The Annals of Probability
2004, Vol. 32, No. 1A, 365–379

CONVERGENCE IN DISTRIBUTION OF NONMEASURABLE RANDOM ELEMENTS

BY PATRIZIA BERTI AND PIETRO RIGO

Universita' di Modena e Reggio-Emilia and Universita' di Pavia

A notion of convergence in distribution for non (necessarily) measurable random elements, due to Hoffmann-Jørgensen, is characterized in terms of weak convergence of finitely additive probability measures. A similar characterization is given for a strengthened version of such a notion. Further, it is shown that the empirical process for an exchangeable sequence can fail to converge, due to the nonexistence of any measurable limit, although it converges for an i.i.d. sequence. Because of phenomena of this type, Hoffmann-Jørgensen's definition is extended to the case of a nonmeasurable limit. In the extended definition, naturally suggested by the main results, the limit is a finitely additive probability measure.

1. Introduction. In some problems, the need arises for a notion of convergence in distribution for non (necessarily) measurable random elements. Examples are in the theory of empirical processes. Accordingly, a few definitions of convergence in distribution have been proposed. See Dudley (1966, 1967), Pyke and Shorack (1968) and Pollard (1984).

One more definition, due to Hoffmann-Jørgensen, has been recently analyzed by van der Vaart and Wellner (1996) and Dudley (1999) and taken as a basis for convergence of empirical processes and asymptotic statistics. Most statistical limit theorems, in fact, are just consequences of convergence of the (general) empirical process.

Hoffmann-Jørgensen's definition (HJ, in what follows) allows a rich and elegant theory and is very handy in applied situations. However, it also has some potential drawbacks and points to be developed. Indeed, (i) the case of a nonmeasurable limit is not covered; (ii) the probabilistic meaning should be deepened; and (iii) the underlying probability spaces of the random elements play a crucial role, since they determine outer expectations, and are to be explicitly declared to make the definition usable [cf. van der Vaart and Wellner (1996), pages 18 and 83].

In this paper, points (i)–(iii) are investigated. Our main result (Theorem 4) deals with (ii). It states that HJ is equivalent to convergence in distribution under *all finitely additive extensions* of the underlying probability measures. A similar characterization is given for a strengthened version of HJ. In view of Theorem 4, in our opinion, the probabilistic meaning of HJ is made more transparent. As to

Received July 2002; revised January 2003.

AMS 2000 subject classifications. 60B10, 60A05.

Key words and phrases. Convergence in distribution, empirical process, exchangeability, extension, finitely additive probability measure, measurability.

point (i), it is shown that measurability of the limit can be a real limitation of HJ, even in simple problems. Indeed, the empirical process for an exchangeable sequence can fail to converge, due to the nonexistence of any measurable limit, although it converges for an i.i.d. sequence; see Example 11. Because of these phenomena, HJ is extended to the case of a nonmeasurable limit. In the extended definition, naturally suggested by Theorem 4, the limit is a finitely additive probability measure. Finally, point (iii) is briefly discussed in Section 2. It turns out that (iii) is more or less unavoidable as far as a definition is asked to work in nonmeasurable problems.

A last, general remark is the following. There are problems in probability and statistics which cannot be solved in the usual countably additive setting, while admitting a finitely additive solution. Examples can be found in conditional probability, statistical inference, filtering, stochastic integration and the first digit problem. Moreover, some widely used statistical procedures, like formal posteriors of improper priors, can be justified in a finitely additive setting only. See Bumby and Ellentuck (1969), Dubins (1975), Heath and Sudderth (1978, 1989), Karandikar (1982, 1988), Kallianpur and Karandikar (1983, 1988), Regazzini (1987), Berti and Rigo (1994, 1996, 1999) and Dubins and Prikry (1995). In a sense, convergence in distribution of nonmeasurable random elements is one more example of this type. Indeed, by next Theorem 4, HJ is basically a finitely additive definition.

2. Hoffmann-Jørgensen's definition. As far as possible, the notation is that of van der Vaart and Wellner (1996). Thus, A is a directed set, $\mathbb{D}$ a metric space, $\mathcal{D}$ the Borel σ-field on $\mathbb{D}$ and $C_b(\mathbb{D})$ the class of real bounded continuous functions on $\mathbb{D}$. For all $\alpha \in A$, $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$ is a probability space and $X_\alpha : \Omega_\alpha \to \mathbb{D}$ an *arbitrary* map. Moreover, $(\Omega, \mathcal{A}, P)$ is a probability space and $X : \Omega \to \mathbb{D}$ a *measurable* map, that is, $X^{-1}(\mathcal{D}) \subset \mathcal{A}$. Because of measurability, the distribution μ of X can be defined as $\mu(B) = P(X \in B)$ for all $B \in \mathcal{D}$, while the distribution of X_α cannot be defined on the whole of $\mathcal{D}$. So, we let

$$\mathcal{D}_\alpha = \{B \in \mathcal{D} : X_\alpha^{-1}(B) \in \overline{\mathcal{A}_\alpha}\}, \qquad \alpha \in A,$$

be the sub-σ-field of $\mathcal{D}$ where the distribution of X_α can be unambiguously defined, $(\Omega_\alpha, \overline{\mathcal{A}_\alpha}, \overline{P_\alpha})$ denoting the completion of $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$. Clearly, $\mathcal{D}_\alpha = \mathcal{D}$ whenever X_α is measurable. For $B \in \mathcal{D}_\alpha$, the "partial" distribution of X_α is just $\mu_\alpha(B) = \overline{P_\alpha}(X_\alpha \in B)$.

In the standard theory, the X_α are measurable and "$X_\alpha \to X$ in distribution" means $Ef(X_\alpha) \to Ef(X)$ for all $f \in C_b(\mathbb{D})$. To extend this definition to the present case, where the X_α need not be measurable, for all $\alpha \in A$ and $Z : \Omega_\alpha \to \overline{\mathbb{R}}$, let us define

$$E^* Z = \inf\{EU : U : \Omega_\alpha \to \overline{\mathbb{R}} \text{ measurable}, EU \text{ exists}, U \geq Z\}$$

and $E_* Z = -E^*(-Z)$. Here, EU exists if at least one of EU^+ and EU^- is finite, and in that case $EU = EU^+ - EU^-$. In this notation, HJ (i.e., Hoffmann-Jørgensen's definition) is

$$X_\alpha \overset{\text{HJ}}{\to} X \quad \iff \quad E^* f(X_\alpha) \to Ef(X) \qquad \text{for all } f \in C_b(\mathbb{D}).$$

As noted in Section 1, despite its utility and elegance, some aspects of HJ need to be developed further. First, the case of a nonmeasurable limit X is not covered. It will be seen in Example 11 how the measurability of X can be a real limitation, even in simple problems. Second, the probabilistic content should still be investigated. Indeed, using outer expectations could appear something like a trick for avoiding measurability problems. Third, to make HJ usable, the underlying probability spaces $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$ are to be explicitly declared, since they determine outer expectations. In van der Vaart and Wellner (1996) and Dudley (1999), for instance, the $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$ are often taken to be the canonical spaces.

This paper focus on the three points above. The third one is dealt with in the rest of this section, while the other two are discussed in Sections 4 and 5.

At first sight, a definition of convergence which ignores $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$, and only depends on $(\mathbb{D}, \mathcal{D}_\alpha, \mu_\alpha)$, could appear preferable. However, in various nonmeasurable situations, the $\mathcal{D}_\alpha$ are very poor or even trivial. If the adopted definition only depends on $(\mathbb{D}, \mathcal{D}_\alpha, \mu_\alpha)$, it is often quite arbitrary whether $X_\alpha \to X$ in distribution or not. We illustrate this by a trivial example.

EXAMPLE 1. Suppose $\mathbb{D} = \{0, 1\}$, $X_\alpha = I_{B_\alpha}$ and $X = I_B$, where $B \in \mathcal{A}$ and $B_\alpha \notin \overline{\mathcal{A}_\alpha}$ for all α. Then $\mathcal{D}_\alpha = \{\varnothing, \mathbb{D}\}$ for all α. Note also that $X_\alpha \overset{\text{HJ}}{\to} X$ if and only if $\lim_\alpha P_\alpha^*(B_\alpha) = \lim_\alpha P_{\alpha *}(B_\alpha) = P(B)$ (P_α^* and $P_{\alpha *}$ denoting outer and inner measure, resp.).

In Example 1, according to any reasonable definition, $X_\alpha \to X$ in distribution should mean $\text{Prob}(X_\alpha = 1) \to P(X = 1)$ for suitable estimates $\text{Prob}(X_\alpha = 1)$ of the nonavailable probabilities of the events $\{X_\alpha = 1\}$. If a definition only depends on $(\mathbb{D}, \mathcal{D}_\alpha, \mu_\alpha)$, there are no sound rules for choosing $\text{Prob}(X_\alpha = 1)$ and, in this sense, convergence is quite arbitrary. Instead, a definition which makes use of the available information on $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$ can work. For instance, suppose that $X_\alpha \to X$ in distribution whenever $\text{Prob}(X_\alpha = 1) \to P(X = 1)$ *for all admissible estimates* $\text{Prob}(X_\alpha = 1)$, where "admissible" means

$$P_{\alpha *}(X_\alpha = 1) \leq \text{Prob}(X_\alpha = 1) \leq P_\alpha^*(X_\alpha = 1).$$

Then convergence amounts to $\lim_\alpha P_\alpha^*(X_\alpha = 1) = \lim_\alpha P_{\alpha *}(X_\alpha = 1) = P(X = 1)$, which is precisely convergence according to HJ.

To sum up, it seems more or less unavoidable that the $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$ are strongly involved in any definition which is asked to work in nonmeasurable problems.

3. Finitely additive probability measures. The finitely additive theory of probability, based on de Finetti's coherence principle and the subsequent work of Dubins and Savage, is well developed and understood by now. It includes the standard (countably additive) theory as a particular case. In this section, we collect for later use a few known facts on finitely additive probabilities and we state two preliminary lemmas.

Given any set $\mathcal{X}$, $\mathcal{P}(\mathcal{X})$ denotes the power set of $\mathcal{X}$. If $T : L \to \mathbb{R}$ is a functional on a linear space L of real functions on $\mathcal{X}$, we let

$$T^*(f) = \inf\{T(\phi) : \phi \in L, \phi \geq f\} \quad \text{and} \quad T_*(f) = -T^*(-f)$$

for each $f : \mathcal{X} \to \mathbb{R}$ (with the usual convention $\inf \varnothing = \infty$). A *finitely additive probability* (f.a.p.), defined on a field $\mathcal{E}$ of subsets of $\mathcal{X}$, is a nonnegative, finitely additive function ν on $\mathcal{E}$ such that $\nu(\mathcal{X}) = 1$. For $H \subset \mathcal{X}$, define $\nu^*(H) = \inf\{\nu(B) : B \in \mathcal{E}, B \supset H\}$ and $\nu_*(H) = 1 - \nu^*(H^c)$. Then, for all $H \subset \mathcal{X}$ and $c \in [\nu_*(H), \nu^*(H)]$, ν can be extended to an f.a.p. ν' on $\mathcal{P}(\mathcal{X})$ such that $\nu'(H) = c$. This follows from Lemma 2, which is also one of our main tools. It is essentially well known and we give a proof just to make the paper self-contained.

LEMMA 2. *Let L and M be linear spaces of real bounded functions on a set $\mathcal{X}$ and let $T : L \to \mathbb{R}$ and $U : M \to \mathbb{R}$ be linear functionals. Suppose L includes the constants, T is positive and $T(1) = 1$. Then there is an f.a.p. β on $\mathcal{P}(\mathcal{X})$ such that*

$$T(f) = \int f \, d\beta \quad \text{and} \quad U(g) = \int g \, d\beta \qquad \text{for all } f \in L \text{ and } g \in M$$

if and only if

$$U(g) \geq T_*(g) \qquad \text{for all } g \in M.$$

In particular, given any bounded function h on $\mathcal{X}$ and constant $c \in [T_*(h), T^*(h)]$, there is an f.a.p. β on $\mathcal{P}(\mathcal{X})$ such that $T(f) = \int f \, d\beta$ for all $f \in L$ and $\int h \, d\beta = c$ [take $M = \{ah : a \in \mathbb{R}\}$ and $U(ah) = ac$ for all $a \in \mathbb{R}$].

PROOF. If T and U are integrals w.r.t. the same f.a.p. β, then

$$U(g) = \int g \, d\beta \geq \sup\left\{ \int \phi \, d\beta : \phi \in L, \phi \leq g \right\} = T_*(g) \qquad \text{for all } g \in M.$$

Conversely, suppose $U \geq T_*$ on M. Since $U(g) \leq -T_*(-g) = T^*(g)$ for all $g \in M$, one has $T = U$ on $L \cap M$. Hence, if $f_1 + g_1 = f_2 + g_2$ for some $f_1, f_2 \in L$ and $g_1, g_2 \in M$, then $T(f_1 - f_2) = T(g_2 - g_1) = U(g_2 - g_1)$ and thus $T(f_1) + U(g_1) = T(f_2) + U(g_2)$. Therefore, it is possible to set

$$V(f + g) = T(f) + U(g) \qquad \text{for all } f \in L \text{ and } g \in M,$$

and V turns out to be a linear functional on $L + M$ extending both T and U. If $f \in L$, $g \in M$ and $f + g \geq 0$, then $g \geq -f$ and thus $U(g) \geq T_*(g) \geq -T(f)$.

It follows that $V(f+g) = T(f) + U(g) \geq 0$, that is, V is positive. By the Hahn–Banach theorem, V can be extended to a linear positive functional V_0 defined on all real bounded functions on $\mathcal{X}$. Let $\beta(E) = V_0(I_E)$ for all $E \subset \mathcal{X}$. Then $V_0(h) = \int h\, d\beta$ for all bounded h on $\mathcal{X}$, and this concludes the proof. $\square$

A well-established theory of weak convergence of f.a.p.'s has been available since Karandikar (1982, 1988); see also Girotto and Holzer (1993). Let $\mathbb{P}$ be the class of all f.a.p.'s on $\mathcal{D}$. Given $\nu_\alpha, \nu \in \mathbb{P}$, where α ranges over A, say that $\nu_\alpha \to \nu$ *weakly* in case $\int f\, d\nu_\alpha \to \int f\, d\nu$ for all $f \in C_b(\mathbb{D})$. If ν is *regular on open sets*, that is, $\nu(G) = \sup\{\nu(F) : F \text{ closed}, F \subset G\}$ for all open $G \subset \mathbb{D}$, then

$$\nu_\alpha \to \nu \text{ weakly} \quad \Longleftrightarrow \quad \limsup_\alpha \nu_\alpha(F) \leq \nu(F) \qquad \text{for all closed } F \subset \mathbb{D}.$$

The implication "$\Leftarrow$" holds even if ν is not regular on open sets. Further, each linear positive functional T on $\mathrm{C}_b(\mathbb{D})$ satisfying $T(1) = 1$ can be represented as $T(f) = \int f\, d\nu$, $f \in C_b(\mathbb{D})$, for some $\nu \in \mathbb{P}$ regular on open sets.

Finally, we give a technical lemma, needed for proving Theorem 4.

LEMMA 3. *Given $\nu \in \mathbb{P}$, suppose that $\nu_\alpha \to \nu$ weakly whenever $\nu_\alpha \in \mathbb{P}$, $\nu_\alpha = \mu_\alpha$ on $\mathcal{D}_\alpha$ and $\nu_\alpha^*(X_\alpha(\Omega_\alpha)) = 1$ for all α. Then $\gamma_\alpha \to \nu$ weakly whenever $\gamma_\alpha \in \mathbb{P}$ and $\gamma_\alpha = \mu_\alpha$ on $\mathcal{D}_\alpha$ [but not necessarily $\gamma_\alpha^*(X_\alpha(\Omega_\alpha)) = 1$] for all α.*

PROOF. Let $\gamma_\alpha \in \mathbb{P}$ be such that $\gamma_\alpha = \mu_\alpha$ on $\mathcal{D}_\alpha$ and let $f \in C_b(\mathbb{D})$. We have to show that $\int f\, d\gamma_\alpha \to \int f\, d\nu$. Denote $C_\alpha = X_\alpha(\Omega_\alpha)$. Since $\mu_\alpha^*(C_\alpha) = 1$, there is an f.a.p. λ_α on $\mathcal{P}(\mathbb{D})$ extending μ_α and satisfying $\lambda_\alpha(C_\alpha) = 1$. Define

$$L_\alpha = \{\phi + bI_{C_\alpha} : \phi \text{ bounded and } \mathcal{D}_\alpha\text{-measurable on } \mathbb{D}, b \in \mathbb{R}\}$$

and $T_\alpha(\psi) = \int \psi\, d\lambda_\alpha$ for all $\psi \in L_\alpha$. Fix $\psi = \phi + bI_{C_\alpha} \in L_\alpha$ such that $\psi \geq f$ and define

$$h = (\phi + b)I_{\{f \leq \phi + b\}} + (\sup f)I_{\{f > \phi + b\}}.$$

Then $\lambda_\alpha(\psi \neq h) = 0$, h is bounded and $h \geq f$. Also, h is $\mathcal{D}_\alpha$-measurable since it is Borel measurable and $h(X_\alpha) = \phi(X_\alpha) + b$ is $\overline{\mathcal{A}_\alpha}$-measurable. By recalling that γ_α extends μ_α, one obtains

$$T_\alpha(\psi) = T_\alpha(h) = \int h\, d\lambda_\alpha = \int h\, d\mu_\alpha = \int h\, d\gamma_\alpha \geq \int f\, d\gamma_\alpha.$$

Hence, $T_\alpha^*(f) \geq \int f\, d\gamma_\alpha$, which in turn implies $\int f\, d\gamma_\alpha \geq -T_\alpha^*(-f) = T_{\alpha *}(f)$. By Lemma 2, there is an f.a.p. β_α on $\mathcal{P}(\mathbb{D})$ such that $\int \psi\, d\beta_\alpha = T_\alpha(\psi) = \int \psi\, d\lambda_\alpha$ for all $\psi \in L_\alpha$ and $\int f\, d\beta_\alpha = \int f\, d\gamma_\alpha$. Call ν_α^f the restriction of β_α to $\mathcal{D}$. Then $\nu_\alpha^f = \mu_\alpha$ on $\mathcal{D}_\alpha$ and $(\nu_\alpha^f)^*(C_\alpha) = 1$, and thus the assumption of the lemma yields

$$\int f\, d\gamma_\alpha = \int f\, d\beta_\alpha = \int f\, d\nu_\alpha^f \to \int f\, d\nu. \qquad \square$$

4. A characterization and a finitely additive extension of HJ. To define the convergence in distribution of X_α to X, say $X_\alpha \xrightarrow{d} X$, it is tempting to declare that $X_\alpha \xrightarrow{d} X$ means $\nu_\alpha \to \mu$ weakly whenever, for each α, ν_α is a possible distribution for X_α. To realize this rough idea, clearly one has to decide what the distribution of a nonmeasurable random element is. In what follows, *a distribution for* X_α is any f.a.p. ν on $\mathcal{D}$ of the form $\nu = Q \circ X_\alpha^{-1}$, where Q is an f.a.p. on $\mathcal{P}(\Omega_\alpha)$ extending P_α. Hence, denoting by $\mathbb{P}_\alpha$ the class of all such distributions for X_α, the above tentative definition becomes

$$(1) \qquad X_\alpha \xrightarrow{d} X \quad \Longleftrightarrow \quad \nu_\alpha \to \mu \qquad \text{weakly whenever } \nu_\alpha \in \mathbb{P}_\alpha \text{ for all } \alpha.$$

In definition (1), using f.a.p.'s seems (to us) quite natural. The main reason is that the partial distribution μ_α of X_α can fail to admit any countably additive extension to $\mathcal{D}$. Instead, in a finitely additive setting, extensions are always available. In particular, it may be that $X_\alpha \xrightarrow{d} X$ [according to (1)] even though no X_α admits a countably additive distribution on $\mathcal{D}$. Situations of this type can occur in very simple problems, as shown in Example 10. Another reason for using f.a.p.'s, even if of the theoretical type, is that the dual of $C_b(\mathbb{D})$ is just a space of bounded finitely additive measures. Finally, as noted in Section 3, the finitely additive theory of probability is sufficiently developed by now.

One more remark is in order. A distribution for X_α has been defined above by first extending P_α to Q and then taking the image measure $Q \circ X_\alpha^{-1}$. Another reasonable possibility is to extend μ_α directly, in such a way that the extension is supported by $X_\alpha(\Omega_\alpha)$, but without passing through image measures. That is, a distribution for X_α could be defined as any element of $\mathbb{T}_\alpha := \{\nu \in \mathbb{P} : \nu = \mu_\alpha$ on $\mathcal{D}_\alpha$ and $\nu^*(X_\alpha(\Omega_\alpha)) = 1\}$. This alternative definition of distribution for X_α is not used in the rest of the paper. It is worth noting, however, that it would induce a very strong notion of convergence—stronger than (1), in particular. In fact, since $\mathbb{P}_\alpha \subset \mathbb{T}_\alpha$, if $\nu_\alpha \to \mu$ weakly whenever $\nu_\alpha \in \mathbb{T}_\alpha$ for all α then $X_\alpha \xrightarrow{d} X$, while the converse is not true (see Example 1).

Our main result (Theorem 4) is that HJ is equivalent to definition (1). Moreover, the above-mentioned condition that $\nu_\alpha \to \mu$ weakly whenever $\nu_\alpha \in \mathbb{T}_\alpha$ for all α amounts to a strengthened version of HJ. To introduce the latter, for all $\alpha \in A$ and bounded Borel $f : \mathbb{D} \to \mathbb{R}$ define

$$(2) \qquad \begin{aligned} E^0 f(X_\alpha) = \inf\{&E\phi(X_\alpha) : \phi : \mathbb{D} \to \mathbb{R} \text{ bounded and Borel}, \\ &\phi(X_\alpha) \text{ is } \overline{\mathcal{A}_\alpha}\text{-measurable}, \phi(X_\alpha) \geq f(X_\alpha)\}. \end{aligned}$$

E^0 is just another type of outer expectation. Since $E^0 f(X_\alpha) \geq E^* f(X_\alpha)$, the condition

$$(3) \qquad E^0 f(X_\alpha) \to Ef(X) \qquad \text{for all } f \in C_b(\mathbb{D})$$

implies $X_\alpha \xrightarrow{\mathrm{HJ}} X$. In this sense, (3) is a strengthening of HJ.

THEOREM 4. *$X_\alpha \overset{\text{HJ}}{\to} X$ if and only if $X_\alpha \overset{d}{\to} X$. Moreover, condition* (3) *holds if and only if $\nu_\alpha \to \mu$ weakly whenever $\nu_\alpha \in \mathbb{T}_\alpha$ for all α.*

PROOF. Suppose $X_\alpha \overset{\text{HJ}}{\to} X$ and fix $\nu_\alpha \in \mathbb{P}_\alpha$ for each α. Denoting by Q_α an f.a.p. on $\mathcal{P}(\Omega_\alpha)$, extending P_α and such that $\nu_\alpha = Q_\alpha \circ X_\alpha^{-1}$, one has

$$E^* f(X_\alpha) \geq \int f(X_\alpha)\, dQ_\alpha = \int f\, d\nu_\alpha \geq E_* f(X_\alpha) \qquad \text{for every } f \in C_b(\mathbb{D}).$$

Since $X_\alpha \overset{\text{HJ}}{\to} X$, $E^* f(X_\alpha)$ and $E_* f(X_\alpha)$ both converge to $\int f\, d\mu$, so that $\int f\, d\nu_\alpha \to \int f\, d\mu$ for every $f \in C_b(\mathbb{D})$. Hence, $\nu_\alpha \to \mu$ weakly, that is, $X_\alpha \overset{d}{\to} X$. Conversely, assume $X_\alpha \overset{d}{\to} X$ and fix $f \in C_b(\mathbb{D})$. For each α, define $T_\alpha(U) = \int U\, dP_\alpha$ on $L_\alpha = \{U : U \text{ bounded and } \mathcal{A}_\alpha\text{-measurable on } \Omega_\alpha\}$ and note that $T_\alpha^*(f(X_\alpha)) = E^* f(X_\alpha)$. By Lemma 2, there is an f.a.p. Q_α on $\mathcal{P}(\Omega_\alpha)$ extending P_α and such that $E^* f(X_\alpha) = \int f(X_\alpha)\, dQ_\alpha$. Define $\nu_\alpha^f = Q_\alpha \circ X_\alpha^{-1}$ on $\mathcal{D}$. Since $X_\alpha \overset{d}{\to} X$ and $\nu_\alpha^f \in \mathbb{P}_\alpha$ for each α, one obtains

$$E^* f(X_\alpha) = \int f(X_\alpha)\, dQ_\alpha = \int f\, d\nu_\alpha^f \to \int f\, d\mu.$$

Since $f \in C_b(\mathbb{D})$ is arbitrary, it follows that $X_\alpha \overset{\text{HJ}}{\to} X$.

Next, the second part of the theorem is just a consequence of the first and the following fact

$$(4) \qquad E^0 f(X_\alpha) = \inf\{E\phi(X_\alpha) : \phi : \mathbb{D} \to \mathbb{R} \text{ bounded and } \mathcal{D}_\alpha\text{-measurable}, \phi \geq f\}.$$

Suppose, in fact, that (4) holds. Let $Z_\alpha = Z = I$, where I denotes the identity on $\mathbb{D}$, and regard Z_α as defined on $(\mathbb{D}, \mathcal{D}_\alpha, \mu_\alpha)$ and Z as defined on $(\mathbb{D}, \mathcal{D}, \mu)$. By (4), $E^* f(Z_\alpha) = E^0 f(X_\alpha)$ for all $f \in C_b(\mathbb{D})$. Hence, the equivalence between HJ and (1) yields

$$\begin{aligned}
&E^0 f(X_\alpha) \to Ef(X) \qquad \text{for all } f \in C_b(\mathbb{D}) \\
&\iff \quad Z_\alpha \overset{\text{HJ}}{\to} Z \\
&\iff \quad Z_\alpha \overset{d}{\to} Z \\
&\iff \quad \gamma_\alpha \to \mu \qquad \text{weakly whenever } \gamma_\alpha \in \mathbb{P} \text{ and } \gamma_\alpha = \mu_\alpha \text{ on } \mathcal{D}_\alpha \text{ for all } \alpha.
\end{aligned}$$

Finally, by Lemma 3, the latter condition holds if and only if $\nu_\alpha \to \mu$ weakly whenever $\nu_\alpha \in \mathbb{T}_\alpha$ for all α.

It remains to prove (4). Fix $\alpha \in A$ and a bounded Borel $f : \mathbb{D} \to \mathbb{R}$. Since $\mu_\alpha^*(X_\alpha(\Omega_\alpha)) = 1$, there is an f.a.p. λ_α on $\mathcal{P}(\mathbb{D})$ extending μ_α and such that $\lambda_\alpha(X_\alpha(\Omega_\alpha)) = 1$. Let us call Φ and Φ_0 the sets of functions ϕ's involved in (2) and in (4), respectively. Since $\Phi_0 \subset \Phi$, to get (4), it is enough to prove that for each $\phi \in \Phi$ there is $\phi_0 \in \Phi_0$ with $E\phi_0(X_\alpha) = E\phi(X_\alpha)$. Given $\phi \in \Phi$, define ϕ_0

by $\phi_0 = \phi$ on $\{f \le \phi\}$ and $\phi_0 = \sup f$ on $\{f > \phi\}$. Then ϕ_0 is bounded, $\phi_0 \ge f$ everywhere and ϕ_0 is $\mathcal{D}_\alpha$-measurable since is Borel measurable and $\phi_0(X_\alpha) = \phi(X_\alpha)$ is $\overline{\mathcal{A}_\alpha}$-measurable. Hence, $\phi_0 \in \Phi_0$. Finally, since $\lambda_\alpha(\phi_0 \ne \phi) = 0$, one obtains

$$E\phi(X_\alpha) = \int \phi \, d\mu_\alpha = \int \phi \, d\lambda_\alpha = \int \phi_0 \, d\lambda_\alpha = \int \phi_0 \, d\mu_\alpha = E\phi_0(X_\alpha). \qquad \square$$

Theorem 4, in our opinion, makes the probabilistic meaning of HJ more transparent.

Let us now turn to another issue. According to HJ, the limit X of the nonmeasurable random elements X_α must be measurable. This fact is perhaps unsuitable from a theoretical point of view. Furthermore, as shown in Example 11, it can be a real limitation even in simple problems. Motivated by phenomena as in Example 11, we introduce an extension of HJ to the case of a nonmeasurable limit. Such an extension is naturally suggested by Theorem 4.

Suppose that

$$\liminf_\alpha E_* f(X_\alpha) \ge \limsup_\alpha E^* f(X_\alpha) \qquad \text{for all } f \in C_b(\mathbb{D}).$$

Then one can define the functional

$$T(f) = \lim_\alpha E_* f(X_\alpha) = \lim_\alpha E^* f(X_\alpha) \qquad \text{for all } f \in C_b(\mathbb{D}),$$

and it is straightforward to see that T is linear and positive with $T(1) = 1$. Hence, T is the integral w.r.t. some $\nu \in \mathbb{P}$, or, equivalently,

$$\int f \, d\nu = \lim_\alpha E^* f(X_\alpha) \qquad \text{for all } f \in C_b(\mathbb{D}). \tag{5}$$

Note that (5) also implies $\int f \, d\nu = \lim_\alpha E_* f(X_\alpha)$ for all $f \in C_b(\mathbb{D})$. Each such ν can be seen as a limit in distribution for X_α.

DEFINITION 5. We say that X_α converges in distribution to ν and we write $X_\alpha \to \nu$, where $\nu \in \mathbb{P}$, if condition (5) holds.

Definition 5 clearly extends HJ. In fact, if $X_\alpha \xrightarrow{\text{HJ}} X$ (X is measurable with distribution μ), then, in particular, $X_\alpha \to \mu$, and μ is the only countably additive element of $\mathbb{P}$ satisfying (5). Moreover, one deeper reason for regarding Definition 5 as an extension of HJ comes from Theorem 4 and the following result.

THEOREM 6. *Let $\nu \in \mathbb{P}$. Then $X_\alpha \to \nu$ if and only if*

$$\nu_\alpha \to \nu \qquad \textit{weakly whenever } \nu_\alpha \in \mathbb{P}_\alpha \textit{ for each } \alpha. \tag{6}$$

Moreover, for $X_\alpha \to \nu$, it is sufficient (necessary and sufficient if ν is regular on open sets) that

$$\limsup_\alpha P_\alpha^*(X_\alpha \in F) \le \nu(F) \qquad \textit{for all closed } F \subset \mathbb{D}. \tag{7}$$

PROOF. Suppose $X_\alpha \to \nu$. For each α, fix $\nu_\alpha \in \mathbb{P}_\alpha$ and take an f.a.p. Q_α on $\mathcal{P}(\Omega_\alpha)$ extending P_α and such that $\nu_\alpha = Q_\alpha \circ X_\alpha^{-1}$. Since

$$E^* f(X_\alpha) \geq \int f(X_\alpha)\, dQ_\alpha = \int f\, d\nu_\alpha \geq E_* f(X_\alpha) \qquad \text{for all } f \in C_b(\mathbb{D}),$$

condition (5) implies $\nu_\alpha \to \nu$ weakly. Conversely, suppose (6) holds and fix $f \in C_b(\mathbb{D})$. Arguing as in the proof of Theorem 4, for each α there is $\nu_\alpha^f \in \mathbb{P}_\alpha$ such that $\int f\, d\nu_\alpha^f = E^* f(X_\alpha)$, and thus (6) yields $\int f\, d\nu = \lim_\alpha \int f\, d\nu_\alpha^f = \lim_\alpha E^* f(X_\alpha)$. Hence, $X_\alpha \to \nu$, and this proves the first part of the theorem. Next, assume (7) holds and fix $\nu_\alpha \in \mathbb{P}_\alpha$ for all α. Since $\nu_\alpha(B) \leq P_\alpha^*(X_\alpha \in B)$ for all $B \in \mathcal{D}$, (7) implies $\limsup_\alpha \nu_\alpha(F) \leq \nu(F)$ for all closed $F \subset \mathbb{D}$. Hence, $\nu_\alpha \to \nu$ weakly, and the first part of the theorem yields $X_\alpha \to \nu$. Finally, suppose $X_\alpha \to \nu$ and ν is regular on open sets. Given a closed $F \subset \mathbb{D}$, for each α take $\nu_\alpha^F \in \mathbb{P}_\alpha$ such that $\nu_\alpha^F(F) = P_\alpha^*(X_\alpha \in F)$. Since $\nu_\alpha^F \to \nu$ weakly and ν is regular on open sets, one obtains

$$\limsup_\alpha P_\alpha^*(X_\alpha \in F) = \limsup_\alpha \nu_\alpha^F(F) \leq \nu(F). \qquad \square$$

The following proposition provides an additional criterion for deciding whether X_α converges in distribution to some limit.

PROPOSITION 7. *In order for $X_\alpha \to \nu$, for some $\nu \in \mathbb{P}$, it is necessary and sufficient that for any finite family $\{F_1, \dots, F_k\}$ of closed subsets of $\mathbb{D}$ there is $\gamma \in \mathbb{P}$ such that*

$$\limsup_\alpha P_\alpha^*(X_\alpha \in F_i) \leq \gamma(F_i) \qquad \textit{for each } i = 1, \dots, k.$$

PROOF. Let $\mathcal{C}$ be the class of closed subsets of $\mathbb{D}$. We first prove sufficiency. By Theorem 6, letting $\mathcal{R}_F = \{\gamma \in \mathbb{P} : \limsup_\alpha P_\alpha^*(X_\alpha \in F) \leq \gamma(F)\}$ for $F \in \mathcal{C}$, it is enough to prove $\bigcap_{F \in \mathcal{C}} \mathcal{R}_F \neq \varnothing$. Let $[0,1]^{\mathcal{D}}$ be the set of functions from $\mathcal{D}$ to $[0,1]$, equipped with product topology. Then $[0,1]^{\mathcal{D}}$ is compact, $\mathcal{R}_F$ is closed in $[0,1]^{\mathcal{D}}$, and, by assumption, $\{\mathcal{R}_F : F \in \mathcal{C}\}$ has the finite intersection property. Hence, $\bigcap_{F \in \mathcal{C}} \mathcal{R}_F \neq \varnothing$. Conversely, suppose $X_\alpha \to \nu$ for some $\nu \in \mathbb{P}$. Since $\int f\, d\nu = \int f\, d\gamma$, $f \in C_b(\mathbb{D})$, for some $\gamma \in \mathbb{P}$ regular on open sets (cf. Section 3), one also has $X_\alpha \to \gamma$. Hence, Theorem 6 implies $\limsup_\alpha P_\alpha^*(X_\alpha \in F) \leq \gamma(F)$ for all $F \in \mathcal{C}$. $\square$

So far, X has denoted a measurable random element. Suppose now that measurability is dropped and $X : \Omega \to \mathbb{D}$ is an arbitrary map. When is it possible to write $X_\alpha \xrightarrow{d} X$? In the spirit of this paper, we say that $X_\alpha \xrightarrow{d} X$ *if X_α converges in distribution*, *according to Definition* 5, *and one of the limits ν is a possible distribution for X*, that is, $\nu = Q \circ X^{-1}$ for some f.a.p. Q on $\mathcal{P}(\Omega)$ extending P. The last result in this section provides a characterization of this fact.

PROPOSITION 8. *Let $\nu \in \mathbb{P}$, $(\Omega, \mathcal{A}, P)$ a probability space and $X : \Omega \to \mathbb{D}$ any map. Then $\nu = Q \circ X^{-1}$ for some f.a.p. Q on $\mathcal{P}(\Omega)$ extending P if and only if*

$$\nu^*(X(\Omega)) = 1 \quad \textit{and} \quad \nu(B) \le P^*(X \in B) \qquad \textit{for all } B \in \mathcal{D}. \tag{8}$$

In particular, $X_\alpha \xrightarrow{d} X$ if and only if there is $\gamma \in \mathbb{P}$ such that

$$X_\alpha \to \gamma, \qquad \gamma^*(X(\Omega)) = 1, \qquad \gamma(B) \le P^*(X \in B) \qquad \textit{for all } B \in \mathcal{D}.$$

PROOF. Suppose (8) holds and let $C = X(\Omega)$. Since $\nu^*(C) = 1$, there is an f.a.p. λ on $\mathcal{P}(\mathbb{D})$ extending ν and such that $\lambda(C) = 1$. If $B_1, B_2 \in \mathcal{D}$ and $\{X \in B_1\} = \{X \in B_2\}$, then $B_1 \cap C = B_2 \cap C$, and thus $\nu(B_1) = \lambda(B_1 \cap C) = \lambda(B_2 \cap C) = \nu(B_2)$. Hence, one can define $P_0(X \in B) = \nu(B)$ for all $B \in \mathcal{D}$ obtaining an f.a.p. P_0 on $\sigma(X) := \{X^{-1}(B) : B \in \mathcal{D}\}$. Fix $H \in \sigma(X)$ and $G \in \mathcal{A}$ with $H \subset G$. Since $H = \{X \in B\}$ for some $B \in \mathcal{D}$, the second part of condition (8) implies

$$P_0(H) = P_0(X \in B) = \nu(B) \le P^*(X \in B) \le P(G).$$

By Theorem 3.6.1 of Bhaskara Rao and Bhaskara Rao (1983), there exists an f.a.p. Q on $\mathcal{P}(\Omega)$ extending both P and P_0. Thus, $\nu = Q \circ X^{-1}$, and this proves the "if" part. Finally, the "only if" part is trivial. $\square$

5. Empirical processes. In this section, we let $A = \mathbb{N}$, $D[0, 1]$ is the space of real cadlag functions on $[0, 1]$, and $l^\infty(S)$ is the space of real bounded functions on the set S. *Both $D[0, 1]$ and $l^\infty(S)$ are equipped with uniform distance*. Among other things, it is shown that the empirical process for an exchangeable sequence can fail to converge, due to the nonexistence of any measurable limit, although it converges for an i.i.d. sequence; see Example 11.

Let $\xi_1, \xi_2, \dots$ be i.i.d. random variables, defined on the probability space $(\Omega, \mathcal{A}, P)$ and taking values in the measurable space $(\mathcal{X}, \mathcal{B})$. Further, let $\mathcal{F}$ be a class of real measurable functions on $\mathcal{X}$ such that $E|f(\xi_1)| < \infty$ for all $f \in \mathcal{F}$ and $\sup_{f \in \mathcal{F}} |f(x) - Ef(\xi_1)| < \infty$ for all $x \in \mathcal{X}$. Then, for each n, the nth empirical process

$$X_n(f) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} f(\xi_i) - Ef(\xi_1)\right), \qquad f \in \mathcal{F}, \tag{9}$$

can be seen as a map $X_n : \Omega \to \mathbb{D} = l^\infty(\mathcal{F})$.

A measurable space $(S, \mathcal{S})$ is said to be *standard* if S is a Borel set in some Polish space and $\mathcal{S}$ the Borel σ-field on S. When $(\Omega, \mathcal{A})$ and $(\mathcal{X}, \mathcal{B})$ are standard spaces, one can conjecture that HJ is equivalent to its strengthened version given by (3). Actually, something more is true, provided $\mathcal{F}$ is rich enough to separate points in a suitable sense. In this case, in fact, $\mathbb{P}_n = \mathbb{T}_n$ for all n, and thus

Theorem 4 implies $X_n \xrightarrow{\text{HJ}} X$ if and only if (3) holds. Precisely, the condition on $\mathcal{F}$ is

$$\begin{aligned}&\text{for all } n, \text{ there is } c_n > 0 \text{ such that} \qquad \sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n f(u_i) - \sum_{i=1}^n f(v_i)\right| \geq c_n \\ &\qquad\qquad \text{whenever } (u_1,\dots,u_n) \text{ is not a permutation of } (v_1,\dots,v_n).\end{aligned} \tag{10}$$

Condition (10) holds, for instance, for various significant classes of indicators.

PROPOSITION 9. *Suppose* $(\Omega,\mathcal{A})$ *and* $(\mathcal{X},\mathcal{B})$ *are standard spaces,* $\mathcal{F}$ *meets* (10), $\mathbb{D} = l^\infty(\mathcal{F})$ *and* X_n *is given by* (9). *Then* $\mathbb{P}_n = \mathbb{T}_n$.

PROOF. Let $\phi = (\xi_1,\dots,\xi_n)$, $C \in \mathcal{A}$ and Π the set of permutations π of $\{1,\dots,n\}$. By (10), $X_n(C)$ is closed in $l^\infty(\mathcal{F})$ and thus $X_n(C) \in \mathcal{D}$. Also, (10) implies

$$X_n^{-1}(X_n(C)) = \bigcup_{\pi\in\Pi}\{(\xi_{\pi(1)},\dots,\xi_{\pi(n)}) \in \phi(C)\}.$$

Since $(\Omega,\mathcal{A})$ and $(\mathcal{X},\mathcal{B})$ are standard spaces, $\phi(C)$ is analytic in $\mathcal{X}^n$, so that $X_n^{-1}(X_n(C))$ is analytic in Ω, which implies $X_n^{-1}(X_n(C)) \in \overline{\mathcal{A}}$. Thus, $X_n(C) \in \mathcal{D}_n$ for all $C \in \mathcal{A}$. Now, fix $\nu \in \mathbb{T}_n$ and $B \in \mathcal{D}$. If $H \in \mathcal{A}$ and $H \supset \{X_n \in B\}$, then

$$\nu(B^c) \geq \nu(X_n(H^c)) = \mu_n(X_n(H^c)) \geq P(H^c),$$

where the equality is due to $X_n(H^c) \in \mathcal{D}_n$ and $\nu = \mu_n$ on $\mathcal{D}_n$. Hence, $\nu(B) \leq P(H)$, and this implies $\nu(B) \leq P^*(X_n \in B)$. By Proposition 8, it follows that $\nu \in \mathbb{P}_n$, and since $\mathbb{P}_n \subset \mathbb{T}_n$ this concludes the proof. $\square$

Note that, provided $(\mathcal{X},\mathcal{B})$ is a standard space and (10) holds, Proposition 9 applies, in particular, when $\{\xi_n\}$ is the sequence of coordinate projections on $(\Omega,\mathcal{A}) = (\mathcal{X}^\infty,\mathcal{B}^\infty)$. Note also that, as far as X_n is defined through (9), Proposition 9 holds under any distributional assumption on $\{\xi_n\}$ (and not only if $\{\xi_n\}$ is i.i.d.).

We next give two examples. The first one shows that, as noted in Section 4, it may be that $X_\alpha \xrightarrow{d} X$ while no X_α admits a countably additive distribution on $\mathcal{D}$.

EXAMPLE 10. Let X_n be given by (9), where $\mathcal{X} = [0,1]$, $\mathcal{B}$ is the Borel σ-field on $[0,1]$, $\mathcal{F} = \{I_{[0,t]} : t \in [0,1]\}$ and $\{\xi_n\}$ is i.i.d. with $P(\xi_1 = t) = 0$ for all t. Clearly, by identifying $X_n(t)$ with $X_n(I_{[0,t]})$, X_n can be seen as a map $X_n : \Omega \to \mathbb{D} = D[0,1]$. It is well known that $X_n \xrightarrow{\text{HJ}} X$, where X is a (scaled) Brownian bridge process, so that $X_n \xrightarrow{d} X$ by Theorem 4. Note also that $\mathcal{F}$ meets (10), and thus, if $(\Omega,\mathcal{A})$ is a standard space, Proposition 9 yields $\nu_n \to \mu$

weakly whenever $\nu_n \in \mathbb{T}_n$ for all n. As a contradiction, suppose now that the partial distribution μ_n of X_n admits a countably additive extension to $\mathcal{D}$, say λ_n. Let

$$J = \{x \in D[0,1] : x \text{ has precisely } n \text{ jumps}\},$$

$$U_B = \left\{x \in D[0,1] : \exists\, 0 < t_1 < \cdots < t_n \le 1 \text{ with } \sum_{i=1}^{n} t_i \in B \text{ and } x(t_i) \neq x(t_i-) \text{ for all } i\right\}$$

where $B \subset (0,n]$.

Since U_B is open, one can define $\gamma_n(B) = \lambda_n(U_B)$ for all $B \subset (0,n]$. On noting that $U_B \cap U_C \cap J = \varnothing$ if $B \cap C = \varnothing$ and

$$\mu_n(J) = \overline{P}(X_n \in J) = P(\xi_1, \ldots, \xi_n \text{ all distinct and strictly positive}) = 1,$$

it follows that γ_n is a countably additive probability on $\mathcal{P}((0,n])$. Further, $\gamma_n(B) = P(\sum_{i=1}^n \xi_i \in B)$ whenever B is a Borel set, and thus $\gamma_n\{v\} = 0$ for all $v \in (0,n]$. Therefore, the existence of a countably additive extension of μ_n to $\mathcal{D}$ implies the existence of a countably additive probability γ_n on $\mathcal{P}((0,n])$ vanishing on singletons. But the latter fact is impossible in various models of the usual ZFC set theory. For instance, by Ulam's theorem, it is impossible under the continuum hypothesis.

Finally, let us turn to the second example, which is one of the basic motivations for extending HJ to the case of a nonmeasurable limit, as done in Section 4.

A *random distribution function* is a process, indexed by $\mathbb{R}$, whose paths are distribution functions. In the rest of the paper, X^F denotes a process, on some probability space $(\Omega_0, \mathcal{A}_0, P_0)$, of the form

$$X_t^F = W_{F(t)}^0, \qquad t \in [0,1],$$

where W^0 is a Brownian bridge process and F a random distribution function independent of W^0. Let $\mathbb{D} = D[0,1]$ and let $\mathcal{D}_b$ denote the σ-field on $D[0,1]$ generated by the balls. Then $X^F : \Omega_0 \to D[0,1]$ is measurable w.r.t. $\mathcal{D}_b$ but can fail to be measurable w.r.t. $\mathcal{D}$. Setting $C(I) = \{x \in D[0,1] : x(t) = x(t-) \text{ for all } t \in I\}$, where $I \subset (0,1]$, a set $B \subset D[0,1]$ is separable if and only if $B \subset C(I)$ for some I such that $(0,1] - I$ is countable. Thus, X^F has separable range whenever the F-paths have jumps within a *fixed* countable set. In this case, X^F is measurable and tight (i.e., its probability distribution is tight). In particular, X^H is measurable for any fixed (i.e., nonrandom) distribution function H. Let $\mu(\cdot, H) = P_0(X^H \in \cdot)$ be the probability distribution of X^H. For later purposes, note that $\mu(C(I), H) = 0$ if $0 < H(t) - H(t-) < 1$ for some $t \in I$. We are now able to give the example.

EXAMPLE 11. Let ξ_n be the nth coordinate projection on $([0,1]^\infty, \mathcal{B}^\infty, P)$, where $\mathcal{B}$ is the Borel σ-field on $[0,1]$, and suppose that $\{\xi_n\}$ is exchangeable under P. By de Finetti's theorem,

$$P(B) = \int P_u(B) P(du) \qquad \text{for all } B \in \mathcal{B}^\infty,$$

where, for all $u \in [0,1]^\infty$, P_u is a law on $\mathcal{B}^\infty$ such that $\{\xi_n\}$ is i.i.d. under P_u. Letting $G(u,t) = P_u(\xi_1 \le t)$ [and using the standard notation $G(t) = G(\cdot, t)$ and $G(u) = G(u, \cdot)$], the nth empirical process can be defined as

$$X_n(t) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n I_{\{\xi_i \le t\}} - G(t)\right), \qquad t \in [0,1].$$

Let the process X^F be defined on $(\Omega_0, \mathcal{A}_0, P_0)$, where the random distribution function F is distributed as G. Then the probability distribution of X^F on $\mathcal{D}_b$, say μ_b, can be expressed as

$$\mu_b(B) = P_0(X^F \in B) = \int \mu(B, G(u)) P(du) \qquad \text{for all } B \in \mathcal{D}_b.$$

By the classical result for i.i.d. sequences, for almost all u, one has $X_n \overset{\text{HJ}}{\to} X^{G(u)}$ under P_u. Hence, for each bounded continuous $\mathcal{D}_b$-measurable function f on $D[0,1]$, one obtains

$$\begin{aligned}\lim_n Ef(X_n) &= \lim_n \int E_u f(X_n) P(du) = \int \lim_n E_u f(X_n) P(du)\\ &= \iint f(x)\mu(dx, G(u)) P(du) = E_0(f(X^F)),\end{aligned} \tag{11}$$

where E_u and E_0 denote expectation w.r.t. P_u and P_0. It follows that $X_n \overset{\text{HJ}}{\to} X^F$ provided $(\Omega_0, \mathcal{A}_0, P_0)$ and X^F can be taken such that X^F is measurable and tight. In particular, this happens if ξ_1 has a discrete distribution or if $P(\xi_1 = \xi_2) = 0$. See Theorem 1.7.2 of van der Vaart and Wellner (1996) and Theorem 4.5 of Berti, Pratelli and Rigo (2004).

Now, let us take P such that $\{\xi_n\}$ is exchangeable, $P(\xi_1 = t) = 0$ for all t and

$$G(u,t) = P_u(\xi_1 \le t) = \tfrac{1}{2}\big(t + I_{[0,t]}(\xi_1(u))\big) \qquad \text{for } t \in [0,1] \text{ and } u \in [0,1]^\infty.$$

If $B \in \mathcal{D}$ is separable, then $B \in \mathcal{D}_b$ and $B \subset C(I)$ for some $I \subset (0,1]$ such that $(0,1] - I$ is countable. Since $G(u)$ has a jump of size $\frac{1}{2}$ at $\xi_1(u)$ if $\xi_1(u) > 0$, one has $\mu(B, G(u)) \le \mu(C(I), G(u)) = 0$ if $\xi_1(u) \in I$, and thus

$$\mu_b(B) = \int_{\{\xi_1 \in I\}} \mu(B, G(u)) P(du) = 0.$$

If μ_b admits a countably additive extension to $\mathcal{D}$, then, since μ_b vanishes on separable Borel sets, $D[0,1]$ has *measurable cardinality*; see Dudley (1999), page 403. But this is impossible in various models of ZFC. For instance, since $\operatorname{card} D[0,1] = \operatorname{card} \mathbb{R}$, it is impossible under the continuum hypothesis. As a consequence, *no measurable random element* X *of* $D[0,1]$, *defined on any probability space*, *satisfies* $X_n \overset{\text{HJ}}{\to} X$. In view of (11), in fact, the distribution of such X would be a countably additive probability on $\mathcal{D}$ which extends μ_b.

In Example 11, X_n can fail to converge according to HJ, due to the nonexistence of any measurable limit, even if $X_n \overset{\text{HJ}}{\to} X^{G(u)}$ under P_u for almost all u. An obvious question is whether $X_n \to \nu$, for some $\nu \in \mathbb{P}$, according to Definition 5. Our conjecture is that the answer is affirmative, and one reason is the following.

Let L be the space of bounded continuous $\mathcal{D}_b$-measurable functions on $D[0,1]$, and for $f \in L$ let $T(f) = E_0(f(X^F)) = \int f\, d\mu_b$. If $B \subset D[0,1]$ and $f \le I_B \le g$ for some $f, g \in L$, then (11) yields

$$T(f) = \lim_n Ef(X_n) \le \limsup_n P^*(X_n \in B) \le \lim_n Eg(X_n) = T(g),$$

so that $T_*(I_B) \le \limsup_n P^*(X_n \in B) \le T^*(I_B)$. This inequality gives some hope of successfully applying Proposition 7. Suppose, in fact, that, for given closed sets $F_1, \dots, F_k$, there is a linear functional U on the linear span M of $I_{F_1}, \dots, I_{F_k}$ such that

$$\begin{aligned} U(g) &\ge T_*(g) && \text{for all } g \in M, \\ \limsup_n P^*(X_n \in F_i) &\le U(I_{F_i}) && \text{for } i = 1, \dots, k. \end{aligned}$$

Then Lemma 2 grants the existence of $\gamma \in \mathbb{P}$ satisfying

$$T(f) = \int f\, d\gamma \qquad \text{for all } f \in L$$

and

$$U(I_{F_i}) = \gamma(F_i) \qquad \text{for } i = 1, \dots, k,$$

and in turn Proposition 7 implies $X_n \to \nu$ for some $\nu \in \mathbb{P}$. The point, thus, is the existence of a certain linear functional U on M. For $k = 1$, such U surely exists. We suspect it exists for any k, but we do not have a proof.

REFERENCES

BERTI, P. and RIGO, P. (1994). Coherent inferences and improper priors. *Ann. Statist.* **22** 1177–1194.

BERTI, P. and RIGO, P. (1996). On the existence of inferences which are consistent with a given model. *Ann. Statist.* **24** 1235–1249.

BERTI, P. and RIGO, P. (1999). Sufficient conditions for the existence of disintegrations. *J. Theoret. Probab.* **12** 75–86.

BERTI, P., PRATELLI, L. and RIGO, P. (2004). Limit theorems for a class of identically distributed random variables. *Ann. Probab.* To appear.

BHASKARA RAO, K. P. S. and BHASKARA RAO, M. (1983). *Theory of Charges*. Academic, New York.

BUMBY, R. and ELLENTUCK, E. (1969). Finitely additive measures and the first digit problem. *Fund. Math.* **65** 33–42.

DUBINS, L. E. (1975). Finitely additive conditional probabilities, conglomerability and disintegrations. *Ann. Probab.* **3** 89–99.

DUBINS, L. E. and PRIKRY, K. (1995). On the existence of disintegrations. *Séminaire de Probabilités XXIX*. *Lecture Notes in Math.* **1613** 248–259. Springer, Berlin.

DUDLEY, R. (1966). Weak convergence of measures on non separable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

DUDLEY, R. (1967). Measures on nonseparable metric spaces. *Illinois J. Math.* **11** 449–453.

DUDLEY, R. (1999). *Uniform Central Limit Theorems*. Cambridge Univ. Press.

GIROTTO, B. and HOLZER, S. (1993). Weak convergence of masses on normal topological spaces. *Sankhyā Ser. A* **55** 188–201.

HEATH, D. and SUDDERTH, W. D. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.* **6** 333–345.

HEATH, D. and SUDDERTH, W. D. (1989). Coherent inference from improper priors and from finitely additive priors. *Ann. Statist.* **17** 907–919.

KALLIANPUR, G. and KARANDIKAR, R. L. (1983). A finitely additive white noise approach to non linear filtering. *Appl. Math. Optim.* **10** 159–185.

KALLIANPUR, G. and KARANDIKAR, R. L. (1988). *White Noise Theory of Prediction*, *Filtering and Smoothing*. Gordon and Breach, London.

KARANDIKAR, R. L. (1982). A general principle for limit theorems in finitely additive probability. *Trans. Amer. Math. Soc.* **273** 541–550.

KARANDIKAR, R. L. (1988). A general principle for limit theorems in finitely additive probability: The dependent case. *J. Multivariate Anal.* **24** 189–206.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

PYKE, R. and SHORACK, G. (1968). Weak convergence of a two sample empirical process and a new approach to Chernoff–Savage theorems. *Ann. Math. Statist.* **39** 755–771.

REGAZZINI, E. (1987). De Finetti's coherence and statistical inference. *Ann. Statist.* **15** 845–864.

VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

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@unimo.it

DIPARTIMENTO DI ECONOMIA
POLITICA E METODI QUANTITATIVI
UNIVERSITA' DI PAVIA
VIA S. FELICE 5
27100 PAVIA
ITALY
E-MAIL: prigo@eco.unipv.it