Probab. Th. Rel. Fields 73, 395–413 (1986)

Probability Theory and Related Fields

Exchangeable Random Variables and the Subsequence Principle

István Berkes[1] and Erika Péter[2]

1 Mathematical Institute of the Hungarian Academy of Sciences, 1053 Budapest, Reáltanoda u. 13–15, Hungary
2 Department of Mathematics, Institute of Building, Dávid Ferenc u. 6, 1113 Budapest, Hungary

Summary. Call a sequence $\{X_n\}$ of r.v.'s ε-exchangeable if on the same probability space there exists an exchangeable sequence $\{Y_n\}$ such that $P(|X_n - Y_n| \geqq \varepsilon) \leqq \varepsilon$ for all n. We prove that any tight sequence $\{X_n\}$ defined on a rich enough probability space contains ε-exchangeable subsequences for every $\varepsilon > 0$. The distribution of the approximating exchangeable sequences is also described in terms of $\{X_n\}$. Our results give a convenient way to prove limit theorems for subsequences of general r.v. sequences. In particular, they provide a simplified way to prove the subsequence theorems of Aldous [1] and lead also to various extensions.

1. Introduction

It has been known for a long time that sufficiently rarified subsequences of every norm-bounded sequence of r.v.'s behave like mixed i.i.d. sequences. A heuristic principle related to this phenomenon was formulated by Chatterji (see [7]):

Subsequence Principle. *Let T be a limit theorem valid for all sequences of i.i.d. r.v.'s belonging to an integrability class L defined by the finiteness of a norm $\| \ \|_L$. Then if $\{X_n\}$ is an arbitrary (dependent) sequence of r.v.'s satisfying $\sup_n \|X_n\|_L < +\infty$ then there exists a subsequence $\{X_{n_k}\}$ satisfying T in a mixed form.*

For example, if $\{X_n\}$ is an arbitrary sequence of r.v.'s with $\sup_n \|X_n\|_1 < +\infty$ then there is a subsequence $\{X_{n_k}\}$ satisfying the strong law of large numbers in a mixed (randomized) form, i.e.

$$\frac{1}{N} \sum_{k=1}^{N} X_{n_k} \to X \qquad \text{a.s.}$$

for some integrable r.v. X (see [13]). If $\sup_n \|X_n\|_2 < +\infty$ then there exists a subsequence $\{X_{n_k}\}$ obeying the central limit theorem and the law of the iterated logarithm, again in a "randomized" form:

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} (X_{n_k} - X) \xrightarrow{\mathcal{D}} N(0, Y)$$

$$\overline{\lim_{N\to\infty}} (2N \log\log N)^{-1/2} \sum_{k=1}^{N} (X_{n_k} - X) = Y^{1/2} \quad \text{a.s.}$$

for some r.v.'s X and $Y \geqq 0$; here $N(0, Y)$ denotes Gaussian distribution with mean zero and (random) variance Y, i.e., $N(0, Y)$ is the distribution of $\xi \cdot Y^{1/2}$ where ξ is an $N(0,1)$ variable independent of Y (see [2, 8, 9, 12]). Several further special cases of the principle have been proved by ad hoc methods, see [12] for an extensive bibliography. Although it is natural to expect a general theorem behind these examples, nothing beyond special cases has been obtained until 1977 when, using a new and powerful method, Aldous showed that the subsequence principle is valid for all distributional and a.s. limit theorems satisfying mild technical conditions. In his paper [1] he gave an interesting analysis of the structure of limit theorems and also gave examples of simple (although artificial) limit theorems T for which the subsequence principle is not valid.

The purpose of the present paper is to prove theorems describing precisely the structure of sparse subsequences of general r.v. sequences. As we shall see, our theorems provide a simplified approach to the subsequence principle and lead to various extensions. To state the first result, call a sequence $\{X_n\}$ of r.v.'s ε-exchangeable if on the same probability space there exists an exchangeable sequence $\{Y_n\}$ such that $P\{|X_n - Y_n| \geqq \varepsilon\} \leqq \varepsilon$ for all n. Then our theorem can be formulated as follows:

Theorem 1. *Let $\{X_n\}$ be a sequence of r.v.'s bounded in probability and let ε_n be a positive numerical sequence tending to zero. Then, if the underlying probability space is large enough, there exists a subsequence $\{X_{n_k}\}$ such that, for all $l \geqq 1$, the sequence $\{X_{n_l}, X_{n_{l+1}}, \ldots\}$ is ε_l-exchangeable.*

Thus, every tight sequence $\{X_n\}$ contains a subsequence $\{X_{n_k}\}$ which is "exchangeable at infinity" in the sense that for any large l, the tail sequence $\{X_{n_l}, X_{n_{l+1}}, \ldots\}$ is a small perturbation of an exchangeable sequence. By De Finetti's theorem every exchangeable sequence is conditionally i.i.d. with respect to its tail field whence it follows easily that limit theorems for i.i.d. r.v.'s continue to hold for exchangeable sequences in a mixed form. (See [1], p. 63 and p. 65 for a formalization and quick proof of this principle for all a.s. and purely distributional limit theorems.) Theorem 1 extends a very large class of these limit theorems for subsequences, thereby establishing a general version of the subsequence principle; this applies also for many limit theorems outside of the class considered by Aldous. Examples will be given in Sect. 3.

The idea to use near exchangeability to derive limit theorems for subsequences is due to Aldous. However, he used distributional exchangeability properties of subsequences (see Lemma 12 of [1]) and applied a rarifying procedure depending on the particular limit theorem we want to prove. The fact that Theorem 1 involves pointwise approximation simplifies the situation considerably and leads directly to limit theorems. It also enables one, as we shall prove in a subsequent paper, to give converses of Aldous' theorem, in particular to characterize domains of attraction for subsequences of r.v.'s.

In [1] Aldous gave an example of a uniformly bounded sequence $\{X_n\}$ such that no subsequence $\{X_{n_k}\}$ and exchangeable sequence $\{Y_k\}$ can satisfy $X_{n_k}-Y_k\xrightarrow{P}0$. Theorem 1 shows, on the other hand, that given any tight sequence $\{X_n\}$ and $\varepsilon>0$, there exists a subsequence $\{X_{n_k}\}$ and an exchangeable sequence $\{Y_k\}$ such that

$$P\{|X_{n_k}-Y_k|\geqq\varepsilon\}\leqq\varepsilon \qquad k=1,2,\dots \tag{1.1}$$

By Aldous' example, the last relation cannot be replaced by

$$P\{|X_{n_k}-Y_k|\geqq\delta_k\}\leqq\delta_k \qquad k=1,2,\dots$$

for any $\delta_k\to 0$ and thus (1.1) is best possible. Of course, to use the above approximation to derive limit theorems for subsequences, we need to know the distribution of the exchangeable sequence $\{Y_k\}$ for every $\varepsilon>0$. Our next theorem provides this information, showing that $\{Y_k\}$ can always be chosen as a finite mixture of i.i.d. sequences with explicitly given distribution functions. To formulate the result, we introduce some terminology.

Definition. A sequence $\{X_n\}$ of r.v.'s on $(\Omega,\mathscr{F},P)$ is determining if it has a limit distribution on each set $A\subset\Omega$ of positive probability.

For a determining sequence $\{X_n\}$ we put

$$F_A(t)=\lim_{n\to\infty} P(X_n<t\,|\,A) \tag{1.2}$$

where the limit exists at continuity points t of F_A. It is not difficult to prove (see e.g. [4]) that every tight sequence $\{X_n\}$ of r.v.'s contains a determining subsequence. Hence we can restrict our attention to determining sequences instead of stochastically bounded ones whenever it is convenient.

We now introduce the notion of "strong exchangeability at infinity", playing a central role in our paper.

Definition. Let $\{\varepsilon_n\}$ be a positive numerical sequence tending to zero. We say that the sequence $\{X_n\}$ of r.v.'s is strongly exchangeable at infinity with speed ε_n if the sequence $\{X_n\}$ is determining, the r.v.'s X_n are all simple (i.e. take only finitely many values), $\sigma\{X_1\}\subset\sigma\{X_2\}\subset\dots$ and the following is true:

For any $k>1$ the sets $A=\{X_{k-1}=c\}$ (where c runs through the range of X_{k-1}) can be divided into two classes Γ_1 and Γ_2 such that

(i) $\sum_{A\in\Gamma_1} P(A)\leqq\varepsilon_k$.

(ii) For any $A\in\Gamma_2$ there exist P_A-independent r.v.'s $\{Y_j^{(A)},\ j=k,k+1,\dots\}$ defined on A with common distribution function F_A such that

$$P_A\{|X_j-Y_j^{(A)}|\geqq\varepsilon_k\}\leqq\varepsilon_k \qquad j=k,k+1,\dots. \tag{1.3}$$

Here P_A denotes conditional probability with respect to A; F_A is defined by (1.2).

Thus, $\{X_n\}$ is strongly exchangeable at infinity if for every large k, the sequence $\{X_k, X_{k+1}, \dots\}$ is a small perturbation of an i.i.d. sequence on 'almost all' sets A of the form $A=\{X_{k-1}=c\}$. Clearly, this is a rather strong structural property; its consequences will be studied in Sect. 3.

We formulate now the main result of our paper.

Theorem 2. *Let $\{X_n\}$ be a sequence of r.v.'s bounded in probability and let $\{\varepsilon_n\}$ be a positive numerical sequence tending to zero. Then, if the underlying probability space is rich enough, there exists a subsequence $\{X_{n_k}\}$ and a sequence $\{Y_k\}$ of r.v.'s such that $\{Y_k\}$ is strongly exchangeable at infinity with speed ε_k and*

$$\sum_{k=1}^{\infty} |X_{n_k} - Y_k| < \infty \quad \text{a.s.}$$

If $\{X_n\}$ is strongly exchangeable at infinity with speed ε_n then the sequence $\{X_k, X_{k+1}, \dots\}$ is obviously $2\varepsilon_k$-exchangeable. Hence Theorem 2 implies Theorem 1, together with a description of the approximating exchangeable sequences. The main advantage of Theorem 2 over Theorem 1 is that it yields an approximation of subsequences $\{X_{n_k}\}$ directly by i.i.d. sequences and thus it implies limit theorems for $\{X_{n_k}\}$ by using the theory of independent r.v.'s, without referring to exchangeability. It is worth mentioning that the i.i.d. approximation given by Theorem 2 for lacunary sequences $\{X_{n_k}\}$ on subsets A of the probability space is generally optimal for each A in the same sense as Theorem 1 is optimal on the whole probability space. By Example 2 of [4] there exists a tight sequence $\{X_n\}$ on a suitable probability space $(\Omega, \mathscr{F}, P\}$ such that no subsequence $\{X_{n_k}\}$, set $A\in\mathscr{F}$ with $P(A)>0$ and i.i.d. sequence $\{Y_k^{(A)}\}$ defined on A can satisfy $X_{n_k} - Y_k^{(A)} \xrightarrow{P} 0$ on A. See [4] for more information on this point, in particular for a characterization of tight sequences $\{X_n\}$ having a subsequence $\{X_{n_k}\}$ allowing the approximation $X_{n_k} - Y_k \xrightarrow{P} 0$ with an i.i.d. resp. exchangeable $\{Y_k\}$.

2. Proof of Theorem 2

In what follows, ρ stands for the Prohorov distance of probability measures i.e. for any two probability measures P and Q on the Borel sets of the real line we put

$$\rho(P, Q) = \inf\{\varepsilon > 0 \colon P(A) \leqq Q(A^\varepsilon) + \varepsilon \text{ and } Q(A) \leqq P(A^\varepsilon) + \varepsilon \text{ for all Borel sets } A \subset R^1\}.$$

Here A^ε denotes the ε-neighbourhood of A i.e. $A^\varepsilon = \{x\in R^1 \colon |x-y| < \varepsilon \text{ for some } y\in A\}$. It is known (see [5], Appendix III) that ρ metrizes the weak convergence of probability measures i.e. $\rho(P_n, P) \to 0$ iff $P_n \to P$ weakly.

Lemma 1. *Let $X_1, X_2, \dots$ be a sequence of simple r.v.'s and denote by $\mu_n^{(X_1, \dots, X_{n-1})}$ the conditional distribution of X_n given $X_1, \dots, X_{n-1}$. Assume that there exist distributions ν_n, $n=1, 2, \dots$ such that*

$$P\{\rho(\mu_n^{(X_1, \dots, X_{n-1})}, \nu_n) \geqq \delta_n\} \leqq \delta_n \qquad n=1, 2, \dots$$

for some constants $\delta_n > 0$. *Then, if the underlying probability space is rich enough, there exist independent r.v.'s* $Y_1, Y_2, \ldots$ *such that the distribution of* Y_n *is* ν_n *and*

$$P(|X_n - Y_n| \geqq 6\delta_n) \leqq 6\delta_n \qquad n = 1, 2, \ldots$$

This lemma is implicit in [3]; its proof is identical with that of Theorems 1 and 2 of the just mentioned paper.

Lemma 2. *Given any* $\varepsilon > 0$, *a r.v.* X *and a finite* σ-*field* $\mathscr{F}$, *there exists a simple r.v.* Y *such that* $\mathscr{F} \subset \sigma\{Y\}$ *and* $P(|X - Y| \geqq \varepsilon) \leqq \varepsilon$.

Proof. Choose first a simple r.v. Z such that $P(|X - Z| \geqq \varepsilon/2) \leqq \varepsilon/2$. Let $z_1, \ldots, z_k$ be the values of Z and $A_i = \{Z = z_i\}$, $1 \leqq i \leqq k$. Then if $B_1, \ldots, B_l$ are the atoms of $\mathscr{F}$, choose numbers $|c_{i,j}| \leqq \varepsilon/2$, $1 \leqq i \leqq k$, $1 \leqq j \leqq l$ such that all the numbers $z_i + c_{i,j}$ are different. Let V be the r.v. taking $c_{i,j}$ on $A_i \cap B_j$. Then obviously $Y = Z + V$ satisfies the requirements.

Lemma 3. *If* $\{X_n\}$ *is determining with limit distribution function* F *then the weak limit*

$$\eta_t = \lim_{n \to \infty} \chi(X_n < t)^1 \tag{2.1}$$

exists for all continuity points t *of* F. *Here* $\chi\{\cdot\}$ *denotes the indicator function of the set in brackets.*

Proof. See [4], Proposition (2.1).

Lemma 4. *Let* μ *and* ν *be probability measures on the real line, let* $x_1 < x_2 < \ldots < x_k$ *and set* $I_0 = (-\infty, x_1)$, $I_j = [x_j, x_{j+1})$ $(1 \leqq j \leqq k-1)$, $I_k = [x_k, +\infty)$. *Assume that*

$$\max_{j=1,\ldots,k-1} (x_{j+1} - x_j) < \varepsilon \tag{2.2}$$

$$\mu(I_0) + \mu(I_k) < \varepsilon \tag{2.3}$$

$$\sum_{j=0}^{k} |\mu(I_j) - \nu(I_j)| < \varepsilon. \tag{2.4}$$

Then

$$\rho(\mu, \nu) \leqq 2\varepsilon.$$

Proof. Let $B \subset R^1$ be a Borel set and let H_B denote the set of those integers $0 \leqq j \leqq k$ such that $B \cap I_j$ is not empty. Then using (2.2)–(2.4) we get

$$\mu(B) \leqq \sum_{j \in H_B} \mu(I_j) \leqq \varepsilon + \sum_{\substack{j \in H_B \\ 1 \leqq j \leqq k-1}} \mu(I_j)$$

$$\leqq 2\varepsilon + \sum_{\substack{j \in H_B \\ 1 \leqq j \leqq k-1}} \nu(I_j) = 2\varepsilon + \sum_{\substack{j \in H_B \\ 1 \leqq j \leqq k-1}} \nu(I_j \cap B^\varepsilon) \leqq 2\varepsilon + \nu(B^\varepsilon)$$

---

[1] For integrable r.v.'s ξ_n, ξ we say $\xi_n \to \xi$ weakly if $E(\xi_n \eta) \to E(\xi\eta)$ for any bounded r.v. η. This notion is not to be confused with the weak convergence of probability measures and distributions. To avoid ambiguity, in the sequel we keep the term "weak" for convergence of r.v.'s; weak convergence of probability measures and distributions will be referred to as distributional convergence.

where the equality in the fourth step follows from the fact that for $1 \leqq j \leqq k-1$ the length of I_j is $<\varepsilon$ and thus if such an I_j contains a point of B (i.e. $j \in H_B$) then $I_j \subset B^\varepsilon$.

The following lemma is a trivial consequence of the Markov inequality.

Lemma 5. *Let $X \geqq 0$ be a r.v. with $EX \leqq \varepsilon$ and let $\{A_i, i=1, \ldots, l\}$ be a partition of the probability space with all A_i's having positive probability. Then the total probability of those A_i's such that*

$$E_{A_i}(X) \leqq \sqrt{\varepsilon}$$

is at least $1-\sqrt{\varepsilon}$. Here E_A denotes contitional expectation given A.

Proof of Theorem 2. Step 1. By Lemma 2 there exist simple r.v.'s Y_k such that

$$P(|X_k - Y_k| \geqq 2^{-k}) \leqq 2^{-k} \quad (k=1,2,\ldots) \quad \text{and} \quad \sigma\{Y_1\} \subset \sigma(Y_2\} \subset \ldots.$$

Hence without loss of generality we may assume that the r.v.'s X_n themselves are all simple and

$$\sigma\{X_1\} \subset \sigma\{X_2\} \subset \ldots. \tag{2.5}$$

As any sequence $\{X_n\}$ bounded in probability contains a determining subsequence (see e.g. [4]) we can also assume without loss of generality that $\{X_n\}$ itself is determining. Finally, there is no loss of generality in assuming that the sequence $\{\varepsilon_n\}$ is decreasing. Let $\mu_{n,A}$ denote the conditional distribution of X_n given A and let $\mu_A = \lim_{n\to\infty} \mu_{n,A}$; denote by $F_A(x)$ the distribution function of μ_A. As $\{X_n\}$ is determining, these quantities are defined for any $A \subset \Omega$ with $P(A) > 0$.

2. We can choose a subsequence $\{X_{n_k}\}$ such that for any $k > 1$,

$$\rho(\mu_{n_l,A}, \mu_A) \leqq \varepsilon_k \quad \text{for any } A \in \sigma\{X_{n_{k-1}}\} \text{ and } l \geqq k \tag{2.6}$$

Indeed, let $X_{n_1} = X_1$ and assume that $X_{n_1}, \ldots, X_{n_{k-1}}$ are already constructed. Let $A_1, A_2, \ldots, A_r$ be the sets of the finite σ-field $\sigma\{X_{n_{k-1}}\}$. Since $\rho(\mu_{n,A_i}, \mu_{A_i}) \to 0$ as $n \to \infty$ for every $1 \leqq i \leqq r$, there exists an integer $m_k > 0$ such that

$$\rho(\mu_{n,A_i}, \mu_{A_i}) \leqq \varepsilon_k \quad \text{for all } n \geqq m_k \text{ and } 1 \leqq i \leqq r.$$

Set $X_{n_k} = X_{m_k}$. Obviously, the sequence $\{X_{n_k}\}$ and all of its subsequences satisfy (2.6).

3. Let $F(x) = F_\Omega(x)$ be the limit distribution function of X_n relative to Ω and let C_F be the set of continuity points of $F(x)$. Choose the number $L_k > 0$ so that $L_k \in C_F$, $-L_k \in C_F$ and

$$F(L_k) - F(-L_k) \geqq 1 - \varepsilon_k. \tag{2.7}$$

We show that the following statement is true:

Let $\{A_i,\ i=1, \ldots, l\}$ be any partition of the probability space with all A_i's having positive probability. Then the total probability of those A_i's for which the inequality

$$\mu_{A_i}([-2L_k, 2L_k]) \geqq 1-\sqrt{\varepsilon_k} \tag{2.8}$$

holds, is at least $1-\sqrt{\varepsilon_k}$.

Proof. By Lemma 3, the weak limits η_t in (2.1) exist for each $t \in C_F$, Obviously $0 \leqq \eta_t \leqq 1$ a.s. and $t<t'$ implies $\eta_t \leqq \eta_{t'}$ a.s. Further, for any $A \subset \Omega$ with $P(A)>0$ (2.1) implies $P(X_n<t \mid A) \to E_A(\eta_t)$ for $t \in C_F$ and thus

$$F_A(t)=E_A(\eta_t) \tag{2.9}$$

for any t which is a continuity point of both F and F_A. As $L_k \in C_F$, $-L_k \in C_F$, (2.7) and (2.9) imply $E(\eta_{L_k}-\eta_{-L_k}) \geqq 1-\varepsilon_k$ and applying Lemma 5 for the non-negative r.v. $1-\eta_{L_k}+\eta_{-L_k}$ we get that the total probability of those A_i's such that

$$E_{A_i}(\eta_{L_k}-\eta_{-L_k}) \geqq 1-\sqrt{\varepsilon_k} \tag{2.10}$$

is at least $1-\sqrt{\varepsilon_k}$. Choose an A_i satisfying (2.10) and let $x \in (L_k, 2L_k)$ be such that x and $-x$ are continuity points of both F and F_{A_i}. Then (2.10) remains valid if $\eta_{L_k}-\eta_{-L_k}$ is replaced by $\eta_x-\eta_{-x}$ and using (2.9) we get

$$\mu_{A_i}([-x, x]) \geqq 1-\sqrt{\varepsilon_k}.$$

The last relation evidently implies (2.8).

4. Let $F(x)$ and C_F be as in step 3; by Lemma 3 the weak limits η_t in (2.1) exist for each $t \in C_F$. Define $\eta_I=\eta_{t_2}-\eta_{t_1}$ for any interval $I=[t_1, t_2)$ with $t_1, t_2 \in C_F$. We allow here also the values $t_i=\pm\infty$ by setting $\eta_{-\infty}=0$, $\eta_\infty=1$. We now construct a subsequence $\{X_{n_k}\}$, together with a sequence $H_1 \subset H_2 \subset \ldots$ of finite subsets of C_F such that setting

$$H_k=\{x_1^{(k)}, \ldots, x_{q_k}^{(k)}\}$$

and

$$U_k=\{I=[a, b): a<b \text{ and } a, b \in H_k \cup \{+\infty\} \cup \{-\infty\}\}$$

the following properties hold:

$$0<x_{v+1}^{(k)}-x_v^{(k)} \leqq \varepsilon_k, \qquad 1 \leqq v \leqq q_k-1 \tag{2.11}$$

$$\sum_{I \in U_k} |P(X_{n_l} \in I \mid X_{n_{k-1}})-E(\eta_I \mid X_{n_{k-1}})| \leqq \varepsilon_k \qquad (l \geqq k) \tag{2.12}$$

$$P\{\sum_{I \in U_{k+1}} |E(\eta_I \mid X_{n_l})-\eta_I| \geqq \varepsilon_k \mid X_{n_{k-1}}\} \leqq \varepsilon_k \qquad (l \geqq k). \tag{2.13}$$

Moreover, with probability $\geqq 1-2\sqrt{\varepsilon_k}$ we have

$$\inf_{l \geqq k} \mathrm{P}(x_1^{(k)}<X_{n_l}<x_{q_k}^{(k)} \mid X_{n_{k-1}}) \geqq 1-2\sqrt{\varepsilon_k}. \tag{2.14}$$

Construction. Let L_k be the numbers defined in step 3 and choose finite sets

$$H_k=\{x_1^{(k)}, \ldots, x_{q_k}^{(k)}\} \qquad k=1, 2, \ldots$$

such that $H_1 \subset H_2 \subset \ldots$, $x_1^{(k)} \leqq -3L_k$, $x_{q_k}^{(k)} \geqq 3L_k$ and (2.11) holds. Since C_F is dense, the H_k's can be chosen to be subsets of C_F. Now we define a subsequence $\{X_{n_k}\}$ by induction as follows. Set $X_{n_1} = X_1$ and assume that $X_{n_1}, \ldots, X_{n_{k-1}}$ are already constructed. Let $c_1, \ldots, c_r$ be the possible values of $X_{n_{k-1}}$ and put $A_i = \{X_{n_{k-1}} = c_i\}$, $1 \leqq i \leqq r$. By step 3 there is a set $\Gamma \subset \{A_1, \ldots, A_r\}$ such that

$$\sum_{A_i \in \Gamma} P(A_i) \geqq 1 - \sqrt{\varepsilon_k} \tag{2.15}$$

and

$$\mu_{A_i}([-2L_k, 2L_k]) \geqq 1 - \sqrt{\varepsilon_k} \quad \text{for } A_i \in \Gamma.$$

As μ_A is the limit distribution of X_n given A, the last inequality implies

$$\liminf_{n \to \infty} P_{A_i}(|X_n| < 3L_k) \geqq 1 - \sqrt{\varepsilon_k} \quad \text{for } A_i \in \Gamma$$

and thus there exists an integer $s_k > 0$ such that

$$P_{A_i}(|X_n| < 3L_k) \geqq 1 - 2\sqrt{\varepsilon_k} \quad \text{for } n \geqq s_k \quad \text{and all } A_i \in \Gamma. \tag{2.16}$$

Note further that by (2.1)

$$\chi(X_n \in I) \to \eta_I \quad \text{weakly} \tag{2.17}$$

for any $I = [a, b)$ where $a < b$ and $a, b \in C_F \cup \{+\infty\} \cup \{-\infty\}$. Since $H_k \subset C_F$, (2.17) implies

$$\sum_{I \in U_k} |P(X_n \in I \mid A) - E(\eta_I \mid A)| \to 0 \quad \text{as } n \to \infty$$

for any $A \subset \Omega$ with $P(A) > 0$. Choosing $A = A_1, \ldots, A_r$ we get that there exists an integer $s_k^* > 0$ such that

$$\sum_{I \in U_k} |P(X_n \in I \mid X_{n_{k-1}}) - E(\eta_I \mid X_{n_{k-1}})| \leqq \varepsilon_k \quad \text{for } n \geqq s_k^*. \tag{2.18}$$

Set $\mathscr{F}^* = \sigma\{X_1, X_2, \ldots\}$, then η_I is $\mathscr{F}^*$ measurable and thus using (2.5) and the martingale convergence theorem we get

$$E(\eta_I \mid X_n) \to E(\eta_I \mid \mathscr{F}^*) = \eta_I \quad \text{a.s. as } n \to \infty$$

for any fixed $I \in U_{k+1}$. This shows that for any fixed $1 \leqq i \leqq r$

$$P_{A_i}\{\sum_{I \in U_{k+1}} |E(\eta_I \mid X_n) - \eta_I| \geqq \varepsilon_k\} \to 0 \quad \text{as } n \to \infty$$

and thus there exists an integer $s_k^{**} > 0$ such that

$$P\{\sum_{I \in U_{k+1}} |E(\eta_I \mid X_n) - \eta_I| \geqq \varepsilon_k \mid X_{n_{k-1}}\} \leqq \varepsilon_k \quad \text{for } n \geqq s_k^{**}. \tag{2.19}$$

Choose $n_k = \max(s_k, s_k^*, s_k^{**})$. This completes the k-th step of induction and thus the construction of the subsequence $\{X_{n_k}\}$ is also completed. Now relations (2.12), (2.13) follow evidently from (2.18), (2.19); further, since the r.v. on the left side of (2.14) is identical with $\inf_{l\geqq k} P_{A_i}(x_1^{(k)}<X_{n_l}<x_{q_k}^{(k)})$ on each set A_i, $1\leqq i\leqq r$, relation (2.14) follows from (2.15), (2.16) and $x_1^{(k)}\leqq -3L_k$, $x_{q_k}^{(k)}\geqq 3L_k$.

5. From now on, let $\{X_{n_k}\}$ denote a subsequence of $\{X_n\}$ satisfying the properties guaranteed in steps 2 and 4. We show that $\{X_{n_k}\}$ satisfies the following statement:

For every $k>1$ the sets $A=\{X_{n_{k-1}}=c\}$ can be divided into two classes Γ_1 and Γ_2 such that

$$\text{(i)} \sum_{A\in\Gamma_1} P(A)\leqq 3\sqrt{\varepsilon_{k-1}}. \tag{2.20}$$

(ii) For each $A\in\Gamma_2$ we have

$$P_A\{\sum_{I\in U_k} |P_A(X_{n_l}\in I\,|\,X_{n_k},\dots,X_{n_{l-1}})-\eta_I|\geqq 2\varepsilon_k\}\leqq\varepsilon_k \tag{2.21}$$

$$P_A\{\sum_{I\in U_k} |P_A(X_{n_l}\in I)-\eta_I|\geqq 2\varepsilon_{k-1}\}\leqq\sqrt{\varepsilon_{k-1}} \tag{2.22}$$

and

$$P_A(x_1^{(k)}<X_{n_l}<x_{q_k}^{(k)})\geqq 1-2\sqrt{\varepsilon_k} \tag{2.23}$$

for every $l>k$.

Proof. Let c be a possible value of $X_{n_{k-1}}$. Then setting $A=\{X_{n_{k-1}}=c\}$ and using the identity $P_A(B\,|\,C)=P(B\,|\,A\,C)$ we get for $l>k$

$$\begin{aligned} &P_A(X_{n_l}\in I\,|\,X_{n_k}=a_k,\dots,X_{n_{l-1}}=a_{l-1})\\ &=P(X_{n_l}\in I\,|\,X_{n_{k-1}}=c,\,X_{n_k}=a_k,\dots,X_{n_{l-1}}=a_{l-1})\\ &=P(X_{n_l}\in I\,|\,X_{n_{l-1}}=a_{l-1})\end{aligned}$$

where in the last step we used (2.5). We thus see that the r.v.'s $P_A(X_{n_l}\in I\,|\,X_{n_k},\dots,X_{n_{l-1}})$ and $P(X_{n_l}\in I\,|\,X_{n_{l-1}})$ are identical on A and thus on A we have, using (2.12), $\varepsilon_k\downarrow$ and $U_1\subset U_2\subset\dots$

$$\begin{aligned} &\sum_{I\in U_k} |P_A(X_{n_l}\in I\,|\,X_{n_k},\dots,X_{n_{l-1}})-\eta_I|\\ &\quad=\sum_{I\in U_k} |P(X_{n_l}\in I\,|\,X_{n_{l-1}})-\eta_I|\\ &\quad\leqq\sum_{I\in U_k} |P(X_{n_l}\in I\,|\,X_{n_{l-1}})-E(\eta_I\,|\,X_{n_{l-1}})|+\sum_{I\in U_k} |E(\eta_I\,|\,X_{n_{l-1}})-\eta_I|\\ &\quad\leqq\varepsilon_k+\sum_{I\in U_k} |E(\eta_I\,|\,X_{n_{l-1}})-\eta_I|=\varepsilon_k+\tau_{k,l},\quad\text{say.}\end{aligned} \tag{2.24}$$

By (2.13), $l>k$ and $U_k\subset U_{k+1}$ we have

$$P_A(|\tau_{k,l}|\geqq\varepsilon_k)\leqq\varepsilon_k$$

and thus the P_A-probability that the first expression of (2.24) exceeds $2\varepsilon_k$ is at most ε_k, proving (2.21). (Note that (2.21) holds for *all* $A=\{X_{n_{k-1}}=c\}$ without exception.) To get (2.22) and (2.23) we note that on A we have, using (2.12),

$$\begin{aligned}\sum_{I\in U_k} |P_A(X_{n_l}\in I)-\eta_I| &= \sum_{I\in U_k} |P(X_{n_l}\in I|X_{n_{k-1}})-\eta_I| \\ &\leqq \sum_{I\in U_k} |P(X_{n_l}\in I|X_{n_{k-1}})-E(\eta_I|X_{n_{k-1}})|+\sum_{I\in U_k} |E(\eta_I|X_{n_{k-1}})-\eta_I| \\ &\leqq \varepsilon_k+\sum_{I\in U_k} |E(\eta_I|X_{n_{k-1}})-\eta_I|=\varepsilon_k+\zeta_k, \quad \text{say.}\end{aligned} \tag{2.25}$$

Integrating (2.13) and setting $l=k$ we get

$$P(\zeta_{k+1}\geqq \varepsilon_k)\leqq \varepsilon_k.$$

Changing k to $k-1$ and applying Lemma 5 to the indicator function of $\{\zeta_k\geqq \varepsilon_{k-1}\}$ it follows that the total probability of those sets $A=\{X_{n_{k-1}}=c\}$ for which the inequality

$$P_A(\zeta_k\geqq \varepsilon_{k-1})\leqq \sqrt{\varepsilon_{k-1}} \tag{2.26}$$

holds, is at least $1-\sqrt{\varepsilon_{k-1}}$. On the other hand, since inequality (2.14) holds with probability $\geqq 1-2\sqrt{\varepsilon_k}$, the total probability of those sets $A=\{X_{n_{k-1}}=c\}$ for which

$$\inf_{l\geqq k} P_A(x_1^{(k)}<X_{n_l}<x_{q_k}^{(k)})\geqq 1-2\sqrt{\varepsilon_k} \tag{2.27}$$

holds, is $\geqq 1-2\sqrt{\varepsilon_k}$. Hence using the monotonicity of ε_k it follows that the events $A=\{X_{n_{k-1}}=c\}$ can be divided into two classes Γ_1 and Γ_2 such that (2.20) holds and for $A\in\Gamma_2$ we have (2.26) and (2.27). It remains now to notice that (2.25) and (2.26) imply (2.22) for $A\in\Gamma_2$.

6. Fix $k>1$, $A\in\Gamma_2$ and consider the sequence $\{X_{n_k}, X_{n_{k+1}}, \ldots\}$ as a sequence of r.v.'s on the probability space $(A, P_A, \mathscr{F}|A)$. We claim that there exist P_A-independent r.v.'s $Y_k^{(A)}, Y_{k+1}^{(A)}, \ldots$ defined on this space, all having distribution μ_A such that

$$P_A(|X_{n_j}-Y_j^{(A)}|\geqq 54\sqrt{\varepsilon_{k-1}})\leqq 54\sqrt{\varepsilon_{k-1}} \qquad j=k, k+1, \ldots \tag{2.28}$$

This implies that the sequence $\{X_{n_k}\}$ is strongly exchangeable at infinity with speed $54\sqrt{\varepsilon_{k-1}}$ which is only notationally different from our theorem since ε_n can be chosen arbitrarily.

To prove our claim above we define, for any $l>k$, the probability measures $\mu_l^{(X_{n_k}, \ldots, X_{n_{l-1}})}$ and μ_l (the first depending on chance) by

$$\mu_l^{(X_{n_k}, \ldots, X_{n_{l-1}})}(G)=P_A(X_{n_l}\in G|X_{n_k}, \ldots, X_{n_{l-1}})$$
$$\mu_l(G)=P_A(X_{n_l}\in G) \qquad (G\subset R^1 \text{ Borel-set})$$

To estimate the Prohorov distance of $\mu_l^{(X_{n_k}, \ldots, X_{n_{l-1}})}$ and μ_l we note that (2.21), (2.22) and $\varepsilon_n\downarrow$ imply

$$\sum_{I\in U_k} |P_A(X_{n_l}\in I|X_{n_k}, \ldots, X_{n_{l-1}})-P_A(X_{n_l}\in I)|\leqq 4\varepsilon_{k-1}$$

with P_A-probability $\geqq 1-2\sqrt{\varepsilon_{k-1}}$. Also by (2.23),

$$P_A(X_{n_l}\notin[x_1^{(k)}, x_{q_k}^{(k)}])\leqq 2\sqrt{\varepsilon_k}.$$

The last two relations, $\varepsilon_k\downarrow$, (2.11) and Lemma 4 imply

$$\rho(\mu_l^{(X_{n_k}, \ldots, X_{n_{l-1}})}, \mu_l) \leqq 8\sqrt{\varepsilon_{k-1}}$$

with P_A-probability at least $1-2\sqrt{\varepsilon_{k-1}}$. Further, as our set $A\in\Gamma_2$ is an atom of $\sigma\{X_{n_{k-1}}\}$, (2.6) implies

$$\rho(\mu_l, \mu_A) \leqq \varepsilon_k \qquad (l\geqq k).$$

By the triangle inequality for ρ,

$$P_A\{\rho(\mu_l^{(X_{n_k}, \ldots, X_{n_{l-1}})}, \mu_A) \geqq 9\sqrt{\varepsilon_{k-1}}\} \leqq 2\sqrt{\varepsilon_{k-1}} \qquad l=k,\ k+1, \ldots$$

and our claim above follows from Lemma 1. Hence the proof of Theorem 2 is completed.

3. Applications

In this section we show that Theorem 2 implies a very large class of limit theorems for subsequences of r.v.'s. As a first application, we derive one of Aldous' general theorems ([1], Theorem 6) stating the validity of the subsequence principle for distributional limit theorems. Then we give examples showing that the method applies for a.s. limit theorems and also for many limit theorems lying outside Aldous' formalization.

To derive limit theorems from Theorem 2, we need a few preliminary remarks. Let $\{X_n\}$ be a determining sequence of r.v.'s with limit distribution function F. By Lemma 3, the weak limits η_t in (2.1) exist for all continuity points t of F. Let H be a dense countable set of continuity points of F. The limits η_t are determined only with probability one and following the method of [14], Lemma 6.1.4. one can construct versions of η_t, $t\in H$ such that for every fixed $\omega\in\Omega$, the function $\eta_t(\omega)$, $t\in H$ extends to a distribution function $F_\omega(t)$. Let μ_ω denote the probability measure corresponding to the distribution function F_ω; we call μ_ω the *limit random distribution of* $\{X_n\}$. This notion is due to Aldous and plays an important role in the investigations of [1]. Obviously, $\mu_\omega(B)$ is a τ-measurable r.v. for any Borel set $B\subset R^1$ where τ is the tail field of $\{X_n\}$.

For any $A\subset\Omega$ with $P(A)>0$, let $\mu_{n,A}$ denote the conditional distribution of X_n given A. Since $\{X_n\}$ is determining, $\mu_{n,A}$ converges, as $n\to\infty$, to a distribution μ_A; let F_A denote the distribution function of μ_A. It is easy to prove the following

Lemma 6. (a) *For any A of positive probability and any real t we have*

$$F_A(t)=E_A(F_\omega(t)). \tag{3.1}$$

(b) *Let $\psi(x)$, $x\in(-\infty, +\infty)$ be any nonnegative, piecewise continuously differentiable function. If $\int_{-\infty}^{+\infty}\psi(x)dF(x)<+\infty$ then $\int_{-\infty}^{+\infty}\psi(x)dF_\omega(x)<+\infty$ a.s. and*

$$E\left(\int_{-\infty}^{+\infty}\psi(x)dF_\omega(x)\right)=\int_{-\infty}^{+\infty}\psi(x)dF(x).$$

Proof. (3.1) is an easy consequence of (2.9); the equality of part (b) follows from (3.1) by integration by parts.

If the r.v.'s X_n are all simple and (2.5) holds, the random measure μ_ω can be obtained as a limit of random measures of simple structure. Put

$$F_\omega^{(n)}(t) = E(F.(t) \mid X_n)(\omega) \qquad t \in R^1 \tag{3.2}$$

where $F.(t)$ denotes the r.v. which takes $F_\omega(t)$ at ω. Denote by $\mu_\omega^{(n)}$ the probability measure corresponding to the distribution function $F_\omega^{(n)}$. (3.2) and part (a) of Lemma 6 imply that if A is an atom of $\sigma\{X_n\}$ then $F_\omega^{(n)}$ is identical with F_A for $\omega \in A$. Hence, for any $n \geqq 1$ the range of the random map $\mu_\omega^{(n)}$ is finite.

Lemma 7. *For almost all ω we have*

$$\mu_\omega^{(n)} \xrightarrow{\mathscr{D}} \mu_\omega \qquad \text{as } n \to \infty. \tag{3.3}$$

Proof. For any fixed rational r, $F_\omega^{(n)}(r) \to F_\omega(r)$ for almost all ω by (3.2) and the martingale convergence theorem. (Note that $F.(r)$ is $\sigma\{X_1, X_2, \ldots\}$ measurable.) Hence, for almost all ω, the relation $F_\omega^{(n)}(r) \to F_\omega(r)$ holds simultaneously for all rational r and thus (3.3) is valid.

With ρ denoting the Prohorov distance as in Sect. 2, (3.3) implies $\rho(\mu_\omega^{(n)}, \mu_\omega) \to 0$ a.s. and thus there exists a numerical sequence $\varepsilon_n \downarrow 0$ such that

$$P\{\omega\colon \rho(\mu_\omega^{(n)}, \mu_\omega) \geqq \varepsilon_n\} \leqq \varepsilon_n \qquad n = 1, 2, \ldots \tag{3.4}$$

If we replace $\{X_n\}$ by a subsequence $\{X_{m_n}\}$ then μ_ω remains the same and $\mu_\omega^{(n)}$ changes to $\mu_\omega^{(m_n)}$. Hence for the sequence $\{X_{m_n}\}$, (3.4) holds with ε_n replaced by ε_{m_n}. Thus we have

Lemma 8. *By passing to a suitable subsequence of $\{X_n\}$, the speed of convergence to zero of ε_n in (3.4) can be made as rapid as desired.*

We are now in a position to derive limit theorems from Theorem 2. We begin with proving a slightly weakened version of Aldous' general "distributional" theorem, Theorem 6 of [1]. To state this result, we need a few definitions from [1]. Let $\mathscr{M}$ denote the set of probability measures on the real line, equipped by the topology generated by the Prohorov metric.

Definition. A weak limit theorem of i.i.d. r.v.'s is a system $T = (f_1, f_2, \ldots, S, \{G_\mu, \mu \in S\})$ where

a) S is a Borel subset of $\mathscr{M}$,

b) For each $k \geqq 1$, $f_k = f_k(x_1, x_2, \ldots, \mu)$ is a real function on $R^\infty \times S$, measurable in the product topology (in R we take the usual topology),

c) For each $\mu \in S$, G_μ is a probability distribution on the real line such that the function $\mu \mapsto G_\mu$ is measurable (with respect to the Borel fields in S and $\mathscr{M}$),

d) If $\mu \in S$ and $X_1, X_2, \ldots$ are independent r.v.'s with common distribution μ then

$$f_k(X_1, X_2, \ldots, \mu) \xrightarrow{\mathscr{D}} G_\mu \qquad \text{as } k \to \infty \tag{3.5}$$

For example, the central limit theorem corresponds to the case S = class of distributions with finite variance, $f_k(x_1, x_2, \ldots, \mu) = k^{-1/2}\,(x_1 + \ldots + x_k - k \cdot E\mu)$,

G_μ = normal distribution with mean zero and the same variance as μ. In general, S describes the class of distributions to which the limit theorem T applies (the "condition" of the theorem); the limit theorem itself is expressed by relation (3.5).

For the functions f_k Aldous makes some additional assumptions ensuring that f_k are continuous in the x_i's and also that for large k, f_k depends weakly on the first few variables x_i. In what follows, we give a slightly strengthened form of these assumptions.

Definition. We call the weak limit theorem $T=(f_1, f_2, \ldots, S; \{G_\mu, \mu\in S\})$ regular if there exist sequences p_k, q_k of positive integers tending to $+\infty$ such that $p_k \leqq q_k$ and

(i) $f_k(x_1, x_2, \ldots, \mu)$ depends only on $x_{p_k}, \ldots, x_{q_k}, \mu$

(ii) f_k satisfies the Lipschitz condition

$$|f_k(x_{p_k}, \ldots, x_{q_k}, \mu) - f_k(x'_{p_k}, \ldots, x'_{q_k}, \mu)| \leqq \sum_{i=p_k}^{q_k} |x_i - x'_i|^\alpha \tag{3.6}$$

for some $0<\alpha\leqq 1$.

For regular limit theorems, (3.5) takes on the form

$$f_k(X_{p_k}, \ldots, X_{q_k}, \mu) \xrightarrow{\mathscr{D}} G_\mu \qquad \text{as } k\to\infty. \tag{3.7}$$

Clearly, the central limit theorem, as given above by the corresponding f_k, S, G_μ, is not regular. However, the validity of the relation $k^{-1/2}(X_1+\ldots+X_k-k\cdot E\mu) \xrightarrow{\mathscr{D}} N(0, D\mu)$ is not affected by deleting the terms $X_1, \ldots, X_{[k^{1/4}]-1}$ from the sum $X_1+\ldots+X_k$ and thus in the central limit theorem we can also choose $f_k=k^{-1/2}(x_{[k^{1/4}]}+\ldots+x_k-k\cdot E\mu)$. Obviously, the theorem becomes regular with this choice.

We need one final definition, namely that of the mixture of probability distributions. Let $(\Omega, \mathscr{F}, P)$ be a probability space and for any $\omega\in\Omega$ let a distribution ν_ω be given such that the map $\omega\to\nu_\omega$ is measurable with respect to the Borel field in $\mathscr{M}$. Then it is easily seen that the set function ν^* defined by

$$\nu^*(B)=E\,\nu_\omega(B)=\int_\Omega \nu_\omega(B)\,dP(\omega)$$

is a probability measure on the Borel sets of the real line. We call ν^* the mixture of the ν_ω's with weight function P; we use the notation

$$\nu^*=\int \nu_\omega\, dP(\omega).$$

We can now formulate our

Theorem 3. *Let $\{X_n\}$ be a determining sequence of r.v.'s with limit random distribution μ_ω. Let $T=(f_1, f_2, \ldots, S; \{G_\mu, \mu\in S\})$ be a regular weak limit theorem and assume that $\mu_\omega\in S$ for almost all ω. Then there exists a subsequence $\{X_{n_k}\}$ such that*

$$f_k(X_{n_1}(\omega), X_{n_2}(\omega), \ldots, \mu_\omega) \xrightarrow{\mathscr{D}} \nu^* \qquad \textit{as } k\to\infty \tag{3.8}$$

where $\nu^*=\int G_{\mu_\omega}\, dP(\omega)$.

Except the slight difference in the technical assumptions made on the functions f_k, Theorem 3 is identical with Theorem 6 of [1]. Considering again the case of the central limit theorem, the assumption $\mu_\omega \in S$ a.s. reduces in this case to the requirement that μ_ω has a finite variance for almost every ω; this guarantees that the r.v. $X(\omega)=E\mu_\omega$ is a.s. finite. Then (3.8) becomes

$$(X_{n_1}+\ldots+X_{n_k}-k\cdot X)/\sqrt{k}\xrightarrow{\mathscr{D}}\nu^*$$

where ν^* is the mixed normal distribution whose characteristic function is $\int_0^\infty \exp(-ct^2/2)\,dH(c)$ where H is the distribution function of the r.v. $D^2\mu_\omega$.

As we shall see below, Theorem 3 is an easy consequence of Theorem 2. To simplify the writing, let $f_k(\mu)$ denote, for any $\mu\in S$, the distribution of the r.v. $f_k(Y_1, Y_2, \ldots, \mu)$ where $Y_1, Y_2, \ldots$ are independent r.v.'s with common distribution μ. (Clearly, this distribution depends only on f_k and μ.) Then (3.5) can be written as

$$f_k(\mu)\xrightarrow{\mathscr{D}} G_\mu \qquad \text{for any } \mu\in S. \tag{3.9}$$

Assume for the sake of simplicity that $f_k(x_{p_k}, \ldots, x_{q_k}, \mu)$ does not depend on μ. Let ε_n tend to zero monotonically and so rapidly that

$$\varepsilon_{p_k-1}^{\alpha}\, q_k \leqq k^{-1} \qquad k=1,2,\ldots \tag{3.10}$$

where p_k, q_k, α are the quantities appearing in (3.6). From Theorem 2 it follows that there exists a subsequence $\{X_{n_k}\}$ and a sequence $\{Y_k\}$ of r.v.'s such that $\{Y_k\}$ is strongly exchangeable at infinity with speed ε_k and $X_{n_k}=Y_k+\tau_k$ where $\sum_{k=1}^\infty |\tau_k|^\alpha<+\infty$ a.s. Hence there is no loss of generality in assuming that $\{X_n\}$ itself is strongly exchangeable at infinity with speed ε_n. Similarly, on the basis of Lemma 8 we can assume without loss of generality that (3.4) holds for the sequence $\{X_n\}$. We show that $\{X_n\}$ satisfies the conclusion of Theorem 3 i.e.

$$f_k(X_{p_k}, \ldots, X_{q_k})\xrightarrow{\mathscr{D}}\int G_{\mu_\omega}\,dP(\omega) \qquad \text{as } k\to\infty. \tag{3.11}$$

The heuristic reason of (3.11) is the following. Let $\mathscr{F}_k=\sigma\{X_{p_k-1}\}$ and denote Γ_1, Γ_2 the two classes of atoms of $\mathscr{F}_k$ guaranteed by the definition of strong exchangeability at infinity for the index p_k. Then for $A\in\Gamma_2$ i.e. on "almost all" atoms of $\mathscr{F}_k$ the sequence $X_{p_k}, \ldots, X_{q_k}$ can be approximated by an i.i.d. sequence $Y_{p_k}^{(A)}, \ldots, Y_{q_k}^{(A)}$ with common distribution μ_A such that the order of magnitude of $X_j-Y_j^{(A)}$ is $\leqq\varepsilon_{p_k}$ for all j; more precisely,

$$P_A(|X_j-Y_j^{(A)}|\geqq\varepsilon_{p_k})\leqq\varepsilon_{p_k} \qquad j=p_k,\ p_k+1, \ldots \tag{3.12}$$

By (3.10), (3.12) and the Lipschitz condition (3.6), the (conditional) distribution of $f_k(X_{p_k}, \ldots, X_{q_k})$ on A is almost equal to the distribution of $f_k(Y_{p_k}^{(A)}, \ldots, Y_{q_k}^{(A)})$ i.e. to $f_k(\mu_A)$ and thus distribution of $f_k(X_{p_k}, \ldots, X_{q_k})$ on the whole probability space is

$$\approx\sum_A f_k(\mu_A)P(A)=\int f_k(\mu_\omega^{(p_k-1)})\,dP(\omega)$$

where the last equality follows from the fact that for $\omega \in A$, $\mu_\omega^{(p_k-1)} = \mu_A$. (See the remark preceding Lemma 7.) Thus the distribution of $f_k(X_{p_k}, \ldots, Y_{q_k})$ is close to the first of the following three distributions:

$$\int f_k(\mu_\omega^{(p_k-1)})\, dP(\omega), \qquad \int f_k(\mu_\omega)\, dP(\omega), \qquad \int G_{\mu_\omega}\, dP(\omega). \tag{3.13}$$

On the other hand, the integrals in (3.13) are close to each other for large k. For the first two integrals this follows from (3.4) while the closeness of the second and third integral follows from the fact that $f_k(\mu_\omega) \xrightarrow{\mathcal{D}} G_{\mu_\omega}$ for almost all ω by (3.9) and $\mu_\omega \in S$ a.s. Hence for large k the distribution of $f_k(X_{p_k}, \ldots, X_{q_k})$ is also close to the third integral in (3.13) i.e. (3.11) holds.

The above heuristic argument can be made precise without any difficulty, we only have to estimate the closeness of the considered distributions. We shall do this in the Prohorov metric by using the following simple remarks:

A) The Prohorov distance of two measures μ and ν is $<\varepsilon$ if and only if on some probability space there exist r.v.'s X and Y with distribution μ and ν such that $P(|X-Y|>\varepsilon)<\varepsilon$.

B) If $\rho(\mu, \nu)<\varepsilon$ then $\rho(f_k(\mu), f_k(\nu)) \leqq \varepsilon^\alpha q_k$ where α and q_k are the quantities appearing in (3.6).

C) Let $\mu_1, \ldots, \mu_r$ and $\nu_1, \ldots, \nu_r$ be probability distributions, let further $p_1, \ldots, p_r$ be nonnegative numbers with $\sum p_i = 1$. Assume that the sum of those p_i's such that $\rho(\mu_i, \nu_i) \geqq \varepsilon$ is at most ε. Then the Prohorov distance of $\sum_{i=1}^r p_i \mu_i$ and $\sum_{i=1}^r p_i \nu_i$ is $\leqq 3\varepsilon$.

D) Let μ_ω and ν_ω be random measures (i.e. measurable maps from a probability space $(\Omega, \mathcal{F}, P)$ to $\mathcal{M}$) such that $P\{\omega: \rho(\mu_\omega, \nu_\omega) \geqq \varepsilon\} \leqq \varepsilon$. Then the Prohorov distance of $\int \mu_\omega\, dP(\omega)$ and $\int \nu_\omega\, dP(\omega)$ is $\leqq 3\varepsilon$.

Statement A) is Strassen's theorem (see [15]); statement B) follows immediately from statement A) and the Lipschitz condition (3.6). Statements C) and D) are almost evident; C) is of course a special case of D).

Now our heuristic argument can be made precise as follows. By (3.12) and the Lipschitz condition on f_k we have

$$P_A\{|f_k(X_{p_k}, \ldots, X_{q_k}) - f_k(Y_{p_k}^{(A)}, \ldots, Y_{q_k}^{(A)})| \geqq \varepsilon_{p_k}^\alpha q_k\} \leqq \varepsilon_{p_k} \cdot q_k, \qquad A \in \Gamma_2$$

and thus $\sum_{A \in \Gamma_1} P(A) \leqq \varepsilon_{p_k}$, (3.10) and statements A) and C) imply that the Prohorov distance of the distributions of $f_k(X_{p_k}, \ldots, X_{q_k})$ and $\sum_A f_k(\mu_A) P(A)$ i.e. the first integral in (3.13) is $\leqq 3\varepsilon_{p_k}^\alpha q_k \leqq 3k^{-1}$. On the other hand, (3.4), (3.10) and statements B) and D) imply that the Prohorov distance of the first two integrals in (3.13) is $\leqq 3\varepsilon_{p_k-1}^\alpha q_k \leqq 3k^{-1}$. Finally, (3.9) and $\mu_\omega \in S$ a.s. imply $\rho(f_k(\mu_\omega), G_{\mu_\omega}) \to 0$ a.s. and thus there exists a numerical sequence $\delta_k \downarrow 0$ such that

$$P\{\omega: \rho(f_k(\mu_\omega), G_{\mu_\omega}) \geqq \delta_k\} \leqq \delta_k \qquad k=1, 2, \ldots$$

Then by statement D) the Prohorov distance of the last two integrals in (3.13) is $\leqq 3\delta_k$. Adding up our estimates, it follows that the Prohorov distance of the distribution of $f_k(X_{p_k}, \ldots, X_{q_k})$ and $\int G_{\mu_\omega} dP(\omega)$ is $\leqq 6k^{-1} + 3\delta_k \to 0$ i.e. (3.11) is valid. Thus Theorem 3 is proved.

We turn now to a.s. limit theorems. Instead of treating this class generally as weak limit theorems above, we illustrate the method on the case of the law of the iterated logarithm (convered also by Aldous' general theorems in [1]). We prove

Theorem 4. *Let $\{X_n\}$ be a determining sequence of r.v.'s with limit random distribution μ_ω. Assume that μ_ω has finite variance for almost all ω and set*

$$X(\omega) = E\mu_\omega = \int x\, d\mu_\omega(x), \qquad Y(\omega) = D\mu_\omega = \{\int x^2\, d\mu_\omega(x) - (\int x\, d\mu_\omega(x))^2\}^{1/2}.$$

Then there exists a subsequence $\{X_{n_k}\}$ such that

$$\overline{\lim_{N\to\infty}}\, (2N \log\log N)^{-1/2} \sum_{k=1}^{N} (X_{n_k} - X) = Y \qquad \text{a.s.}$$

A sufficient condition for the a.s. finiteness of $D\mu_\omega$ is $\sup_n EX_n^2 < +\infty$ (cf. part b) of Lemma 6). This condition is not necessary: there exists e.g. an exchangeable sequence $\{X_n\}$ with $EX_i^2 = +\infty$ such that $D\mu_\omega < +\infty$ a.s.

Theorem 4 follows from Theorem 2 essentially in the same way as Theorem 3, just we have to reformulate the law of the iterated logarithm in a "regular" form. This is given by the following simple lemma where $L(n)$ denotes $(2n \log\log n)^{1/2}$.

Lemma 9. *Let $Y_1, Y_2, \ldots$ be a sequence of r.v.'s satisfying* $\sup_n E|Y_n| < +\infty$; put $S_n = Y_1 + \ldots + Y_n$, $S_{k,l} = Y_{k+1} + \ldots + Y_l$ $(k<l)$. Then $\overline{\lim_{n\to\infty}}\, S_n/L(n) = 1$ *a.s. if and only if for any $\varepsilon > 0$ there exists an increasing sequence $\{m_k\}$ of integers such that $m_k \geqq 5^k$ and*

$$P\{\max_{m_k \leqq j \leqq m_{k+1}} S_{k,j}/L(j) > 1 + \varepsilon\} \leqq 2^{-k} \qquad (k \geqq k_0) \tag{3.14}$$

and

$$P\{\max_{m_k \leqq j \leqq m_{k+1}} S_{k,j}/L(j) < 1 - \varepsilon\} \leqq 2^{-k} \qquad (k \geqq k_0). \tag{3.15}$$

Proof. Assume first that (3.14) and (3.15) hold for some $m_k \geqq 5^k$. By $\sup E|Y_n| < +\infty$ we have

$$P\{|S_k| \geqq \sqrt{m_k}\} \leqq 2^{-(k+1)}, \qquad k \geqq k_0. \tag{3.16}$$

Using (3.14), (3.15), (3.16) and the Borel-Cantelli lemma we get

$$1 - 2\varepsilon \leqq \overline{\lim_{n\to\infty}}\, S_n/L(n) \leqq 1 + 2\varepsilon \quad \text{a.s.}$$

proving the "if" part of Lemma 9. To prove the converse part, assume $\overline{\lim_{n\to\infty}}\, S_n/L(n) = 1$ a.s. We construct a sequence $\{m_k\}$ of integers such that $m_k \geqq 5^k$ and

$$P\{\sup_{j \geqq m_k} S_j/L(j) > 1 + \varepsilon/2\} \leqq 2^{-(k+1)} \tag{3.17}$$

$$P\{\max_{m_k \leqq j \leqq m_{k+1}} S_j/L(j) < 1-\varepsilon/2\} \leqq 2^{-(k+1)} \tag{3.18}$$

for all $k \geqq 1$. Choose m_1 so large that $m_1 \geqq 5$ and (3.17) holds for $k=1$. Assume that $m_1 < m_2 < \ldots < m_N$ are already constructed such that $m_k \geqq 5^k$ $(1 \leqq k \leqq N)$ holds and (3.17) and (3.18) are valid for $k \leqq N$ and $k \leqq N-1$, respectively. Then choose m_{N+1} so large that $m_{N+1} > \max(5^{N+1}, m_N)$ and (3.17) and (3.18) hold for $k=N+1$ and $k=N$, respectively. (This is possible since $\overline{\lim}\, S_n/L(n)=1$ a.s.) Obviously, the so constructed sequence $\{m_k\}$ satisfies (3.17) and (3.18) for all $k \geqq 1$ and thus, in view of (3.16), relations (3.14) and (3.15) are also valid. Hence Lemma 9 is proved.

Obviously, the r.v.

$$\max_{m_k \leqq j \leqq m_{k+1}} S_{k,j}/L(j)$$

is a function of the r.v.'s $Y_{k+1}, \ldots, Y_{m_{k+1}}$ and thus (3.14) and (3.15) can be written in the form

$$P\{f_k(Y_{k+1}, \ldots, Y_{m_{k+1}}) > 1+\varepsilon\} \leqq 2^{-k} \quad (k \geqq k_0) \tag{3.19}$$

$$P\{f_k(Y_{k+1}, \ldots, Y_{m_{k+1}}) < 1-\varepsilon\} \leqq 2^{-k} \quad (k \geqq k_0) \tag{3.20}$$

with suitable (smooth) functions f_k. Relations (3.19) and (3.20) are very similar to (3.7), the only difference is that weak convergence in (3.7) is replaced by a sequence of probability inequalities for the same r.v.'s. As one can easily check, the argument proving Theorem 3 remains valid in this new situation with inessential changes and Theorem 4 follows.

Closer examination shows that a large class of a.s. limit theorems can be reformulated as a sequence of probability inequalities similar to (3.19), (3.20). (As a matter of fact, the convergence relation (3.7) can also be written in such a form.) Instead of elaborating on this point, however, we show that a similar reformulation applies actually to many limit theorems lying outside of the class of weak and a.s. limit theorems and thus the subsequence principle holds for such limit theorems as well. As an example, we show the reformulation of two "refined" distributional limit theorems, namely that of the central limit theorem with remainder and Cramér's large deviation theorem.

Example 1. Let $\{X_n\}$ be a sequence of r.v.'s satisfying $\sup_n E|X_n|^p < +\infty$ for some $p > 2$. Then, as a trivial calculation shows, the relations

$$\sup_t |P\{(X_1 + \ldots + X_n)/\sqrt{n} < t\} - \Phi(t)| = \mathcal{O}(n^{-\alpha}) \tag{3.21}$$

and

$$\sup_t |P\{(X_{[\log n]} + \ldots + X_n)/\sqrt{n} < t\} - \Phi(t)| = \mathcal{O}(n^{-\alpha}) \tag{3.22}$$

are equivalent provided $0 < \alpha < p/(2p+2)$ and thus for such α relation (3.21) can be reformulated as a sequence of inequalities

$$\Phi(t) - Kn^{-\alpha} \leqq P\{(X_{[\log n]} + \ldots + X_n)/\sqrt{n} < t\} \leqq \Phi(t) + Kn^{-\alpha} \tag{3.23}$$

where the n-th inequality contains only the r.v.'s $X_{[\log n]}, \ldots, X_n$.

Since (3.23) is the same type as (3.19), (3.20), our method applies and yields a version of the central limit theorem with remainder for subsequences of norm-bounded sequences. Note that there is a loss of accuracy caused by the regularization procedure (i.e. discarding the r.v.'s $X_1, \ldots, X_{[\log n]-1}$ from (3.21)): the remainder term obtained in this way cannot be better than $\mathcal{O}(n^{-p/(2p+2)})$ for r.v.'s with uniformly bounded p-th moments. As the classical remainder term in the i.i.d. case is $\mathcal{O}(n^{-(p-2)/2})$ for $2<p\leqq 3$, we do not lose anything for p lying close to 2 but e.g. for $p=3$ we get only a remainder term $\mathcal{O}(n^{-3/8+\varepsilon})$ instead of $\mathcal{O}(n^{-1/2})$. For $p\to\infty$ the remainder term approaches the optimal value $\mathcal{O}(n^{-1/2})$.

Example 2. Let $\{X_n\}$ be a sequence of r.v.'s satisfying $\sup_n E(\exp|X_n|^\alpha)<+\infty$ for some $\alpha>1$. Then the relations

$$P\{(X_1+\ldots+X_n)/\sqrt{n}>x\}\sim 1-\Phi(x) \quad \text{for } x=o(n^{1/6}) \tag{3.24}$$

and

$$P\{(X_{[\log n]}+\ldots+X_n)/\sqrt{n}>x\}\sim 1-\Phi(x) \quad \text{for } x=o(n^{1/6}) \tag{3.25}$$

are equivalent; (3.25) can again be written as a sequence of inequalities analogous to (3.19), (3.20) and our method applies. The same argument holds for the more general form of the large deviation theorem valid for $x=o(n^{1/2})$ (see e.g. [11], p. 520).

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. Berkes, I.: The law of the iterated logarithm for subsequences of random variables. Z. Wahrscheinlichkeitstheor. Verw. Geb. **30**, 209–215 (1974)
3. Berkes, I., Philipp, W.: Approximation theorems for independent and weakly dependent random vectors. Ann. Probab. **7**, 29–54 (1979)
4. Berkes, I., Rosenthal, H.P.: Almost exchangeable sequences of random variables. Z. Wahrscheinlichkeitstheor. Verw. Geb. **70**, 473–507 (1985)
5. Billingsley, P.: Convergence of probability measures. New York: Wiley 1968
6. Chatterji, S.D.: A general strong law. Invent. Math. **9**, 235–245 (1970)
7. Chatterji, S.D.: Un principe de sous-suites dans la théorie des probabilités. Séminaire de probabilités VI, Strasbourg, Lecture Notes in Mathematics **258**, pp. 72–89. Berlin-Heidelberg-New York: Springer 1972
8. Chatterji, S.D.: A principle of subsequences in probability theory: The central limit theorem. Adv. Math. **13**, 31–54 (1974)
9. Chatterji, S.D.: A subsequence principle in probability theory II. The law of the iterated logarithm. Invent. Math. **25**, 241–251 (1974)
10. Dacunha-Castelle, D.: Indiscernability and exchangeability in L^p-spaces. In: Proceedings of the seminar on random series, convex sets and geometry of Banach spaces, pp. 50–56. Aarhus Universitet Matematisk Institut 1975
11. Feller, W.: An introduction to probability theory and its applications, Vol. II. New York: Wiley 1966
12. Gaposhkin, V.F.: Lacunary series and independent functions. (In Russian). Uspehi Mat. Nauk **21/6**, 3–82 (1966). English translation: Russian Math. Surveys **21/6**, 3–82 (1966)

13. Komlós, J.: A generalisation of a problem of Steinhaus. Acta Math. Acad. Sci. Hung. **18**, 217–229 (1967)
14. Révész, P.: The laws of large numbers. New York-London: Academic Press 1967
15. Strassen, V.: The existence of probability measures with given marginals. Ann. Math. Statist. **36**, 423–439 (1965)

Received December 28, 1982; in revised form October 4, 1983;
in final form November 1, 1985