Check for
updates

Some Skorohod-type results

Luca Pratelli[1] · Pietro Rigo[2]

Abstract

Let S be a metric space, $g : S \to \mathbb{R}$ a Borel function, and $(\mu_n : n \geq 0)$ a sequence of tight probability measures on $\mathcal{B}(S)$. If $\mu_n = \mu_0$ on $\sigma(g)$, there are S-valued random variables X_n, all defined on the same probability space, such that $X_n \sim \mu_n$ and $g(X_n) = g(X_0)$ for all $n \geq 0$. Moreover, $X_n \xrightarrow{a.s.} X_0$ if and only if $E_{\mu_n}(f \mid g) \xrightarrow{\mu_0 - a.s.} E_{\mu_0}(f \mid g)$ for each $f \in C_b(S)$. This result, proved in Pratelli and Rigo (J Theoret Probab 36:372-389, 2023) , is the starting point of this paper. Three types of contributions are provided. First, $\sigma(g)$ is replaced by an arbitrary sub-σ-field $\mathcal{G} \subset \mathcal{B}(S)$. Second, the result is applied to some specific frameworks, including equivalence couplings, total variation distances, and the decomposition of cadlag processes with finhite activity. Third, following Hansen et al. (Tempered Bayesian analysis, Unpublished manuscript, 2024), the result is extended to models and kernels. This extension has a fairly natural interpretation in terms of decision theory, mass transportation and statistics.

Keywords Coupling · Duality · Kernel · Optimal transport · Skorohod representation theorem

Mathematics Subject Classification 60B10 · 60E05 · 49N15 · 49Q22

JEL Classification: C02 · C011 · C61

✉ Pietro Rigo
    pietro.rigo@unibo.it

    Luca Pratelli
    luca.pratelli@icloud.com

[1] Naval Academy of Livorno, Livorno, Italy

[2] Department of Statistics, University of Bologna, Bologna, Italy

 Springer

1 Introduction

Let S be a metric space and $(\mu_n : n \geq 0)$ a sequence of probability measures on $\mathcal{B}(S)$. (Throughout, for any topological space T, we let $\mathcal{B}(T)$ denote the Borel σ-field on T). We say that $(X_n : n \geq 0)$ is a *coupling* of (μ_n) if

- The X_n are S-valued random variables, all defined on the same probability space, such that $X_n \sim \mu_n$ for each $n \geq 0$.

The *Skorohod representation theorem* (SRT) states that, if $\mu_n \to \mu_0$ weakly and μ_0 has a separable support, there is a coupling (X_n) of (μ_n) such that $X_n \xrightarrow{a.s.} X_0$. This version of SRT is due to Wichura (Wichura 1970) who reworked the previous versions by Skorohod (Skorohod 1956) and Dudley (Dudley 1968). We refer to [Dudley (1999), p. 130] and [van der Vaart and Wellner (1996), p. 77] for historical notes, and to Berti et al. (2013) for the case where μ_0 does not have a separable support. Some other related references are (Berti et al. 2007, 2011, 2015; Blackwell and Dubins 1983; Chau and Rasonyi 2017; Cortissoz 2007; Dumav and Stinchombe 2016; Fernique 1988; Hernandez-Ceron 2010; Jakubowski 1998; Sethuraman 2002).

We aim at getting some new results in the spirit of SRT. Our starting point is the following version of SRT, recently proved in Pratelli and Rigo (2023).

Theorem 1 *Let T be a separable metric space, $g : S \to T$ a Borel function, and*

$$\sigma(g) = \left\{ g^{-1}(B) : B \in \mathcal{B}(T) \right\}.$$

Suppose

$$\mu_n \text{ is tight and } \mu_n = \mu_0 \text{ on } \sigma(g) \text{ for every } n \geq 0.$$

Then, on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, there is a coupling (X_n) of (μ_n) such that

$$g(X_n) = g(X_0) \text{ for all } n \geq 0. \tag{1}$$

In addition to (1), one also obtains $X_n \xrightarrow{\mathbb{P}-a.s.} X_0$ if and only if

$$E_{\mu_n}(f \mid g) \xrightarrow{\mu_0-a.s.} E_{\mu_0}(f \mid g) \quad \text{for each bounded continuous } f : S \to \mathbb{R}.$$

Here and in the sequel, for any probability ν on $\mathcal{B}(S)$, the notation $E_\nu(f \mid g)$ stands for the conditional expectation of f given $\sigma(g)$ in the probability space $(S, \mathcal{B}(S), \nu)$. Note that, when g is constant, $\sigma(g)$ reduces to the trivial σ-field and $E_\nu(f \mid g) = E_\nu(f) = \int f \, d\nu$. Hence, if g is constant and the μ_n are tight, Theorem 1 reduces to SRT.

This paper provides some extensions of Theorem 1 and investigates some of its consequences. Our results are of three types.

(i) In Theorem 3, $\sigma(g)$ is replaced by an arbitrary sub-σ-field $\mathcal{G} \subset \mathcal{B}(S)$. In this case, the coupling (X_n) of (μ_n) only satisfies

$$\mathbb{P}(X_n \in A, \ X_0 \notin A) = 0 \quad \text{for all } A \in \mathcal{G} \text{ and } n \geq 0.$$

However, in the special case $\mathcal{G} = \sigma(g)$ with g as in Theorem 1, the above condition is equivalent to $g(X_n) = g(X_0)$ a.s. Hence, Theorem 3 actually extends Theorem 1.

(ii) In Examples 3 and 4, Theorem 1 is applied to some specific frameworks. Example 3 deals with a sequence $(U_n : n \geq 0)$ of cadlag processes with finite activity. Let U_n^* be the continuous part of U_n. It is shown that, if $U_n^* \sim U_0^*$ for all n, the U_n admit a common decomposition. Precisely,

$$U_n \sim I + J_n \quad \text{for all } n \geq 0,$$

where the processes I and J_n are defined on the same probability space, I has continuous paths and J_n is a pure jump process. Example 4 is concerned with optimal transport. It is shown that Theorem 1 implies (and slightly improves) a recent duality result on equivalence couplings and total variation distances; see (Jaffe 2023).

(iii) In Sect. 4, we deal with models and kernels. Let $(\Theta, \mathcal{H})$ and $(\mathcal{X}, \mathcal{E})$ be measurable spaces. A model is a collection $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, indexed by Θ, of probability measures P_θ on $\mathcal{E}$. A kernel is a model which satisfies a certain measurability condition. Those non-atomic kernels such that

$$P_\theta\left(h^{-1}\{\theta\}\right) = 1, \quad \theta \in \Theta,$$

for some measurable function $h : \mathcal{X} \rightarrow \Theta$, have been recently characterized. Such a characterization, obtained in Hansen et al. (2024), is reported in Theorem 4. Our contribution consists in two versions of Theorem 4. One extends Theorem 4 from kernels to models, while the other is in the spirit of Theorem 1. Unlike Theorem 4, both versions admit a straightforward proof.

Obviously, models and kernels are fundamental in probability theory (just think of conditional distributions and Markov processes). But models and kernels play a role in many other frameworks. For instance, in decision theory, a model $\mathcal{P}$ can be regarded as the collection of probability distributions of a state-contingent payoff conditional on a parameter θ. Or else, in statistical inference, $\mathcal{P}$ may be viewed as the class of possible probability distributions on the data. Accordingly, Theorem 4 and its two versions can be attached some interpretation. In Sect. 4, this interpretation is discussed and various examples are given.

2 Preliminaries

We briefly recall some well known definitions and results. To this end, we let $(\mathcal{X}, \mathcal{E}, \mu)$ denote any probability space.

The measurable space $(\mathcal{X}, \mathcal{E})$ is said to be a *standard Borel space* if $\mathcal{X}$ is a Borel subset of a Polish space and $\mathcal{E} = \mathcal{B}(\mathcal{X})$. Similarly, $(\mathcal{X}, \mathcal{E})$ is a *Radon space* if $\mathcal{X}$ is a metric space, $\mathcal{E} = \mathcal{B}(\mathcal{X})$, and each probability measure on $\mathcal{E}$ is tight. A standard Borel space is a Radon space but not conversely. For instance, if $\mathcal{X}$ is a universally measurable, non-Borel subset of a Polish space, then $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is not a standard Borel space but every probability measure on $\mathcal{B}(\mathcal{X})$ is tight.

A μ-*atom* is a set $A \in \mathcal{E}$ such that $\mu(A) > 0$ and $\mu(\cdot \mid A)$ is 0-1 valued. We say that $(\mathcal{X}, \mathcal{E}, \mu)$ is a *non-atomic* probability space, or that μ is non-atomic, if μ has no atoms. If $\mathcal{X}$ is a separable metric space and $\mathcal{E} = \mathcal{B}(\mathcal{X})$, then μ is non-atomic if and only if $\mu\{x\} = 0$ for all $x \in \mathcal{X}$.

Let $\mathcal{F} \subset \mathcal{E}$ be a sub-σ-field. A *regular conditional distribution* for μ given $\mathcal{F}$ is a collection $\gamma = \{\gamma(x) : x \in \mathcal{X}\}$ such that:

− $\gamma(x)$ is a probability measure on $\mathcal{E}$ for each $x \in \mathcal{X}$;
− $\gamma(\cdot)(A)$ is a version of $E_\mu(1_A \mid \mathcal{F})$ for each $A \in \mathcal{E}$.

If $(\mathcal{X}, \mathcal{E})$ is a Radon space, a regular conditional distribution for μ given $\mathcal{F}$ exists and is μ-a.s. unique.

Finally, to prove forthcoming Theorem 3, we report the following version of SRT; see (Blackwell and Dubins 1983) and [Hernandez-Ceron (2010), p. 52–54] for a detailed proof.

Theorem 2 (Blackwell and Dubins) *Let m be the Lebesgue measure on $\mathcal{B}((0, 1))$ and Λ the collection of probability measures on $\mathcal{B}(S)$. If S is Polish, there is a Borel map $\Phi : (0, 1) \times \Lambda \to S$ such that*

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

It is easily seen that Theorem 2 is still true if S is a Borel subset of a Polish space (but not necessarily a Polish space).

3 Theorem 1 and its consequences

This section includes three applications of Theorem 1, outlined in the form of examples, as well as an extension of Theorem 1. We begin with the latter.

Any σ-field $\mathcal{G}$ over S can be written as $\mathcal{G} = \sigma(g)$ for a suitable function g on S. More precisely, the following result is available.

Lemma 1 *For each σ-field $\mathcal{G}$ over S, there are a measurable space $(T, \mathcal{C})$ and a function $g : S \to T$ such that*

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

Proof For each $x \in S$, let $H(x)$ be the $\mathcal{G}$-atom including the point x, that is

$$H(x) = \{y \in S : 1_B(y) = 1_B(x) \text{ for each } B \in \mathcal{G}\};$$

see e.g. (Berti and Rigo 2007) and (Blackwell and Dubins 1975). Define

$$T = \big\{ H(x) : x \in S \big\}.$$

Then, T is a partition of S and every element of $\mathcal{G}$ is a union of elements of T. For any $C \subset T$, define $C^* = \big\{ x \in S : H(x) \in C \big\}$. Then, it suffices to let

$$\mathcal{C} = \big\{ C \subset T : C^* \in \mathcal{G} \big\} \quad \text{and} \quad g(x) = H(x) \quad \text{for every } x \in S.$$

$\square$

Based on Lemma 1, it is tempting to extend Theorem 1 to an arbitrary sub-σ-field $\mathcal{G} \subset \mathcal{B}(S)$. This is impossible, however, if Theorem 1 is stated as above.

Example 1 Suppose $\mu_n\{x\} = \mu_0\{x\} = 0$ for all $x \in S$ and take $\mathcal{G}$ to be the collection of countable and co-countable subsets of S. In this case, $\mu_n = \mu_0$ on $\mathcal{G}$. However, since $\mathcal{G}$ includes the singletons, any function g such that $\mathcal{G} = \sigma(g)$ is injective, so that $g(X_n) = g(X_0)$ amounts to $X_n = X_0$. Hence, (μ_n) admits a coupling (X_n) satisfying condition (1) if and only if $\mu_n = \mu_0$ on all of $\mathcal{B}(S)$.

The next result is motivated by the previous comments. In the sequel, for any topological space T, we denote by $C_b(T)$ the collection of real bounded continuous functions on T.

Theorem 3 *Fix a sub-σ-field $\mathcal{G} \subset \mathcal{B}(S)$ and suppose*

$$\mu_n \text{ is tight and } \mu_n = \mu_0 \text{ on } \mathcal{G} \text{ for every } n \geq 0.$$

Then, on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, there is a coupling (X_n) of (μ_n) such that

$$\mathbb{P}(X_n \in A, \ X_0 \notin A) = 0 \text{ for all } A \in \mathcal{G} \text{ and } n \geq 0. \tag{2}$$

In addition to (2), one also obtains $X_n \stackrel{\mathbb{P}-a.s.}{\longrightarrow} X_0$ if and only if

$$E_{\mu_n}(f \mid \mathcal{G}) \stackrel{\mu_0-a.s.}{\longrightarrow} E_{\mu_0}(f \mid \mathcal{G}) \quad \text{for each } f \in C_b(S). \tag{3}$$

Proof We just give a sketch of the proof, for it is quite similar to that of Theorem 1.

Since all the μ_n are tight, S can be assumed to be a Borel subset of a Polish space. Hence, Theorem 2 applies. Moreover, for each $n \geq 0$, we can fix a regular conditional distribution for μ_n given $\mathcal{G}$, say $\gamma_n = \{\gamma_n(x) : x \in S\}$; see Sect. 2.

Let m be the Lebesgue measure on $\mathcal{B}((0, 1))$ and $\Phi : (0, 1) \times \Lambda \to S$ the Borel map involved in Theorem 2. Define

$$\Omega = (0, 1) \times (0, 1), \quad \mathcal{A} = \mathcal{B}\big((0, 1) \times (0, 1)\big), \quad \mathbb{P} = m \times m.$$

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

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

The X_n are S-valued random variables on $(\Omega, \mathcal{A}, \mathbb{P})$. Arguing as in the proof of Theorem 1, it can be shown that (X_n) is a coupling of (μ_n) and condition (3) is equivalent to $X_n \xrightarrow{\mathbb{P}-a.s.} X_0$. Finally, we prove (2). Fix $A \in \mathcal{G}$ and note that

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

Since $A \in \mathcal{G}$, then $\gamma_n(x)(A) = 1_A(x)$ for μ_n-almost all $x \in S$. Since $\mu_n = \mu_0$ on $\mathcal{G}$, it follows that

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

Therefore, since $m \circ \phi^{-1} = \mu_0$, one obtains

$$\mathbb{P}(X_n \in A, \ X_0 \notin A) = \int_0^1 m\Big\{\beta : X_n(\alpha, \beta) \in A, \ X_0(\alpha, \beta) \notin A\Big\} d\alpha$$
$$\leq \int_0^1 \min\Big\{\gamma_n[\phi(\alpha)](A), \ \gamma_0[\phi(\alpha)](A^c)\Big\} d\alpha$$
$$= \int \min\Big\{\gamma_n(x)(A), \ \gamma_0(x)(A^c)\Big\} \mu_0(dx)$$
$$= \int \min\Big\{1_A(x), \ 1_{A^c}(x)\Big\} \mu_0(dx) = 0.$$

$\square$

Theorem 3 extends Theorem 1 to an arbitrary sub-σ-field $\mathcal{G} \subset \mathcal{B}(S)$. In fact, if g is as in Theorem 1, then $g(X_n) = g(X_0)$ a.s. if and only if $\mathbb{P}(X_n \in A, \ X_0 \notin A) = 0$ for all $A \in \sigma(g)$.

One more remark on Theorem 3 is in order. If X and Y are S-valued random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ such that

$$\mathbb{P}(X \in A, \ Y \notin A) = 0 \quad \text{for all } A \in \mathcal{G}, \tag{4}$$

then

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A, \ Y \in A) = \mathbb{P}(Y \in A) \quad \text{for all } A \in \mathcal{G}.$$

Therefore, for any tight probability measures μ and ν on $\mathcal{B}(S)$, Theorem 3 yields

$$\mu = \nu \text{ on } \mathcal{G} \quad \Leftrightarrow \quad \text{Condition (4) holds for some } X \text{ and } Y$$
$$\text{such that } X \sim \mu \text{ and } Y \sim \nu.$$

We now turn to some applications of Theorem 1. We begin with an example which is not new but may be useful to make clear the scope of Theorem 1.

Example 2 (**Corollary 2 of** Pratelli and Rigo (2023)) For each $n \geq 0$, let U_n and V_n be random variables on a probability space $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$. Suppose U_n is S_1-valued and V_n is S_2-valued, where S_1 and S_2 are metric spaces and S_1 is separable. Suppose also that $U_n \sim U_0$ and (U_n, V_n) has a tight probability distribution. Under these conditions, Theorem 1 applies to $S = S_1 \times S_2$ and $g(x, y) = x$. It follows that, on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, there are random variables U and V_n^* such that $(U, V_n^*) \sim (U_n, V_n)$ for all $n \geq 0$. Moreover, $V_n^* \xrightarrow{a.s.} V_0^*$ if and only if $E_{\mu_n}(f \mid g) \xrightarrow{\mu_0 - a.s.} E_{\mu_0}(f \mid g)$ for each $f \in C_b(S_2)$, where μ_n denotes the probability distribution of (U_n, V_n).

In a nutshell, Example 2 may be summarized as follows. If $U_n \sim U_0$ for all n, the random variables (U_n, V_n) can be replaced by (U, V_n^*). In addition to satisfying $(U, V_n^*) \sim (U_n, V_n)$, the new random variables (U, V_n^*) are all defined on the same probability space and they all have the same first coordinate (that is, U). Using (U, V_n^*) instead of (U_n, V_n) may be useful in various settings, such as mass transportation and stochastic control.

The next example deals with a sequence $U_0, U_1, \ldots$ of cadlag processes indexed by $[0, \infty)$. Using Theorem 1 we prove that, if the continuous part of U_n is distributed as that of U_0 for every n, then $U_0, U_1, \ldots$ can be coupled so as to have exactly the same continuous part.

Example 3 (Decomposition of cadlag processes with finite activity) Let D be the set of real cadlag functions on $[0, \infty)$, equipped with the Skorohod topology. Define

$$S = \left\{ x \in D : \sum_{0 < s \leq t} |\Delta x(s)| < \infty \text{ for each } t > 0 \right\} \quad \text{and}$$

$$g(x)(t) = x(t) - \sum_{0 < s \leq t} \Delta x(s) \quad \text{for each } x \in S \text{ and } t \geq 0,$$

where $\Delta x(s) = x(s) - x(s-)$ is the jump of x at the point s. In financial econometrics, a cadlag function is said to have *finite activity* if it has only finitely many jumps on any bounded interval. Hence, in particular, S includes all elements of D with finite activity. In turn, the function g associates every $x \in S$ with its continuous part $g(x)$. It can be shown that $g : S \to C$ is a Borel map, where C denotes the set of continuous functions on $[0, \infty)$ (we omit the calculations). Moreover, since D is Polish and $S \in \mathcal{B}(D)$, each probability measure on $\mathcal{B}(S)$ is tight.

For each $n \geq 0$, let U_n be a process with paths in S. Suppose $g(U_n) \sim g(U_0)$ for each $n \geq 0$, namely, the continuous parts of the U_n are identically distributed. Then, there are processes I and J_n such that:

- I and the J_n are all defined on the same probability space;
- $I + J_n \sim U_n$ for all $n \geq 0$;
- I has continuous paths while J_n is a pure jump process.

The existence of I and J_n follows from Theorem 1. It suffices to take μ_n as the probability distribution of U_n and to let

$$I = g(X_0) \quad \text{and} \quad J_n = X_n - g(X_n) = X_n - g(X_0).$$

Note also that $I + J_n \xrightarrow{a.s.} I + J_0$ (in the Skorohod topology) if and only if

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

Our next example deals with a notion of duality recently introduced by Jaffe (2023). In addition to be theoretically intriguing, this notion is potentially useful in various frameworks, including mathematical finance, decision theory, mass transportation and probability theory.

Example 4 (Equivalence couplings and total variation) To keep the notation easier, in this example, we write $\mathcal{B}$ instead of $\mathcal{B}(S)$. Let $E \subset S \times S$ be a measurable equivalence relation. This means that $E \in \mathcal{B} \otimes \mathcal{B}$ and the relation on S defined as

$$x \sim y \quad \Leftrightarrow \quad (x, y) \in E$$

is reflexive, symmetric and transitive. Say that E is *strongly dualizable* if there is a sub-σ-field $\mathcal{C} \subset \mathcal{B}$ such that

$$\min_{P \in \Gamma(\mu,\nu)} (1 - P(E)) = \sup_{A \in \mathcal{C}} |\mu(A) - \nu(A)| \tag{5}$$

for all probability measures μ and ν on $\mathcal{B}$. Here, $\Gamma(\mu, \nu)$ is the collection of probability measures on $\mathcal{B} \otimes \mathcal{B}$ with marginals μ and ν, and the notation "min" asserts that the infimum is actually achieved.

Various conditions for E to be strongly dualizable are in Jaffe (2023); see also (Pratelli and Rigo 2024). One of such conditions is the following. Define the sub-σ-field

$$\mathcal{U} = \{A \in \mathcal{B} : 1_A(x) = 1_A(y) \text{ for all } (x, y) \in E\}.$$

Then, E is strongly dualizable provided $E \in \mathcal{U} \otimes \mathcal{U}$ and $(S, \mathcal{B})$ is a standard Borel space; see [Jaffe (2023), Theo. 3.13] and [Pratelli and Rigo (2024), Cor. 6]. This result is a consequence of Theorem 1, however, as we now prove. Moreover, the assumption that $(S, \mathcal{B})$ is standard Borel can be weakened.

Suppose $E \in \mathcal{U} \otimes \mathcal{U}$ and $(S, \mathcal{B})$ is a Radon space. Since $E \in \mathcal{U} \otimes \mathcal{U}$,

$$E \in \sigma(A_1 \times B_1, A_2 \times B_2, \ldots)$$

for some A_n, $B_n \in \mathcal{U}$, $n \geq 1$. Define $\mathcal{G} = \sigma(A_1, B_1, A_2, B_2, \ldots)$. Since $\mathcal{G}$ is a countably generated sub-σ-field of $\mathcal{B}$, there is a Borel function $g : S \to \mathbb{R}$ such that $\mathcal{G} = \sigma(g)$. Moreover, since $E \in \mathcal{G} \otimes \mathcal{G}$, one obtains

$$\big\{(x, y) : g(x) = g(y)\big\} \subset E.$$

Next, fix two probability measures μ and ν on $\mathcal{B}$ such that $\mu = \nu$ on $\mathcal{U}$. Since $\mathcal{G} \subset \mathcal{U}$ and $(S, \mathcal{B})$ is a Radon space, μ and ν are tight and $\mu = \nu$ on $\mathcal{G}$. Because of Theorem 1, applied with $\mu_0 = \mu$ and $\mu_n = \nu$ for $n > 0$, there are S-valued random variables X_0 and X_1 such that $X_0 \sim \mu$, $X_1 \sim \nu$ and $g(X_0) = g(X_1)$. Denoting by P the probabilty distribution of (X_0, X_1), it follows that

$$P \in \Gamma(\mu, \nu) \quad \text{and} \quad P(E) \geq P\big\{(x, y) : g(x) = g(y)\big\} = 1.$$

Therefore, letting $\mathcal{C} = \mathcal{U}$, equation (5) holds provided $\mu = \nu$ on $\mathcal{U}$. This concludes the proof. In fact, if $\mathcal{C} = \mathcal{U}$, equation (5) holds for all μ and ν if and only if it holds for those μ and ν such that $\mu = \nu$ on $\mathcal{U}$; see e.g. [Jaffe (2023),Prop. 3.9].

4 Kernels versus models

Let $(\Theta, \mathcal{H})$ and $(\mathcal{X}, \mathcal{E})$ be measurable spaces. To avoid trivialities, we assume

$$\text{card}\,(\mathcal{X}) > 1.$$

A *model* is a collection

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$$

where each P_θ is a probability measure on $\mathcal{E}$. A model $\mathcal{P}$ is non-atomic if P_θ is a non-atomic probability measure on $\mathcal{E}$ for each $\theta \in \Theta$. Moreover, $\mathcal{P}$ is measurable if the real valued map $\theta \mapsto P_\theta(A)$ is $\mathcal{H}$-measurable for fixed $A \in \mathcal{E}$. A measurable model is usually called *a kernel*.

One more definition is needed. Suppose $\mathcal{H}$ includes the singletons. Then, a model $\mathcal{P}$ is said to be *orthogonal* if there is a measurable function $h : \mathcal{X} \to \Theta$ such that

$$P_\theta\Big(h^{-1}\{\theta\}\Big) = 1 \quad \text{for all } \theta \in \Theta.$$

Here, measurability of h is meant as $h^{-1}(\mathcal{H}) \subset \mathcal{E}$. Orthogonal kernels are investigated in Mauldin et al. (1983) and (Weis 1984). They are involved in many contexts, including ergodic decompositions, Gibbs states, disintegrations and extremal models; see e.g. (Berti and Rigo 2007; Blackwell and Dubins 1975; Dynkin 1978; Farrell 1962; Fölmer 1975; Lauritzen 1974; Maitra 1977). The next example, even if obvious, is useful to frame orthogonal kernels.

Example 5 *(An orthogonal kernel)* For any real random variables U and V, there is an orthogonal version of the conditional distribution of (U, V) given U. Take in fact $(\Theta, \mathcal{H}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $(\mathcal{X}, \mathcal{E}) = (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ and define the function $h(u, v) = u$ for all $(u, v) \in \mathbb{R}^2$. Also, denote by π the marginal distribution of U. Any kernel $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ satisfying the equation

$$\text{Prob}\big(U \in A, \ V \in B\big) = \int_A P_\theta(\mathbb{R} \times B) \, \pi(d\theta), \quad \text{for all } A, \ B \in \mathcal{B}(\mathbb{R}),$$

is a version of the conditional distribution of (U, V) given U. If $\mathcal{P}$ is one such version, it is well known that

$$P_\theta\left(h^{-1}\{\theta\}\right) = P_\theta\big(\{\theta\} \times \mathbb{R}\big) = 1 \quad \text{for } \pi\text{-almost all } \theta \in \mathbb{R};$$

see e.g. (Berti and Rigo 2007) and (Blackwell and Dubins 1975). Hence, up to modifying $\mathcal{P}$ on a π-null set, one obtains a kernel $\mathcal{Q} = \{Q_\theta : \theta \in \Theta\}$ such that

$$\pi\big\{\theta : Q_\theta \neq P_\theta\big\} = 0 \quad \text{and} \quad Q_\theta\left(h^{-1}\{\theta\}\right) = 1 \quad \text{for all } \theta \in \mathbb{R}.$$

Such a $\mathcal{Q}$ is an orthogonal version of the conditional distribution of (U, V) given U.

In this section, we focus on the following result from (Hansen et al. 2024).

Theorem 4 (Hansen, Maccheroni, Marinacci, Sargent) Let $(\Theta, \mathcal{H})$ and $(\mathcal{X}, \mathcal{E})$ be standard Borel spaces and $\mathcal{P}$ a kernel. Then, $\mathcal{P}$ is non-atomic and orthogonal if and only if, for any other kernel $\mathcal{Q} = \{Q_\theta : \theta \in \Theta\}$, there is a measurable function $f : \mathcal{X} \to \mathcal{X}$ such that

$$Q_\theta = P_\theta \circ f^{-1} \quad \text{for each } \theta \in \Theta.$$

In Theorem 4, measurability of f is meant as $f^{-1}(\mathcal{E}) \subset \mathcal{E}$ and $P_\theta \circ f^{-1}$ denotes the probability on $\mathcal{E}$ defined as $P_\theta \circ f^{-1}(A) = P_\theta\big(f^{-1}(A)\big)$ for all $A \in \mathcal{E}$.

Essentially, Theorem 4 states that a kernel $\mathcal{P}$ is non-atomic and orthogonal if and only if any other kernel $\mathcal{Q}$ is a push forward of $\mathcal{P}$, in the sense that $Q_\theta = P_\theta \circ f^{-1}$ for all θ and a suitable function f. This characterization may be useful in every framework where kernels play a role, and the list of such frameworks is very long. In probability theory, for instance, kernels are obviously a basic ingredient: just think of conditional distributions or Markov processes. In Bayesian statistical inference, a kernel $\mathcal{P}$ may be viewed as the collection of the distributions on the data conditional on a parameter. In decision theory, $\mathcal{P}$ can be regarded as the collection of the distributions of a state-contingent payoff conditional on a parameter; see e.g. (Hansen et al. 2024). In weak optimal transport, each P_θ provides information about how the mass taken at θ is distributed over $\mathcal{X}$; see e.g. (Chone and Kramarz 2021; Chone et al. 2023; Galichon et al. 2014) and references therein. In each of these frameworks, thus, Theorem 4 has some motivation.

The previous remarks are still valid if kernels are replaced by models. In fact, there are several problems where measurability of a kernel is superfluous. We support this claim by three examples.

Example 6 *(Classical statistical inference)* According to the classical approach to statistics, the two basic ingredients of an inferential problem are a measurable space $(\mathcal{X}, \mathcal{E})$ and a model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. The set $\mathcal{X}$ is regarded as the sample space and P_θ is the probability distribution of the data when the value of the parameter is θ. Importantly, the parameter is viewed as an unknown but fixed constant, and there is no reason to integrate over it. Hence, the σ field $\mathcal{H}$ is superfluous and measurability of $\mathcal{P}$ is not required. In the language of this paper, $\mathcal{P}$ is a model but not a kernel.

Example 7 *(Disintegrations)* For any model $\mathcal{P}$, let $\sigma(\mathcal{P})$ denote the σ-field over Θ generated by the maps $\theta \mapsto P_\theta(A)$ for all $A \in \mathcal{E}$. One of the main reasons for requiring measurability of a kernel is the need of defining a probability on $\mathcal{E}$ as

$$\mu_\pi(A) = \int P_\theta(A) \, \pi(d\theta), \quad A \in \mathcal{E}, \tag{6}$$

where π is a given probability on $\mathcal{H}$. Such μ_π cannot be defined if $\mathcal{P}$ is a model but not a kernel. In Bayesian inference, for instance, $\mathcal{P}$ is asked to be a kernel and π is the prior distribution. This procedure assumes that the σ-field $\mathcal{H}$ is fixed before than $\mathcal{P}$. However, these two steps could be reverted. Precisely, one first selects a model $\mathcal{P}$ and then takes $\mathcal{H} = \sigma(\mathcal{P})$. This actually happens as regards non-measurable disintegrations. To illustrate, suppose we are given a probability P on $\mathcal{E}$ and a partition $\{A_\theta : \theta \in \Theta\}$ with $A_\theta \in \mathcal{E}$ for all θ. A (non-measurable) disintegration for P is a pair $(\mathcal{P}, \pi)$ where $\mathcal{P}$ is a model, π a probability on $\sigma(\mathcal{P})$, and

- $P_\theta(A_\theta) = 1$ for all $\theta \in \Theta$;
- $P(A) = \int P_\theta(A) \, \pi(d\theta)$ for all $A \in \mathcal{E}$.

A disintegration is said to be measurable if Θ is equipped with a σ-field $\mathcal{H}$ and $\mathcal{P}$ is a kernel. Obviously, the conditions for having a non-measurable disintegration are much more general than those for a measurable disintegation; see e.g. (Berti et al. 2020) and references therein.

Example 8 *(Orthogonality preserving models)* As noted in Example 7, if $\mathcal{P}$ is a kernel and π a probability on $\mathcal{H}$, one can define a probability μ_π on $\mathcal{E}$ via equation (6). A kernel $\mathcal{P}$ is orthogonality preserving if μ_{π_1} and μ_{π_2} are singular whenever π_1 and π_2 are singular probabilities on $\mathcal{H}$. It is straightforward to prove that an orthogonal kernel is orthogonality preserving; see (Mauldin et al. 1983). This implication is still valid if kernels are replaced by models. Indeed, in Proposition 7, we will show that a weakly orthogonal model (as defined below) is orthogonality preserving in a suitable sense.

We now extend Theorem 4 from kernels to models. Unlike Theorem 4, the extended version admits a straightforward proof. Moreover, the notion of orthogonality can be weakened.

For any model $\mathcal{P}$, define the σ-field

$$\mathcal{E}_{\mathcal{P}} = \bigcap_{\theta \in \Theta} \overline{\mathcal{E}}^{P_\theta}$$

where $\overline{\mathcal{E}}^{P_\theta}$ is the completion of $\mathcal{E}$ with respect to P_θ. Given a function $f : \mathcal{X} \to \mathcal{X}$, we say that f is *measurable* if $f^{-1}(\mathcal{E}) \subset \mathcal{E}$ and that f is $\mathcal{P}$-*measurable* if $f^{-1}(\mathcal{E}) \subset \mathcal{E}_{\mathcal{P}}$. Note that f is $\mathcal{P}$-measurable if and only if it is measurable with respect to P_θ for every $\theta \in \Theta$. We also say that $\mathcal{P}$ is *weakly orthogonal* if there is a partition $\{A_\theta : \theta \in \Theta\}$ of $\mathcal{X}$ such that

$$A_\theta \in \mathcal{E}_{\mathcal{P}} \quad \text{and} \quad P_\theta(A_\theta) = 1 \quad \text{for each } \theta \in \Theta. \tag{7}$$

Here, with a slight abuse of notation, the only extension of P_θ to $\mathcal{E}_{\mathcal{P}}$ is still denoted by P_θ. In this notation, the following result is available.

Theorem 5 *Suppose card $(\Theta) \leq$ card $(\mathcal{X})$ and $(\mathcal{X}, \mathcal{E})$ is a Radon space. Then, a model $\mathcal{P}$ is non-atomic and weakly orthogonal if and only if, for any other model $\mathcal{Q}$, there is a $\mathcal{P}$-measurable function $f : \mathcal{X} \to \mathcal{X}$ such that*

$$Q_\theta = P_\theta \circ f^{-1} \quad \text{for each } \theta \in \Theta. \tag{8}$$

Proof If $\mathcal{E}$ does not support non-atomic probability measures, non-atomic models do not exist and condition (8) certainly fails for some choice of $\mathcal{Q}$. Hence, $\mathcal{E}$ can be assumed to support a non-atomic probability measure.

Suppose $\mathcal{P}$ is non-atomic and weakly orthogonal. Fix a model $\mathcal{Q}$ and a partition $\{A_\theta : \theta \in \Theta\}$ of $\mathcal{X}$ satisfying condition (7). Given $\theta \in \Theta$, since Q_θ is tight and P_θ is a non-atomic probability measure, there is a measurable function $f_\theta : \mathcal{X} \to \mathcal{X}$ such that $Q_\theta = P_\theta \circ f_\theta^{-1}$; see [Berti et al. (2007), Theo. 3.1]. For each $x \in \mathcal{X}$, denote by θ_x the unique $\theta \in \Theta$ such that $x \in A_\theta$. Define a function $f : \mathcal{X} \to \mathcal{X}$ as

$$f(x) = f_{\theta_x}(x) \quad \text{for every } x \in \mathcal{X}.$$

Fix $\theta \in \Theta$ and $A \in \mathcal{E}$. Then,

$$\{f \in A\} = \{f_\theta \in A, \ f = f_\theta\} \cup \{f \in A, \ f \neq f_\theta\}.$$

Since $\{f \neq f_\theta\} \subset A_\theta^c$ and $P_\theta(A_\theta^c) = 0$, both the sets $\{f = f_\theta\}$ and $\{f \in A, \ f \neq f_\theta\}$ belong to $\overline{\mathcal{E}}^{P_\theta}$. Since f_θ is measurable, $\{f_\theta \in A\} \in \mathcal{E}$. It follows that

$$\{f \in A\} \in \overline{\mathcal{E}}^{P_\theta}.$$

Therefore, f is $\mathcal{P}$-measurable. Furthermore,

$$Q_\theta = P_\theta \circ f_\theta^{-1} = P_\theta \circ f^{-1} \quad \text{for each } \theta \in \Theta.$$

Conversely, suppose that, for any model $\mathcal{Q}$, there is a $\mathcal{P}$-measurable function $f : \mathcal{X} \to \mathcal{X}$ satisfying condition (8). Fix $\theta \in \Theta$ and a non-atomic probability measure ν on $\mathcal{E}$. Taking $\mathcal{Q}$ such that $Q_\theta = \nu$, condition (8) implies $P_\theta \circ f^{-1} = \nu$ for some f. Hence, P_θ is non-atomic since ν is non-atomic and $P_\theta \circ f^{-1} = \nu$. We next prove that $\mathcal{P}$ is weakly orthogonal. Since $\mathrm{card}\,(\Theta) \leq \mathrm{card}\,(\mathcal{X})$, there is an injective function $\phi : \Theta \to \mathcal{X}$. Letting $Q_\theta = \delta_{\phi(\theta)}$ for each $\theta \in \Theta$, condition (8) yields

$$P_\theta\big(f^{-1}\{\phi(\theta)\}\big) = \delta_{\phi(\theta)}\{\phi(\theta)\} = 1$$

for some $\mathcal{P}$-measurable function $f : \mathcal{X} \to \mathcal{X}$. Define $B_\theta = f^{-1}\{\phi(\theta)\}$ and

$$D = \left(\bigcup_{\theta \in \Theta} B_\theta\right)^c.$$

The sets B_θ belong to $\mathcal{E}_\mathcal{P}$ and are pairwise disjoint since ϕ is injective. Moreover, $D \in \mathcal{E}_\mathcal{P}$ since $D \subset B_\theta^c$ and $P_\theta(B_\theta^c) = 0$ for all $\theta \in \Theta$. Hence, fixed any point $\theta_0 \in \Theta$, condition (7) holds with $A_\theta = B_\theta$ for $\theta \neq \theta_0$ and $A_{\theta_0} = B_{\theta_0} \cup D$. □

We do not know whether the assumption $\mathrm{card}\,(\Theta) \leq \mathrm{card}\,(\mathcal{X})$ can be dropped. Such an assumption, instead, is superfluous in Theorem 4. In fact, Theorem 4 is trivially true if $\mathcal{X}$ is countable. Otherwise, if $\mathcal{X}$ is uncountable, $\mathrm{card}\,(\Theta) \leq \mathrm{card}\,(\mathcal{X})$ follows from $(\Theta, \mathcal{H})$ and $(\mathcal{X}, \mathcal{E})$ are standard Borel spaces.

As noted above, the heuristic interpretation of kernels can be attached to models as well. Thus, Theorem 5 has essentially the same motivations as Theorem 4.

Our next result is actually a mixture of Theorems 1, 4 and 5. Let $\mathcal{P}$, $\mathcal{Q}_0$, $\mathcal{Q}_1, \ldots$ be models with $\mathcal{P}$ non-atomic and weakly orthogonal. By Theorem 5, for each $n \geq 0$, there is a $\mathcal{P}$-measurable function $f_n : \mathcal{X} \to \mathcal{X}$ such that $P_\theta \circ f_n^{-1} = Q_{n,\theta}$ for all θ. We now prove that, if $Q_{n,\theta} = Q_{0,\theta}$ on $\sigma(g)$ for all θ and a suitable function g, then f_n can be taken such that $g(f_n) = g(f_0)$. Moreover, we give conditions for $f_n \xrightarrow{P_\theta - a.s.} f_0$, as $n \to \infty$, for fixed $\theta \in \Theta$.

Theorem 6 *Let $(\mathcal{X}, \mathcal{E})$ be a Radon space, $\mathcal{Y}$ a separable metric space, and $g : \mathcal{X} \to \mathcal{Y}$ a Borel function. Let $\mathcal{P}$ and $\mathcal{Q}_n$ be models, where $n \geq 0$. Suppose $\mathcal{P}$ is non-atomic and weakly orthogonal and*

$$Q_{n,\theta} = Q_{0,\theta} \quad \text{on } \sigma(g) \text{ for all } n \geq 0 \text{ and } \theta \in \Theta.$$

Then, there are $\mathcal{P}$-measurable functions $f_n : \mathcal{X} \to \mathcal{X}$ such that

$$P_\theta \circ f_n^{-1} = Q_{n,\theta} \quad \text{and} \quad g(f_n) = g(f_0) \text{ for all } n \geq 0 \text{ and } \theta \in \Theta. \tag{9}$$

In addition to (9), for fixed $\theta \in \Theta$, one obtains

$$f_n \xrightarrow{P_\theta - a.s.} f_0$$

whenever

$$E_{Q_{n,\theta}}(\varphi \mid g) \xrightarrow{Q_{0,\theta}-a.s.} E_{Q_{0,\theta}}(\varphi \mid g) \quad \text{for each } \varphi \in C_b(\mathcal{X}). \tag{10}$$

Proof Fix $\theta \in \Theta$. By Corollary 4 of Pratelli and Rigo (2023), since $(\mathcal{X}, \mathcal{E})$ is Radon and $(\mathcal{X}, \mathcal{E}, P_\theta)$ is a non-atomic probability space, there are measurable functions $f_{n,\theta} : \mathcal{X} \to \mathcal{X}$ such that

$$Q_{n,\theta} = P_\theta \circ f_{n,\theta}^{-1} \quad \text{and} \quad g(f_{n,\theta}) = g(f_{0,\theta}) \quad \text{for all } n \geq 0.$$

Moreover, under condition (10), one also obtains $f_{n,\theta} \xrightarrow{P_\theta-a.s.} f_{0,\theta}$.

Next, since $\mathcal{P}$ is weakly orthogonal, there is a partition $\{A_\theta : \theta \in \Theta\}$ of $\mathcal{X}$ satisfying condition (7). For all $n \geq 0$ and $x \in \mathcal{X}$, define

$$f_n(x) = f_{n,\theta_x}(x)$$

where θ_x denotes the unique $\theta \in \Theta$ such that $x \in A_\theta$. Then, it is obvious that $g(f_n) = g(f_0)$ for all n. Moreover, arguing as in the proof of Theorem 5, the f_n are $\mathcal{P}$-measurable and $Q_{n,\theta} = P_\theta \circ f_n^{-1}$ for all n and θ. Finally, since $P_\theta(f_n = f_{n,\theta}) = 1$, condition (10) implies $f_n \xrightarrow{P_\theta-a.s.} f_0$.

Incidentally, we note that $\mathrm{card}(\Theta) \leq \mathrm{card}(\mathcal{X})$ under the assumptions of Theorem 6. In fact, $\mathrm{card}(\Theta) \leq \mathrm{card}(\mathcal{X})$ follows from $\mathcal{P}$ being weakly orthogonal.

We close the paper by proving a claim made in Example 8.

Proposition 7 *Let $\mathcal{P}$ be a model and $\sigma(\mathcal{P})$ the σ-field defined in Example 7. If $\mathcal{P}$ is weakly orthogonal and π_1 and π_2 are singular probabilities on $\sigma(\mathcal{P})$, then*

$$\mu_{\pi_1}(A) = \mu_{\pi_2}(A^c) = 1 \quad \text{for some } A \in \mathcal{E}_\mathcal{P}.$$

Proof Let $\{A_\theta : \theta \in \Theta\}$ be a partition of $\mathcal{X}$ satisfying condition (7). Since π_1 and π_2 are singular, there is $B \in \sigma(\mathcal{P})$ such that $\pi_1(B) = \pi_2(B^c) = 1$. Define

$$A = \bigcup_{\theta \in B} A_\theta.$$

Then, $A \supset A_\theta$ for $\theta \in B$ and $A \subset A_\theta^c$ for $\theta \in B^c$. Since $P_\theta(A_\theta) = 1$ for all $\theta \in \Theta$, it follows that $A \in \mathcal{E}_\mathcal{P}$. Moreover,

$$\mu_{\pi_1}(A) = \int P_\theta(A) \, \pi_1(d\theta) = \int_B P_\theta(A) \, \pi_1(d\theta) = \int_B 1 \, d\pi_1 = \pi_1(B) = 1,$$

and similarly $\mu_{\pi_2}(A^c) = 1$. □

Declarations

Conflict of interest The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

References

Berti, P., Rigo, P.: 0–1 laws for regular conditional distributions. Ann. Probab. **35**, 649–662 (2007)

Berti, P., Pratelli, L., Rigo, P.: Skorohod representation on a given probability space. Prob. Theo. Relat. Fields **137**, 277–288 (2007)

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

Berti, P., Pratelli, L., Rigo, P.: A Skorohod representation theorem without separability. Electr. Commun. Probab. **18**, 1–12 (2013)

Berti, P., Pratelli, L., Rigo, P.: Gluing lemmas and Skorohod representations. Electr. Commun. Probab. **20**, 1–11 (2015)

Berti, P., Dreassi, E., Rigo, P.: A notion of conditional probability and some of its consequences. Decisions Econ. Financ. **43**, 3–15 (2020)

Blackwell, D., Dubins, L.E.: On existence and non-existence of proper, regular, conditional distributions. Ann. Probab. **3**, 741–752 (1975)

Blackwell, D., Dubins, L.E.: An extension of Skorohod's almost sure representation theorem. Proc. Am. Math. Soc. **89**, 691–692 (1983)

Chau, H.N., Rasonyi, M.: Skorohod's representation theorem and optimal strategies for markets with frictions. SIAM J. Control. Optim. **55**, 3592–3608 (2017)

Chone, P., Kramarz, F.: Matching workers' skills and firms' technologies: From bundling to unbundling, *Working Papers 2021-10 CREST* (2021)

Chone, P., Gozlan, N., Kramarz, F.: Weak optimal transport with unnormalized kernels. SIAM J. Math. Anal. **55**(6), 6039–6092 (2023)

Cortissoz, J.: On the Skorokhod representation theorem. Proc. Amer. Math. Soc. **135**, 3995–4007 (2007)

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

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

Dumav, M., Stinchombe, M.B.: Skorohod's representation theorem for sets of probabilities. Proc. Am. Math. Soc. **144**, 3123–3133 (2016)

Dynkin, E.B.: Sufficient statistics and extreme points. Ann. Probab. **6**, 705–730 (1978)

Farrell, R.H.: Representation of invariant measures. Ill. J. Math. **6**, 447–467 (1962)

Fernique, X.: Un modele presque sur pour la convergence en loi. C.R Acad. Sci. Paris Ser. I **306**, 335–338 (1988)

Fölmer, H.: Phase transition and Martin boundary, Sem. Prob. IX, Springer. Lect. Notes in Math. **465**, 305–317 (1975)

Galichon, A., Henry-Labordere, P., Touzi, N.: A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. Ann. Appl. Probab. **24**, 312–336 (2014)

Hansen, L.P., Maccheroni, F., Marinacci, M., Sargent, T.J.: Tempered Bayesian analysis, *Unpublished manuscript* (2024)

Hernandez-Ceron, N.: Extensions of Skorohod's almost sure representation theorem, *Master of Science Thesis, University of Alberta*, https://era.library.ualberta.ca/items/78ddb981-ca1c-4945-a6e9-8298217d6be6 (2010)

Jaffe, A.Q.: A strong duality principle for equivalence couplings and total variation. Electr. J. Probab. **28**, 1–33 (2023)

Jakubowski, A.: The almost sure Skorokhod representation for subsequences in nonmetric spaces. Theo. Probab. Appl. **42**, 167–174 (1998)

Lauritzen, S.L.: Sufficiency, prediction and extreme models. Scand. J. Stat. **1**, 128–134 (1974)

Maitra, A.: Integral representations of invariant measures. Trans. Am. Math. Soc. **229**, 209–225 (1977)

Mauldin, R.D., Preiss, D., Weizsacker, H.V.: Orthogonal transition kernels. Ann. Probab. **11**, 970–988 (1983)

Pratelli, L., Rigo, P.: A strong version of the Skorohod representation theorem. J. Theoret. Probab. **36**, 372–389 (2023)

Pratelli, L., Rigo, P.: Some duality results for equivalence couplings and total variation. Electr. Commun. Probab. **29**, 1–12 (2024)

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

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

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

Weis, L.: On the representation of order continuous operators by random measures. Trans. Amer. Math. Soc. **285**, 535–563 (1984)

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)