The Annals of Probability
1986, Vol. 14, No. 3, 902–915

PRINCIPLE OF CONDITIONING IN LIMIT THEOREMS FOR SUMS OF RANDOM VARIABLES

By Adam Jakubowski

Nicholas Copernicus University

Let $\{X_{nk}: k \in \mathbb{N}, n \in \mathbb{N}\}$ be a double array of random variables adapted to the sequence of discrete filtrations $\{\{\mathscr{F}_{nk}: k \in \mathbb{N} \cup \{0\}\}: n \in \mathbb{N}\}$.

It is proved that for every weak limit theorem for sums of independent random variables there exists an analogous limit theorem which is valid for the system $(\{X_{nk}\}, \{\mathscr{F}_{nk}\})$ and obtained by conditioning expectations with respect to the past. Functional extensions and connections with the Martingale Invariance Principle are discussed.

1. Principle of Conditioning. The aim of this paper is to describe the "Principle of Conditioning," which is a way to derive limit theorems for arrays of arbitrary random variables, when limit theorems for arrays of independent random variables are given.

Let us consider a double array $\mathbb{X} = \{X_{nk}: k \in \mathbb{N}, n \in \mathbb{N}\}$ of random variables defined on some probability space $(\Omega, \mathscr{F}, P)$ and a double array $\mathbb{F} = \{\mathscr{F}_{nk}: k \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}, n \in \mathbb{N}\}$ of σ-subalgebras of $\mathscr{F}$. Suppose that in each row (i.e., for each $n \in \mathbb{N}$) the σ-algebras $\{\mathscr{F}_{nk}: k \in \mathbb{N}_0\}$ form a filtration: $\mathscr{F}_{n,k-1} \subset \mathscr{F}_{nk}$, $k \in \mathbb{N}$. The pair $(\mathbb{X}, \mathbb{F})$ is "adapted" if X_{nk} is $\mathscr{F}_{nk}$-measurable for all $n, k \in \mathbb{N}$.

For each $n \in \mathbb{N}$ let $\sigma_n: (\Omega, \mathscr{F}, P) \to \mathbb{N}_0$ be an $\{\mathscr{F}_{nk}: k \in \mathbb{N}_0\}$-stopping time. If $\mathbb{S} = \{\sigma_n: n \in \mathbb{N}\}$, then the triple $(\mathbb{X}, \mathbb{F}, \mathbb{S})$ will be called an adapted system.

Given $(\mathbb{X}, \mathbb{F}, \mathbb{S})$, define a sequence of row-sums

(1.1) $$S_n(\sigma_n)(\omega) = \sum_{1 \le k \le \sigma_n(\omega)} X_{nk}(\omega), \qquad n \in \mathbb{N}.$$

The following theorem is due to Brown (1971) for normed martingales and Brown and Eagleson (1971) in the general setting of martingale difference arrays (in fact in both papers nonrandom σ_n were considered).

1.0 Theorem. *Let $(\mathbb{X}, \mathbb{F})$ be a martingale difference array, i.e., for every $n, k \in \mathbb{N}$, $E|X_{nk}| < +\infty$ and*

(1.2) $$E(X_{nk} | \mathscr{F}_{n,k-1}) = 0.$$

Received September 1984; revised October 1985.

AMS 1980 subject classifications. 60F05, 60F17.

Key words and phrases. Weak limit theorems for sums of random variables, martingale difference arrays, Martingale Invariance Principle, processes with independent increments, random measures, tightness.

Suppose that for the adapted system $(\mathbb{X}, \mathbb{F}, \mathbb{S})$ *the following conditions* (1.3) *and* (1.4) *are satisfied*:

$$\sum_{1 \le k \le \sigma_n} E\left(X_{nk}^2 | \mathscr{F}_{n,k-1}\right) \to_P 1, \tag{1.3}$$

$$\sum_{1 \le k \le \sigma_n} E\left(X_{nk}^2 I(|X_{nk}| > \varepsilon) | \mathscr{F}_{n,k-1}\right) \to_P 0, \qquad \varepsilon > 0; \tag{1.4}$$

then for the row-sums $\{S_n(\sigma_n): n \in \mathbb{N}\}$ *the Central Limit Theorem holds*:

$$S_n(\sigma_n) \to_{\mathscr{D}} N(0,1).$$

Clearly, Theorem 1.0 contains the sufficiency part of the classical Lindeberg–Feller theorem: Consider an array $\{X_{nk}\}$ of independent random variables in rows and set $\mathscr{F}_{n0} = \{\phi, \Omega\}$, $\mathscr{F}_{nk} = \sigma(X_{n1}, X_{n2}, \dots, X_{nk})$, $\sigma_n = k_n =$ const., $n, k \in \mathbb{N}$.

Conversely, if Lindeberg–Feller CLT is given, then conditions (1.2)–(1.4) in Theorem 1.0 formally arise according to the heuristic rule that can be called the

PRINCIPLE OF CONDITIONING. *Given any limit theorem for sums of independent random variables we obtain a limit theorem for* $(\mathbb{X}, \mathbb{F}, \mathbb{S})$ *by replacing*:

(a) *the expectations of functions of summands by conditional expectations with respect to the past*;
(b) *the summation to constants by summations to stopping times*;
(c) *the convergence of numbers by convergence in probability of the random variables described in* (a) *and* (b).

Using the above terminology one can say that Brown and Eagleson (1971) showed the Principle of Conditioning for Lindeberg–Feller CLT. But the Principle of Conditioning is valid for more general limit theorems: The successive extensions were proved by Dvoretzky [(1971), CLT for summands without finite second moment], Brown and Eagleson [(1971), convergence to infinitely divisible laws with finite variation], Kłopotowski [(1977), convergence of random vectors to a general infinitely divisible distribution], Walk [(1977), Brown's Theorem in Hilbert space], Jakubowski [(1980), Hilbert case for conditionally infinitesimal summands]. The last mentioned paper contains a simple idea of the proof of the Principle of Conditioning while in all the earlier papers some more or less particular cases of the Principle of Conditioning were obtained by means of specially designed methods. However, in order to make the Principle of Conditioning a mathematical theorem, one must formalize the notion of limit theorem, restrict attention to conditions of certain types, etc. This has been done [see Jakubowski (1982)] but is too formal to be presented here. Instead we shall precisely describe the essential step (Theorem 1.1 below) and then prove a few conditional limit theorems by reducing them to Theorem 1.1.

Let $(\mathbb{X}, \mathbb{F}, \mathbb{S})$ be an adapted system. For $k, n \in \mathbb{N}$, let μ_{nk}: $\mathscr{B}^1 \times \Omega \to [0,1]$ be a regular version of the conditional distribution of X_{nk} given $\mathscr{F}_{n,k-1}$. Define the pointwise row-convolutions of μ_{nk}:

$$\mu_n(\sigma_n)(\cdot,\omega) = \mu_{n1}(\cdot,\omega) * \cdots * \mu_{n,\sigma_n(\omega)}(\cdot,\omega). \tag{1.5}$$

Then $\mu_n(\sigma_n)$ are random probability measures, i.e., measurable mappings of $(\Omega, \mathscr{F}, P)$ into the space $\mathscr{P}(R^1)$ of probability distributions on R^1, equipped with the topology of weak convergence. Note that $\mathscr{P}(R^1)$ can be considered as a complete and separable metric space [see Parthasarathy (1967)], hence the convergence in probability in $\mathscr{P}(R^1)$ (denoted by $\Rightarrow_P$) has the usual sense.

1.1 THEOREM. *Let* $(\mathbb{X}, \mathbb{F}, \mathbb{S})$ *be an adapted system and let the sequence* $\{\mu_n(\sigma_n)\colon n \in \mathbb{N}\}$ *be defined by* (1.5).

If for some (nonrandom!) $\mu \in \mathscr{P}(R^1)$ *the convergence*

$$\mu_n(\sigma_n) \Rightarrow_P \mu \tag{1.6}$$

holds, then

(i) *the sequence* $\{S_n(\sigma_n)\colon n \in \mathbb{N}\}$ *is tight.*

(ii) *If along some subsequence* $\{n'\} \subset \{n\}$, $S_{n'}(\sigma_{n'}) \to_{\mathscr{D}} \nu$, *then, necessarily,* ν *satisfies the equation*

$$\nu * \mu = \mu * \mu. \tag{1.7}$$

(iii) *If the equation* (1.7) *has the unique solution* $\nu = \mu$, *then*

$$S_n(\sigma_n) \to_{\mathscr{D}} \mu.$$

1.2 LEMMA [Jakubowski (1980) and Beśka, Kłopotowski, and Słomiński (1982) with summation to stopping times]. *If* $z \neq 0$ *and*

$$\prod_{1 \le k \le \sigma_n} E\left(e^{i\theta X_{nk}} | \mathscr{F}_{n,k-1}\right) \to_P z,$$

then $Ee^{i\theta S_n(\sigma_n)} \to z$.

PROOF OF THEOREM 1.1. For a random probability measure $\mu(\cdot,\omega)$ denote by $\hat{\mu}(\theta,\omega)$ its characteristic function taken in $\theta \in R^1$:

$$\hat{\mu}(\theta,\omega) = \int e^{i\theta x}\mu(dx,\omega).$$

It is clear that assumption (1.6) implies

$$\widehat{\mu_n(\sigma_n)}(\theta,\cdot) = \prod_{1 \le k \le \sigma_n} \widehat{\mu_{nk}}(\theta,\cdot) \to_P \hat{\mu}(\theta), \qquad \theta \in R^1. \tag{1.8}$$

By Lemma 1.2

$$Ee^{i\theta S_n(\sigma_n)} \to \hat{\mu}(\theta) \tag{1.9}$$

provided $\hat{\mu}(\theta) \neq 0$. Since $\hat{\mu}(\theta) \neq 0$ at least in some neighborhood of zero, the sequence $\{S_n(\sigma_n)\colon n \in \mathbb{N}\}$ is tight.

If ν is any limit distribution of the sequence $\{S_n(\sigma_n)\}$ and $\theta \in R^1$ is such that $\hat{\mu}(\theta) \neq 0$, then by (1.9) $\hat{\nu}(\theta) = \hat{\mu}(\theta)$. In any case

$$\hat{\nu}(\theta) \cdot \hat{\mu}(\theta) = (\hat{\mu}(\theta))^2,$$

i.e., equation (1.7) holds. □

Consider Brown's Theorem 1.0 again. In fact, (1.2)–(1.4) are equivalent to a countable number of conditions only. Hence in every subsequence $\{n'\}$ one can find a further subsequence $\{n''\} \subset \{n'\}$ for which (1.3) and (1.4) hold a.s. The application of Lindeberg–Feller CLT for each ω separately shows that $\mu_{n''}(\sigma_{n''}) \Rightarrow N(0,1)$ a.s. This implies $\mu_n(\sigma_n) \Rightarrow_P N(0,1)$, since $\mathscr{P}(R^1)$ is a metric space. Thus Theorem 1.0 is a corollary of the limit theorem for independent random variables and Theorem 1.1.

The next two corollaries are examples of application of the Principle of Conditioning to limit theorems which deal with noninfinitesimal summands.

Let Φ be the distribution function of the standard normal law $N(0,1)$. By convention, let $\Phi(x/0) = 1$ if $x > 0$ and $\Phi(x/0) = 0$ if $x < 0$.

1.3 COROLLARY. *Let $(\mathbb{X}, \mathbb{F})$ be a martingale difference array which satisfies $EX_{nk}^2 < +\infty$, $k, n \in \mathbb{N}$. Define $s_{nk}^2 = s_{nk}^2(\omega) = E(X_{nk}^2 | \mathscr{F}_{n,k-1})(\omega)$. The following conditions*:

$$\sum_k s_{nk}^2 \to_P 1, \tag{1.10}$$

$$\max_k |P(X_{nk} < x | \mathscr{F}_{n,k-1}) - \Phi(x/s_{nk})| \to_P 0 \tag{1.11}$$

for $x \in D$, where D is a dense subset of R^1, $0 \notin D$,

$$\sum_k \left[E\big(X_{nk}^2 I(|X_{nk}| > \varepsilon) | \mathscr{F}_{n,k-1}\big) - s_{nk}^2 \int_{\{|x| s_{nk} > \varepsilon\}} x^2 \Phi(dx) \right] \to_P 0, \quad \varepsilon > 0, \tag{1.12}$$

imply the convergence $S_n(\sigma_n) \to_{\mathscr{D}} N(0,1)$. [As usual, the summation and maximum are taken over the set $\{k: 1 \le k \le \sigma_n(\omega)\}$.]

PROOF. By the Principle of Conditioning it suffices to prove the corollary for independent random variables only. But in this case conditions (1.10)–(1.12) are equivalent to the assumptions of the second version of the CLT due to Zolotarev (1967). □

Note that a different proposition of a conditioned version of Zolotarev's theorem is given in Adler and Scott (1975).

In both the examples given above it is not difficult to identify the limit, since there is only one solution of (1.7) for $\mu = N(0,1)$. The same holds for infinitely divisible μ (since $\hat{\mu}$ is nonvanishing) and for μ concentrated on $R^+ = [0, +\infty)$.

1.4 COROLLARY. *Suppose that $X_{nk} = I_{A_{nk}}$ a.s., $k, n \in \mathbb{N}$, where $A_{nk} \in \mathscr{F}$. Let $\lambda \ge 0$ and the numbers p_k, $1 \ge p_k \ge 0$, $k \in \mathbb{N}$, form a convergent series*:

$$\sum_{k=1}^{\infty} p_k < +\infty. \tag{1.13}$$

Define $P_{nk} = P(A_{nk}|\mathscr{F}_{n,k-1})$. *Then the conditions*

$$\sum_k P_{nk} \to_P \lambda + \sum_{k=1}^{\infty} p_k, \tag{1.14}$$

$$\sum_k (P_{nk})^j \to_P \sum_{k=1} (p_k)^j, \qquad j \in \mathbb{N}, \tag{1.15}$$

imply the convergence $S_n(\sigma_n) \to_{\mathscr{D}} \mu$, *where* μ *is the convolution of* 0–1 *distributions with parameters* p_k *and the Poisson distribution with parameter* λ:

$$\mu = \text{Poiss}(\lambda) * \prod_{k=1}^{\infty} B(p_k).$$

PROOF. Suppose that the present corollary is proved for independent random variables, which can be done straightforwardly if we use the Laplace transform.

Let $B = \{p\delta_1 + (1-p)\delta_0 \colon 0 \le p \le 1\} \subset \mathscr{P}(R^1)$. For every $A \in \mathscr{F}$, the regular conditional distribution of I_A given any $\mathscr{F}' \subset \mathscr{F}$ is a random element with values in B. Hence condition (1.6) in Theorem 1.1 can be verified by application of independent version of the present corollary. □

Let us note, that the above procedure was implicitly used also in the proof of Theorem 1.0 and Corollary 1.3, where instead of B the class $\{\mu \in \mathscr{P}(R^1)\colon \int x\mu(dx) = 0, \int x^2\mu(dx) < +\infty\}$ was considered.

2. Some remarks on the Principle of Conditioning.

2.1 GENERALIZATIONS TO HIGHER DIMENSIONS. The Principle of Conditioning holds for adapted systems of finite dimensional random vectors. It can be obtained by a change of notation in the proof of Theorem 1.1 only. But Theorem 1.1 is valid also in the infinite dimensional case: When the adapted system $(\mathbb{X}, \mathbb{F}, \mathbb{S})$ consists of random elements taking values in a real and separable Hilbert space [the author's paper (1980) in the conditionally infinitesimal case and (1982) in full generality].

It should be noted that Theorem 1.1 cannot be generalized to the case of an arbitrary Banach space. Suitable counterexample and further discussion can be found in the paper by Rosiński (1981).

2.2 THE EQUATION $\nu * \mu = \mu * \mu$. In Theorem 1.1(iii) the uniqueness assumption of the solution of the equation $\nu * \mu = \mu * \mu$ cannot be completely omitted if the convergence in distribution to μ is to hold. This follows from the fact that the random measure $\mu_n(\sigma_n)$ does not determine the distribution of $S_n(\sigma_n)$ even if $\mu_n(\sigma_n)$ is nonrandom.

2.3 EXAMPLE [Kwapień (1983)]. Let μ and $\nu \neq \mu$ be probability measures on R^1 having the property (1.7), i.e., $\nu * \mu = \mu * \mu$. Then there necessarily exists an open subset $A \subset R^1$ such that for $\theta \in A$

$$\hat{\mu}(\theta) = 0 \quad \text{and} \quad \hat{\nu}(\theta) \neq 0.$$

Moreover, for every $C > 0$, $\mu(|x| > C) > 0$. Choose the constant C such that $\mu(|x| \le C) > 0$. Define

$$X_1 = X, \qquad X_2 = YI(|X| > C) + ZI(|X| \le C),$$
$$\mathscr{F}_0 = \{\phi, \Omega\}, \qquad \mathscr{F}_1 = \sigma(X), \qquad \mathscr{F}_2 = \sigma(X, Y, Z),$$

where X, Y, Z are independent, $\mathscr{L}(X) = \mathscr{L}(Y) = \mu$, $\mathscr{L}(Z) = \nu$. Then by the definition

$$\begin{aligned} \mu_n(\sigma_n) &= \mu * [\mu \cdot I(|X| > C) + \nu \cdot I(|X| \le C)] \\ &= \mu * \mu, \end{aligned}$$

which does not depend on ω. But

$$\begin{aligned} E \exp i\theta(X_1 + X_2) = {} & P(|X| > C) E(\exp i\theta X \| X| > C) \cdot \hat{\mu}(\theta) \\ & + P(|X| \le C) E(\exp i\theta X \| X| \le C) \cdot \hat{\nu}(\theta). \end{aligned}$$

Since $\hat{\nu}(\theta) \ne 0$, $P(|X| \le C) > 0$, and the characteristic function

$$E(\exp i\theta X \| X| \le C)$$

is analytic on R^1, the characteristic function $E \exp i\theta(X_1 + X_2)$ cannot be identically equal to zero on the set A. Hence $\mathscr{L}(X_1 + X_2) \ne \mu * \mu$.

2.4 A REMARK ON ANOTHER WAY OF CONDITIONING. Dvoretzky (1971) has suggested that the Principle of Conditioning holds (at least for those limit theorems which are the solutions of the Central Limit problem) if we consider conditioning with respect to the previous sum, i.e., $\mathscr{F}_{n,k-1} = \sigma(S_n(k-1))$. However, the counterexample of Kłopotowski (1980a) shows that this form of the Principle of Conditioning cannot be true.

3. Adapted systems and conditionally independent random variables. There exists a simple interpretation of the predictable random measures $\mu_n(\sigma_n)$ considered in the previous two sections.

Suppose that $(\mathbb{X}, \mathbb{F}, \mathbb{S})$ is an adapted system and consider an accompanying array of random variables $\mathbb{X}^* = \{X^*_{nk}\colon k \in \mathbb{N},\ n \in \mathbb{N}\}$ defined on a suitable extension of $(\Omega, \mathscr{F}, P)$ with the following properties:

(3.1) for each $n \in \mathbb{N}$, $\{X^*_{nk}\colon k \in \mathbb{N}\}$ are conditionally independent over $\mathscr{F}$,

the regular conditional distribution of X^*_{nk} given $\mathscr{F}$ coincides with μ_{nk}:

$$\begin{aligned} P(X^*_{nk}|\mathscr{F})(A, \omega) &= \mu_{nk}(A, \omega) \\ &= P(X_{nk}|\mathscr{F}_{n,k-1})(A, \omega), \qquad A \in \mathscr{B}^1, \quad \omega \in \Omega. \end{aligned} \tag{3.2}$$

Define

$$S^*_n = S^*_n(\sigma_n) = \sum_{1 \le k \le \sigma_n} X^*_{nk}. \tag{3.3}$$

By the definition of $\mathbb{X}^*$, the regular conditional distribution of $S^*_n(\sigma_n)$ given $\mathscr{F}$ satisfies

$$\begin{aligned} P(S^*_n(\sigma_n)|\mathscr{F})(\cdot, \omega) &= \mu_{n1}(\cdot, \omega) * \cdots * \mu_{n\sigma_n(\omega)}(\cdot, \omega) \\ &= \mu_n(\sigma_n)(\cdot, \omega). \end{aligned} \tag{3.4}$$

Our basic Theorem 1.1 asserts that for a "good" limit distribution μ, the convergence $P(S_n^*(\sigma_n)|\mathscr{F})(\cdot,\omega) \Rightarrow_P \mu$ implies $S_n(\sigma_n) \to_{\mathscr{D}} \mu$.

One may expect that admitting nonconstant random measures as limits for $P(S_n^*(\sigma_n)|\mathscr{F})(\cdot,\omega)$ or even replacing the convergence in probability by the convergence in distribution, new limit theorems can be derived with limit laws for $S_n(\sigma_n)$ of the form $E\mu(\cdot,\omega)$ [for the random measure $\mu(\cdot,\omega)$ its expectation $E\mu \in \mathscr{P}(R^1)$ is defined by the formula $(E\mu)(A) = E\mu(A,\cdot)$, $A \in \mathscr{B}^1$]. This is not true as has been shown for example by Dvoretzky (1971) and Hall and Heyde (1980). However, such a theorem can be proved for adapted systems obtained by scaling a single sequence.

3.1 THEOREM. *Suppose that the adapted pair* $(\mathbb{X},\mathbb{F})$ *is obtained from a single sequence* $\{Y_k\colon k \in \mathbb{N}\}$ *adapted to a filtration* $\{\mathscr{F}_k\colon k \in \mathbb{N}_0\}$

$$X_{nk} = B_n^{-1}\cdot X_k, \qquad \mathscr{F}_{nk} = \mathscr{F}_k, \tag{3.5}$$

where $B_n \nearrow +\infty$ *is a sequence of positive numbers.*

Suppose that for the system $(\mathbb{X},\mathbb{F},\mathbb{S})$ *the condition*

$$P\big(S_n^*(\sigma_n)|\mathscr{F}\big) \Rightarrow_P \mu(\cdot,\omega) \tag{3.6}$$

holds. If the random measure $\mu(\cdot,\omega)$ *has the property*

$$P(\hat{\mu}(\theta,\omega)=0) = 0, \qquad \theta \in D \subset R^1, \tag{3.7}$$

where D *is a fixed dense subset of* R^1, *then*

$$S_n(\sigma_n) \to_{\mathscr{D}} E\mu(\cdot,\omega),$$

or more precisely, there exists a nondecreasing sequence $\{k_n\} \subset \mathbb{N}$, $k_n \nearrow +\infty$ *such that*

$$P\big(S_n(\sigma_n)|\mathscr{F}_{k_n}\big)(\cdot,\omega) \Rightarrow_P \mu(\cdot,\omega). \tag{3.8}$$

PROOF. We need an improved version of Lemma 1.2.

3.2 LEMMA [Jakubowski (1982) and Jakubowski and Słomiński (1986)]. *Let* $(\mathbb{X},\mathbb{F})$ *be an arbitrary adapted pair. Fix* $n \in \mathbb{N}$. *Then for each* $\varepsilon > 0$, $\theta \in R^1$, *and every* $\{\mathscr{F}_{nk}\colon k \in \mathbb{N}_0\}$*-stopping time* σ_n, *there holds the inequality*

$$\begin{aligned} &\Big|E_{n0}\big(e^{i\theta S_n(\sigma_n)}\big) - E_{n0}\big(\widehat{\mu_n(\sigma_n)}(\theta)\big)\Big| \\ &\quad\le 4P_{n0}\Big(\big|\widehat{\mu_n(\sigma_n)}(\theta)\big| \le \varepsilon\Big) + \varepsilon^{-1}E_{n0}\Big|\widehat{\mu_n(\sigma_n)}(\theta) - E_{n0}\big(\widehat{\mu_n(\sigma_n)}(\theta)\big)\Big|, \end{aligned} \tag{3.9}$$

where $E_{n0}(\cdot) = E(\cdot|\mathscr{F}_{n0})$ *and* $P_{n0}(\cdot) = E_{n0}(I(\cdot))$.

In order to prove (3.8) it suffices to find a sequence $k_n \nearrow +\infty$, independent of θ and such that

$$E_{k_n}\big(e^{i\theta S_n(\sigma_n)}\big) \to_P \hat{\mu}(\theta,\cdot), \qquad \theta \in R^1.$$

Since $B_n \nearrow +\infty$, for each fixed k, $\mu_{nk}(\cdot, \omega) \Rightarrow \delta_0$ a.s. Hence there exists $k_n \nearrow +\infty$ with the property

$$\prod_{1 \le k \le k_n}^{*} \mu_{nk}(\cdot, \omega) \Rightarrow \delta_0 \quad \text{a.s.}$$

By Theorem 1.1 $S_n(k_n) \to_P 0$. Define a new system $X'_{nk} = X_{n,k+k_n}$, $\mathscr{F}'_{nk} = \mathscr{F}_{k+k_n}$, $\sigma'_n = \sigma_n - k_n \wedge \sigma_n$, and observe that $S_n(\sigma_n) - S'_n(\sigma'_n) \to_P 0$, $\widehat{\mu_n(\sigma_n)}(\theta) - \widehat{\mu'_n(\sigma'_n)}(\theta) \to 0$ a.s., and, since $k_n \to +\infty$, $E(\hat{\mu}(\theta, \cdot)|\mathscr{F}_{k_n}) \to \hat{\mu}(\theta, \cdot)$ a.s. By Lemma 3.2 and property (3.7),

$$E\left|E_{k_n}\left(e^{i\theta S_n(\sigma_n)}\right) - \hat{\mu}(\theta, \cdot)\right| \le E\left|e^{i\theta(S_n(\sigma_n) - S'_n(\sigma'_n))} - 1\right| + E\left|\widehat{\mu'_n(\sigma'_n)}(\theta) - \hat{\mu}(\theta, \cdot)\right|$$

$$+ E\left|E'_{n0}\left(e^{i\theta S'_n(\sigma'_n)}\right) - E'_{n0}\left(\widehat{\mu'_n(\sigma'_n)}(\theta)\right)\right| \to 0.$$

3.3 Remark. By condition (3.8), Theorem 3.1 is stable in the Rényi sense—see Aldous and Eagleson (1978).

3.4 Central Limit theorem for martingales with stationary differences. It is clear that by virtue of Theorem 3.1 one can construct another Principle of Conditioning, which admits mixtures of probability distributions as limit laws and is valid for the systems $(\mathbb{X}, \mathbb{F}, \mathbb{S})$ obtainable by scaling a single adapted sequence only.

For example, it follows readily from this new Principle of Conditioning, Lindeberg CLT and Pointwise Ergodic theorem, that for every two-sided strictly stationary sequence $\{Y_k\colon k \in \mathscr{Z}\}$ such that $E(Y_1 - Y_0|\sigma(Y_j\colon j \le 0)) = 0$ and $EY_1^2 < +\infty$, the convergence in law $n^{-1/2}\sum_{1 \le j \le n} Y_j \to EN(0, \sigma^2(\omega))$ holds. Here $\sigma^2(\omega) = E(Y_1^2|\mathscr{I})(\omega)$ and $\mathscr{I}$ is a σ-algebra of invariant subsets for the sequence Y_k. This is a generalization of the well-known CLT for martingales with stationary ergodic differences due to Billingsley (1961) and Ibragimov (1963).

3.5 Other generalizations. Theorem 3.1 is an example of how the basic inequality (3.9) works in the case when some additional information on the system $(\mathbb{X}, \mathbb{F}, \mathbb{S})$ is given. It should be noted that inequality (3.9) may be explored in some other directions [see the results of Eagleson (1975), Rootzén (1977a), Hall (1977), and Kłopotowski (1980b), which can be obtained from (3.9)].

3.6 Tightness of $S_n^*(\sigma_n)$ implies tightness of $S_n(\sigma_n)$. Although it is not possible to describe the asymptotic behaviour of $\mathscr{L}(S_n(\sigma_n))$ given $\mathscr{L}(S_n^*(\sigma_n))$ for general adapted systems $(\mathbb{X}, \mathbb{F}, \mathbb{S})$, partial information is accessible. Namely, if the sequence $\{S_n^*(\sigma_n)\}$ is tight, then $\{S_n(\sigma_n)\}$ is tight [Jakubowski (1982)]. In fact, there exists a more detailed description of connections of this type.

3.7 THEOREM. *Suppose that there exist $\mathscr{F}$-measurable random variables A_n such that the sequence $\{S_n^*(\sigma_n) - A_n: n \in \mathbb{N}\}$ is tight. Then there exist random variables C_{nk} such that*

(a) *for every $n, k \in \mathbb{N}$, C_{nk} is $\mathscr{F}_{n,k-1}$-measurable, i.e.,*

$$C_n(k) = \sum_{1 \le j \le k} C_{nj}, \qquad k \in \mathbb{N}, \tag{3.10}$$

is a predictable sequence with respect to $\{\mathscr{F}_{nk}: k \in \mathbb{N}_0\}$;

(b) *the sequence $\{S_n^*(\sigma_n) - C_n(\sigma_n): n \in \mathbb{N}\}$ is tight;*

(c) *the sequence $\{S_n(\sigma_n) - C_n(\sigma_n): n \in \mathbb{N}\}$ is tight.*

The proof is based on some tightness criteria for random measures and is rather arduous [refer to Jakubowski (1985)].

4. Principle of Conditioning in functional limit theorems. In this section we would like to suggest a formalism leading to the functional version of Theorem 1.1.

4.1 FUNCTIONAL PRINCIPLE OF CONDITIONING. Let $(\mathbb{X}, \mathbb{F})$ be an adapted pair. Let $\mathbb{TS} = \{\Sigma_n: n \in \mathbb{N}\}$ be a sequence of time-scales with respect to $\mathbb{F}$, i.e., for each n the family $\Sigma_n = \{\sigma_n(t): t \in R^+\}$ satisfies the following conditions:

$$\sigma_n(0) = 0, \tag{4.1}$$

(4.2) for each $t \in R^+$, $\sigma_n(t)$: $(\Omega, \mathscr{F}, P) \to \mathbb{N}_0$ is an $\{\mathscr{F}_{nk}: k \in \mathbb{N}_0\}$-stopping time,

(4.3) for each $\omega \in \Omega$, the trajectory $t \mapsto \sigma_n(t)(\omega)$ is nondecreasing, right continuous, and increases only by jumps of size 1,

$$\lim_{t \to \infty} \sigma_n(t)(\omega) = +\infty \quad \text{for each } \omega \in \Omega. \tag{4.4}$$

For example, $\sigma_n(t) = [nt]$ defines a time scale.

Given the triple $(\mathbb{X}, \mathbb{F}, \mathbb{TS})$ one can construct a sequence of stochastic processes $\{X_n = \{X_n(t): t \in R^+\}: n \in \mathbb{N}\}$ by the summation of $\mathbb{X}$ with respect to $\mathbb{TS}$:

$$X_n(t) = S_n(\sigma_n(t)) = \sum_{1 \le k \le \sigma_n(t)} X_{nk}, \qquad t \in R^+, \quad n \in \mathbb{N}. \tag{4.5}$$

Fix $n \in \mathbb{N}$. Clearly, the trajectories $t \mapsto X_n(t, \omega)$ of the process X_n belong to the space $D(R^+: R^1)$ of functions f: $R^+ \to R^1$ which are right-continuous and admit left-hand limits. It follows that the stochastic process X_n can be considered as a random element in the space $D(R^+: R^1)$ equipped with the Skorokhod topology [see Billingsley (1968) and Lindvall (1973)].

For $0 \le s < t$ define

$$\mu_n(s, t, \omega) = \prod_{\sigma_n(s) < k \le \sigma_n(t)}^{*} \mu_{nk}(\cdot, \omega). \tag{4.6}$$

Of course, $\mu_n(0, t, \omega) = \mu_n(\sigma_n(t))$ by definition (1.5). Fix $\omega \in \Omega$ and consider the family of distributions $\{\mu_n(s, t, \omega): 0 \le s < t\}$. If random variables $Y^{\omega}_{n1}, Y^{\omega}_{n2}, \dots$ are independent and have distributions given by

$$\mathscr{L}(Y^{\omega}_{nk}) = \mu_{nk}(\cdot, \omega), \tag{4.7}$$

then the stochastic process $Y^{\omega}_n = \{Y^{\omega}_n(t): t \in R^+\}$,

$$Y^{\omega}_n(t) = \sum_{1 \le k \le \sigma_n(t)} Y^{\omega}_{nk}, \qquad t \in R^+, \tag{4.8}$$

has independent increments and the distributions of its increments are given by $\mathscr{L}(Y^{\omega}_n(t) - Y^{\omega}_n(s)) = \mu_n(s, t, \omega)$. Define a random measure

$$\mu_n(\cdot): (\Omega, \mathscr{F}, P) \to \Pi \subset \mathscr{P}(D(R^+: R^1)) \tag{4.9}$$

by setting as $\mu_n(\omega)$ the law of the process $Y^{\omega}_n = \{Y^{\omega}_n(t): t \in R^+\}$. [Here Π denotes the set of laws of processes with independent increments and trajectories belonging to $D(R^+: R^1)$.]

Suppose that

(a) $\mu_n(\cdot)$ converges in probability to some measure $\mu_\infty = \mathscr{L}(X_\infty) \in \Pi$:

$$\mu_n(\cdot) \Rightarrow_P \mu_\infty, \tag{4.10}$$

(b) the process X_∞ corresponding to μ_∞ has no fixed points of discontinuity.

If $s < t$, then (4.10) implies

$$\mu_n(s, t, \cdot) \Rightarrow_P \mu_\infty(s, t) = \mathscr{L}(X_\infty(t) - X_\infty(s)) \tag{4.11}$$

and by (b) the measure $\mu_\infty(s, t)$ is infinitely divisible. Hence by Theorem 1.1 $X_n(t) - X_n(s) \to_{\mathscr{D}} X_\infty(t) - X_\infty(s)$, and this statement can be easily strengthened to the convergence of finite dimensional distributions.

On the other hand, (4.10) implies

$$\mu_n(\tau_n, \tau_n + \delta_n, \cdot) \Rightarrow_P \delta_0 \tag{4.12}$$

for any bounded sequence $\{\tau_n: n \in \mathbb{N}\}$ of nonnegative random variables and any sequence of numbers $\delta_n \searrow 0$. Then subsequent application of Theorem 1.1 gives

$$X_n(\tau_n + \delta_n) - X_n(\tau_n) \to_P 0 \tag{4.13}$$

for any sequence $\{\tau_n: n \in \mathbb{N}\}$ of *discrete stopping times* with respect to the natural filtrations $\{\mathscr{F}_n(t) = \sigma(X_n(s): s \le t): t \in R^+\}$.

Now, application of the Aldous (1978a) criterion gives an especially clear proof of tightness of the sequence $\{X_n: n \in \mathbb{N}\}$ and ends the proof of the functional convergence

$$X_n \to_{\mathscr{D}} X_\infty. \tag{4.14}$$

The limit theorem proved above is the most important case of the Functional Principle of Conditioning. Indeed, the situation is just the same as in Section 1, because it is possible to express the condition $\mu_n(\cdot) \Rightarrow_P \mu_\infty$ in terms of convergence of predictable characteristics of processes X_n [e.g., using the general limit theorem for processes with independent increments due to Jacod (1983)].

A suitable example will be given in the next section (Theorem 5.1). Here we note only that by "conditioning" of the well-known Donsker Invariance Principle we get Brown's (1971) Invariance Principle for Martingales, and that certain other [e.g., Durrett and Resnick (1978)] results of this type can be also obtained following the line presented above.

Whether the Functional Principle of Conditioning is true for the general limit process X_∞ is still an open question. However, the results of Jacod, Kłopotowski, and Memin (1982) show that the answer should be positive.

4.2 THE FUNCTIONAL PRINCIPLE OF CONDITIONING HOLDS FOR SEMIMARTINGALES. The construction of the predictable random measure $\mu_n(\cdot)$ was very simple for stochastic processes arising by the summation of random variables according to a given time scale. There is a more extensive class of processes, semimartingales, for which an analogous random measure can be defined. An explicit expression for such a measure can be found in Jacod et al. (1982). Moreover, limit theorems proved in that paper, as well as in the earlier paper by Liptser and Shiryayev (1980), show (implicitly, of course) that the Functional Principle of Conditioning also holds for semimartingales.

4.3 FUNCTIONAL PRINCIPLE OF CONDITIONING GIVES EXTENDED CONVERGENCE. The quantities appearing in conditions of any limit theorem derived by the Principle of Conditioning, contain information about the connections between the processes X_n and filtrations $\{\mathscr{F}_{n,\sigma_n(t)}\colon t \in R^+\}$. Such connections can be described, for example, by the notion of extended convergence of processes with filtration, introduced by Aldous (1978b).

One can prove that the assumption $\mu_n(\cdot) \Rightarrow_P \mu_\infty$ is almost equivalent to the extended convergence of respective processes with filtration—see Jakubowski and Słomiński (1983). Similar results belong to Helland (1980) and Grigelionis, Kubilius, and Mikulevičius (1982). The case, when the limit process does not have independent increments is considered in Słomiński (1984) and Kubilius (1985).

5. Principle of Conditioning and martingale central limit theorems. There are several methods of derivation of martingale CLT or Martingale Invariance Principles. They are exhaustively described in the well-known book by Hall and Heyde (1980).

Here we shall describe another one via the Principle of Conditioning.

We begin with the limit theorem that is a particular case of Theorem A from the paper by Lipster and Shiryayev (1980). From our point of view, this theorem is derived by the Principle of Conditioning from the most general Invariance Principle for processes obtained by summation of independent random variables.

All notations below are taken from Section 4. The Wiener Process is denoted by $W = \{W(t)\colon t \in R^+\}$.

5.1 THEOREM. *Let* $(\mathbb{X}, \mathbb{F}, \mathbb{TS})$ *be an adapted system. Denote* $Y_{nk} = X_{nk} I(|X_{nk}| \le 1)$. *Suppose that the following conditions* (5.1)–(5.3) *hold*:

$$\sup_{s \le t} \Big| \sum_{1 \le k \le \sigma_n(s)} E(Y_{nk} | \mathscr{F}_{n,k-1}) \Big| \to_P 0, \qquad t \in R^+, \tag{5.1}$$

$$\sum_{1 \le k \le \sigma_n(t)} E(Y_{nk}^2 | \mathscr{F}_{n,k-1}) - [E(Y_{nk} | \mathscr{F}_{n,k-1})]^2 \to_P t, \qquad t \in R^+, \tag{5.2}$$

$$\sum_{1 \le k \le \sigma_n(t)} P(|X_{nk}| > \varepsilon | \mathscr{F}_{n,k-1}) \to_P 0, \qquad \varepsilon > 0, \quad t \in R^+. \tag{5.3}$$

Then $X_n \to_{\mathscr{D}} W$.

We shall prove that conditions (5.1)–(5.3) in the above theorem *are implied* by assumptions (5.4) and (5.5) in the following Invariance Principle for martingale difference arrays. This invariance principle is due to McLeish (1974) but here is presented in the form improved by Gänssler and Häusler (1979).

5.2 THEOREM. *Let* $(\mathbb{X}, \mathbb{F})$ *be a martingale difference array. Suppose that for the sequence* $\{\Sigma_n: n \in \mathbb{N}\}$ *of time scales the following two conditions hold*:

$$E \max_{1 \le k \le \sigma_n(t)} |X_{nk}| \to 0, \qquad t \in R^+, \tag{5.4}$$

$$\sum_{1 \le k \le \sigma_n(t)} X_{nk}^2 \to_P t, \qquad t \in R^+. \tag{5.5}$$

Then $X_n \to_{\mathscr{D}} W$.

The proof is contained in two lemmas below.

5.3 LEMMA [Helland (1982)]. *Suppose that* $(\mathbb{X}, \mathbb{F})$ *is a martingale difference array and for* $\{\Sigma_n: n \in \mathbb{N}\}$ *condition* (5.4) *holds. Then*

$$\sum_{1 \le k \le \sigma_n(t)} E(|X_{nk}| I(|X_{nk}| > 1) | \mathscr{F}_{n,k-1}) \to_P 0, \qquad t \in R^+, \tag{5.6}$$

and, in particular, (5.1) *holds.*

5.4 LEMMA [Liptser and Shiryayev (1980)]. *Let* $(\mathbb{X}, \mathbb{F}, \mathbb{TS})$ *be an adapted system.*

(a) *Condition* (5.3) *and*

$$\max_{1 \le k \le \sigma_n(t)} |X_{nk}| \to_P 0, \qquad t \in R^+, \tag{5.7}$$

are equivalent.

(b) *Under condition* (5.3), *condition* (5.2) *and*

$$\sum_{1 \le k \le \sigma_n(t)} (X_{nk} - E(Y_{nk} | \mathscr{F}_{n,k-1}))^2 \to_P t, \qquad t \in R^+, \tag{5.8}$$

are equivalent.

If, in addition, (5.6) *is satisfied, then* (5.2) *is equivalent to* (5.5).

5.5 REMARK. It is well known, that under condition (5.1) conditions (5.2) and (5.3) are equivalent to $X_n \to W$. Similarly, if for a martingale difference array $(\mathbb{X}, \mathbb{F})$ the sequence $\{\max_{1 \le k \le \sigma_n(t)} |X_{nk}|: n \in \mathbb{N}\}$ is uniformly integrable for each $t \in R^+$, then $X_n \to_{\mathscr{D}} W$ if and only if (5.7) [$\equiv$ (5.3)] and (5.5) are satisfied. See Rootzén (1977b), Gänssler and Häusler (1979), and Liptser and Shiryayev (1981).

This is the contrary to nonfunctional martingale CLT where there is no hope for necessary and sufficient conditions (excluding the case of independent random variables).

Acknowledgment. The author would like to express his gratitude to Professor Stanisław Kwapień for his valuable remarks and encouragement.

REFERENCES

ADLER, R. J. and SCOTT, D. J. (1975). Martingale central limit theorems without U.A.N. conditions. *Bull. Austral. Math. Soc.* **13** 45–55.

ALDOUS, D. J. (1978a). Stopping times and tightness. *Ann. Probab.* **6** 335–340.

ALDOUS, D. J. (1978b). A concept of weak convergence for stochastic processes viewed in the Strassbourg manner. Statist. Laboratory, Cambridge. Unpublished.

ALDOUS, D. J. and EAGLESON, G. K. (1978). On mixing and stability of limit theorems. *Ann. Probab.* **6** 325–331.

BEŚKA, M., KŁOPOTOWSKI, A. and SŁOMIŃSKI, L. (1982). Limit theorems for random sums of dependent d-dimensional random vectors. *Z. Wahrsch. verw. Gebiete* **61** 43–57.

BILLINGSLEY, P. (1961). The Lindeberg–Lévy theorem for martingales. *Proc. Amer. Math. Soc.* **12** 788–792.

BILLINGSLEY, P. (1968). *Weak Convergence of Probability Measures.* Wiley, New York.

BROWN, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42** 59–66.

BROWN, B. M. and EAGLESON, G. K. (1971). Martingale convergence to infinitely divisible laws with finite variances. *Trans. Amer. Math. Soc.* **162** 449–453.

DURRETT, R. and RESNICK, S. I. (1978). Functional limit theorems for dependent random variables. *Ann. Probab.* **6** 829–846.

DVORETZKY, A. (1971). Asymptotic normality for sums of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* 513–535 Univ. California Press.

EAGLESON, G. K. (1975). Martingale convergence to mixtures of infinitely divisible laws. *Ann. Probab.* **3** 557–562.

GÄNSSLER, P. and HÄUSLER, E. (1979). Remarks on the functional central limit theorem for martingales. *Z. Wahrsch. verw. Gebiete* **50** 237–243.

GRIGELIONIS, B., KUBILIUS, K. and MIKULEVIČIUS, R. (1982). Martingale approach to functional limit theorems. *Russian Math. Surveys* **37/6/228** 39–51.

HALL, P. (1977). Martingale invariance principles. *Ann. Probab.* **5** 876–887.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic, New York.

HELLAND, I. S. (1980). On weak convergence to Brownian motion. *Z. Wahrsch. verw. Gebiete* **52** 251–265.

HELLAND, I. S. (1982). Central limit theorem for martingales with discrete or continuous time. *Scand. J. Statist.* **9** 79–94.

IBRAGIMOV, I. A. (1963). Central limit theorem for certain dependent random variables. *Theory Probab. Appl.* **8** 89–94.

JACOD, J. (1983). Processus a accroissements independants: une condition necessaire et suffisante de convergence en loi. *Z. Wahrsch. verw. Gebiete* **63** 109–136.

JACOD, J., KŁOPOTOWSKI, A. and MEMIN, J. (1982). Théorème de la limite centrale et covergence fonctionelle vers un processus à accroissements indépendants: la méthode des martingales. *Ann. Inst. H. Poincaré Sect. B (N.S)* **18** 1–45.

JAKUBOWSKI, A. (1980). On limit theorems for sums of dependent Hilbert space valued random variables. *Lecture Notes in Statist.* **2** 178–187. Springer, Berlin.

JAKUBOWSKI, A. (1982). Limit theorems for sums of dependent random variables with values in a Hilbert space. Thesis, Toruń. (In Polish.)

JAKUBOWSKI, A. (1985). Tightness criteria for random measures with application to Principle of Conditioning in Hilbert space. Preprint. To appear in *Probab. and Math. Statist.*

JAKUBOWSKI, A. and SŁOMIŃSKI, L. (1986). Extended convergence to continuous in probability processes with independent increments. *Probab. Theory Rel. Fields* **72** 55–82.

KŁOPOTOWSKI, A. (1977). Limit theorems for sums of dependent random vectors in R^d. *Dissertationes Math.* **151** 1–62.

KŁOPOTOWSKI, A. (1980a). A remark on the conditioning in limit theorems for dependent random vectors in R^d. Banach Center Publ., *Math. Statist.* **6** 175–177.

KŁOPOTOWSKI, A. (1980b). Mixtures of infinitely divisible distributions as limit laws for sums of dependent random variables. *Z. Wahrsch. verw. Gebiete* **51** 101–113.

KUBILIUS, K. (1985). Extended convergence of stochastic processes. Preprint.

KWAPIEŃ, S. (1983). Oral communication.

LINDVALL, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0,\infty)$. *J. Appl. Probab.* **10** 109–121.

LIPTSER, R. and SHIRYAYEV, A. (1980). Functional limit theorem for semimartingales. *Theory Probab. Appl.* **25** 683–703.

LIPTSER, R. and SHIRYAYEV, A. (1981). Necessary and sufficient conditions for the functional central limit theorem for semimartingales. *Theory Probab. Appl.* **26** 132–137.

MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** 620–628.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces.* Academic, New York.

ROOTZÉN, H. (1977a). A note on convergence to mixtures of normal distributions. *Z. Wahrsch. verw. Gebiete* **38** 211–216.

ROOTZÉN, H. (1977b). On the functional central limit theorem for martingales. *Z. Wahrsch. verw. Gebiete* **38** 199–210.

ROSIŃSKI, J. (1981). The central limit theorems for dependent random vectors in Banach spaces. Preprint 241, Inst. Math. of Polish Acad. Sci.

SŁOMIŃSKI, L. (1984). Necessary and sufficient conditions for extended convergence of semimartingales. To appear in *Probab. and Math. Statist.*

WALK, H. (1977). An invariance principle for the Robbins–Monro process in a Hilbert space. *Z. Wahrsch. verw. Gebiete* **39** 135–150.

ZOLOTAREV, V. M. (1967). A generalisation of the Lindeberg–Feller theorem. *Theory Probab. Appl.* **12** 608–618.

INSTITUTE OF MATHEMATICS
NICHOLAS COPERNICUS UNIVERSITY
UL. CHOPINA 12 / 18
87-100 TORUŃ
POLAND