)$, there are S-valued random variables Z_n and Z such that $Z_n \sim \mu_n$, $Z \sim \mu$, and $Z_n \to Z$ a.s. Let

$$\gamma(B) = Q\big((Z,Z_1,Z_2,\ldots) \in B\big) \quad \text{for all } B \in \mathcal{B}^\infty.$$

Then γ is separable, since its marginals $\mu, \mu_1, \mu_2, \ldots$ are separable, and thus γ can be extended to a separable probability ν on the Borel σ-field of S^∞. Moreover, ν is tight if and only if $\mu, \mu_1, \mu_2, \ldots$ are tight. Thus, Theorem 3.1 applies. Precisely, if μ and the μ_n are tight, Theorem 3.1 yields

$$Y = (X, \widetilde{X}_1, \widetilde{X}_2, \ldots) \sim \nu$$

for some S^∞-valued random variable Y on $(\Omega, \mathcal{A}, P)$. Otherwise, if P is perfect, $Y \sim \nu$ for some S^∞-valued random variable Y on $(\Omega, \sigma(\mathcal{A} \cup \{H\}), P_H)$, where $H \subset \Omega$ and $P^*(H) = 1$. □

In Theorem 3.2, unlike Skorohod theorem, the μ_n are asked to be separable. We recall that it is consistent with the usual axioms of set theory (i.e., with the ZFC set theory) that nonseparable probability measures on $\mathcal{B}$ do not exist; see [4], p. 403, and [8], p. 24.

To apply Theorems 3.1 and 3.2, conditions for nonatomicity of P are useful.

Lemma 3.3 *For P to be nonatomic, it is enough that $(\Omega, \mathcal{A}, P)$ supports a separable S-valued random variable X such that $P(X = x) = 0$ for all $x \in S$.*

Proof Suppose A is a P-atom and Z a separable S-valued random variable on $(\Omega, \mathcal{A}, P)$. Let $\nu(B) = P(Z \in B \mid A)$, $B \in \mathcal{B}$. Then, ν is separable and 0-1 valued, so that $\nu\{x\} = 1$ for some $x \in S$. Thus, $P(Z = x) \geq P(A, Z = x) = P(A) > 0$. □

Theorems 3.1 and 3.2 provide answers to questions (a) and (b), though under some assumptions on P. In Sect. 5, these assumptions are weakened or even dropped. To this end, we need to show that some subsequence X_{n_k} also converges in distribution under $P(\cdot \mid A)$, for each possible P-atom A. This naturally leads to *stable convergence*.

4 Stable convergence

Given a probability ν on $\mathcal{B}$, say that X_n *converges in distribution to* ν in case $E^*f(X_n) \to \int f \, d\nu$ for all bounded continuous functions $f : S \to \mathbb{R}$, where E^* denotes outer expectation; see [4,8]. Such a definition, due to Hoffmann–Jørgensen, reduces to the usual one if the X_n are measurable. Say also that X_n *converges stably* in case X_n converges in distribution under $P(\cdot \mid H)$ for each $H \in \mathcal{A}$ with $P(H) > 0$. Stable convergence has been introduced by Renyi in [7] and subsequently investigated by various authors (in case the X_n are measurable). We refer to [3,5] for more on stable convergence.

Theorem 4.1 *If μ is separable and X_n converges in distribution to μ, then X_{n_k} converges stably for some subsequence (n_k).*

Proof We first suppose that μ is tight and the X_n are measurable with separable range. As X_n converges in distribution to a tight limit, X_n is asymptotically tight; see Lemma 1.3.8 of [8]. Thus, X_n is also asymptotically tight under

$P(\cdot \mid H)$ whenever $H \in \mathcal{A}$ and $P(H) > 0$. Moreover, $\sigma(X_1, X_2, \ldots)$ is a countably generated sub-σ-field of $\mathcal{A}$, due to the X_n being measurable with separable range. Let $\mathcal{G}$ be a countable field such that $\sigma(\mathcal{G}) = \sigma(X_1, X_2, \ldots)$. Since $\mathcal{G}$ is countable, by Prohorov's theorem (cf. Theorem 1.3.9 of [8]) and a diagonalizing argument, there is a subsequence (n_k) such that

$$X_{n_k} \text{ converges in distribution, under } P(\cdot \mid G), \text{ for each } G \in \mathcal{G} \text{ with } P(G) > 0.$$

Next, fix $A \in \sigma(X_1, X_2, \ldots)$ with $P(A) > 0$. Given $\epsilon > 0$ and a bounded continuous function $f : S \to \mathbb{R}$, there is $G \in \mathcal{G}$ with $P(G) > 0$ and $2\sup|f| P(A \Delta G) < \epsilon P(A)$. Thus,

$$\begin{aligned} \limsup_{j,k} \Big| E(f(X_{n_j}) \mid A) - E(f(X_{n_k}) \mid A) \Big| &\leq \frac{2\sup|f| P(A \Delta G)}{P(A)} \\ &+ \frac{P(G)}{P(A)} \limsup_{j,k} \Big| E(f(X_{n_j}) \mid G) - E(f(X_{n_k}) \mid G) \Big| < \epsilon. \end{aligned}$$

Therefore, $E(f(X_{n_k}) \mid A)$ converges to a limit for each bounded continuous f. By Alexandrov's theorem, this implies that X_{n_k} converges in distribution under $P(\cdot \mid A)$. Next, let $H \in \mathcal{A}$ with $P(H) > 0$, and let V_H be a bounded version of $E(I_H \mid X_1, X_2, \ldots)$. Given a bounded continuous function f on S, $E\big(f(X_{n_k}) I_A\big)$ converges to a limit for each $A \in \sigma(X_1, X_2, \ldots)$. Since V_H is the uniform limit of some sequence of simple functions in $\sigma(X_1, X_2, \ldots)$, it follows that

$$E\big(f(X_{n_k}) I_H\big) = E\big(f(X_{n_k}) V_H\big)$$

also converges to a limit. Once again, Alexandrov's theorem implies that X_{n_k} converges in distribution under $P(\cdot \mid H)$. Thus, X_{n_k} converges stably.

Let us now consider the general case (μ separable and the X_n arbitrary functions). Since X_n converges in distribution to a separable limit, there are maps $Z_n : \Omega \to S$, all measurable with finite range, such that $P^*\big(d(X_n, Z_n) \geq \epsilon\big) \to 0$ for all $\epsilon > 0$; see Proposition 1.10.12 of [8] and its proof. Fix a separable set $S_1 \in \mathcal{B}$ with $\mu(S_1) = 1$ and let $S_0 = S_1 \cup (\cup_n Z_n(\Omega))$. As in the proof of Theorem 3.1, define $C = [0,1]^\infty$ and

$$h(x) = \big(d(x, x_1) \wedge 1, d(x, x_2) \wedge 1, \ldots\big), \quad x \in S,$$

where $\{x_1, x_2, \ldots\} \subset S_0$ is dense in S_0. Since Z_n converges in distribution to μ and $h : S \to C$ is continuous, $h(Z_n)$ converges in distribution to $\mu \circ h^{-1}$. Also, $\mu \circ h^{-1}$ is tight (due to C being Polish) and the $h(Z_n)$ are measurable with separable range. Thus, $h(Z_{n_k})$ converges stably for some subsequence (n_k). Since $d(X_n, Z_n) \to 0$ in outer probability, X_{n_k} converges stably if and only if Z_{n_k} converges stably. Hence, it suffices proving that Z_{n_k} converges stably.

Let $Y_k = h(Z_{n_k})$. For each $H \in \mathcal{A}$ with $P(H) > 0$, let γ_H denote the limit in distribution of Y_k under $P(\cdot \mid H)$. Then $\gamma_\Omega = \mu \circ h^{-1}$, since $h(Z_n)$ converges in distribution to $\mu \circ h^{-1}$ (under P), so that

$$\gamma_\Omega^*(h(S_0)) = (\mu \circ h^{-1})^*(h(S_0)) \geq \mu(S_0) = 1.$$

As $\gamma_\Omega = P(H)\gamma_H + P(H^c)\gamma_{H^c}$ whenever $0 < P(H) < 1$, one obtains $\gamma_H^*(h(S_0)) = 1$ for all $H \in \mathcal{A}$ with $P(H) > 0$. Fix one such H. Then, $Y_k : \Omega \to h(S_0) \subset C$ and, under $P(\cdot \mid H)$, Y_k converges in distribution to γ_H as a random element of C. Since $\gamma_H^*(h(S_0)) = 1$, Y_k also converges in distribution as a random element of $h(S_0)$. Since h is an homeomorphism as a map $h : S_0 \to h(S_0)$, it follows that $Z_{n_k} = h^{-1}(Y_k)$ converges in distribution under $P(\cdot \mid H)$. This concludes the proof. □

5 A Skorohod representation

In Sect. 3, under the assumption that P is nonatomic, questions (a) and (b) have been answered. Here, nonatomicity of P is dropped. Instead, as in Sect. 3, perfectness of P is retained in the separable case while it is superfluous in the tight case. Let us begin with counterexamples.

Example 5.1 (Question (a) can have a negative answer even if S is separable) Let P be nonatomic and perfect and let $S \subset (0, 1)$ be such that $m_*(S) = 0 < 1 = m^*(S)$. If equipped with the relative topology, S is a separable metric space and $\mathcal{B} = \{B \cap S : B \in \mathcal{U}\}$. Define $\mu(B \cap S) = m^*(B \cap S)$, $B \in \mathcal{U}$, and take discrete probabilities μ_n on $\mathcal{B}$ such that $\mu_n \to \mu$ weakly. For each n, since P is nonatomic and μ_n is tight (it is even discrete), Theorem 3.1 yields $X_n \sim \mu_n$ for some S-valued random variable X_n on $(\Omega, \mathcal{A}, P)$. Suppose now that $X \sim \mu$ for some measurable $X : \Omega \to S$. Since P is perfect and X is also a measurable map $X : \Omega \to \mathbb{R}$, there is $B \in \mathcal{U}$ such that $B \subset X(\Omega) \subset S$ and $\mu(B) = P(X \in B) = 1$. It follows that μ is tight, which is a contradiction since $\mu(K) = 0$ for each compact $K \subset S$. Thus, no S-valued random variable X on $(\Omega, \mathcal{A}, P)$ meets $X \sim \mu$.

Example 5.2 (Question (b) can have a negative answer even if S is Polish) Let $\Omega = (0, 1)$, $\mathcal{A} = \mathcal{U}$, $P((0, x)) = x$ for $0 < x < 1/6$, $P\{a\} = 1/2$, and $P\{b\} = 1/3$, where $\frac{1}{6} < a < b < 1$. Define $S = \mathbb{R}$ and

$$X_n(a) = 1, \quad X_n(b) = 2, \quad X_n(x) = \frac{4}{\pi} \arctan(nx) \text{ for } 0 < x < \frac{1}{6}, \text{ if } n \text{ is even,}$$
$$X_n(a) = 2, \quad X_n(b) = 1, \quad X_n(x) = \frac{2}{\pi} \arctan(nx) \text{ for } 0 < x < \frac{1}{6}, \text{ if } n \text{ is odd.}$$

Then, X_n converges in distribution to $\mu = (\delta_1 + \delta_2)/2$. If $\widetilde{X}_n$ is a real random variable on $(\Omega, \mathcal{A}, P)$ such that $\widetilde{X}_n \sim X_n$, then $\widetilde{X}_n(a) = 1$ if n is even and $\widetilde{X}_n(a) = 2$ if n is odd. Thus, $\widetilde{X}_n$ does not converge a.s. (or even in probability).

As suggested by Example 5.2, even if μ and the X_n are nice, question (b) can have a negative answer in case P has atoms. However, Example 5.2 also suggests that a.s. convergence of suitable subsequences can be obtained. Next result shows that this is actually true, independently of P having atoms or not.

Theorem 5.3 *Let μ be a probability measure on $\mathcal{B}$ and (X_n) a sequence of S-valued random variables on $(\Omega, \mathcal{A}, P)$. Suppose μ and the X_n are separable and X_n converges in distribution to μ. Then:*

(i) *If μ and the X_n are tight, there are S-valued random variables $\widetilde{X}_n$ and X on $(\Omega, \mathcal{A}, P)$ such that*

$$\widetilde{X}_n \sim X_n, \quad X \sim \mu, \quad \widetilde{X}_{n_k} \to X \text{ a.s. for some subsequence } (n_k). \quad (1)$$

(ii) *If P is perfect, there are $H \subset \Omega$ with $P^*(H) = 1$ and S-valued random variables $\widetilde{X}_n$ and X on $(\Omega, \sigma(\mathcal{A} \cup \{H\}), P_H)$ such that condition (1) holds.*

Proof By Theorem 3.2, P can be assumed to have atoms. Let $A_1, A_2, \dots$ be pairwise disjoint P-atoms such that either $P(A_0) = 0$ or $P(\cdot \mid A_0)$ is nonatomic, where $A_0 = (\cup_{j \ge 1} A_j)^c$. We assume $P(A_0) > 0$. (If $P(A_0) = 0$, the proof given below can be repeated by just neglecting A_0). By Theorem 4.1, there is a subsequence (n_k) such that X_{n_k} converges in distribution under $P(\cdot \mid A_j)$ for all $j \ge 0$. Fix $j > 0$ and let $\nu_{kj}(\cdot) = P(X_{n_k} \in \cdot \mid A_j)$. Then, ν_{kj} is 0–1 valued and separable (since A_j is a P-atom and X_{n_k} is separable). Hence, $\nu_{kj} = \delta_{x(k,j)}$ for some point $x(k,j) \in S$. Since ν_{kj} converges weakly (as $k \to \infty$), one also obtains $x(k,j) \to x(j)$ for some point $x(j) \in S$. Next, let μ_0 be the limit in distribution of X_{n_k} under $P(\cdot \mid A_0)$.

Suppose μ and the X_n are tight. Then, $P(\cdot \mid A_0)$ is nonatomic and μ_0 is tight (due to μ being tight). By Theorem 3.2, there are S-valued random variables V_{n_k} and V on $(\Omega, \mathcal{A}, P(\cdot \mid A_0))$ such that

$$V_{n_k} \sim X_{n_k}, \quad V \sim \mu_0, \quad V_{n_k} \to V \text{ a.s.,} \qquad \text{under } P(\cdot \mid A_0).$$

Thus, to get (1), it suffices to let $\widetilde{X}_n = X_n$ if $n \neq n_k$ for all k, and

$$X = V \text{ and } \widetilde{X}_{n_k} = V_{n_k} \text{ on } A_0, \quad X = x(j) \text{ and } \widetilde{X}_{n_k} = X_{n_k} \text{ on } A_j \text{ for } j > 0.$$

Finally, suppose P is perfect. Then, $P(\cdot \mid A_0)$ is nonatomic and perfect and μ_0 is separable (due to μ being separable). By Theorem 3.2, there are $M \subset \Omega$ with $P^*(M \mid A_0) = 1$ and S-valued random variables V_{n_k} and V on $(\Omega, \sigma(\mathcal{A} \cup \{M\}), Q)$ such that

$$V_{n_k} \sim X_{n_k}, \quad V \sim \mu_0, \quad V_{n_k} \to V \text{ a.s.,} \qquad \text{under } Q,$$

where Q is the only extension of $P(\cdot \mid A_0)$ satisfying $Q(M) = 1$. Thus, it suffices to let $H = (A_0 \cap M) \cup A_0^c$ and to define $\widetilde{X}_n$ and X as above. □

As a corollary, Theorem 5.3 implies that question (a) admits a positive answer whenever μ is tight. Next result is analogous to Theorem 3.1. Now, $(\Omega, \mathcal{A}, P)$ is not assumed nonatomic, but it supports a sequence of (arbitrary) functions which converges in distribution to μ.

Corollary 5.4 *Let $X_n : \Omega \to S$ be arbitrary maps. Suppose μ is separable and X_n converges in distribution to μ. Then:*

(i) *If μ is tight, $X \sim \mu$ for some S-valued random variable X on $(\Omega, \mathcal{A}, P)$.*
(ii) *If P is perfect, $X \sim \mu$ for some S-valued random variable X defined on $(\Omega, \sigma(\mathcal{A} \cup \{H\}), P_H)$ where $H \subset \Omega$ and $P^*(H) = 1$.*

Proof Just note that, as in the proof of Theorem 4.1, there are maps $Z_n : \Omega \to S$, all measurable with finite range, such that $d(X_n, Z_n) \to 0$ in outer probability. Thus, it suffices applying Theorem 5.3 with Z_n in the place of X_n. □

A particular case of Corollary 5.4 (S Polish and X_n measurable) is contained in Lemma 2 of [2].

Next, we give conditions for question (b) to have a positive answer.

Corollary 5.5 *In the notation and under the assumptions of Theorem 5.3, suppose also that $\mu\{x\} = 0$ for all $x \in S$, or that $P(X_n = x) = 0$ for some $n \geq 1$ and all $x \in S$. Then, in both (i) and (ii), one has $\tilde{X}_n \to X$ a.s.*

Proof By Theorem 3.2, it suffices proving that P is nonatomic. By Lemma 3.3, this is obvious if $P(X_n = x) = 0$ for some n and all x, and thus assume $\mu\{x\} = 0$ for all x. If μ is tight, P is nonatomic by Lemma 3.3 and Corollary 5.4. If P is perfect, Lemma 3.3 and Corollary 5.4 imply that P_H is nonatomic, and this in turn implies nonatomicity of P. □

Finally, we apply Corollaries 5.4 and 5.5 to empirical processes.

Example 5.6 (Empirical processes) Let (ξ_n) be an i.i.d. sequence of random variables, defined on $(\Omega, \mathcal{A}, P)$ and taking values in some measurable space $(\mathcal{X}, \mathcal{F})$, and let F be an uniformly bounded class of real measurable functions on $\mathcal{X}$. Define $S = l^\infty(F)$, the space of real bounded functions on F equipped with *uniform distance*, and

$$X_n(f) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} f(\xi_i) - Ef(\xi_1)\Big), \quad f \in F.$$

The (nonmeasurable) map $X_n : \Omega \to l^\infty(F)$ is called empirical process. To the best of our knowledge, all existing conditions for X_n to converge in distribution entail tightness of the limit law μ; see [4,8]; see also [1,9] for empirical processes based on nonindependent sequences of random variables or on diffusion processes. Under anyone of these conditions, by Corollary 5.4, $X_n \xrightarrow{d} X$ for some $l^\infty(F)$-valued random variable X on $(\Omega, \mathcal{A}, P)$. Indeed, relying on Theorem 4.1

and Corollary 5.4 together, a little bit more is true: *Under anyone of such conditions, there are a subsequence* (n_k) *and measurable maps* $X_H : \Omega \to l^\infty(F)$, *where* $H \in \mathcal{A}$ *and* $P(H) > 0$, *such that*

$$X_{n_k} \xrightarrow{d} X_H, \text{under } P(\cdot \mid H), \text{ for all } H \in \mathcal{A} \text{ with } P(H) > 0.$$

Example 5.7 (More on empirical processes) Sometimes, the X_n take values in a subset $D \subset l^\infty(F)$ admitting a Polish topology. If the X_n are also measurable and converge in distribution under such topology, something more can be said. To be concrete, suppose $\mathcal{X} = [0,1]$ and $F = \{I_{[0,t]} : 0 \le t \le 1\}$. Let D be the set of real cadlag functions on $[0,1]$, $\mathcal{D}$ the ball σ-field on D with respect to uniform distance, and

$$X_n(t) := X_n(I_{[0,t]}) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} I_{\{\xi_i \le t\}} - P(\xi_1 \le t)\Big), \quad t \in [0,1].$$

Then, $X_n : \Omega \to D$ and $X_n^{-1}(\mathcal{D}) \subset \mathcal{A}$. If D is equipped with Skorohod topology, the Borel σ-field on D is $\mathcal{D}$ and X_n converges in distribution to a probability measure μ on $\mathcal{D}$. Since D is Polish under Skorohod topology and $\mu\{x\} = 0$ for all $x \in D$ (unless ξ_1 has a degenerate distribution, in which case everything is trivial), Corollary 5.5 applies with $S = D$. Accordingly, there are measurable maps $\widetilde{X}_n : \Omega \to D$ and $X : \Omega \to D$ such that $\widetilde{X}_n \sim X_n$ and $\widetilde{X}_n \to X$ a.s. with respect to Skorohod topology. Further, convergence is actually uniform whenever $P(\xi_1 = t) = 0$ for all t, since in this case almost all paths of X are continuous. Finally, the assumption $\mathcal{X} = [0,1]$ can be generalized into $\mathcal{X} = \mathbb{R}$ provided D is taken to be the space of real cadlag functions on $\mathbb{R}$ with finite limits at $\pm\infty$; see [2], proof of Theorem 3. To sum up: *If the* ξ_n *are real i.i.d. random variables with a continuous distribution function, there are D-valued maps* $\widetilde{X}_n$ *and* X *on* $(\Omega, \mathcal{A}, P)$ *such that* $\widetilde{X}_n^{-1}(\mathcal{D}) \subset \mathcal{A}$, $X^{-1}(\mathcal{D}) \subset \mathcal{A}$, *and*

$$P(\widetilde{X}_n \in \cdot) = P(X_n \in \cdot) \text{ on } \mathcal{D}, \quad \sup_t \Big|\widetilde{X}_n(t) - X(t)\Big| \to 0 \text{ a.s.}$$

References

1. Berti, P., Pratelli, L., Rigo, P.: Limit theorems for a class of identically distributed random variables. Ann. Probab. **32**, 2029–2052 (2004)
2. Berti, P., Pratelli, L., Rigo, P.: Asymptotic behaviour of the empirical process for exchangeable data. Stoch. Proc. Appl. **116**, 337–344 (2006)
3. Crimaldi, I., Letta, G., Pratelli, L.: A strong form of stable convergence. Sem Probab (to appear) (2005)
4. Dudley, R.: Uniform central limit theorems. Cambridge University Press, Cambridge (1999)
5. Hall, P., Heyde, C.C.: Martingale limit theory and its applications. Academic, New York (1980)
6. Letta, G., Pratelli, L.: Le théorème de Skorohod pour des lois de Radon sur un espace métrisable. In: Rendiconti Accademia Nazionale delle Scienze detta dei XL, vol. 115, pp. 157–162 (1997)

7. Renyi, A.: On stable sequences of events. Sankhya A **25**, 293–302 (1963)
8. van der Vaart, A., Wellner, J.A.: Weak convergence and empirical processes. Springer, Berlin Heidelberg New York (1996)
9. van der Vaart, A., van Zanten, H.: Donsker theorems for diffusions: necessary and sufficient conditions. Ann. Probab. **33**, 1422–1451 (2005)