Z. Wahrscheinlichkeitstheorie verw. Gebiete
70, 473-507 (1985)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Almost Exchangeable Sequences of Random Variables*

Istvan Berkes and Haskell P. Rosenthal**

Hungarian Academy of Sciences, Department of Mathematics, Budapest, Hungary
The University of Texas, Department of Mathematics, Austin, TX 78712, USA

Introduction

A sequence of (real valued) random variables (X_n) defined on some probability space is said to be almost exchangeable if it is a small perturbation of an exchangeable sequence after a suitable enlargement of the probability space. That is, if, after enlarging the space, there exists an exchangeable sequence (Y_n) with $\Sigma|X_n - Y_n| < \infty$ almost surely. We give here necessary and sufficient intrinsic conditions for a sequence of random variables to have an almost exchangeable subsequence (Theorem 2.4). We also give several examples of sequences of random variables bounded in probability with no almost exchangeable subsequences (see the end of Sect. 3).

As special cases of our results, we obtain that a sequence of random variables has an almost i.i.d. subsequence if and only if there is a distribution μ and a subsequence whose distributions relative to any set of positive measure converge to μ. It follows for example that any subsequence of $(\sin 2\pi n x)$ has a further almost i.i.d. subsequence. We also obtain that a sequence of random variables has an almost exchangeable subsequence provided it has an atomic tail field.

The above results were submitted for publication in 1976 in an earlier version of this paper. Our theorems are closely related to the results of Aldous' paper [1] obtained at the same time. In his paper, Aldous showed that every tight sequence has a subsequence with distributional exchangeability properties and used this fact to verify the so-called "subsequence principle" formulated by Chatterji. He also established the special cases of our results mentioned above, and raised the question of when the distributional exchangeability properties of subsequences can be improved to a pointwise approximation with exchangeable sequences (called "property B" in his terminology). Our main result (Theorem 2.4) answers Aldous' question and our examples in Sect. 3 give

* Part of this research was done at the University of Illinois at Urbana-Champaign
** The research of the second named author was partially supported by NSF Grant numbers MPS 71-02839-A04, MCS-8002393 and MCS-8303534

additional information concerning the counterexample in [1], p. 81. We mention that in [4] the best possible exchangeability condition valid for subsequences of general tight sequences is determined, together with applications to the subsequence principle.

While our results are closely connected with subsequence behavior, much of our work was originally motivated by Banach space theory, namely by the discoveries of Dacunha-Castelle and Krivine [6] related to the famous problem: does every infinite dimensional reflexive subspace X of L^1 contain a subspace isomorphic to l^p for some $1<p\leqq 2$? Recently Aldous in a profound study, showed that the answer to this question is affirmative [2]. By the results in [6], this shows that every such X has a sequence of random variables satisfying the conditions of our Corollary 3.5. (Incidentally, 3.5 gives another way of looking at the remarkable discovery in [6]: if a weakly null sequence in L^1 has l^p isometrically as a spreading model, it has a subsequence almost isometrically equivalent to the l^p basis.) For generalizations of the discovery of Aldous, see the recent work of J.L. Krivine and B. Maurey [11]; see also [7] for another exposition. Also see [16] and [17] for further connections between probability theory and the Banach space structure of subspaces of L^p.

To formulate the conditions for a sequence to have an almost exchangeable subsequence, we introduce the following notion: a sequence of random variables is said to be *determining* if the sequence of its distributions relative to any set of positive probability converges completely.[1] We show in Theorem 2.2 that if the distributions of a sequence are tight, then the sequence has a determining subsequence. Determining sequences of randomized stopping times are studied by Baxter and Chacon in [3]. Some work of Dacunha-Castelle and Krivine [5] may be reformulated as asserting that to a determining sequence (X_n) there corresponds a unique (up to distribution) exchangeable sequence (Y_n) so that (X_n) has a subsequence which behaves like (Y_n) at infinity; in particular the subsequence is "exchangeable at infinity."

Let (X_n) be a determining sequence, let μ be the limit distribution of the distributions of the X_n's in the sense of complete convergence, and let $T=\{t\in\mathbb{R}\colon \mu(\{t\})=0\}$. For each $t\in T$ it follows that $(I_{[X_n\leqq t]})$ converges weakly to a random variable denoted $F(t)$; we define the *limit tail field* of (X_n) to be the σ-field $\mathfrak{A}$ generated by the random variables $F(t)$, $t\in T$. ($t\to F(t)$ is defined as the *limit conditional distribution function* of (X_n).) Our main result, Theorem 2.4, then asserts that a sequence has an almost exchangeable subsequence if and only if it has a determining subsequence (X_n) satisfying the following condition:

(*) $\mathscr{P}([X_n\leqq t]\cap S)$ *converges strongly for every* $t\in T$ *and measurable* S, *where* $\mathscr{P}$ *denotes conditional probability with respect to the limit tail field* $\mathfrak{A}$.

(At the end of Sect. 3, we give an example of a determining sequence (X_n) with no almost exchangeable subsequence so that $\mathscr{P}([X_n\leqq t])$ converges

[1] To avoid ambiguity, ordinary (weak) convergence of probability distributions will be called complete convergence, in order to distinguish it from weak convergence of random variables and weak convergence of conditional distributions (cf. page 10)

strongly for every $t \in T$, thus answering in the negative an open question raised in an earlier version of the paper.) Let us say that a determining sequence (X_n) is *strongly conditionally convergent in distribution* (s.c.c.d.) provided it satisfies (*). It follows by standard arguments (see Lemma 2.12 and the preceding remarks) that an almost exchangeable sequence is s.c.c.d. The essential new ingredients are contained in our proof that an s.c.c.d. sequence has an almost exchangeable subsequence. We accomplish this by first establishing two results (Lemmas 2.13 and 2.14) which yield that if the conditional distribution of a random variable Y is close to that of a variable X (in probability), then if the probability space is large enough, there is a random variable Z with the same conditional distribution as X, with Z close to Y itself in probability. The proof is then completed at the end of Sect. 2, using several preliminary results developed in the first two sections. (The version of Lemma 2.14 for convergence of ordinary distributions lies deeper than the usual equivalents for convergence in distribution, and can be alternatively deduced from results of Strassen [19].)

In Sect. 1 we develop the needed machinery for conditional distributions. We do not use regular conditional probabilities, but rather regard a conditional distribution as a vector-valued measure defined on the Borel subsets of $\mathbb{R}$, the real line. We present a streamlined proof of a theorem of de Finetti, as improved by Dacunha-Castelle and Krivine [5], in Theorem 1.1. In Theorem 1.3 we prove Maharam's Lemma [15]; this result shows that if $(\Omega, \mathcal{S}, P)$ is a probability space and $\mathcal{S}$ is atomless over $\mathcal{A}$, a sub-σ-algebra of $\mathcal{S}$ then any $\mathcal{A}$-measurable function h with $0 \leqq h \leqq 1$ is equal to the conditional probability of some $S \in \mathcal{S}$. In Theorem 1.5 we show that if $\mathcal{S}$ is atomless over $\mathcal{A}$, then there exists a random variable on Ω with a prescribed conditional distribution (relative to $\mathcal{A}$). We then introduce the notions of weak and strong convergence of sequences of conditional distributions, and prove the compactness result (Theorem 1.7): every tight sequence of conditional distributions has a weakly convergent subsequence. In the first part of Sect. 2 we present the concepts of determining sequences of random variables, limit tail field, and limit conditional distribution. We then draw some simple consequences of our main result in Corollaries 2.6, 2.10 and 2.11 before passing to the proof of the main result outlined above. For example 2.11 yields that a sequence (X_j) has an almost exchangeable subsequence provided it is conditionally identically distributed with respect to its tail field.

Section 3 consists of complements to the results in the previous section. After treating the case of almost i.i.d. sequences, we consider the case of sequences of random variables almost exchangeable after a change of density. We show in Lemma 3.3 that a determining sequence remains so after a change of density, and determine the form of the limit conditional characteristic function of the sequence after the density change. We give in Theorem 3.4 a necessary and sufficient condition for a sequence to have an almost i.i.d. subsequence after a change of density, and then deduce the result concerning "l^p-sequences in L^1" in Corollary 3.5. The next three results were discovered after our original version of this work was completed. Lemma 3.6 shows that exchangeable sequences remain so after a tail-measurable density change;

Lemma 3.7 shows that if (X_j) is an integrable sequence of random variables with $|X_j| \to 1$ weakly so that (X_j) is almost exchangeable after a change of density, then (X_j) is already almost exchangeable. Theorem 3.8 solves the problem of determining the appropriate density change as follows: suppose (X_j) is a uniformly integrable determining sequence of random variables with $\int |X_j| \, dP = 1$ for all j. Then if (X_j) has a subsequence almost exchangeable after a density change, it has one almost exchangeable after the change of density φ, where $|X_j|$ tends to φ weakly as j tends to infinity.

The motivation for studying this question and indeed almost exchangeability in general derives from Banach space theory. Suppose that (X_j) is a normalized weakly null sequence in L^1 so that (X_j) is almost exchangeable after a change of density. Then (X_j) is a small norm-perturbation of an isometrically symmetric sequence in the Banach-space sense, and hence (X_j) is itself a symmetric basic sequence (see [14] for definitions and related results; also see [2], [10] and [18] for related results).

Section 3 concludes with three examples of determining sequences with no almost exchangeable subsequence. (Aldous also gives an example of such a sequence in [1]). The first is actually i.i.d. *after* a change of density, while the second consists of a sequence of indicator functions with no subsequence almost exchangeable after a change of density. The third example, new with this version of our results and considerably more elaborate than the previous two, produces a sequence of indicator functions which is conditionally indentically distributed with respect to its limit tail field (but definitely *not* conditionally identically distributed with respect to its tail field in view of Corollary 2.11). We note finally some complements of our work as yet unpublished. W. Henson has proved that any uncountable family of integrable random variables contains an infinite sequence which is almost exchangeable. The second-named author has shown that if X is a subspace of L^1 spanned by an exchangeable sequence and isomorphic to Hilbert space, then every norm-bounded sequence in X has an almost exchangeable subsequence.

§ 1. Conditional Distributions

The purpose of the present section is to give some basic definitions and to prove preliminary results for conditional distributions which will be needed in the proof of our main results in Sect. 2.

Let $(\Omega, \mathfrak{A}, Q)$ be a probability space. Slightly changing the standard definition, an n-dimensional conditional distribution with respect to Q (i.e., with respect to $(\Omega, \mathfrak{A}, Q)$) will be meant in the sequel as a map $\vec{\mu}$ from $\mathfrak{B}(R^n)$ (the Borel subsets of R^n) to $L^1(\Omega, \mathfrak{A}, Q)$ such that $\vec{\mu}(R^n) = 1$ a.s., $\vec{\mu}(B) \geqq 0$ a.s. for any $B \in \mathfrak{B}(R^n)$ and $\vec{\mu}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \vec{\mu}(B_i)$ a.s. for any disjoint $B_1, B_2, \ldots$ in $\mathfrak{B}(R^n)$. Thus, we shall not assume regularity which will not have any importance for our purposes. Given an n-dimensional conditional distribution $\vec{\mu}$, we define its associated (ordinary) distribution μ by $\mu(B) = E\vec{\mu}(B) = \int_{\Omega} \vec{\mu}(B) \, dQ$. We refer to a 1-dimensional (conditional) distribution simply as a (conditional) distribution.

From now on, let $(\Omega, \mathscr{S}, P)$ denote a fixed probability space. Given $\mathfrak{A}$, a σ-subfield of $\mathscr{S}$, let $\mathscr{E}_{\mathfrak{A}}$ and $\mathscr{P}_{\mathfrak{A}}$ denote conditional expectation resp. conditional probability with respect to $\mathfrak{A}$. For random variables $X_1, \ldots, X_n$ on Ω, we define $\vec{\mu} = c \cdot (\mathfrak{A})$ $\operatorname{dist}(X_1, \ldots, X_n)$ by $\vec{\mu}(B) = \mathscr{P}_{\mathfrak{A}}[(X_1, \ldots, X_n) \in B]$ for all $B \in \mathfrak{B}(\mathbb{R}^n)$. Obviously, $\vec{\mu}$ is a conditional distribution with respect to $Q = P|\mathfrak{A}$ (i.e., with respect to $(\Omega, \mathfrak{A}, P|\mathfrak{A})$). If $\vec{\mu} = c \cdot (\mathfrak{A})$ $\operatorname{dist}(X_1, \ldots, X_n)$ and μ is the associated distribution of $\vec{\mu}$ then $\mu = \operatorname{dist}$ $(X_1, \ldots, X_n)$, the ordinary distribution of $(X_1, \ldots, X_n)$. If $\mathfrak{A}$ is understood, we denote $c \cdot (\mathfrak{A})$ $\operatorname{dist}(X_1, \ldots, X_n)$ by $c \cdot \operatorname{dist}(X_1, \ldots, X_n)$ and also drop the index $\mathfrak{A}$ from $\mathscr{E}_{\mathfrak{A}}$ and $\mathscr{P}_{\mathfrak{A}}$. We let $\mathscr{S}^+$ denote the family of $S \in \mathscr{S}$ with $P(S) > 0$. Given $S \in \mathscr{S}^+$, we define the probability $P|S$ on $\mathscr{S}$ by

$$(P|S)(D) = P(S \cap D)/P(S) \qquad \text{for all } D \in \mathscr{S}.$$

Given an n-dimensional distribution $\vec{\mu}$ with respect to $(\Omega, \mathfrak{A}, Q)$ and $A \in \mathfrak{A}^+$, we define $R(A)\vec{\mu}$ to be $i \circ \vec{\mu}$ where i: $L^1(Q) \to L^1(Q|A)$ is the natural identity injection. In other words, $(R(A)\vec{\mu})(B) = I_A \vec{\mu}(B)$ where I_A is the indicator function of A. Clearly, $R(A)\vec{\mu}$ is an n-dimensional conditional distribution with respect to $Q|A$. Finally, given a σ-subfield $\mathfrak{A} \subset \mathscr{S}$, $S \in \mathscr{S}^+$ and random variables $X_1, \ldots, X_n$ on Ω, we set $A = \operatorname{supp} \mathscr{P}_{\mathfrak{A}}(S)$ $(= [\mathscr{P}_{\mathfrak{A}}(S) \neq 0])$ and define $\vec{v} = c \cdot (\mathfrak{A})$ $\operatorname{dist}(X_1, \ldots, X_n)|S$ as the n-dimensional conditional distribution with respect to $P|\mathfrak{A}|A$ given by

$$(c \cdot (\mathfrak{A}) \operatorname{dist}(X_1, \ldots, X_n)|S)(B) = \frac{\mathscr{P}_{\mathfrak{A}}([(X_1, \ldots, X_n) \in B] \cap S)}{\mathscr{P}_{\mathfrak{A}}(S)} \qquad \text{for all } B \in \mathfrak{B}(\mathbb{R}^n).$$

If $\mathfrak{A}$ is understood, we again drop the indices $\mathfrak{A}$ on the right-hand side and write $c \cdot \operatorname{dist}(X_1, \ldots, X_n)|S$ on the left-hand side. We call $\vec{v}$ the conditional distribution of $(X_1, \ldots, X_n)$ relative to S (with respect to $\mathfrak{A}$). Since $\mathscr{P}_{\mathfrak{A}}(S) > 0$ a.s. with respect to $P|A$, $1/\mathscr{P}_{\mathfrak{A}}(S)$ in the above formula is well defined; it is also worth noting that $I_S \leqq I_A$ a.s. where $A = \operatorname{supp} \mathscr{P}_{\mathfrak{A}}(S)$.

Given sequences (X_j) and (Y_j) of random variables, each defined on a fixed but possibly different probability space, we say that $\operatorname{dist}(X_j) = \operatorname{dist}(Y_j)$ if dist $(X_1, \ldots, X_n) = \operatorname{dist}$ $(Y_1, \ldots, Y_n)$ for all n.

We give first a streamlined proof of the fundamental result of de Finetti (cf. [12]) as extended by Dacunha-Castelle and Krivine [5].

Theorem 1.1. *Let (X_j) be a sequence of random variables with tail field $\mathscr{A}$. Assume that for all positive integers k and $j_1, \ldots, j_k$ with $j_1 < j_2 < \ldots < j_k$, $\operatorname{dist}(X_1, \ldots, X_k) = \operatorname{dist}(X_{j_1}, \ldots, X_{j_k})$. Then (X_j) is conditionally i.i.d. with respect to $\mathscr{A}$; consequently (X_j) is exchangeable.*

Proof. For each n, let $\mathscr{A}_n = \sigma(\{X_j \colon j = n, n+1, n+2, \ldots\})$ and $\mathscr{E}_n = \mathscr{E}_{\mathscr{A}_n}$; also $\mathscr{E} = \mathscr{E}_{\mathscr{A}}$ as usual. Say that $Y_n \to Y$ strongly if $\int |Y_n - Y| \, dP \to 0$. We first need an elementary lemma, the first assertion of which follows from standard martingale results (or Schauder decomposition theorems in Banach space theory) and the other assertions of which each follow easily from the preceding assertion.

Lemma 1.2. *Let (φ_n) be a uniformly bounded sequence of random variables and g an integrable random variable. Then*

(a) $\mathscr{E}_n g \to \mathscr{E} g$ *strongly.*
(b) $(\mathscr{E}_n g)\,\varphi_n - (\mathscr{E} g)\,\varphi_n \to 0$ *strongly.*
(c) $\mathscr{E}[(\mathscr{E}_n g \cdot \varphi_n] - (\mathscr{E} g)(\mathscr{E} \varphi_n) \to 0$ *strongly.*

In the following, we let $f_1, f_2, \ldots$ range over bounded continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

We first prove that the joint conditional distributions of any q of the X_j's in increasing order, are the same. (In particular, this shows the X_j's are conditionally identically distributed). Analytically, this is equivalent to the following assertion:

For all positive integers q, $j_1 < \ldots < j_q$ and $f_1, \ldots, f_q$,

$$\mathscr{E}(f_1(X_1) \cdot \ldots \cdot f_q(X_q)) = \mathscr{E}(f_1(X_{j_1}) \cdot \ldots \cdot f_q(X_{j_q})). \tag{1}$$

To see this let $n > j_q$. By (a) of the lemma it suffices to show that

$$\mathscr{E}_n(f_1(X_1) \cdot \ldots \cdot f_q(X_q)) = \mathscr{E}_n(f_1(X_{j_1}) \cdot \ldots \cdot f_q(X_{j_q})). \tag{2}$$

By a standard approximation argument, to show (2) it suffices to show that for any k and φ: $\mathbb{R}^{k+1} \to \mathbb{R}$ bounded continuous,

$$\begin{aligned} &\int f_1(X_1) \cdot \ldots \cdot f_q(X_q)\, \varphi(X_n, \ldots, X_{n+k})\, dP \\ &\quad = \int f_1(X_{j_1}) \cdot \ldots \cdot f_q(X_{j_q})\, \varphi(X_n, \ldots, X_{n+k})\, dP. \end{aligned} \tag{3}$$

But (3) follows immediately from the assumption that $\operatorname{dist}(X_1, \ldots, X_q, X_n, \ldots, X_{n+k}) = \operatorname{dist}(X_{j_1}, \ldots, X_{j_q}, X_n, \ldots, X_{n+k})$.

To complete the proof, it suffices to prove that the following statement holds for all positive integers q: For any $f_1, \ldots, f_q$,

$$\mathscr{E}(f_1(X_1) \cdot \ldots \cdot f_q(X_q)) = \mathscr{E} f_1(X_1) \cdot \ldots \cdot \mathscr{E} f_q(X_q). \tag{4}$$

Suppose $q > 1$ and (4) has been proved for all $q' < q$. Fixing $f_1, \ldots, f_q$, we have by (1) that for all positive integers n,

$$\begin{aligned} \mathscr{E}(f_1(X_1) \cdot \ldots \cdot f_q(X_q)) &= \mathscr{E}(f_1(X_1) \cdot f_2(X_{2+n}) \cdot \ldots \cdot f_q(X_{q+n})) \\ &= \mathscr{E}(\mathscr{E}_n(f_1(X_1) \cdot f_2(X_{2+n}) \cdot \ldots \cdot f_q(X_{q+n})) \\ &= \mathscr{E}([\mathscr{E}_n f_1(X_1)] \cdot f_2(X_{2+n}) \cdot \ldots \cdot f_q(X_{q+n})). \end{aligned} \tag{5}$$

Now setting $g = f_1(X_1)$ and $\varphi_n = f_2(X_{2+n}) \cdot \ldots \cdot f_q(X_{q+n})$, we have by (5) and (c) of the lemma that

$$\mathscr{E}(f_1(X_1) \cdot \ldots \cdot f_q(X_q)) - (\mathscr{E} g)(\mathscr{E} \varphi_n) \to 0 \tag{6}$$

strongly. But

$$\mathscr{E} \varphi_n = \mathscr{E} f_2(X_2) \cdot \ldots \cdot \mathscr{E} f_q(X_q) \tag{7}$$

for all n by (1) and the induction hypothesis. (6) and (7) show that (4) holds, completing the proof.

Our next result is a generalization of a lemma of Maharam [15]. *We say that $\mathscr{S}$ is atomless over a sub-σ-algebra $\mathscr{A}$ (with respect to P) if for every $B \in \mathscr{S}$*

of positive measure there exists an $F \in \mathscr{S}$ *with* $F \subset B$ *so that there is no* A *in* $\mathscr{A}$ *with* $F = B \cap A$ *(i.e. precisely,* $I_F = I_{B \cap A}$ *a.e.).* $\mathscr{S}$ is called atomless provided $\mathscr{S}$ it atomless over the trivial algebra. (The utility of this concept is found in the following elementary fact: Let $(\Omega, \mathscr{D}, P)$ be an atomless probability space and $(\Omega', \mathscr{A}, P')$ another probability space; let $\mathscr{A} \times \Omega = \{A \times \Omega\colon A \in \mathscr{A}\}$. Then in the product probability space, the product σ-algebra $\mathscr{A} \times \mathscr{D}$ is atomless over $\mathscr{A} \times \Omega$.)

Theorem 1.3. *(The Maharam Lemma). Let* $\mathscr{A}$ *be a* σ*-subalgebra of* $\mathscr{S}$*. Let* $\mathscr{S}$ *be atomless over* $\mathscr{A}, S \in \mathscr{S}$*, and* f *an* $\mathscr{A}$*-measurable function so that* $0 \leqq f \leqq \mathscr{P}(S)$ *a.e. Then there exists a* $D \in \mathscr{S}$ *with* $D \subset S$ *so that* $\mathscr{P}(D) = f$ *a.e.*

Before giving the proof, we note an immediate consequence.

Corollary 1.4. *Let* $\mathscr{S}$ *be atomless over* $\mathscr{A}$ *and* $S \in \mathscr{S}$*. Let* k *be a positive integer and* $f_1, \ldots, f_k$ *be non-negative* $\mathscr{A}$*-measurable functions with* $\mathscr{P}S \geqq \sum_{i=1}^{k} f_i$ *a.e. Then there exist disjoint subsets* $S_1, \ldots, S_k$ *of* S *consisting of elements of* $\mathscr{S}$ *with* $\mathscr{P}S_i = f_i$ *for all* i*.*

Proof. We do this by induction. The case $k=1$ follows from the Maharam lemma. Suppose the result proved for k and let $f_1, \ldots, f_{k+1}$ be non-negative $\mathscr{A}$-measurable with $\mathscr{P}S \geqq \sum_{i=1}^{k+1} f_i$. By induction hypothesis, we may choose disjoint subsets $S_1, \ldots, S_{k-1}$, $\tilde{S}$ of S, all belonging to $\mathscr{S}$, with $\mathscr{P}S_i = f_i$ for $1 \leqq i \leqq k-1$ and $\mathscr{P}\tilde{S} = f_k + f_{k+1}$. By the Maharam Lemma, we may choose an $S_k \in \mathscr{S}$ with $S_k \subset \tilde{S}$ and $\mathscr{P}S_k = f_k$. We then simply set $S_{k+1} = \tilde{S} \sim S_k$.

Proof of the Maharam Lemma. We give a functional analytic proof inspired by an argument in [13]. Let K denote the set of all measurable functions g supported on S with $0 \leqq g \leqq 1$ and $\mathscr{E}g = f$. Evidently $\frac{f}{\mathscr{P}(S)} \cdot I_S$ belongs to K, hence K is non-empty. K is a weak* compact subset of $L^\infty(P)$ (the weak*-topology refers to the topology induced in $L^\infty(P)$ by $L^1(P)$). Thus by the Krein-Milman theorem there exists an extreme point g of K. We need only prove that $g = I_D$ for some measurable set D. If not, there exists a set $B \subset S$ of positive measure and a $\delta > 0$ so that $\delta \leqq g \leqq 1 - \delta$ on B. Since $\mathscr{S}$ is atomless over $\mathscr{A}$, we may choose a measurable $F \subset B$ so that there is no A in $\mathscr{A}$ with $I_F = I_{B \cap A}$ a.e. In turn, this implies that $\mathscr{P}(F) I_{B \sim F} \neq 0$. Indeed, if $\mathscr{P}(F) I_{B \sim F} = 0$ a.e., $\mathscr{P}(F) \mathscr{P}(B \sim F) = 0$ which yields that $\mathscr{P}(F)$ and $\mathscr{P}(B \sim F)$ are disjointly supported. But then letting $A = \operatorname{supp} \mathscr{P}(F)$ we have that $\mathscr{P}(F) = I_A \mathscr{P}(B)$ a.e. which implies $F = B \cap A$.

Now set $\varphi = \frac{\delta}{2}[\mathscr{P}(F) I_{B \sim F} - \mathscr{P}(B \sim F) \cdot I_F]$. We have that φ is a non-zero element of L^∞ with $\mathscr{E}\varphi = 0$. Since $|\varphi| \leqq \delta$ and φ is supported on B, $g + \varphi$ and $g - \varphi$ both belong to K, contradicting the fact that g is an extreme point of K.

Our next result shows that any conditional distribution is the conditional distribution of some random variable provided $\mathscr{S}$ is atomless over the conditioning subfield $\mathscr{A}$. It is a natural generalization of (and implies) the result that on an atomless measure space, any distribution is the distribution of some random variable defined on the space.

Theorem 1.5. *Let $\mathscr{A}$ be a σ-subalgebra of $\mathscr{S}$. Let $\mathscr{S}$ be atomless over $\mathscr{A}$, $S \in \mathscr{S}$ of positive measure, $A = \operatorname{supp} \mathscr{P}(S)$ and $\vec{\mu}: \mathscr{B} \to L^1(P|\mathscr{A}|A)$ be a conditional distribution. Then there exists a random variable X supported on S with $c \cdot \operatorname{dist} X|S = \vec{\mu}$.*

Proof. We shall construct a sequence (X_n) of random supported on S with the following properties:

There exists a random variable X so that $X_n \to X$ in probability. (8)

For each integer j and non-negative integer k,

$$\mathscr{P}\left(\left(\frac{j}{2^k} < X_n \leqq \frac{j+1}{2^k}\right] \cap S\right) = \vec{\mu}\left(\left(\frac{j}{2^k}, \frac{j+1}{2^k}\right]\right) \mathscr{P}(S) \tag{9}$$

for all n sufficiently large.

It then follows that

$$c \cdot \operatorname{dist} X|S = \vec{\mu}. \tag{10}$$

Indeed, (9) implies that for any integer j and non-negative integer k,

$$\mathscr{P}\left(\left[X_n \leqq \frac{j}{2^k}\right] \cap S\right) \to \vec{\mu}\left(\left(-\infty, \frac{j}{2^k}\right]\right) \mathscr{P}(S) \tag{11}$$

strongly. Now let t be a real number so that $P([X=t])=0$. Then (8) implies that $\mathscr{P}([X_n \leqq t] \cap S) \to \mathscr{P}([X \leqq t] \cap S)$ strongly. It follows, using (9), that if $\frac{j}{2^k} \leqq t$, then

$$\vec{\mu}\left(\left(-\infty, \frac{j}{2^k}\right]\right) \mathscr{P}(S) \leqq \mathscr{P}([X \leqq t] \cap S)$$

while if $t \leqq \frac{j}{2^k}$, then $\vec{\mu}\left(\left(-\infty, \frac{j}{2^k}\right]\right) \mathscr{P}(S) \geqq \mathscr{P}([X \leqq t) \cap S)$. Thus if t is also such that $\vec{\mu}(\{t\})=0$, then $\vec{\mu}((-\infty, t]) \mathscr{P}(S) = \mathscr{P}([X \leqq t] \cap S)$. Since the set of t's satisfying both properties is dense in $\mathbb{R}$ and $t \to \vec{\mu}(-\infty, t] \mathscr{P}(S)$ is a right continuous function, $\vec{\mu}(-\infty, t] = \frac{\mathscr{P}([X \leqq t] \cap S}{\mathscr{P}(S)}$ a.e. for all real t, which implies (10).

To achieve the construction of (X_n) we build a family of measurable subsets

$$\{S_j^n: -2^{2n} \leqq j \leqq 2^{2n}-1, n=0,1,2,\ldots\}$$

of S so that for all n and $-2^{2n} \leqq j \leqq 2^{2n}-1$,

(a) $\mathscr{P}(S_j^n) = \vec{\mu}\left(\left(\frac{j}{2^n}, \frac{j+1}{2^n}\right]\right) \mathscr{P}(S)$.

(b) $S_j^n \cap S_{j'}^n = \emptyset$ for all $j' \neq j$.

(c) $S_j^n = S_{2j}^{n+1} \cup S_{2j+1}^{n+1}$.

We then set $X_n=\sum_{j=-2^{2n}}^{2^{2n}-1}\frac{j+1}{2^n}I_{S_j^n}$ for all n. It is evident that the X_n's satisfy (9). To verify (8), it suffices to show that the sequence (X_n) is Cauchy in probability. Set $S_n=\bigcup_{j=-2^{2n}}^{2^{2n}-1}S_j^n$ for all n. Fix $k<n$, and let μ be the distribution associated with $\vec{\mu}$. Then

$$P(S_n\sim S_k)\leqq\mu(\{t\colon |t|\geqq 2^k\}). \tag{12}$$

Now on the set S_j^k, $X_k=\frac{j+1}{2^k}$ while $\frac{j}{2^k}<X_n\leqq\frac{j+1}{2^k}$. This implies that $|X_n-X_k|\leqq\frac{1}{2^k}$ on S_k; hence by (12),

$$P\left(\left[|X_n-X_k|>\frac{1}{2^k}\right]\right)\leqq\mu(\{t\colon|t|\geqq 2^k\}). \tag{13}$$

This of course implies (8).

We construct the S_j^n's by induction on n. By the corollary to the Maharam Lemma, we may choose S_{-1}^0 and S_0^0 disjoint measurable subsets of S with $\mathscr{P}(S_{-1}^0)=\vec{\mu}((-1,0])\mathscr{P}(S)$ and $\mathscr{P}(S_0^0)=\vec{\mu}((0,1])\mathscr{P}(S)$. Suppose S_j^n have been constructed satisfying (a) and (b) for all $-2^{2n}\leqq j\leqq 2^{2n}-1$. Fix $-2^{2n}\leqq j\leqq 2^{2n}-1$ and choose (again using the corollary) disjoint measurable subsets S_{2j}^{n+1} and S_{2j+1}^{n+1} of S_j^n with

$$\mathscr{P}(S_{2j}^{n+1})=\vec{\mu}\left[\left(\frac{2j}{2^{n+1}},\frac{2j+1}{2^{n+1}}\right]\right)\mathscr{P}(S)\quad\text{and}\quad\mathscr{P}(S_{2j+1}^{n+1})=\vec{\mu}\left(\left(\frac{2j+1}{2^{n+1}},\frac{2j+2}{2^{n+1}}\right]\right)\mathscr{P}(S).$$

Since

$$\vec{\mu}\left(\left(\frac{2j}{2^{n+1}},\frac{2j+1}{2^{n+1}}\right]\right)\mathscr{P}(S)+\vec{\mu}\left(\left(\frac{2j+1}{2^{n+1}},\frac{2j+2}{2^{n+1}}\right]\right)\mathscr{P}(S)=\vec{\mu}\left(\frac{j}{2^n},\frac{j+1}{2^n}\right]\mathscr{P}(S)=\mathscr{P}(S_j^n)$$

by induction hypothesis, it follows that (c) holds.

Now set $S_n=\bigcup_{j=-2^{2n}}^{2^{2n}-1}S_j^n$. We have by induction hypothesis that $\mathscr{P}(S_n)=\vec{\mu}((-2^n,2^n])\mathscr{P}(S)$. Let $G=\{k\colon k\notin[-2^{2n+1},2^{2n+1}),\ k\in[-2^{2n+2},2^{2n+2}),\ k$ an integer$\}$. Then

$$\sum_{k\in G}\vec{\mu}\left(\frac{k}{2^{n+1}},\frac{k+1}{2^{n+1}}\right]\mathscr{P}(S)\leqq\vec{\mu}(\mathbb{R}\sim(-2^n,2^n])\mathscr{P}(S)=\mathscr{P}(S\sim S_n).$$

Hence by the corollary to the Maharam Lemma, we may choose subsets S_k^{n+1} of $S\sim S_n$ for all $k\in G$ with $S_k^{n+1}\cap S_{k'}^{n+1}=\emptyset$ for all $k'\neq k$ and

$$\mathscr{P}(S_k^{n+1})=\vec{\mu}\left(\frac{k}{2^{n+1}},\frac{k+1}{2^{n+1}}\right]\mathscr{P}(S)$$

for all $k\in G$. It follows that the family $\{S_k^{n+1}\colon -2^{2n+2}\leqq k<2^{2n+2}\}$ satisfies (a) and (b) for "$n+1$" replacing "n" This completes the construction of the S_j^n's; it follows by induction that (a)–(c) are satisfied. Q.E.D.

We present now a useful characterization of conditional distributions. Its proof is rather routine and will be omitted.

Proposition 1.6. *Let $(\Omega, \mathfrak{A}, Q)$ be a probability space and let T: $C_0(\mathbb{R}) \to L^\infty(Q)$ be a bounded linear operator satisfying the following conditions:*

(a) *T is positive of norm one.*

(b) $\|i \circ T\| = 1$, *where $i: L^\infty(Q) \to L^1(Q)$ is the canonical injection. Then there is a unique conditional distribution $\vec{\mu}$ with respect to Q so that*

$$\int \varphi \, d\vec{\mu} = T\varphi \qquad \text{for all } \varphi \in C_0(\mathbb{R}). \tag{14}$$

Conversely if $\vec{\mu}$ is a given conditional distribution and T is defined by (14), *then T satisfies* (a) *and* (b).

Let $\vec{\mu}$ be a conditional distribution with respect to $(\Omega, \mathfrak{A}, Q)$ and let F: $R \to L^1(Q)$ be its associated distribution function, i.e., $F(x) = \vec{\mu}((-\infty, x])$. Let T be the operator associated to $\vec{\mu}$ defined by (14) and μ the associated ordinary distribution of $\vec{\mu}$. We define the associated field of $\vec{\mu}$, denoted by $\mathscr{A}_{\vec{\mu}}$, to be $\sigma(\{\vec{\mu}(B)\colon B \in \mathscr{B}\})$. It is easily seen that $\mathscr{A}_{\vec{\mu}} = \sigma(\{T\varphi\colon \varphi \in C_0(\mathbb{R})\}) = \sigma(\{F(r)\colon r \in D\})$ for any dense subset D of $\mathbb{R}$. Evidently $\vec{\mu}$ can be regarded as a conditional distribution with respect to $Q|\mathscr{A}_{\vec{\mu}}$. We also note that for $x \in \mathbb{R}$, $\vec{\mu}\{x\} = 0$ a.s. if and only if $\mu([x]) = 0$; i.e. x is a point of continuity of the conditional distribution function F if and only if it is a point of continuity of its associated ordinary distribution function.

We pass now to the crucial concepts of *weak* and *strong* convergence of conditional distributions. Let $(\Omega, \mathscr{A}, Q)$ be a fixed probability space. Given f and (f_j) in $L^1(Q)$, recall that we say $f_j \to f$ strongly if $E|f - f_j| \to 0$. We say $f_j \to f$ *weakly* if $f_j \to f$ weakly in $L^1(Q)$; i.e. if $Egf_j \to Egf$ for all $g \in L^\infty$. It is worth pointing out that if the f_n's are uniformly integrable, then $f_n \to f$ strongly if and only if $f_n \to f$ in probability. Moreover, if the f_n's are uniformly bounded, then $f_n \to f$ weakly if and only if $f_n \to f$ weak* in $L^\infty(Q)$ (with respect to $L^1(Q)$).

Definition. *Let $\vec{\mu}, \vec{\mu}_1, \vec{\mu}_2, \ldots$ be conditional distributions with respect to Q. We say that the sequence $(\vec{\mu}_n)$ converges strongly (resp. weakly) to $\vec{\mu}$ if $\vec{\mu}_n(-\infty, x] \to \vec{\mu}(-\infty, x]$ strongly (resp. weakly) for all real x with $\vec{\mu}(\{x\}) = 0$* a.s. (Of course when the $\vec{\mu}_j$'s are ordinary distributions, this coincides with the notion of complete convergence.)

We say that a sequence $(\vec{\mu}_j)$ of conditional distributions is *tight* if its associated sequence of ordinary distributions is tight. It is evident that a sequence of conditional distributions is tight provided it converges weakly. We require the following fundamental compactness result:

Theorem 1.7. *Let $(\vec{\mu}_j)$ be a tight sequence of conditional distributions with respect to $(\Omega, \mathscr{A}, Q)$. Then $(\vec{\mu}_j)$ has a weakly convergent subsequence.*

Proof. We use the classical facts that a tight sequence (ν_j) of ordinary distributions converges completely provided $\lim_{j \to \infty} \nu_j(-\infty, x]$ exists for all $x \in D$, where D is some dense subset of $\mathbb{R}$, and that a uniformly bounded sequence of random variables has a weakly convergent subsequence. By a standard diagonalization argument, we may choose a subsequence $(\vec{\mu}'_j)$ of $(\vec{\mu}_j)$ so that for every rational number r, $(\vec{\mu}'_j(-\infty,r])$ converges weakly in $L^1(Q)$. It follows that the ordinary distributions associated to $(\vec{\mu}'_j)$ converge completely to some distribution denoted μ. Let $D=\{x: \mu\{x\}=0\}$. Now fix $S\in\mathscr{A}^+$ and set $\nu_j(B) = \frac{1}{Q(S)} EI_S \vec{\mu}'_j(B)$ for all $B\in\mathscr{B}$. Then (ν_j) is a tight sequence of distributions, so by the classical fact mentioned above, there is a distribution ν so that $\nu_j\to\nu$ completely. Now if $x\in D$, then $\nu\{x\}=0$. Indeed, let $\varepsilon>0$ and choose $\delta>0$ so that $\mu([x-\delta, x+\delta])\leqq\varepsilon$. Then $\nu_j((x-\delta, x+\delta))\leqq\frac{1}{Q(S)}E\vec{\mu}'_j \quad ((x-\delta, x+\delta))= \frac{1}{Q(S)}\mu_j((x-\delta, x+\delta))$ for all j and $\overline{\lim_{j\to\infty}}\ \mu_j((x-\delta, x+\delta))\leqq\mu([x-\delta, x+\delta])\leqq\varepsilon$ since $\mu_j\to\mu$ completely. Thus $\nu\left(x-\frac{\delta}{2}, x+\frac{\delta}{2}\right)\leqq\overline{\lim_{j\to\infty}}\ \nu_j((x-\delta, x+\delta))\leqq\frac{\varepsilon}{Q(S)}$. Since $\varepsilon>0$ is arbitrary, $\nu(\{x\})=0$. *Hence* $\lim_{j\to\infty} \nu_j((-\infty,x])$ exists. Thus $\lim_{j\to\infty} EI_S\vec{\mu}'_j(-\infty,x]$ exists. Since $S\in\mathscr{A}^+$ was arbitrary and $(\vec{\mu}'_j(-\infty,x])$ is a uniformly bounded sequence of random variables, $(\vec{\mu}'_j(-\infty,x])$ converges weakly to some random variable, denoted $F(x)$.

It follows easily that $F\colon D\to L^1(Q)$ is an increasing function with $0\leqq F(x)\leqq 1$ and $EF(x)=\mu(-\infty,x]$ for all $x\in D$. The latter equality shows that $\lim_{\substack{x\to-\infty\\ x\in D}} F(x) =0$ and $\lim_{\substack{x\to+\infty\\ x\in D}} F(x)=1$ strongly. Moreover this equality yields that F is right continuous on the set D. We now define $\tilde{F}(x)=F(x+)$ for all $x\in\mathbb{R}$. I.e. $\tilde{F}(x) =\lim_{\substack{d\downarrow x\\ d\in D}} F(d)$ for all $x\in\mathbb{R}$. The usual standard arguments show that $\tilde{F}|D=F$ and $\tilde{F}$ has all the properties stated above for F, but without the restriction $x\in D$. Using the existence of a regular version of $\tilde{F}$, it is easily seen that there is a conditional distribution $\vec{\mu}$ with $\tilde{F}$ as its corresponding conditional distribution function, and then, of course, μ is the ordinary distribution associated with $\vec{\mu}$ and $\vec{\mu}'_j(-\infty,x]\to\vec{\mu}(-\infty,x]$ for all $x\in D$. As noted above, $\vec{\mu}\{x\}=0$ if and only if $\mu\{x\}=0$, so $\vec{\mu}'_j\to\vec{\mu}$ weakly, completing the proof.

Remark. It is possible to give an alternate proof based on Proposition 1.6. Thus, one chooses Y a countable subset of $C_0(\mathbb{R})$ with linear span Z dense in $C_0(\mathbb{R})$ and a subsequence $(\vec{\mu}'_j)$ of $(\vec{\mu}_j)$ so that $(\int\varphi\, d\vec{\mu}'_j)$ converges weakly in $L^1(Q)$ for all $\varphi\in Y$. It follows that $\int\varphi\, d\vec{\mu}'_j$ converges weakly to an element denoted $T\varphi$ for all $\varphi\in Z$. One then verifies that $T\colon Z\to L^\infty(Q)$ is a norm-one linear operator, hence has a unique linear extension (also denoted T) to all of $C_0(\mathbb{R})$. Then T satisfies the conditions of 1.6. Employing standard arguments, it follows that $(\vec{\mu}'_j)$ converges weakly to the conditional distribution $\vec{\mu}$ corresponding to T as in 1.6.

We conclude this section with a summary of some equivalent formulations of weak and strong conditional convergence. The proofs are routine extensions of the classical equivalences for complete convergence, and shall be omitted.

Proposition 1.8. *Let $\vec{\mu}$, $\vec{\mu}_1$, $\vec{\mu}_2,\ldots$ be conditional distributions with respect to Q. Then the following are equivalent:*

(a) $\vec{\mu}_n \to \vec{\mu}$ *strongly (resp. weakly).*

(b) $\vec{\mu}_n(-\infty, x] \to \vec{\mu}(-\infty, x]$ *strongly (resp. weakly) for all* $x \in D$, *where* D *is a dense subset of* $\mathbb{R}$ *with* $\vec{\mu}\{x\}=0$ *for all* $x \in D$.

(c) *Let* $\mathcal{G}$ *be the smallest algebra of subsets of* $\mathbb{R}$ *containing all finite open intervals* (a, b) *with* $\vec{\mu}\{a\}=\vec{\mu}\{b\}=0$ *and all singletons* $\{x\}$ *with* $\mu\{x\}=0$. *Then* $\vec{\mu}_n(G) \to \vec{\mu}(G)$ *strongly (resp. weakly) for all* $G \in \mathcal{G}$.

(d) $\int \varphi\, d\vec{\mu}_n \to \int \varphi\, d\vec{\mu}$ *strongly (resp. weakly) for every* $\varphi \in C_0(\mathbb{R})$.

(e) *The same as* (c), *except "φ" ranges over all bounded continuous real valued functions.*

(f) $\int e^{itx}\, d\vec{\mu}_n(x) \to \int e^{itx}\, d\vec{\mu}(x)$ *strongly (resp. weakly) for every real* t.

Remarks. 1. If $(\Omega, \mathfrak{A}, Q)$ is a probability space and the field $\mathfrak{A}$ is atomic, then by a standard result in functional analysis, weak and strong sequential convergence coincide in $L^1(Q)$. Hence in this case weak and strong convergence of sequences of conditional distributions with respect to Q coincide. *This is the only case in which we have this general coincidence.*

2. Let us say that a family F of conditional distributions with respect to Q is strongly conditionally compact if every sequence in F has a strongly convergent subsequence. It follows easily from our discussion that F is strongly conditionally compact if and only if F is tight and for every $\varphi \in C_0(\mathbb{R})$ and sequence $(\vec{\mu}_j)$ in F, $(\int \varphi\, d\vec{\mu}_j)$ has a strongly convergent subsequence. Also, if $(\vec{\mu}_j)$ is a sequence of conditional distributions whose associated distributions converge completely to some distribution μ, then $F=\{\vec{\mu}_1, \vec{\mu}_2, \ldots\}$ is strongly conditionally compact if and only if for some dense subset D of $\mathbb{R}$ with $\mu\{x\}=0$ for all $x \in D$, $(\vec{\mu}'_j(-\infty, x])$ has a strongly convergence subsequence for all $x \in D$ and subsequences $(\vec{\mu}'_j)$ or $(\vec{\mu}_j)$.

3. Let $\vec{\mu}$, $(\vec{\mu}_j)$ be conditional distributions with respect to $(\Omega, \mathcal{S}, P)$ and let $\mathcal{A}$ be a σ-sub-algebra of $\mathcal{S}$. Then $\mathcal{E}_{\mathcal{A}}\vec{\mu}$ is a conditional distribution with respect to $P|\mathcal{A}$. Suppose $\vec{\mu}_j \to \vec{\mu}$ weakly (resp. strongly). Then $\mathcal{E}_{\mathcal{A}}\vec{\mu}_j \to \mathcal{E}_{\mathcal{A}}\vec{\mu}$ weakly (resp. strongly). It may happen, of course, that $\mathcal{E}_{\mathcal{A}}\vec{\mu}_j \to \mathcal{E}_{\mathcal{A}}\vec{\mu}$ strongly even though $\vec{\mu}_j \not\to \vec{\mu}$ strongly. It is natural to study the question of when $\mathcal{E}_{\mathcal{A}}\vec{\mu}_j \to \mathcal{E}_{\mathcal{A}}\vec{\mu}$ *in the case where* $\mathcal{A}$ *is the field associated with* $\vec{\mu}$, (so $\mathcal{E}_{\mathcal{A}}\vec{\mu}=\vec{\mu}$). We draw the consequences of this phenomena in a fairly general setting in the next section.

§ 2. The Main Result

The main object of this section is to formulate and prove the basic criterion for almost exchangeability discussed in the Introduction. Throughout, we let $(\Omega, \mathcal{S}, P)$ denote a fixed probability space. We first require the following fundamental concept:

Definition. *A sequence* (X_n) *of random variables defined on* Ω *is said to be determining if* $\operatorname{dist}(X_n|S)$ *converges completely for every* $S \in \mathcal{S}^+$.

Evidently a determining sequence of random variables must be bounded in probability. We shall see momentarily that any sequence of variables bounded in probability has a determining subsequence. We first present some equivalent formulations of this concept, phrased in terms of the results of Sect. 1.

Proposition 2.1. *Let (X_n) be a sequence of random variables defined on Ω and assume that the sequence* $(\operatorname{dist} X_n)$ *converges completely to a distribution μ. Let D be a dense subset of $\mathbb{R}$ so that $\mu(\{x\})=0$ for all $x\in D$. The following are equivalent:*

(a) *(X_n) is determining.*

(b) *$(I_{[X_n\leqq r]})$ converges weakly for every $r\in D$.*

(c) *There is a conditional distribution $\vec{\mu}$ with respect to P so that $c\cdot(\mathscr{S})\operatorname{dist} X_n\to\vec{\mu}$ weakly.*

Proof. This follows easily from the arguments and results of the previous section. Let $S\in\mathscr{S}^+$. By definition there is a distribution μ_S so that $\dfrac{P([X_n\leqq r)\cap S)}{P(S)}\to\mu_S(-\infty,r]$ when r is such that $\mu_S\{r\}=0$. It follows easily that $\mu\{r\}=0$ implies $\mu_S\{r\}=0$, whence $r\in D$ implies that $(P([X_n\leqq r)\cap S))$ converges, hence $I_{[X_n\leqq r]}$ converges weakly. Thus (a)$\Rightarrow$(b). Now let $\vec{\mu}=c\cdot(\mathscr{S})\operatorname{dist} X_j$ for all j. (Of course $c\cdot(\mathscr{S})\operatorname{dist} X_j(B)=I_{[X_j\in B]}$ for all $B\in\mathscr{B}$.) Then $(\vec{\mu}_j)$ is a tight sequence. It follows from the proof of Theorem 1.7 that there is a conditional distribution $\vec{\mu}$ with respect to P so that $\vec{\mu}_j\to\vec{\mu}$ weakly. Thus (b)$\Rightarrow$(c). Assuming (c), then of course, if $\tilde{\mu}$ is the distribution associated with $\vec{\mu}$, $\operatorname{dist} X_n\to\tilde{\mu}$, so $\tilde{\mu}=\mu$. If $r\in D$ and $S\in\mathscr{S}^+$, then by definition, $I_{[X_n\leqq r]}\to\vec{\mu}(-\infty,r]$ weakly, hence $\lim\limits_{n\to\infty} E(I_{[X_n\leqq r]}\cdot I_S)=\lim\limits_{n\to\infty} P([X_n\leqq r]\cap S)=E(I_S\vec{\mu}(-\infty,r])$. Since D is a dense subset of $\mathbb{R}$ and (X_n) is bounded in probability, so is $(X_n|S)$, hence setting $\mu_S(B)=\dfrac{1}{P(S)}E\,I_S\vec{\mu}(B)$ we have that $\operatorname{dist} X_n|S\to\mu_S$ completely. This completes the proof.

The next result (noted first by Baxter-Chacon [3]) is an immediate consequence of Proposition 2.1 and Theorem 1.7.

Theorem 2.2. *Every sequence of random variabes that is bounded in probability has a determining subsequence.*

Remark. Of course all the other equivalent conditions of Proposition 1.8 apply to determining sequences of random variables. It follows from 1.8 and the proof of 1.7, for example, that (X_n) is determining if and only if $(\varphi(X_n))$ converges weakly for every bounded continuous φ (resp. every $\varphi\in C_0(\mathbb{R})$). We focus on condition (b) of Proposition 2.1 because it provides a constructive procedure for finding a determining subsequence. Thus, if (X_n) is a sequence of random variables and μ is a distribution with $\operatorname{dist} X_n\to\mu$; we let D be a countable dense subset of $\{x\in\mathbb{R}\colon \mu(\{x\})=0\}$ and choose a subsequence (X'_n) so that $(I_{[X'_n\leqq r]})$ converges weakly for all $r\in D$. Then (X'_n) is determining.

We arrive now at the crucial concepts enabling us to provide the desired criterion for almost-exchangeable subsequences.

Definition. *Let (X_n) be a determining sequence of random variables and let $\vec{\mu}$ be the conditional distribution so that $c\cdot(\mathscr{S})\operatorname{dist} X_n\to\vec{\mu}$ weakly. We call $\vec{\mu}$ the limit* *conditional distribution of* (X_n). *Let* $\mathscr{A}$ *be the* σ*-field associated with* $\vec{\mu}$. *We call* $\mathscr{A}$ *the limit tail field of* (X_n).

We may "compute" the limit tail field of (X_n) as follows: let F be the limit conditional distribution function of (X_n), i.e. $F(r)=\vec{\mu}(-\infty,r]$ for all $r\in\mathbb{R}$. Let μ be the limit distribution of the sequence of distributions of the X_n's and D be a dense subset of $\{x\in\mathbb{R}: \mu(\{x\})=0\}$. Then $\mathscr{A}=\sigma(\{F(r): r\in D\})$. Of course $F(r) = \lim_{n\to\infty} I_{[X_n\leqq r]}$ for all $r\in D$, the sequence $I_{[X_n\leqq r]}$ converging weakly. If we set $T\varphi = \lim_{n\to\infty} \varphi(X_n)$ for all $\varphi\in C_b(\mathbb{R})$ ($C_b(\mathbb{R})$ denoting the set of bounded continuous real-valued functions), it is easily seen that also

$$\mathscr{A}=\sigma\{T\varphi\colon \varphi\in C_b(\mathbb{R})\}=\sigma\{T\varphi\colon \varphi\in C_0(\mathbb{R})\}$$

(and of course $T\varphi=\int\varphi\,d\vec{\mu}$ for all $\varphi\in C_b(\mathbb{R})$).

We summarize some of the permanence properties of the limit conditional distributions and limit tail field in the next result.

Proposition 2.3. *Let* (X_n) *be a determining sequence of random variables with limit tail field* $\mathscr{A}$ *and limit conditional distribution* $\vec{\mu}$. *Then* $\mathscr{A}$ *is contained in the tail field of* (X_n). *Let* (Y_n) *be a sequence of random variables so that there is a subsequence* (X'_n) *of* (X_n) *with* $X'_n-Y_n\to 0$ *in probability. Then* (Y_n) *is also determining with limit conditional distribution* $\vec{\mu}$ *(and consequently also* $\mathscr{A}=$ *the limit tail field of* (Y_n)*).*

Proof. Let F be the limit conditional distribution of (X_n) and r be a point of continuity of F. Then $F(r)=$weak limit $(I_{[X_n\leqq r]})$ is contained in the tailfield of (X_n), hence $\mathscr{A}$ is contained in the tail field of (X_n). Since $(I_{[X'_n\leqq r]})$ also converges weakly, (X'_n) is of course determining with F as its limit conditional distribution function so (X'_n) has the same limit conditional distribution as (X_n). If $\varphi\in C_0(\mathbb{R})$ and $S\in\mathscr{S}$, then since $X'_n-Y_n\to 0$ in probability,

$$\int_S [\varphi(X'_n)-\varphi(Y_n)]\,dP\to 0.$$

Thus $\varphi(Y_n)\to T\varphi$ weakly (where $T\varphi=\int\varphi\,d\vec{\mu}$ for all $\varphi\in C_0(\mathbb{R})$). This shows that (Y_n) is determining with the same limit conditional distribution as $\vec{\mu}$.

Remark. It is easily seen that if (X_j) is determining and $(\bar{X}_j)$ is another sequence of random variables with $\operatorname{dist}(\bar{X}_j)=\operatorname{dist}(X_j)$, then $(\bar{X}_j)$ is determining. Indeed, we have that (X_j) is determining if and only if

$$\lim_{n\to\infty} E(\varphi_1(X_1)\ldots\varphi_k(X_k)\,\varphi_{k+1}(X_n))$$

exists for all k and $\varphi_1,\ldots,\varphi_{k+1}$ bounded continuous functions.

Definition. *A sequence* (X_n) *of random variables is almost exchangeable if there exist sequences of random variables* $(\bar{X}_n)$ *and* (Y_n) *defined on some (possibly different) probability spaces so that*

(a) $\operatorname{dist}(X_n)=\operatorname{dist}(\bar{X}_n)$

(b) Y_n *is exchangeable*

and

(c) $\Sigma|\bar{X}_n-Y_n|<\infty$ a.e.

We may now formulate our main result.

Theorem 2.4. *A sequence of random variables has an almost exchangeable subsequence if and only if it has a determining subsequence whose conditional distributions (with respect to the limit tail field of the subsequence), relative to any set of positive measure, converge strongly.*

In the sequel, unless stated otherwise, we shall consider conditional distributions of a determining sequence as taken with respect to its limit tail field; $\mathscr{E}$ (resp. $\mathscr{P}$) shall denote conditional expectations (resp. conditional probability) with respect to this field.

Remark. Let us say that a sequence (X_j) is trivially almost exchangeable if there exists a random variable X with $\Sigma|X_j - X| < \infty$ a.e. Evidently, (X_j) has a trivially almost exchangeable subsequence if and only if (X_j) has a subsequence converging in probability. Suppose (X_j) is a determining sequence of random variables with limit conditional distribution $\vec{\mu}$. Let us say that $\vec{\mu}$ is trivial if for all Borel B, there is an A in the limit tail field of (X_j) with $\vec{\mu}(B) = I_A$ a.e. It is evident that *if* (X_j) converges in probability, $\vec{\mu}$ is trivial. *The converse is also true; thus* (X_j) *has a trivially almost exchangeable subsequence if and only if* $\vec{\mu}$ *is trivial.* First note that e.g. by the proof of Theorem 1.5 if $\vec{\mu}$ is a trivial conditional distribution, there is a random variable X with $\vec{\mu}(B) = I_{[X \in B]}$ a.e. for all Borel sets B. The fact that (X_j) converges in probability to X if $\vec{\mu}$ is trivial, now follows immediately from the following elementaty result.

Fact. *Let $X, X_1, X_2, \ldots$ be random variables on Ω such that $I_{[X_n \leqq t]} \to I_{[X \leqq t]}$ weakly for every t such that $P([X = t]) = 0$. Then $X_n \to X$ in probability.*

Proof. We first observe that if S_n, and S are measurable subsets of Ω such that $I_{S_n} \to I_S$ weakly, then $I_{S_n} \to I_S$ strongly i.e. in probability. For then $P(S_n) \to P(S)$ and also $P(S_n \cap S) \to P(S)$, hence $P(S_n \triangle S) \to 0$. Now let $\varepsilon > 0$; we may choose real numbers $a_0 < a_1 < \ldots < a_m$ so that $P[X \notin (a_0, a_m]] < \varepsilon$, $P([X = a_i]) = 0$ and $a_i - a_{i-1} < \varepsilon$ for all i (resp. $1 \leqq i$). Now choose M so that $n \geqq M$ implies

$$P([(a_{i-1} < X_n \leqq a_i]] \triangle [(a_{i-1} < X \leqq a_i]]) < \frac{\varepsilon}{m} \quad \text{for all } 1 \leqq i \leqq m.$$

Since

$$[|X_n - X| > \varepsilon] \subset [X \notin (a_0, a_m]] \cup \bigcup_{i=1}^{m} [(a_{i-1} < X_n \leqq a_i]] \triangle [(a_{i=1} < X \leqq a_i]],$$

$P([|X_n - X| > \varepsilon]) < 2\varepsilon$. This proves the Fact.

Henceforth, for determining sequences (X_n) with limit tail field $\mathfrak{A}$, conditional distributions, expectations and probabilities shall be taken with respect to $\mathfrak{A}$.

Before passing to the proof of 2.4, we wish to draw a number of simple consequences. Let us say that a determining sequence (X_n) is *strongly conditionally convergent in distribution* (*abbreviated* s.c.c.d.) *if* $(c \cdot \operatorname{dist} X_n | S)$ *converges strongly for any set S of positive measure.* Thus 2.4 may be rephrased:

(X_n) has an almost exchangeable subsequence if and only if (X_n) has an s.c.c.d. subsequence. It is easily seen that a determining sequence is already weakly conditionally convergent in distribution. Indeed, we have the following simple result:

Proposition 2.5. *Let (X_n) be a determining sequence with limit conditional distribution $\vec{\mu}$ and limit tail field $\mathscr{A}$. Then for any S of positive measure,*

$$c\cdot\operatorname{dist} X_n|S \to R(A)\,\vec{\mu} \tag{15}$$

weakly where $A=\operatorname{supp}\mathscr{P}(S)$.

Proof. Let r be such that $\vec{\mu}\{r\}=0$. Then $I_{[X_n\leqq r]}\to F(r)$ weakly, where $F(r)=\vec{\mu}(-\infty,r]$. Thus also $I_{[X_n\leqq r]}\cdot I_S\to F(r)\cdot I_S$ weakly. Hence $\mathscr{P}([X_n\leqq r]\cap S)=\mathscr{E}(I_{[X_n\leqq r]}\cdot I_S)\to\mathscr{E}(F(r)\cdot I_S)=F(r)\,\mathscr{E}I_S=F(r)\,\mathscr{P}(S)$ weakly. This implies

$$\frac{\mathscr{P}([X_n\leqq r]\cap S)}{\mathscr{P}(S)}\to I_A F(r)$$

with respect to $P|\mathscr{A}|A$.

Since weak and strong convergence coincide with respect to atomic fields, the next result is an immediate corollary of the two preceeding ones.

Corollary 2.6. *Let a determining sequence have an atomic limit tail field. Then the sequence has an almost exchangeable subsequence.*

Remarks. Evidently, if the tail field of a sequence is atomic, then so is the limit tail field of any subsequence. We thus obtain immediately the following result of D. Aldous [1]: *A sequence (X_n) of random variables has an almost-exchangeable subsequence provided it is bounded in probability and has an atomic tail field.*

The following notion derives its motivation from Banach space theory. We say that an integrable sequence (X_n) of random variables is *norm-almost exchangeable* if there exist sequences of random variables $(\bar{X}_n)$ and (Y_n) on some probability spaces so that

(a) $\operatorname{dist}(X_n)=\operatorname{dist}(\bar{X}_n)$

(b) (Y_n) is exchangeable

and

(c) $\Sigma E(|\bar{X}_n-Y_n|)<\infty$.

It is easily seen that if (X_n) is norm-almost exchangeable, then (X_n) is weakly convergent. If then $X_n\to 0$ weqkly but $E|X_n|\geqq\delta>0$ for some $\delta>0$ and all n, there is an N so that $(X_n)_{n=N}^{\infty}$ is a *symmetric basic sequence*, in Banach space terminology. The following result shows that the study of norm-almost exchangeable sequences reduces easily to that of almost exchangeable sequences.

Proposition 2.7. *A sequence of random variables has a norm-almost exchangeable subsequence if and only if it has a uniformly integrable almost-exchangeable subsequence.*

Proof. Suppose first that (X_n) is norm-almost exchangeable, and let $(\bar{X}_n)$, (Y_n) as in the above definition. Then $E(\Sigma|\bar{X}_n - Y_n|) < \infty$ by Beppo-Levi's theorem, hence $\Sigma|\bar{X}_n - Y_n| < \infty$ a.e., so (X_n) is almost exchangeable. (Y_n) is uniformly integrable since (Y_n) is identically distributed and Y_1 is integrable. Since $E|\bar{X}_n - Y_n| \to 0$, $(\bar{X}_n)$ is also uniformly integrable, hence so is (X_n). Suppose now that (X_n) is uniformly integrable almost exchangeable; let $(\bar{X}_n)$ and (Y_n) be as in the definition of almost-exchangeability. Then since $\operatorname{dist}(X_n) = \operatorname{dist}(\bar{X}_n) \to \operatorname{dist} Y_1$, Y_1 is integrable. Thus $(\bar{X}_n - Y_n)$ is uniformly integrable and $\bar{X}_n - Y_n \to 0$ a.e. Hence $E(|\bar{X}_n - Y_n|) \to 0$. Choose then $n_1 < n_2 < \dots$ with $\Sigma E|\bar{X}_{n_i} - Y_{n_i}| < \infty$. Then (X_{n_i}) is the desired norm-almost exchangeable subsequence.

The following result is thus an immediate consequence of Theorem 2.4.

Corollary 2.8. *(X_n) has a norm-almost exchangeable subsequence if and only if (X_n) has a subsequence that is uniformly integrable and s.c.c.d.*

We present next some further criteria for the s.c.c.d. condition.

Proposition 2.9. *Let (X_n) be a determining sequence of random variables with limit conditional distribution $\vec{\mu}$. Let D be a dense subset of $\mathbb{R}$ so that $\vec{\mu}\{d\} = 0$ for all $d \in D$. Let $\mathscr{G}$ be as in Proposition 1.8 (c). Let $\mathscr{T}$ be the tail field of (X_n) and also $\mathscr{H}$ a subalgebra of $\mathscr{S}$ with $\sigma(\mathscr{H}) = \mathscr{S}$. Then the following are equivalent:*

(1) *(X_j) is s.c.c.d.*
(2) *$\mathscr{P}([X_j \in G] \cap H) \to \vec{\mu}(G)\mathscr{P}(H)$ strongly for all $H \in \mathscr{H}$ and $G \in \mathscr{G}$.*
(3) *$c \cdot \operatorname{dist} X_j | S \to \vec{\mu}$ strongly for all S in $\mathscr{T}$ with $\mathscr{P}S > 0$* a.e.

Proof. (1)$\Leftrightarrow$(2) follows easily from Proposition 2.5 and evident approximation arguments. Of course (1)$\Rightarrow$(3) so it remains to check that (3)$\Rightarrow$(1). So we assume (3). We first note that if $S \in \mathscr{T}^+$, then $c \cdot \operatorname{dist} X_j | S \to R(A)\vec{\mu}$ strongly where $A = \operatorname{supp} \mathscr{P}(S)$. Indeed, let $G \in \mathscr{P}$. Then

$$\mathscr{P}([X_j \in G] \cap S) + \mathscr{P}[[X_j \in G] \cap \sim A] \to \vec{\mu}(G)\mathscr{P}S + \vec{\mu}(G) \cdot I_{\sim A} \tag{16}$$

strongly, since A is $\mathscr{A}$-measurable ($\mathscr{A}$ being the limit tail field of (X_j)). Since

$\operatorname{supp} \mathscr{P}[X_j \in G] \cap S) \subset \operatorname{supp} \mathscr{P}(S) = A$

for all j, it follows that $\mathscr{P}([X_j \in G] \cap S) \to \vec{\mu}(G)\mathscr{P}(S)$ strongly, whence $c \cdot \operatorname{dist} X_j | S \to R(A)\vec{\mu}$ strongly. We next note by standard approximation arguments that

$$\mathscr{E}(I_{[X_j \in G]} \cdot f) \to \vec{\mu}(G)\mathscr{E} f \text{ strongly} \tag{17}$$

for any $\mathscr{T}$-measurable integrable random variable f, any $G \in \mathscr{G}$. Finally let g be an arbitrary integrable random variable and let $\mathscr{E}_j$ denote conditional expectation with respect to $\sigma\{X_n\colon n = j; j+1, \dots\}$. Also let $f = \mathscr{E}_{\mathscr{T}} g$. Then

$$\mathscr{E}_n g \to f \text{ strongly.} \tag{18}$$

Now let $G \in \mathscr{G}$ and $\varphi_j = I_{[X_j \in G]}$ for all j.

Evidently it suffices to show that

$$\mathscr{E}(\varphi_n \cdot g) \to \vec{\mu}(G)\mathscr{E} g \text{ strongly.} \tag{19}$$

Now by (17), $\mathscr{E}(\varphi_n \cdot f) \to \vec{\mu}(G)\mathscr{E}f = \vec{\mu}(G)\mathscr{E}g$. Thus to show (19) it suffices to show that

$$\mathscr{E}[\varphi_n g - \varphi_n f] \to 0 \text{ strongly.} \tag{20}$$

Since $\mathscr{A} \subset \mathscr{T}$, it suffices to show

$$\mathscr{E}_{\mathscr{T}}[\varphi_n g - \varphi_n f] \to 0 \text{ strongly.} \tag{21}$$

But fixing n,

$$\mathscr{E}_{\mathscr{T}}[\varphi_n g - \varphi_n f] = \mathscr{E}_{\mathscr{T}}[\mathscr{E}_n(\varphi_n g - \varphi_n f)] = \mathscr{E}_{\mathscr{T}}[\varphi_n(\mathscr{E}_n g - f)].$$

Hence (21) follows from (18) and the fact that (φ_n) is uniformly bounded.

We now draw further immediate consequences of Theorem 2.4 and the above.

Corollary 2.10. *Let (X_n) and $\mathscr{T}$ as in 2.9 with $\mathscr{A}$ the limit tail field of (X_n). Suppose $(c \cdot \operatorname{dist} X_n)$ converges strongly and every set in $\mathscr{T}$ differs by a null set from a set in $\mathscr{A}$. Then (X_j) has an almost exchangeable subsequence.*

Corollary 2.11. *Suppose (X_j) is conditionally identically distributed with respect to its tail field. Then (X_j) has an almost exchangeable subsequence.*

Remarks. 1. In the third example at the end of Sect. 3, we construct a determining sequence (X_j) with no almost-exchangeable subsequence so that (X_j) is conditionally identically distributed with respect to its *limit tail field;* thus in particular $(c \cdot \operatorname{dist} X_j)$ converges strongly. This answers a question posed in a previous version of this paper, and shows that it is essential to consider $(c \cdot \operatorname{dist} X_j | S)$ for a suitable class of sets S, in order to discover if (X_j) has an almost exchangeable subsequence.

2. It follows easily from 2.5 and 2.9 that if (X_j) is determining and not s.c.c.d., then (X_j) has a subsequence (X'_j) so that (X'_j) has no almost exchangeable subsequence. Indeed let r be such that $\vec{\mu}\{r\} = 0$ and $S \in \mathscr{S}^+$ so that $\mathscr{P}([X_j \leqq r] \cap S) \nrightarrow \mathscr{P}(S)\vec{\mu}(-\infty, r]$ strongly. Now $\mathscr{P}([X_j < r] \cap S) \to \mathscr{P}(S)\vec{\mu}(-\infty, r)$ weakly. Hence we may choose a subsequence (X'_j) and a $\delta > 0$ so that $E|\mathscr{P}([X'_j \leqq r] \cap S) - \mathscr{P}([X'_k \leqq r] \cap S) > \delta$ for all $j \neq k$. It follows that $(\mathscr{P}([X'_j \leqq r] \cap S))$ has no strongly convergent subsequence, hence (X'_j) has no almost exchangeable subsequence. We thus obtain a Ramsey-type dichotomy: every sequence of random variables has a subsequence which is either almost exchangeable or has no further almost exchangeable subsequences.

3. It follows easily from the equivalences in 2.9 and the proof of 2.3 that if (X_j) is s.c.c.d. and (Y_j) is such that $X'_j - Y_j \to 0$ in probability, for some subsequence (X'_j) of (X_j), then (Y_j) is s.c.c.d.; also if $\operatorname{dist}(\tilde{X}_j) = \operatorname{dist}(X_j)$, then $(\tilde{X}_j)$ is s.c.c.d.

We pass now to the proof of Theorem 2.4. It follows rather easily from our above results that *if* (X_j) is almost exchangeable, (X_j) is s.c.c.d. Indeed by Remark 3 above it suffices to prove that if (Y_n) is an exchangeable sequence, (Y_n) is s.c.c.d. We prove this "fairly" standard result for the sake of completeness.

Lemma 2.12. *Let (Y_n) be an exchangeable sequence of random variables. Then (Y_n) is determining and its limit tail field coincides with its tail field up to null sets. $(c\cdot\operatorname{dist} Y_n|S)$ converges strongly to $c\cdot\operatorname{dist} Y_1$ for every set S of positive measure, where $A=\operatorname{supp}\mathscr{P}(S)$.*

To prove 2.12, let $\mathscr{T}$ denote the tail field of (Y_n) and $\mathscr{T}_n=\sigma(\{Y_j\colon j=n, n+1,\dots\})$ for all $n=1,2,\dots$. Let $\mathscr{E}_n=\mathscr{E}_{\mathscr{T}_n}$ and $\mathscr{P}_n=\mathscr{P}_{\mathscr{T}_n}$. Let φ be a bounded continuous function and S a set of positive measure. Then as observed above, since $\mathscr{P}_n(S)-\mathscr{P}_{\mathscr{T}}(S)\to 0$ strongly,

$$\mathscr{E}_{\mathscr{T}}[\varphi(Y_n)\mathscr{P}_n S-\varphi(Y_n)\mathscr{P}_{\mathscr{T}}(S)]\to 0 \tag{22}$$

strongly.

But for each n,

$$\mathscr{E}_{\mathscr{T}}[\varphi(Y_n)I_S]=\mathscr{E}_{\mathscr{T}}\mathscr{E}_n(\varphi(Y_n)I_S)=\mathscr{E}_{\mathscr{T}}[\varphi(Y_n)\mathscr{E}_n I_S]=\mathscr{E}_{\mathscr{T}}[\varphi(Y_n)\mathscr{P}_n S]. \tag{23}$$

Moreover since the Y_n's are conditionally identically distributed with respect to $\mathscr{T}$ by de Finetti's theorem (see Sect. 1, Theorem 1.1)

$$\mathscr{E}_{\mathscr{T}}[\varphi(Y_n)\mathscr{P}_{\mathscr{T}}S]=\mathscr{P}_{\mathscr{T}}(S)\,\mathscr{E}_{\mathscr{T}}\varphi(Y_n)=\mathscr{P}_{\mathscr{T}}(S)\,\mathscr{E}_{\mathscr{T}}\varphi(Y_1) \tag{24}$$

for all n. (22)–(24) imply that $\mathscr{E}_{\mathscr{T}}[\varphi(Y_n)I_S]\to\mathscr{P}_{\mathscr{T}}(S)\,\mathscr{E}_{\mathscr{T}}\varphi(Y_1)$ strongly.

Of course this establishes that (Y_n) is determining and in fact since $\mathscr{A}$, its limit tail field, is contained in $\mathscr{T}$, we obtain that automatically $c\cdot\operatorname{dist} Y_n|S\to R(A)\ c\cdot\operatorname{dist} Y_1$ strongly for any set S of positive measure with $A=\operatorname{supp}\mathscr{P}(S)$ (the conditioning now being with respect to $\mathscr{A}$). Thus the fact that $\mathscr{A}=\mathscr{T}$ up to null sets isn't really needed for the proof; for the sake of completeness, we sketch an argument.

We first observe that for any n and bounded continuous $\varphi\colon \mathbb{R}^n\to\mathbb{R}$, $\mathscr{E}_{\mathscr{T}}\varphi(y_1,,\dots,y_n)$ is measurable with respect to $\mathscr{A}$. Indeed, it is enough to prove this assertion provided $\varphi(r_1,\dots,r_n)=\varphi_1(r_1)\cdot\ldots\cdot\varphi_n(r_n)$ where $\varphi_1,\dots,\varphi_n$ are bounded continuous functions defined on $\mathbb{R}$. Since (Y_n) is exchangeable, for any $m_1<m_2<\ldots<m_n$.

$$\mathscr{E}_{\mathscr{T}}[\varphi_1(Y_1)\cdot\ldots\cdot\varphi_n(Y_n)]=\mathscr{E}_{\mathscr{T}}[\varphi_1(Y_{m_1})\cdot\ldots\cdot\varphi_n(Y_{m_n})]. \tag{25}$$

But for any $\varphi\colon\mathbb{R}\to\mathbb{R}$ bounded continuous, there is an $\mathscr{A}$-measurable function denoted as $T\varphi$ so that $\varphi(Y_j)\to T\varphi$ weakly; this implies also that $h\cdot\varphi(Y_j)\to h\cdot T\varphi$ weakly for any bounded random variable h. Hence

$$\lim_{m_1\to\infty}\ldots\lim_{m_n\to\infty}[\varphi_1(Y_{m_1})\cdot\ldots\cdot\varphi_n(Y_{m_n})]=T\varphi_1\cdot\ldots\cdot T\varphi_n$$

weakly. But then

$$\mathscr{E}_{\mathscr{T}}[\varphi_1(Y_1)\cdot\ldots\cdot\varphi_n(Y_n)]=\mathscr{E}_{\mathscr{T}}(T\varphi_1\cdot\ldots\cdot T\varphi_n)=T\varphi_1\cdot\ldots\cdot T\varphi_n \tag{26}$$

since $T\varphi_1\cdot\ldots\cdot T\varphi_n$ is $\mathscr{A}$, and hence $\mathscr{T}$-measurable.

Finally, given $D\in\mathscr{T}$, we may choose a sequence of bounded continuous functions $\varphi_n\colon\mathbb{R}^n\to\mathbb{R}$ with $\varphi_n(Y_1,\dots,Y_n)\to I_D$ strongly. Hence also $\mathscr{E}_{\mathscr{T}}\varphi_n(Y_1,\dots,Y_n)\to I_D$ strongly, whence since $\mathscr{E}_{\mathscr{T}}\varphi_n(Y_1,\dots,Y_n)$ is $\mathscr{A}$-measurable for all n, I_D differs from an $\mathscr{A}$-measurable function by a null-function. This completes the proof.

We now procedd to the proof of the direct assertion of 2.4; that is, if (X_j) is s.c.c.d. then (Y_j) has an almost exchangeable subsequence. We require two preliminary results which show essentially that *if* the conditional distribution of Y is close to the conditional distribution of X (in probability), then there is a random variable Z with the same conditional distribution as X, with Z close to Y itself in probability.

Lemma 2.13. *Let $\mathscr{A}$ be a σ-subalgebra of $\mathscr{S}$ with $\mathscr{S}$ atomless over $\mathscr{A}$, $\Omega_0 \in \mathscr{S}$, $(F_1, \ldots, F_m)$ and $(E_1, \ldots, E_m)$ measurable partitions of Ω_0 and $\eta > 0$. Assume that $E|\mathscr{P}(E_i) - \mathscr{P}(F_i)| \leqq \eta$ for all i (where $\mathscr{P}$ denotes conditional probability with respect to $\mathscr{A}$). Then there exists a measurable partition $(H_1, \ldots, H_m)$ of Ω_0 so that $\mathscr{P}(H_i) = \mathscr{P}(E_i)$ and $P(F_i \triangle H_i) \leqq \eta$ for all i.*

Proof. For functions f and g, $f \wedge g$ denotes the minimum of f and g. First observe that for $f, g \geqq 0$,

$$|f-g| = f - (f \wedge g) + g - (f \wedge g). \tag{27}$$

Now by the Maharam Lemma (Theorem 1.5), choose for each i a measurable set $D_i \subset F_i$ with $\mathscr{P}(D_i) = \mathscr{P}(F_i) \wedge \mathscr{P}(E_i)$. Since $\bigcup_i E_i \sim \bigcup_i D_i = \Omega_0 \sim \bigcup_i D_i$,

$$\sum_i \mathscr{P}(E_i) - \mathscr{P}(D_i) = \mathscr{P}(\Omega_0 \sim \bigcup_i D_i). \tag{28}$$

Thus by a consequence of Maharam's Lemma, Corollary 1.6, we may choose a measurable partition $(B_1, \ldots, B_m)$ of $\Omega_0 \sim \cup D_i$ with $\mathscr{P}(B_i) = \mathscr{P}(E_i) - \mathscr{P}(D_i)$ for all i. Now set $H_i = D_i \cup B_i$ for all i. Fix i. Since D_i and B_i are disjoint, $\mathscr{P}(H_i) = \mathscr{P}(D_i) + \mathscr{P}(B_i) = \mathscr{P}(E_i)$. $F_i \triangle H_i \subset (F_i \sim D_i) \cup B_i$. Hence

$$\begin{aligned} \mathscr{P}(F_i \triangle H_i) &\leqq \mathscr{P}(F_i) - \mathscr{P}(D_i) + \mathscr{P}(B_i) \\ &= \mathscr{P}(F_i) - (\mathscr{P}(F_i) \wedge \mathscr{P}(E_i)) + \mathscr{P}(E_i) - (\mathscr{P}(F_i) \wedge \mathscr{P}(E_i)) \\ &= |\mathscr{P}(F_i) - \mathscr{P}(E_i)| \quad \text{by (27)}. \end{aligned}$$

Thus $P(F_i \triangle H_i) = E\mathscr{P}(F_i \triangle H_i) \leqq \eta$ by our hypothesis.

Lemma 2.14. *Let $\mathscr{A}$ and $\mathscr{S}$ be as in the previous result and $S \in \mathscr{S}$ of positive probability; set $A = \operatorname{supp} \mathscr{P}(S)$. Let $X, X_1, X_2, \ldots$ be random variables on Ω so that $c\cdot(\mathscr{A}) \operatorname{dist} X_j | S \to R(A)\ c \cdot \operatorname{dist} X$ strongly and let $\varepsilon > 0$. Then there exists a simple Borel function $\varphi = \varphi_\varepsilon : \mathbb{R} \to \mathbb{R}$, depending only on ε and $\operatorname{dist} X$ (neither on the sequence (X_n) nor on S) so that $P[|X - \varphi(X)| > \varepsilon] < \varepsilon$ and such that for all n sufficiently large there exists a simple random variable h_n with*

$$\begin{aligned} &h_n \text{ supported on } S \\ &c \cdot (\mathscr{A}) \operatorname{dist} h_n | S = R(A)\, c \cdot \operatorname{dist} \varphi \circ X \text{ and} \\ &P[|X_n I_S - h_n| > \varepsilon] < \varepsilon. \end{aligned} \tag{29}$$

Proof. We take our conditional distributions and probabilities with respect to $\mathscr{A}$.

Let $\delta=\frac{\varepsilon}{2}$. Choose m and real numbers $c_1<c_2<\ldots<c_m$ so that

$$P[X=c_i]=0 \quad \text{for all } i, c_{i+1}-c_i<\delta \quad \text{for all } 1\leqq i\leqq m-1$$

and

$$P[X\notin[c_1,c_m)]<\delta. \tag{30}$$

Now set $G_i=[c_i,c_{i+1})$ for $1\leqq i\leqq m-1$ and $G_m=\mathbb{R}\sim[c_1,c_m)$; then set

$$\varphi=\sum_{j=1}^{m-1} c_j I_{G_j}. \tag{31}$$

Evidently $[|X-\varphi(X)|>\varepsilon]\subset[X\notin[c_1,c_m)]$ so $P[|X-\varphi(X)|>\varepsilon]<\varepsilon$.

Now it follows from our assumptions (c.f. Proposition 1.8) that $\mathscr{P}([X_n\in G_i]\cap S)\to\mathscr{P}([X\in G_i])\,\mathscr{P}(S)$ in L^1-norm for all i (where $\mathscr{P}$ denotes $\mathscr{P}_{\mathscr{A}}$).

Thus for all n sufficiently large and all $1\leqq i\leqq m$,

$$E|\mathscr{P}([X_n\in G_i]\cap S)-\mathscr{P}([X\in G_i])\,\mathscr{P}(S)|\leqq\eta \tag{32}$$

where $\eta=\frac{\varepsilon}{2m}$.

By Corollary 1.4 we may choose $(E_1,\ldots,E_m)$ a measurable partition of S with $\mathscr{P}(E_i)=\mathscr{P}[X\in G_i]\,\mathscr{P}(S)$ for all $1\leqq i\leqq m$. Let $F_i=[X_n\in G_i]\cap S$ for all i. Then by (32), $(E_1,\ldots,E_m)$ and $(F_1,\ldots,F_m)$ satisfy the hypotheses of the previous lemma. Hence we may choose a measurable partition $(H_1,\ldots,H_m)$ of S satisfying the conclusion of Lemma 2.13. Now set $h_n=\sum_{j=1}^{m-1} c_j I_{H_j}$. Since $\mathscr{P}(E_i)=\mathscr{P}(H_i)$ for all i, $c\cdot\operatorname{dist} h_n|S=R(A)\ c\cdot\operatorname{dist}\varphi\circ X$.

If $i<m$, then

$$[|X_n-h_n|>\delta]\cap F_i\subset F_i\sim H_i. \tag{33}$$

Since $\mathscr{P}[X\in G_m]<\delta$,

$$P([|X_n-h_n|>\delta]\cap F_m)\leqq\mathscr{P}(F_m)<\delta+\eta \tag{34}$$

by (32). Hence since $P(F_i\triangle H_i)\leqq\eta$ for all i, $P([|X_n-h_n|>\delta)\cap S)<(m-1)\eta+\delta+\eta=\varepsilon$ by (33) and (34), proving (29).

Remarks. Of course the conclusion of the Lemma implies the hypotheses. In the case of $\mathscr{A}$ the trivial algebra, this lies deeper than the usual equivalences for convergence in distribution, and may also be deduced from results of Strassen [19].

Proof of the Main Result (Theorem 2.4): We now assume that (X_j) satisfies the hypotheses of 2.4; in view of our later observations, this means that (X_j) is s.c.c.d. As always, our probability space is denoted $(\Omega,\mathscr{S},P)$; $\mathscr{A}$ denotes the limit tail field of the X_j's. After enlarging the space if necessary, we shall construct a sequence (W_j) of random variables conditionally i.i.d. with respect to $\mathscr{A}$, and a subsequence (X'_j) of (X_j) with $\Sigma|X'_j-W_j|<\infty$ a.e. By enlarging $\mathscr{S}$ if necessary, we may assume to begin with that $\mathscr{S}$ is atomless over $\mathfrak{A}$. (Actually, if we let Γ be an uncountable index set and simply take $\Omega=\{0,1\}^\Gamma$ endowed with the Γ-product Lebesgue measure on the Γ-product of the Lebesgue mesurable sets, then $\mathscr{S}$ is automatically atomless over any countably generated σ-subalgebra of $\mathscr{S}$. We could just work with X_j's on this measure space; then the W_j's could be produced with no enlargement.)

Now let $\vec{\mu}$ denote the limit conditional distribution of the X_j's. It follows by Theorem 1.5 that we may choose a random variable X on Ω with $c\cdot\operatorname{dist} X=\vec{\mu}$. *Hence $X, X_1, X_2, \ldots$ satisfy the hypotheses of the previous lemma for any $S\in\mathscr{S}$ of positive probability.*

Now let (ε_k) be a sequence of positive numbers with $\Sigma\varepsilon_k<\infty$. We first construct an increasing sequence (n_k) of positive integers, a sequence (ψ_k) of Borel simple functions, and a sequence (Y_k) of simple random variables so that (Y_k) is conditionally independent (with respect to $\mathscr{A}$) so that for all k,

$$P[|X_{n_k}-Y_k|>\varepsilon_k]<\varepsilon_k, \qquad P[|\psi_k\circ X-X|>\varepsilon_k]<\varepsilon_k$$

and

$$c\cdot\operatorname{dist} Y_k=c\cdot\operatorname{dist}\psi_k\circ X. \tag{35}$$

Let $\psi_1=\varphi_{\varepsilon_1}$ of Lemma 2.14. Letting $S=\Omega$, choose n_1 so that $n=n_1$ satisfies the conclusion of 2.14; then let $Y_1=h_{n_1}$. Now suppose $k>1$, $n_1,\ldots,n_{k-1}$ and simple variables $Y_1,\ldots,Y_{k-1}$ have been chosen. Let $S_1,\ldots,S_m$ be the atoms of $\sigma(Y_1,\ldots,Y_{k-1})$. Applying Lemma 2.14 separately to each of the sets S_i and then taking $n=n_k$ large enough, we obtain the existence of an $n_k>n_{k-1}$, a Borel simple function ψ_k and random variables f_i supported on S_i so that for all i, $P[|X_{n_k}I_{S_i}-f_i|>\varepsilon_k]<\varepsilon_k P(S_i)$, $P[|\psi_k\circ X-X|>\varepsilon_k]<\varepsilon_k$ and $c\cdot\operatorname{dist} f_i|S_i=R(A_i)$ $c\cdot\operatorname{dist}\psi_k\circ X$ where $A_i=\operatorname{supp}\mathscr{P}(S_i)$. (Precisely, $\psi_k=\varphi_\varepsilon$ where $\varepsilon=\min_i P(S_i)\,\varepsilon_k$). Now set $Y_k=\sum_{i=1}^m f_i$. Fixing B a Borel set we have that $\mathscr{P}([Y_k\in B]\cap S_i)$ $=\mathscr{P}[\psi_k\circ X\in B]\,\mathscr{P}(S_i)$ for all i, whence $c\cdot\operatorname{dist} Y_k=c\cdot\operatorname{dist}\psi_k\circ X$ since $\sum_{i=1}^m\mathscr{P}(S_i)$ $=1$. Since $[|X_{n_k}-Y_k|>\varepsilon_k]\subset\bigcup_{i=1}^m[|X_{n_k}I_{S_i}-f_i|>\varepsilon_k]$, (35) holds. It also follows that Y_k is conditionally independent of $\sigma(Y_1,\ldots,Y_{k-1})$; indeed for all i and Borel sets B,

$$\mathscr{P}([Y_k\in B]\cap S_i)=\mathscr{P}([f_i\in B)\cap S_i)=\mathscr{P}[\psi_k\circ X\in B]\,\mathscr{P}(S_i).$$

This completes the construction of (n_k), (Y_k) and (ψ_k) by induction; (35) holds for all k and (Y_k) is conditionally independent (c.f. Proposition 1.1). Now (35) implies that $\Sigma|X_{n_k}-Y_k|<\infty$ a.e. Hence to prove that (X_{n_k}) is almost exchangeable it suffices to show that (Y_k) is almost exchangeable. Again by enlarging the probability space if necessary, we may choose $Z_1, Z_2, \ldots$ i.i.d. uniformly distributed variables with $\sigma(Z_j)$ independent of $\sigma(Y_k)$. (The enlargement can be accomplished by simply taking the product of Ω with the unit interval.)

Fix k and let $\mathscr{S}_k$ denote the σ-field generated by $\mathscr{A}$, Y_k and Z_k. Since $\mathscr{A}\subset\sigma(Y_j)$, Z_k is independent of $\mathscr{A}$ and hence since Z_k is uniformly distributed, $\mathscr{S}_k$ is atomless over $\mathscr{A}$. Let $\psi=\psi_k$, $c_1,\ldots,c_n$ the distinct values of ψ, $D_i=[Y_k=c_i]$ and $E_i=[\psi\circ X=c_i]$. (We may and shall assume that $P(E_i)>0$ for all i.) Now since $c\cdot\operatorname{dist} Y_k=c\cdot\operatorname{dist}\psi\circ X$ by (35),

$$\mathscr{P}(D_i)=\mathscr{P}(E_i) \qquad \text{for all } 1\leqq i\leqq n. \tag{36}$$

By Theorem 1.5 we may choose an $\mathscr{S}_k$-measurable function q_i supported on D_i with

$$c\cdot\operatorname{dist} q_i|D_i = c\cdot\operatorname{dist} X|E_i \qquad \text{for all } i. \tag{37}$$

Now set $W_k=\sum_{i=1}^{n} q_i$. Then of course by (37), $c\cdot\operatorname{dist} W_k = c\cdot\operatorname{dist} X$. Moreover $\operatorname{dist}(\psi\circ X, X)=\operatorname{dist}(Y_k, W_k)$; indeed $\operatorname{dist} W_k|[Y_k=c_i]=\operatorname{dist} X|[\psi\circ X=c_i]$ for all i by (36) and (37). Hence $P(|W_k-Y_k|>\varepsilon_k)=P[|X-\psi_k\circ X|>\varepsilon_k)<\varepsilon_k$.

We have thus established that (W_k) is conditionally identically distributed with respect to $\mathscr{A}$ and $\Sigma|W_k-Y_k|<\infty$ a.e. To complete the proof we need to show that (W_k) is conditionally independent with respect to $\mathscr{A}$. Now each W_k may be obtained as a pointwise limit a.e. of a sequence of linear combinations of functions of the form $h_k=u_k f_k(Y_k)\, g_k(Z_k)$ where u_k is bounded $\mathscr{A}$-measurable and f_k and g_k are bounded Borel functions. But it is evident that any such sequence (h_k) satisfies $\mathscr{E}(h_1\cdot\ldots\cdot h_k)=\mathscr{E}_{h_1}\cdot\ldots\cdot\mathscr{E}_{h_k}$ for all k. The conditional independence of (W_k) now follows by routine approximation arguments, completing the proof.

§ 3. Complements

We begin by discussing the special case of almost exchangeability which motivated this work. We say that a *sequence (X_n) of random variables is almost independent-identically-distributed* (*almost* i.i.d.) *if there exists an independent identically distributed sequence (Y_n) defined on the same probability space with* $\Sigma|X_n-Y_n|<\infty$ a.e.

Theorem 3.1. *A sequence of random variables on an atomless probability space has an almost i.i.d. subsequence if and only if it has a subsequence whose distributions relative to any set of positive mesure, converge to the same limit. That is, (X_n) has an almost i.i.d. subsequence if and only if (X_n) has a subsequence (X'_n) such that there exists a distribution μ with*

$$\operatorname{dist} X'_n|S\to\mu \qquad \textit{for every } S\in\mathscr{S} \quad \textit{with} \quad P(S)>0.$$

Remark. The "if" part of this theorem was proved independently by D. Aldous [1].

Proof. Suppose first that (Y_n) is i.i.d. and $X_n-Y_n\to 0$ in probability. Then of course (X_n) is tight; letting $Y=Y_1$, we shall show that $\operatorname{dist} X_n|S\to\operatorname{dist} Y$ for all S of positive probability.

Fix S of positive probability. Let $\mathscr{A}_n=\sigma\{Y_n, Y_{n+1},\ldots\}$; then $\bigcap_{n=1}^{\infty}\mathscr{A}_n$, the tail field of (Y_n), is trivial by the zero-one law, and

$$\mathscr{E}_{\mathscr{A}_n} I_S\to P(S) \quad \text{a.e.} \tag{38}$$

Now let φ be a continuous function on $\mathbb{R}$ vanishing at infinity; then by (38),

$$E(\varphi(Y_n)[\mathscr{E}_{\mathscr{A}_n} I_S-P(S)])\to 0. \tag{39}$$

But $E\varphi(Y_n)=E\varphi(Y)$ and $E(\varphi(Y_n)I_S)=E(\varphi(Y_n)\mathscr{E}_{\mathscr{A}_n}I_S)$. Thus by (39),

$$E(\varphi(Y_n)I_S)\to P(S)E\varphi(Y). \tag{40}$$

Since $X_n-Y_n\to 0$ in probability, $\varphi(X_n)-\varphi(Y_n)\to 0$ in probability also, whence $E((\varphi(X_n)-\varphi(Y_n))I_S)\to 0$. Thus by (40), we obtain that $E(\varphi(X_n)I_S)\to P(S)E\varphi(Y)$, which shows that dist $X_n|S\to$ dist Y.

Suppose now that there is a distribution μ so that dist $X_n|S\to\mu$ for all S of positive probability. It follows that for each r with $\mu\{r\}=0$ that $I_{[X_n\leq r]}\to \mu(-\infty,r]\cdot 1$ weakly. Hence (X_n) is determining and in fact the limit tail field $\mathscr{A}$ of (X_n) is trivial; thus (X_n) is s.c.c.d. Now our proof of Theorem 2.4 shows that after a suitable enlargement of the probability space, there exists a subsequence (X'_n) of (X_n) and a sequence (W_n) of random variables conditionally i.i.d. with respect to $\mathscr{A}$ so that $\Sigma|X'_n-W_n|=\infty$ a.e. Since $\mathscr{A}$ is trivial, (W_n) is of course i.i.d. A simple modification of the proof shows, however, that no enlargement of the probability space is necessary. Indeed, in the first part of the proof, we obtain (X'_j) and a sequence (Y_j) of *independent* simple random variables with $\Sigma|X'_j-Y_j|<\infty$ a.e. If infinitely many of the Y_j's are constant, evidently (X'_j) has an almost i.i.d. subsequence.

Suppose only finitely many of the Y_j's are constant. Let $V_1, V_2, \ldots$ a sequence of infinite disjoint subsets of positive integers and for each j let $\mathscr{A}_j=\sigma\{Y_i: i\in V_j\}$. Also let (ε_j) be a sequence of positive numbers with $\Sigma\varepsilon_j<\infty$. Then, $(\Omega,\mathscr{A}_j,P|\mathscr{A}_j)$ is an atomless probability space for all j. Since dist $Y_k\to\mu$, standard arguments (and also our proof of Theorem 2.4) show that we may choose $\mathrm{m}_1<m_2<\ldots$ so that for all j, $m_j\in V_j$, and there is a variable W_j which is $\mathscr{A}_j$ measurable and satisfies dist $W_j=\mu$ and $P[|W_j-Y_{m_j}|>\varepsilon_j]<\varepsilon_j$. Since (Y_k) is independent, (W_j) is i.i.d., hence (X'_{m_j}) is almost i.i.d.

Remark. It follows easily from the above arguments and our main results that a determining sequence of random variables has an almost i.i.d. subsequence if and only if it has a trivial limit tail field. We mention the following consequence of this, (also discovered independently by D. Aldous): *If a sequence of random variables is bounded in probability and has a trivial tail field, it has an almost i.i.d. subsequence.*

Our next result shows that any subsequence of $(\sin 2\pi nx)$ has an almost i.i.d. subsequence (relative to $[0,1]$, i.e. the standard atomless probability space). This perhaps explains why $(\sin 2\pi n_k x)$ behaves like a sequence of i.i.d. variables for (n_k) thin enough. It would be desirable to give effective criteria to insure that a sequence $(\sin 2\pi n_k x)$ is almost i.i.d.; or simply that it satisfies the weaker condition that for some i.i.d. sequence (Y_k), $\sin 2\pi n_k x-Y_k\to 0$ in probability.

Corollary 3.2. *Let h be a period-one Borel measurable function on the real line and let (X_n) be defined on $[0,1]$ by $X_n(t)=h(nt)$ for all t. Then any subsequence of (X_n) has an almost i.i.d. subsequence.*

Proof. It suffices to show that dist $X_n|S\to$ dist h for any S of positive measure, by Theorem 3.1. By the definitions involved it suffices to show that

$$\int_S \varphi(X_n)\,dt \to P(S) \int_0^1 \varphi(h)\,dt \quad \text{for any } S \text{ of positive measure and bounded continuous } \varphi. \tag{41}$$

But fixing φ, (41) is equivalent to the assertion that $\varphi(X_n)$ tends to the constant $\int_0^1 \varphi(h)\,dt$ in the weak*-topology of $L^\infty[0,1]$ (with respect to $L^1[0,1]$). It follows that it suffices to prove

$$\int_a^b \varphi(X_n)\,dt \to (b-a) \int_0^1 \varphi(h)\,dt \quad \text{for all real } a, b \quad \text{with} \quad 0 \leqq a < b \leqq 1. \tag{42}$$

(So far these conditions are completely general; a sequence (Z_n) on $[0,1]$ has an almost i.i.d. subsequence if and only if it has a subsequence (X_n) so that for some variable h on $[0,1]$, (42) holds for all φ.) But by just changing notation, (42) is simply the assertion that for any bounded measurable period one function g,

$$\int_a^b g(nt)\,dt \to (b-a) \int_0^1 g(t)\,dt \quad \text{for all } 0 \leqq a < b \leqq 1.$$

Fix n and let $m = m(n)$ be the largest integer less than or equal to $n(b-a)$. Since $na + m \leqq nb < na + m + 1$ we have that

$$\begin{aligned}\int_a^b g(nt)\,dt &= \frac{1}{n} \int_{na}^{na+m} g(x)\,dx + \frac{1}{n} \int_{na+m}^{nb} g(x)\,dx \\ &= \frac{m}{n} \int_0^1 g(t)\,dt + \varepsilon_n \quad \text{where} \quad |\varepsilon_n| \leqq \frac{1}{n} \sup_{0 \leqq x \leqq 1} |g(x)|.\end{aligned}$$

But $b - a - \frac{1}{n} \leqq \frac{m}{n} \leqq b - a$, hence (42) follows. Q.E.D.

We consider next the problem of when a sequence of random variables has an almost-exchangeable subsequence after a change of density. As we show in the first example at the end of this section, there exists a sequence of random variables with no almost exchangeable subsequence, yet the sequence is in fact i.i.d. after a change of density; thus the "true" nature of the distribution of an infinite sequence may only be revealed after a change of density. The precise formulation is as follows: we say that a sequence of random variables (X_j) on $(\Omega, \mathscr{S}, P)$ is almost exchangeable after change of density if there exists a strictly positive probability density φ on Ω so that defining the probability Q by $dQ = \varphi\,dP$, then (X_j/φ) is almost exchangeable with respect to Q. We shall show that if (X_j) is uniformly integrable and determining, there is a canonical change of density which works, if anyone does; after making this change of density, we obtain essentially that $|X_j| \to c$ weakly for some constant c. We show that if $|X_j| \to c$ weakly, then if (X_j) is s.c.c.d. after a change of density, it is already s.c.c.d.

We first employ yet one more equivalence for a sequence (X_j) of random variables to be determining: *(e^{itX_j}) converges weakly for all real t.* When this occurs, we define $\vec{h}(t)=h(t,\omega)=\lim_{j\to\infty} e^{itX_j}$ and call $\vec{h}$ the *limit conditional characteristic function* of (X_j). If the probability space is large enough, that is, if $\mathscr{S}$ is atomless over $\mathscr{A}$, the limit tail field of (X_j), then we know that there is a random variable X with $c\cdot\operatorname{dist} X=\vec{\mu}$ where $\vec{\mu}$ is the limit conditional distribution of (X_j) (see Theorem 1.5 and the definitions preceding Proposition 2.3). Then $\vec{h}$ is simply the conditional characteristic function of X with respect to $\mathscr{A}$: $\vec{h}(t)=\mathscr{E}_{\mathscr{A}} e^{itX}$ for all real t.

Lemma 3.3. *Let (X_j) be a determining sequence with limit conditional characteristic function $h(t,\omega)$ and let φ be a strictly positive density; let Q be the probability with $dQ=\varphi\,dP$. Then (X_j/φ) is determining with respect to Q, with limit conditional characteristic function $h(t/\varphi(\omega),\omega)=h(t/\varphi)$.*

Proof. Since (X_j) is bounded in probability so is (X_j/φ) with respect to φ (i.e. with respect to Q). We thus need only establish that

$$e^{itX_j/\varphi}\to h(t/\varphi(\omega),\omega) \tag{43}$$

weakly with respect to φ, for all real t.

Now if φ is a simple density, this is evident. Indeed, suppose $\varphi=\sum_{j=1}^{n} c_j I_{S_j}$ with $c_j>0$ for all j and $S_1,\ldots,S_n$ is a measurable partition of Ω. For each j, we have that $e^{itX_n/\varphi} I_{S_j}=e^{itX_n/c_j} I_{S_j}\to\vec{h}(t/c_j) I_{S_j}$ weakly, hence $e^{itX_n/\varphi}\to\sum_{j=1}^{n}\vec{h}(t/c_j) I_{S_j}$ $=h(t/\varphi(\omega),\omega)$ weakly.

We next claim that for φ arbitrary strictly positive, there exists a sequence $(\tilde{\varphi}_k)$ of strictly positive simple densities with the following properties:

$$\tilde{\varphi}_k\to\varphi \quad \text{a.e. and in } L^1(P) \tag{44}$$

and

$$\text{for all } t,\ [\exp(itX_n/\tilde{\varphi}_k)]\,\tilde{\varphi}_k-[\exp(itX_n/\varphi)]\,\varphi\to 0 \text{ in } L^1(P), \tag{45}$$

uniformly in n.

Delaying the proof of this claim for the moment, to show (43) it suffices to show (using (45) and the fact that (43) holds for simple densities) that for every measurable S and real t,

$$\int_S h(t/\tilde{\varphi}_k(\omega),\omega)\,\tilde{\varphi}_k(\omega)\,dP(\omega)\to\int_S h(t/\varphi(\omega),\omega)\,\varphi(\omega)\,dP(\omega). \tag{46}$$

(46) in turn follows from (44), the dominated convergence theorem, the fact that $|h(t,\omega)|\leq 1$ for all t and the fact that for almost all $\omega, t\to h(t,\omega)$ is continuous in t. The latter can be proved e.g. by using regular conditional probabilities to see that for almost all ω, $(\mathscr{E}_{\mathscr{A}} e^{itX})(\omega)$ is a characteristic function (where X is as given before the statement of 3.3).

We pass finally to the proof of the existence of $(\tilde{\varphi}_k)$ satisfying (44) and (45). Evidently (45) holds trivially if $t=0$ by (44) so suppose $t\neq 0$.

Let $\varepsilon>0$. Choose $0<\eta<\varepsilon$ so that

$$F \text{ measurable and } P(F)<\eta \text{ implies } E(\varphi I_F)<\frac{\varepsilon}{4}. \tag{47}$$

Next choose $1<K<\infty$ so that

$$P[|X_j|>K]<\eta \quad \text{for all } j \tag{48}$$

This is possible since a determining sequence is automatically bounded in probability.

Finally choose $0<\delta<\varepsilon$ so that

$$|x-y|<\delta \Rightarrow |e^{ix}-e^{iy}|<\varepsilon. \tag{49}$$

Now choose $\tilde{\varphi}=\tilde{\varphi}_\varepsilon$ a simple strictly positive density and G a measurable set satisfying

$$P(\sim G)<\eta \qquad \left(\text{hence } E(\varphi I_{\sim G})<\frac{\varepsilon}{4} \quad \text{by (47)}\right), \tag{50}$$

$$E(\tilde{\varphi} I_{\sim G})<\frac{\varepsilon}{4}, \tag{51}$$

$$\left|\frac{1}{\varphi}-\frac{1}{\tilde{\varphi}}\right|<\frac{\delta}{|\varepsilon|K} \quad \text{on } G, \tag{52}$$

and

$$|\varphi-\tilde{\varphi}|<\varepsilon \quad \text{on } G. \tag{53}$$

Then

$$E|\varphi-\tilde{\varphi}|<\varepsilon+\frac{\varepsilon}{4}+\frac{\varepsilon}{4}=\frac{3\varepsilon}{2} \tag{54}$$

by (50), (51) and (53).

Now fix n and let

$$\beta_n=\exp(itX_n/\tilde{\varphi}), \qquad \beta=\exp(itX_n/\varphi) \quad \text{and} \quad F=[|X_n|>K].$$

Then $\beta_n\tilde{\varphi}-\beta\varphi=\beta_n(\tilde{\varphi}-\varphi)+(\beta_n-\beta)\varphi$, so

$$E|\beta_n\tilde{\varphi}-\beta\varphi|\leqq E|\tilde{\varphi}-\varphi|+E|(\beta_n-\beta)\varphi|<\tfrac{3}{2}\varepsilon+E|(\beta_n-\beta)\varphi| \quad \text{by (54)}. \tag{55}$$

Now by (48), (49), (52) and the definition of F, if $|t|<\dfrac{1}{\varepsilon}$,

$$|\beta_n-\beta|<\varepsilon \quad \text{on} \quad \sim F. \tag{56}$$

By (47) and (56),

$$E|(\beta_n-\beta)\varphi|=E|(\beta_n-\beta)\varphi I_{\sim F}|+E|(\beta_n-\beta)\varphi I_F|<\varepsilon+2\frac{\varepsilon}{4}=\frac{3}{2}\varepsilon.$$

Hence

$$E|\beta_n\tilde{\varphi}-\beta\varphi|<3\varepsilon \quad \text{by (55)}. \tag{57}$$

Evidently we thus obtain (44) and (45) by (54) and (57) if we simply let $\Sigma\varepsilon_k<\infty$ and then set $\tilde{\varphi}_k=\tilde{\varphi}_{\varepsilon_k}$ for all k. This completes the proof of Lemma 3.3.

Let us say that a sequence (X_n) of random variables in *widely almost* i.i.d. if there exists a set S of positive probability so that $\Sigma|X_n|(\omega)<\infty$ for almost all $\omega\notin S$ and $(X_n|S)_{n=1}^\infty$ is almost i.i.d. relative to $P|S$. The above lemma and Theorem 3.1 easily yield the following result; we leave the details of verification to the reader (note that by 3.1, a determining sequence (X_j) has an almost i.i.d. subsequence if and only if its limit conditional characteristic function $h(t,\omega)$ is independent of ω, and hence simply the characteristic function of some random variable).

Theorem 3.4. *Let (X_j) be a determining sequence of random variables on an atomless probability space. Then (X_j) has a subsequence which is almost (resp. widely almost) i.i.d. after change of density if and only if there is a characteristic function h of some random variable and a strictly positive (resp. positive) density u so that $e^{itX_n} \to h(tu)$ weakly for all real t.*

Our next result derives its motivation from Banach space theory. Let $1 \leqq p \leqq 2$, $K < \infty$, and (b_j) a sequence in some Banach space B. (b_j) is said to be K-equivalent to the usual l^p-basis if there are positive numbers a and b with $\frac{b}{a} \leqq K$ so that for all n and scalars $c_1, \ldots, c_n$,

$$a\left(\sum_{i=1}^{n} |c_i|^p\right)^{1/p} \leqq \left\| \sum_{i=1}^{n} c_i b_i \right\| \leqq b\left(\sum_{i=1}^{n} |c_i|^p\right)^{1/p}.$$

l^p is sad to *embed* in B if there exists a sequence in B which is K-equivalent to the usual l^p-basis, for some $K < \infty$.

Corollary 3.5. *Let (X_j) be a uniformly integrable sequence on some probability space, $1 < p \leqq 2$ and u a density so that*

$$e^{itX_n} \to e^{-|t|^p u} \quad \textit{weakly for all real } t. \tag{58}$$

Then for every $\varepsilon > 0$, (X_n) has a subsequence which is $1+\varepsilon$-equivalent in $L^1(P)$ to the usual l^p-basis.

Remarks. This result is essentially proved in [6]. It is also shown there that if (X_n) is an integrable sequence so that for all n and $\varepsilon > 0$, there exists an N so that for all $N < m_1 < m_2 < \ldots < m_n$, all $c_1, \ldots, c_n$,

$$a\left(\sum_{i=1}^{n} |c_i|^p\right)^{1/p} \leqq \int |\Sigma c_i X_{m_i}|\, dP \leqq b\left(\sum_{i=1}^{n} |c_i|^p\right)^{1/p}$$

with $\frac{b}{a} \leqq 1+\varepsilon$, *then* (X_n) satisfies (58) and hence has an l^p-subsequence. That is, in Banach space terminology, *if a sequence in L^1 has l^p as its spreading model isometrically), it has an l^p-subsequence.*

Proof. Theorem 3.4 shows that essentially some subsequence of (X_n), after change of density, is a small perturbation of a symmetric p-stable i.i.d. sequence, which proves the result. Now we elaborate.

Recall that a random variable X is symmetric p-stable if it has a characteristic function h of the form $h(t) = e^{-c|t|^p}$ for some positive c. If (Y_j) is an i.i.d. p-stable sequence for $1 < p \leqq 2$, then, as is well known (cf. [16]) Y_1 is integrable and (Y_j) is 1-equivalent in L^1 to the usual l^p-basis.

Now we may assume without loss of generality that the X_n's are defined on an atomless probability space. Then Theorem 3.4 yields that there is a subsequence (X_{n_j}) of (X_n), a strictly positive density φ, a subset Ω_0 of Ω of positive measure, and a p-stable i.i.d. sequence of random variables (Y_j) defined on the probability space $(\Omega_0, \mathcal{S}|\Omega_0, \Omega)$ where

$$dQ = (\varphi\, dP|\Omega_0) / \int_{\Omega_0} \varphi\, dP$$

so that

$$\Sigma|(X_{n_j}/\varphi) - Y_j| < \infty \quad \text{a.e. on } \Omega_0$$

and

$$\Sigma|X_{n_j}| < \infty \quad \text{a.e. off } \Omega_0.$$

The uniform integrability of the X_j's implies there are subsequences (X'_j) of (X_{n_j}) and (Y'_j) of (Y_j) so that

$$\sum_j \int_{\Omega_0} |(X'_j/\varphi) - Y'_j|\, \varphi\, dP < \infty \quad \text{and} \quad \sum_j \int_{\sim\Omega_0} |X'_j|\, dP < \infty.$$

It follows from standard perturbation arguments that for each $\varepsilon > 0$, there is a k so that $\{X'_j/\varphi : j = k, k+1, \ldots\}$ is $1+\varepsilon$-equivalent to the usual l^p basis in $L^1(\varphi\, dP)$, which implies immediately that $\{X_{n_j} : j = k, k+1, \ldots\}$ is $1+\varepsilon$-equivalent to the usual l^p-basis in $L^1(dP)$. Q.E.D.

We need two more preliminaries before obtaining our general result concerning almost exchangeability after a change of density.

Lemma 3.6. *Let (X_j) be an exchangeable sequence of random variables with tail field $\mathcal{A}$ and φ a strictly positive $\mathcal{A}$-measurable d nsity. Then (X_j/φ) is exchangeable with respect to φ.*

The proof of this lemma is routine and is left to the reader.

For the next preliminary, we need the following observation. Suppose that (X_j) is a uniformly integrable determining sequence of random variables. Then we know that $(\varphi(X_j))$ converges weakly for all bounded continuous φ. The uniform integrability of (X_j) implies that $\varphi(X_j))$ converges weakly for all $\varphi : \mathbb{R} \to \mathbb{R}$ continuous so that for some $K < \infty$, $|\varphi(x)| \leqq K|x|$ all x. *In particular, $(|X_j|)_{j=1}^{\infty}$ converges weakly.*

Lemma 3.7. *Let (X_j) be a uniformly integrable determining sequence of random variables so that $|X_j| \to 1$ weakly. If (X_j) is almost exchangeable after a change of density, then (X_j) is already almost exchangeable.*

Proof. Let φ be a strictly positive density so that (X_j/φ) is almost exchangeable with respect to φ, i.e. with respect to Q where $dQ = \varphi\, dP$, and let $\mathcal{A}$ be the limit tail field of (X_j/φ). By the proof of our main result, after a suitable enlargement of the probability space, there exists a sequence (Y_j) of random variables exchangeable with respect to Q with $\mathcal{A}$ equal to the tail field of (Y_j) so that $\Sigma|X_j/\varphi - Y_j| < \infty$ a.e. Now, of course, (X_j/φ) is uniformly integrable with respect to φ and determining by Lemma 3.3, hence we obtain that $|X_j|/\varphi \to \dfrac{1}{\varphi}$

weakly, so φ is $\mathscr{A}$-measurable. Now $\frac{1}{\varphi}$ is a strictly positive $\mathscr{A}$-measurable density with respect to Q. Thus by the preceding result, setting $Z_j = Y_j \Big/ \frac{1}{\varphi}$ for all j, then (Z_j) is exchangeable with respect to $\varphi\, dQ = dP$. Of course Q and P have the same null sets, so $\Sigma|X_j - Z_j| < \infty$ a.e., showing that (X_j) is almost exchangeable.

The following result, in combination with Theorem 2.4, solves the problem of when a sequence or random variable has an almost exchangeable subsequence after a density change.

Theorem 3.8. *Let (X_j) be a uniformly integrable determining sequence of random variables and let φ equal the weak limit of $(|X_j|)$. If $\varphi = 0$ a.e. (X_j) has a subsequence converging to zero almost everywhere and hence an almost exchangeable subsequence. If not, let $A = \operatorname{supp} \varphi$, Q the probability on $\mathscr{S} \cap A$ defined by $Q(S) = \int_S \varphi\, dP (\int \varphi\, dP)^{-1}$ for all $S \in \mathscr{S}$, $S \subset A$, and $\tilde{X}_j = X_j/\varphi$ for all j. Then (X_j) has a subsequence almost exchangeable after change of density if and only if $(\tilde{X}_j)$ has an almost exchangeable subsequence with respect to Q.*

Proof. We have that $X_j I_{\sim A} \to 0$ in probability since $|X_j| I_{\sim A} \to 0$ weakly. Thus by passing to a subsequence if necessary, we may assume that $X_j(\omega) \to 0$ for almost $\omega \notin A$, thus proving the first assertion. So assume $\int \varphi\, dP \neq 0$ and suppose without loss of generality that f is a strictly positive density so that (X_j/f) is itself almost exchangeable with respect to f. It follows that the exchangeable perturbation constructed in the proof of 2.4 must vanish on $\sim A$ and hence $(X_j/f) \cdot I_{\sim A}$ is also almost exchangeable. In fact, $|X_j|/f \to \varphi/f$ weakly, $A = \operatorname{supp} \varphi/f$, and so A is measurable with respect to the limit tail field of (X_j/f). Now $(\tilde{X}_j)$ is a determining uniformly integrable sequence so that $|\tilde{X}_j| \to 1$ weakly with respect to Q, and $(\tilde{X}_j)$ is almost exchangeable after the change of density $f = \dfrac{f/\varphi \cdot I_A}{\int_A f/\varphi\, dP}$. Hence by Lemma 3.7, $(\tilde{X}_j)$ is almost exchangeable. Suppose conversely that $(\tilde{X}_j)$ is almost exchangeable, and set $c = \int \varphi\, dP$. If $P(A) = 1$, $\frac{\varphi}{c}$ is already a strictly positive density. If not, let $f = \frac{1}{2c} \varphi + \frac{1}{2(1 - P(A))} I_{\sim A}$. If follows easily that (X_j/f) is almost exchangeable with respect to f, completing the proof.

Remarks. 1. Of course Lemma 3.6 shows that if (X_j) is almost exchangeable, then (X_j/φ) is almost exchangeable with respect to φ for any limit-tail measurable strictly positive density φ.

2. If (X_j) is an integrable sequence of random variables, and φ is a strictly positive density, then the map $X \to X/\varphi$ is an isometry from the closed linear span of the X_j's in $L^1(P)$ to the closed linear span of the X_j/φ's in $L^1(\varphi\, dP)$. Another natural isometry is provided by multiplying all the random variables by some "change of sign"; i.e. by a measurable real valued function α with $|\alpha(\omega)| = 1$ for all ω. But if (X_j) is a s.c.c.d. sequence, so is (αX_j) since then $(c \cdot \operatorname{dist} X_j | S)$ converges strongly for all S, in particular for $S \subset [\alpha = 1]$ and $S \subset [\alpha = -1]$. (In fact, by approximating with simple functions, we see that $(g X_j)$ is

s.c.c.d. for any measurable g.) Thus if (X_j) has an almost exchangeable subsequence after a change of sign, it has one without a change of sign.

We conclude with three different examples of determining sequences of random variables with no almost exchangeable subsequence.

Example 1. Let (Y_n) be a sequence of independent standard normal random variables defined on $[0,1]$. Let Q be the Borel probability measure such that $dQ(x)=2x\,dx$ and let $X_n(x)=Y_n(x)/2x$ for all n and $0\leqq x\leqq 1$. Then the sequence (X_n) has no almost exchangeable subsequence with respect to Q. (Of course (X_n) is i.i.d. after a change of density.)

Indeed, $e^{itY_n}\to e^{-t^2/2}$ weakly for any real t. Hence by Lemma 3.3, (X_n) is determining and $e^{itX_n}\to e^{-t^2/8\omega^2}$ weakly in $L^1(Q)$, for all real t. Taking $t=1$ and letting $\mathscr{A}$ denote the limit tail field of (X_n), we have that the function $\omega\to e^{-1/8\omega^2}$ is $\mathscr{A}$-measurable. Hence $\mathscr{A}=\mathscr{B}[0,1]$. Thus if (X_n) had an almost exchangeable subsequence, $(\varphi(X'_n))$ would converge in probability for every bounded continuous φ. This in turn implies that (X'_n) itself converges in probability, which yields that (Y_n) has a subsequence converging in probability; this is absurd.

Example 2. Let (S_n) be a sequence of Borel measurable subsets of $[0,1]$ such that $I_{S_n}\to f$ weakly where $f(x)=x$ for all $0\leqq x\leqq 1$. Then (I_{S_n}) has no subsequence which is almost exchangeable after a change of density. Suppose (S'_n) is a subsequence of (S_n) and φ is a strictly positive probability density so that $(I_{S'_n}/\varphi)$ is almost exchangeable with respect to Q, where $dQ=\varphi\,dx$. We shall show that $1-f$ is measurable with respect to $\mathscr{A}$, the limit tail field of $(I_{S'_n}/\varphi)$. Let $\varepsilon>0$ and choose K so that $\int_{[\varphi\geqq K]}\varphi\,dP<\varepsilon$. Choose τ a continuous function on the reals with $0\leqq\tau\leqq 1$, $\tau(0)=1$, and $\tau(x)=0$ if $|x|\geqq 1/K$. Let $g=\lim_{n\to\infty}\tau\circ(I_{S'_n}/\varphi)$, the weak limit in $L^1(Q)$; g is $\mathscr{A}$-measurable. Now for each n, the set where the functions $\tau\circ(I_{S'_n}/\varphi)$ and $1-I_{S'_n}$ differ is contained in the set where φ is larger than K. Hence $\|(\tau\circ(I_{S'_n}/\varphi))-(1-I_{S'_n})\|_{L^1(Q)}<\varepsilon$. It follows since $1-I_{S'_n}\to 1-f$ weakly, that $\|g-(1-f)\|_{L^1(Q)}\leqq\varepsilon$. The implies that $1-f$ is $\mathscr{A}$-measurable, which in turn yields that $\mathscr{A}=\mathscr{B}|[0,1]$. But then $(I_{S'_n}/\varphi)$ must converge in probability since $\beta\circ(I_{S'_n}/\varphi)$ converges in probability for each bounded continuous β; in turn this yields that $(I_{S'_n})$ itself converges in probability. Since $I_{S'_n}\to f$ weakly, $I_{S'_n}\to f$ in probability, hence f must be an indicator function, which is absurd. (It is easily seen that (I_{S_n}) is determining. What we have actually proved is that for any strictly positive density φ, the limit tail field of (I_{S_n}/φ) coincides with $\mathscr{B}|[0,1]|$.)

Such a sequence (S_n) may be constructed as follows: fix n and choose a measurable subset $A_{j,n}$ of $\left[\frac{j-1}{2^n},\frac{j}{2^n}\right)$ so that $\int I_{A_{j,n}}\,dx=\int_{(j-1)/2^n}^{j/2^n}x\,dx$; then set $S_n=\bigcup_{j=1}^{2^n}A_{j,n}$. It follows that for any fixed k and $1\leqq j\leqq 2^k$,

$$\lim_{n\to\infty}\int_{(j-1)/2^k}^{j/2^k}I_{S_n}\,dx=\int_{(j-1)/2^k}^{j/2^k}x\,dx$$

which implies that $I_{S_n}\to f$ weakly.

Example 3. This is considerably more involved than the previous two examples, and answers an open question raised in an earlier version of this paper. We construct a determining sequence (X_n) of random variables with no almost exchangeable subsequence, so that (X_n) is conditionally i.i.d. with respect to its limit tail field. Thus trivially we have that $(c \cdot \operatorname{dist} X_n)$ converges strongly; this shows that the condition "relative to any set of positive measure" in the statement of our main result Theorem 2.4, cannot be deleted in general (except in special cases, such as the limit tail field coinciding with the tail field). We shall take our X_n's of the form I_{E_n} for some sequence (E_n) of measurable sets. It is easily seen that such a sequence is determining if and only if it converges weakly. Suppose then (X_n) converges weakly to f; the limit tail field $\mathscr{A}$ of (X_n) is simply $\sigma(f)$, the field generated by f. (X_n) is then s.c.c.d. if and only if the sequence $(\mathscr{P}(E_n \cap S))$ converges strongly for all sets S of positive measure.

We now take the unit square $[0,1] \times [0,1]$ with Lebesque measure on its Borel sets as our probability space. Let $f(x,y)=x$ for all $0 \leqq x, y \leqq 1$, we shall construct a sequence of Borel measurable subsets (E_n) of the square so that

$$I_{E_n} \to f \text{ weakly.} \tag{59}$$

This shows that $\mathscr{A}$, the limit tail field of (I_{E_n}), is simply the field of vertical sets; i.e. $\mathscr{A} = \{S \times [0,1] \colon S \text{ is a Borel subset of } [0,1]\}$. We shall construct the E_n's to have two further properties: first,

$$\int_0^1 I_{E_n}(x,y)\,dy = x \qquad \text{for almost all } x, \qquad \text{all } n. \tag{60}$$

(60) means that $\mathscr{P}E_n = f$ a.e. for all n; that is, (I_{E_n}) is conditionally i.i.d. with respect to $\mathscr{A}$.

Second, let $S = \{(x,y) \colon 0 \leqq x, y \leqq \frac{1}{2}\}$. We shall insure that

$$(\mathscr{P}(E_n \cap S)) \text{ has no subsequence convergent in probability.} \tag{61}$$

Thus (59)-(61) imply that (X_n) has the desired properties, where $X_n = I_{E_n}$ for all n.

For each n and $1 \leqq j, k \leqq 2^n$ let $B^n_{j,k} = \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right] \times \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right]$. To obtain (59) it suffices to have

$$P(E_n \cap B^n_{j,k}) = \iint_{B^n_{j,k}} x\,dx\,dy. \tag{62}$$

Indeed, it then follows that also

$$P(E_r \cap B^n_{j,k}) = \iint_{B^n_{j,k}} x\,dx\,dy$$

for all $r \geqq n$, hence $I_{E_n} \to f$ weakly. For $1 \leqq j \leqq 2^{n-1}$, let $\beta_j = \sqrt{j^2 - j + 1/2}$. Then $\frac{\beta_j}{2^n} = \sqrt{\frac{a^2+b^2}{2}}$ where $a = \frac{j-1}{2^n}$ and $b = \frac{j}{2^n}$. Let $G_n = \bigcup_{j=1}^{2^{n-1}} \left[\frac{j-1}{2^n}, \frac{\beta_j}{2^n}\right)$. To obtain (61) it suffices to have

$$\mathscr{P}(E_n \cap S) = f I_{G_n}. \tag{63}$$

Indeed, if $F_n=\bigcup_{j=1}^{2^{n-1}}\left[\frac{j-1}{2^n},\frac{j-\frac{1}{2}}{2^n}\right]$, then relative to $[0,\frac{1}{2})$, (F_n) is an independent sequence of sets each of probability $\frac{1}{2}$; hence $(I_{F_n}\cdot f)$ has no subsequence convergent in probability. But for each n,

$$F_n\subset G_n \quad\text{and}\quad P(G_n\sim F_n)=\frac{1}{2^n}\sum_{j=1}^{2^{n-1}}\left(\sqrt{j^2-j+\tfrac{1}{2}}-(j-\tfrac{1}{2})\right)$$
$$=\frac{1}{2^n}\frac{1}{4}\sum_{j=1}^{2^{n-1}}\frac{1}{\sqrt{j^2-j+\frac{1}{2}}+j-\frac{1}{2}}\sim\frac{n}{2^n}$$

which implies that $\Sigma\|(I_{F_n}-I_{G_n})f\|_1<\infty$, so $(I_{F_n}-I_{G_n})f\to 0$ ae. and hence $(f\,I_{G_n})$ has no subsequence convergent in probability.

We now define E_n by specifying

$$E_{j,k}^n=E_n\cap B_{j,k}^n \qquad\text{for all } 1\leqq j,k\leqq 2^n.$$

Suppose first that $j\geqq 2^{n-1}+1$. Define a, b, c, d by

$$a=\frac{j-1}{2^n},\qquad b=\frac{j}{2^n},\qquad c=\frac{k-1}{2^n}\quad\text{and}\quad d=\frac{k}{2^n}.$$

Define $E_{j,k}^n$ by

$$E_{j,k}^n=\left\{(x,y): c\leqq y\leqq\frac{1}{2^n}x+c,\ a\leqq x<b\right\}. \tag{64}$$

Suppose next that $j\leqq 2^{n-1}$. *If* $k\leqq 2^{n-1}$ *also*, letting a, b, c, d as before (64), define $E_{j,k}^n$ by

$$E_{j,k}^n=\left\{(x,y): c\leqq y\leqq\frac{x}{2^{n-1}}+c,\ a\leqq x<\sqrt{\frac{a^2+b^2}{2}}\right\}. \tag{65}$$

Finally, if $k>2^{n-1}$, set

$$E_{j,k}^n=\left\{(x,y): c\leqq y\leqq\frac{x}{2^{n-1}}+c,\ \sqrt{\frac{a^2+b^2}{2}}\leqq x<b\right\}. \tag{66}$$

We now set $E_n=\cup\{E_{j,k}^n: 1\leqq j,k\leqq 2^n\}$. It remains only to verify the conditions (60), (62) and (63). Suppose first that $2^n\geqq j\geqq 2^{n-1}+1$. Then by (64), $\int I_{E_{j,k}}^n(x,y)\,dy=\frac{x}{2^n}$ if $\frac{j-1}{2^n}\leqq x\leqq\frac{j}{2^n}$. But then for such x, $I_{E_n}(x,y)=\sum_{k=1}^{2^n}I_{E_{j,k}^n}(x,y)$ hence (60) follows if $\frac{1}{2}\leqq x<1$. Also by (64), for any $1\leqq k\leqq 2^n$, $P(E_{j,k}^n)=\frac{1}{2^n}\int_a^b x\,dx$ $=\iint_{B_{j,k}^n}x\,dx\,dy$ where $a=\frac{j-1}{2^n}$, $b=\frac{j}{2^n}$, hence (62) holds.

Suppose now that $j\leqq 2^{n-1}$. If $\frac{j-1}{2^n}\leqq x<\frac{\beta_j}{2^n}$, then by (65), $\int I_{E_{j,k}^n}(x,y)\,dy$ $=\frac{x}{2^{n-1}}$ if $k\leqq 2^{n-1}$; if $k>2^{n-1}$, $\int I_{E_{j,k}^n}(x,y)\,dy=0$. Thus $\int_0^1 I_{E_n}(x,y)\,dy=x$, i.e. (60) holds; also (63) holds on G_n. If $k\leqq 2^{n-1}$,

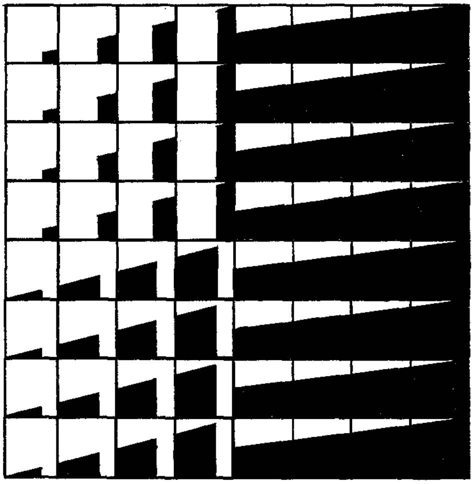

Fig. 1. E_4 is the shaded part of the square

$$P(E_n \cap B^n_{j,k}) = P(E^n_{j,k}) = \frac{1}{2^{n-1}} \int_a^{\sqrt{\frac{a^2+b^2}{2}}} x\,dx = \frac{1}{2^n}\,\frac{b^2-a^2}{2} = \frac{1}{2^n}\int_a^b x\,dx = \iint_{B^n_{j,k}} x\,dx\,dy$$

(where a,b are as before (64)), hence (62) holds, Finally if $\frac{\beta_j}{2^n} \leqq x < \frac{j}{2^n}$, $I_{E^n_{j,k}}(x,y) = 0$ if $k \leqq 2^{n-1}$ and again $\int I_{E^n_{j,k}}(x,y)\,dy = \frac{x}{2^{n-1}}$ if $k > 2^{n-1}$, so (60) holds here too. Since $E_n \cap S = \bigcup_{j,k=1}^{2^{n-1}} E^n_{j,k}$, we also obtain that $\int_0^1 I_{E_n}(x,y) \cap S\,dy = 0$ if $x \notin G_n$, hence (63) holds. At last if $k > 2^{n-1}$,

$$P(E_n \cap B^n_{j,k}) = \frac{1}{2^{n-1}} \int_{\sqrt{\frac{a^2+b^2}{2}}}^{b} x\,dx = \frac{1}{2^n}\,\frac{b^2-a^2}{2} = \iint_{B^n_{j,k}} x\,dx\,dy,$$

so (62) holds in this case also.

This completes the proof that the example has the desired properties. A picture is surely far more intelligible than the above analytical details; we indicate a sketch of the set E_4 in Fig. 1 above.

References

1. Aldous, D.J.: Limit theorems for subsequences of arbitrarily-dependent sequences of random variables. Z. Wahrscheinlichkeitstheor. Verw. Geb. **40**, 59–82 (1977)
2. Aldous, D.J.: Subspaces of L_1 via random measures. Trans. Am. Math. Soc. **267**, 445–463 (1981)
3. Baxter, J.R., Chacon, R.V.: Compactness of stopping times. Z. Wahrscheinlichkeitstheor. Verw. Geb. **40**, 169–182 (1977)

4. Berkes, I., Péter, E.: Exchangeable r.v.'s and the subsequence principle. Z. Wahrscheinlichkeitstheor. Verw. Geb. [To appear]
5. Dacunha-Castelle, D.: Indiscernability and exchangeability in L^p-spaces. Proc. Aarhus Seminar on random series, convex sets and geometry of Banach spaces. Aarhus Universitet, various publications **25**, 50–56 (1975)
6. Dacunha-Castelle, D., Krivine, J.L.: Sous-espace de L^1. Isr. J. Math. **26**, 320–351 (1977)
7. Garling, D.J.H.: Stable Banach spaces, random measures an Orlicz function spaces. Probability Measures on Groups (H. Heyer, ed.). Lect. Notes Math., no. **928.** Berlin-Heidelberg-New York: Springer 1982
8. Henson, W.: (in preparation).
9. Kadec, M.Ĭ, Pełczynški, A.: Bases, lacunary sequences and complemented subspaces in the spaces L_p. Stud. Math. **21**, 161–176 (1962)
10. Krivine, J.L.: Sous-espaces de dimension finie des espaces de Banach reticulés. Ann. Math. **104**, 1–29 (1976)
11. Krivine, J.L., Maurey, B.: Espaces de Banach stables. Isr. J. Math. **39**, 273–295 (1981)
12. Loève, M.: Probability Theory. Princeton: D. Van Nostrand 1963
13. Lindenstrauss, J.: A short proof of Liapounoff's convexity theorem. J. Math. Mech. **15**, 971–972 (1966)
14. Lindenstrauss, J., Tzafriri, L.: Classical Banach Spaces; Lect. Notes Math. no. **338**. Berlin-Heidelberg-New York: Springer 1973
15. Maharam, D.: On homogeneous measure algebras. Proc. Natl. Acad. Sci USA **28**, 108–111 (1942)
16. Rosenthal, H.P.: On the span in L^p of sequences of independent random variables #II. Proc. 6th Berkeley Symposium on Math. Stat. and Probab. **2**, 149–167 (1972)
17. Rosenthal, H.P.: On subspaces of L^p. Ann. Math. **97**, 344–373 (1973)
18. Rosenthal, H.P.: On a theorem of J.L. Krivine concerning block finite-representability of l^p in general Banach spaces. J. Func. Anal. **28**, 197–225 (1978)
19. Strassen, V.: The existence of probability measures with given marginals. Ann. Math. Stat **36**, 423–429 (1965)

Received December 15, 1976; in revised form January 22, 1981 and in final form March 19, 1985