Holger Rootzén

Central Limit Theory for Martingales via Random Change of Time

Preprint
September
1983
5

Institute of Mathematical Statistics
University of Copenhagen

Holger Rootzén*

CENTRAL LIMIT THEORY FOR MARTINGALES VIA RANDOM CHANGE OF TIME

Preprint 1983 No. 5

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

September 1983

CENTRAL LIMIT THEORY FOR MARTINGALES
VIA RANDOM CHANGE OF TIME

by

Holger Rootzén
University of Copenhagen
and
University of North Carolina at Chapel Hill

Abstract

This paper contains an exposition of the by now rather complete central limit theory for discrete parameter martingales providing new and efficient proofs. The basic idea is to start by proving a central limit theorem under quite restrictive conditions (that the summands tend uniformly to zero and that the sums of squares converge uniformly) and then to obtain the most general results by random change of time and truncation. The emphasis is on the sums of squares (or squared variation process), and Burkholder's square function inequality plays a crucial role in the development. In particular, this approach leads to a very short and direct proof of tightness. In the proofs we make much use of a result (Lemma 2.5) which is believed to be new and which binds together convergence to zero of sums and of sums of conditional expectations. In the final section, the results are extended to several dimensions, to mixing convergence, and to convergence to mixtures of normal distributions.

Research supported in part by AFOSR Grant No. F49620 82 C 0009.

0. Introduction

The main thesis of the present paper is to view the martingale central limit theorem as basically concerning summands which tend uniformly to zero, and with squared variation (sum of squares) converging uniformly, and then to reduce the most general situation to this case by (random) change of timescale and by truncation. We think that this both appeals to the intuition and leads to quite efficient proofs. The purpose of the paper thus is to give a selfcontained exposition of the basic martingale central limit theory, using this point of view, providing as simple and efficient proofs as possible. A second assertion we would like to make is the usefulness of stochastic processes point of view: that it is the functional limit results which are important, and not only their one-dimensional versions. There is, of course, a cost associated with this: one has to learn at least some elements on convergence of distributions on function spaces, but the reward then is both better understanding and easier and shorter proofs. One further feature of our development below is an emphasis on the squared variation process and a systematic use of Burkholder's square function inequality. In particular this makes possible a very easy proof of tightness, which in other approaches often requires the main effort.

The central limit theorem for discrete parameter martingales represents one important stage in the development of central limit theory and has in the last few years reached what seems to be essentially its final form and has also proved its value in many applications to statistics and applied probability. The theory has also been recast into the language of "the general theory of processes" of the Strasbourg school and been extended to the continuous parameter case by the work of Rebolledo [15,16], Lipster and Shirayev [13,14] and others. This has led to a very satisfying formulation of the results and a rather complete extension of the theory. Nevertheless it may perhaps also be said that the essential difficulties are present already in the discrete case and that the basic continuous parameter results are rather easy to obtain from the corresponding results for discrete time, as shown by Helland [9].

Except for the multidimensional result, Theorem 4.3, which is only a small step away from the one-dimensional results (although it seems quite useful for applications), none of the theorems of this paper is new. However, almost all the proofs are new (the main idea was mentioned briefly by the present author in [19] and was developed in some detail in mimeographed lecture notes from the Department of Mathematical Statistics, Copenhagen University). In particular we would like to point out Lemma 2.5, which is a versatile tool and which is new formulated in the present generality, although various special cases have been used by many authors.

A related exposition, which starts, however, by assuming known a basic central limit theorem for bounded martingale differences is given by Helland [9]. A further rather different exposition which uses the Skorokhod embedding is contained in the recent book by Hall and Heyde [8]. Both expositions contain extensive lists of references and accounts of the development of the subject, to which we refer the reader. Later papers of interest include work by Klopotowski and colleagues [2,12], the articles by Lipster and Shiryaev mentioned above, and a series of papers by Jeganathan [10,11]. The approach in [11], which in turn was partly inspired by Rosén [20], is somewhat related to this paper and was made independently of it.

The plan of the paper is as follows. Section 1 contains some notation, and in Section 2 the results we need from other areas (functional limit theory and martingales) are collected, and the basic truncation lemmas (Lemmas 2.5-2.7) are obtained. The functional central limit theorem for martingales is then proved in Section 3, starting from scratch, and finally Section 4 contains a somewhat briefer discussion of one direction of extension of the results, to several dimensions and to convergence to mixtures of normal distributions.

1. Notation

Throughout, we will consider doubly indexed arrays $\{X_{n,j}, \mathcal{B}_{n,j}; j\geq 1, n\geq 1\}$ where the $X_{n,j}$'s are random variables or, in Section 4, random vectors, and for each n, $\{\mathcal{B}_{n,j}\}_{j=1}^{\infty}$ is an increasing sequence of sigma-algebras, i.e., $\mathcal{B}_{n,j}\subseteq\mathcal{B}_{n,j+1}$. We will never assume that the $X_{n,j}$'s are obtained by linearly renormalizing a single sequence of random variables since that is not the case in many of the applications but will sometimes assume that the sigma-algebras are *nested*, i.e. that

$$\mathcal{B}_{n,j} \subseteq \mathcal{B}_{n+1,j} \quad , \text{ for } n,j\geq 1 \ ,$$

which seems to hold in most cases of interest. Possibly by going over to a product space, we will assume that all $X_{n,j}$'s and $\mathcal{B}_{n,j}$'s are defined on the same probability space $(\Omega,\mathcal{B},P)$. The array is said to be *adapted* if $X_{n,j} \in \mathcal{B}_{n,j}$ for $j\geq 1$, $n\geq 1$, and it is a *martingale difference array* (m.d.a.) if in addition $\{X_{n,j}, \mathcal{B}_{n,j}; j=1,2,\ldots\}$ is a sequence of martingale differences, i.e., if $E|X_{n,j}| < \infty$ and $E(X_{n,j+1} \| \mathcal{B}_{n,j}) = 0$ for $j\geq 1$.

A stochastic process $\{\tau(t)\}$, defined for t in some interval I is a *time-scale* if it is nondecreasing, has left limits and is right continuous. A sequence $\{\tau_n\}$ of time scales is *adapted* (to $\{\mathcal{B}_{n,j}\}$) if for each n and $t\in I$, $\tau_n(t)$ is a stopping time with respect to $\mathcal{B}_{n,1}, \mathcal{B}_{n,2}\ldots$.

Let $S_n(t) = \sum_{j=1}^{[t]} X_{n,j}$ be the $[t]^{\text{th}}$ partial sum in the n^{th} row, and let

$\{B(t);\ t\in I\}$ be a standard Brownian motion. The problem we are concerned with is convergence in distribution of time-scaled row-sums $S_n \circ \tau_n$ to a Brownian motion, or more generally a time-scaled Brownian motion. Here of course $S_n \circ \tau_n$ is defined by $S_n \circ \tau_n(t) = S_n(\tau_n(t))$. For brevity of notation we will usually write $S \circ \tau_n$ for $S_n \circ \tau_n$, and $E_i(\cdot)$ for $E(\cdot \| B_{n,i})$ (with $E_0 = E$) when taking expectation of variables in the n^{th} row.

A *partition* of an interval $[0,T]$ is a finite set of points, $0=t_0<t_1<\ldots<t_k=T$. For a given partition, we will write

$$(1.1)\quad \Delta(\ell) = t_\ell - t_{\ell-1} \quad , \quad \Delta = \max_{1\le\ell\le k} \Delta(\ell)$$

$$\Delta S \circ \tau_n(\ell) = S \circ \tau_n(t_\ell) - S \circ \tau_n(t_{\ell-1}) \ ,$$

$$v S \circ \tau_n(\ell) = \sup_{\tau_n(t_{\ell-1})<k\le\tau_n(t_\ell)} \left| \sum_{j=\tau_n(t_{\ell-1})+1}^{k} X_{n,j} \right| .$$

Further, indicator functions will be written as 1_B or $1\{\ \}$, i.e. $1_B(\omega)$ is one if $\omega\in B$ and zero if $\omega\in B^c$, and similarly $1\{\ \}$ is one if the event in curly brackets occurs and zero otherwise. Finally, sums with upper limits which are not integers are defined by $\sum_{j=1}^{x} = \sum_{j=1}^{[x]}$, i.e. summation is up to the greatest integer which does not exceed the upper limit.

2. Prerequisites: Functional Limit Theorems, Martingales, Approximation

For easy reference and because one purpose of this paper is to make a complete exposition of the martingale central limit theorem, we will in this section list the results on convergence in distribution and on martingales which are needed for the proofs. The section furthermore contains three lemmas which are essential for the truncation and approximation procedures we will use.

Let X, $\{X_n\}$ be random variables with values in a complete separable metric space (S,ρ). With standard terminology and notation, X_n converges in distribution to X, $X_n \stackrel{d}{\to} X$ in (S,ρ), if $h(X_n) \stackrel{d}{\to} h(X)$ in $\mathbb{R}$, for all functions $h\colon S \to \mathbb{R}$ which are bounded and continuous almost surely with respect to the distribution of X, and X_n converges in probability to X, $X_n \stackrel{P}{\to} X$ if $\rho(X_n,X) \stackrel{d}{\to} 0$. If X has a standard normal distribution in $\mathbb{R}^d$ we also write convergence in distribution as $X_n \stackrel{d}{\to} N_d(0,I)$ and for $d = 1$, as $X_n \stackrel{d}{\to} N(0,1)$. Besides $\mathbb{R}$, $\mathbb{R}^d$, we will be interested in the metric spaces $D[0,T]$ and $D[0,\infty)$ of functions on $[0,T]$ and on $[0,\infty)$ -- which have left limits and are right continuous, with metrics described in [3,21] and in the subset $D_0[0,1]$ of nondecreasing functions in $D[0,1]$. Convergence in distribution of vectors will throughout be with respect to the relevant product metric.

The first result is a criterion for convergence in $D[0,\infty)$. It can be obtained as an easy special case of the results of [21, Theorem 2.8] and [3, Theorem 15.5 amended with the argument of Theorem 8.3].

Proposition 2.1. With the notation of (1.1), suppose the following two conditions hold:

(i) (tightness) for each positive T and ε, there exists a function f such that for any partition of $[0,T]$,

$$\limsup_{n\to\infty} \sum_{\ell=1}^{k} P(\mathrm{v}S\circ\tau_n(\ell) \geq \varepsilon) \leq f(\Delta) ,$$

where $f(\Delta) \to 0$, as $\Delta \to 0$, and

(ii) (finite dimensional convergence) if $\{k_n\}_{n=1}^{\infty}$ is a sequence of integers and X a continuous stochastic process with $S\circ\tau_{k_n} \stackrel{d}{\to} X$ in $D[0,\infty)$ as $k \to \infty$, it follows that X has the same finite-dimensional distributions as B.

Then $$S\circ\tau_n \stackrel{d}{\to} B \quad \text{in} \quad D[0,\infty), \text{ as } n \to \infty .$$

The next results, on "random change of time" and approximation are phrased in terms of general processes X_n, Y_n and Y in $D[0,\infty)$ or $D[0,1]$ and timescales $\{\tau_n\}$.

Proposition 2.2. (i) Suppose $Y_n \xrightarrow{d} Y$ in $D[0,\infty)$ and that $\{\tau_n(t);\ t\in[0,1]\}$ are timescales such that

$$(2.1) \qquad \tau_n(t) \xrightarrow{P} \tau(t) \quad , \quad \text{as } n \to \infty ,$$

for each $t\in[0,1]$, where τ is a non-random continuous function. Then $Y_n \circ \tau_n \xrightarrow{d} Y \circ \tau$.

(ii) Suppose $Y_n \xrightarrow{d} Y$ in $D[0,1]$ and that $\{X_n\}$ are random variables in $D[0,1]$, such that

$$\sup_{0\le t\le 1} |X_n(t) - Y_n(t)| \xrightarrow{P} 0 \quad , \quad \text{as } n \to \infty .$$

Then $X_n \xrightarrow{d} Y$ in $D[0,1]$.

Proof: (i) Since pointwise convergence of increasing functions to a continuous limit implies uniform convergence, (2.1) implies that $\tau_n \xrightarrow{d} \tau$ in $D_0[0,1]$ and the result then follows from [21, Theorem 3.1].

(ii) See [3, Theorem 4.1]. □

From martingale theory we will use some simple consequences of the optional sampling theorem (Proposition 2.3 below), the extension of Kolmogorov's inequality to martingales (Proposition 2.4(i)) and one half of Burkholder's square function inequality (Proposition 2.4(ii)). All of this belongs to the standard fare from a first encounter with martingale theory, except perhaps the square function inequality. An elementary (albeit pedestrian) proof of this latter result is sketched in the appendix for the special case we shall need--an elegant proof for the general case is given in [4].

Proposition 2.3. Let $\{X_j, \mathcal{B}_j;\ j\ge 1\}$ be a sequence of martingale differences, let $\tau\le\tau'$ be stopping times, and write $\mathcal{B}_\tau$ for the pre-τ-sigma-algebra. Suppose that $E\sum_{j=1}^{\tau'} X_j^2 < \infty$. Then

$$E\ \{\sum_{j=\tau+1}^{\tau'} X_j \,||\, \mathcal{B}_\tau\} = 0$$

and

$$E\{(\sum_{j=\tau+1}^{\tau'} X_j)^2 \| \mathcal{B}_\tau\} = E\{\sum_{j=\tau+1}^{\tau'} X_j^2 \| \mathcal{B}_\tau\} .$$

Proposition 2.4. Let $\{X_j, \mathcal{B}_j\}$ be a martingale difference sequence and let $\tau \le \tau'$ be stopping times. Then

(i) for any integer n and real numbers $p \ge 1$, $\varepsilon > 0$,

$$P(\max_{\tau<k\le\tau'\wedge n} |\sum_{j=\tau+1}^{k} X_j| \ge \varepsilon) \le \frac{1}{\varepsilon^p} E|\sum_{j=\tau+1}^{\tau'\wedge n} X_j|^p ,$$

(ii) for $p>1$ and C a constant which only depends on p,

$$E|\sum_{j=\tau+1}^{\tau'} X_j|^p \le CE(\sum_{j=\tau+1}^{\tau'} X_j^2)^{p/2} ,$$

and

(iii) $$P(\max_{\tau<k\le\tau'} |\sum_{j=\tau+1}^{\tau'} X_j| \ge \varepsilon) \le \frac{C}{\varepsilon^p} E(\sum_{j=\tau+1}^{\tau'} X_j^2)^{p/2} .$$

Proof: (i) is the extension of Kolmogorov's inequality applied to the martingale $\{\sum_{j=\tau+1}^{\tau'\wedge k} X_j\}_{k=1}^{n}$ and (ii) is one of Burkholder's square function inequalities. Further, combining (i) and (ii) we have that

$$P(\max_{\tau<k\le\tau'\wedge n} |\sum_{j=\tau+1}^{k} X_j| \ge \varepsilon) \le \frac{C}{\varepsilon^p} E(\sum_{j=\tau+1}^{\tau'\wedge n} X_j^2)^{p/2} ,$$

and (iii) follows by letting $n\to\infty$. □

As will be seen, it is convenient to have an easy means of comparing the size of a sum of positive variables with the sum of their conditional expectation. In the present context, special cases of the following result has been used by several authors, but the result itself--and its easy proof--is believed to be new.

Lemma 2.5. Suppose $\{Z_{n,i}, \mathcal{B}_{n,i}\}$ is an adapted array of positive random variables and that τ_n is a stopping time with respect to $\mathcal{B}_{n,1}, \mathcal{B}_{n,2}, \ldots$, for each n. Then

(i) $$\sum_{j=1}^{\tau_n} E_{j-1}(Z_{n,j}) \xrightarrow{P} 0 \Rightarrow \sum_{j=1}^{\tau_n} Z_{n,j} \xrightarrow{P} 0 ,$$

and

(ii) if $\{\max_{1\le j\le\tau_n} Z_{n,j}\}_{n=1}^{\infty}$ is uniformly integrable then

$$\sum_{j=1}^{\tau_n} Z_{n,j} \overset{P}{\to} 0 \Rightarrow \sum_{j=1}^{\tau_n} E_{j-1}(Z_{n,j}) \overset{P}{\to} 0 .$$

Proof: (i) For any stopping time ν_n, letting N tend to infinity in the identity

$$E\sum_{j=1}^{\nu_n\wedge N} Z_{n,j} = \sum_{j=1}^{N} E1\{j\le\nu_n\}Z_{n,j}$$

$$= \sum_{j=1}^{N} E\{1\{j\le\nu_n\}E(Z_{n,j}\,\|\,\mathcal{B}_{n,j-1})\}$$

$$= E\sum_{j=1}^{\nu_n\wedge N} E(Z_{n,j}\,\|\,\mathcal{B}_{n,j-1}) ,$$

shows that

$$E\sum_{j=1}^{\nu_n} Z_{n,j} = E\sum_{j=1}^{\nu_n} E_{j-1}(Z_{n,j}) .$$

Let

$$\nu_n' = \inf\{k\ge1;\ \sum_{j=1}^{k} E_{j-1}(Z_{n,j})\ge1\}$$

$$\nu_n = (\nu_n'-1)\wedge\tau_n ,$$

and note that $\nu_n'-1$, and hence ν_n, is a stopping time, and clearly $P(\nu_n\ne\tau_n)\to0$ since $\sum_{j=1}^{\tau_n} E_{j-1}(Z_{n,j}) \overset{P}{\to} 0$. Further, since $\nu_n\le\tau_n$,

$$\sum_{j=1}^{\nu_n} E_{j-1}(Z_{n,j}) \overset{P}{\to} 0 , \quad \text{as } n\to\infty$$

and $0 \le \sum_{j=1}^{\nu_n} E_{j-1}(Z_{n,j})\le1$, since $\nu_n \le \nu_n'-1$, so the sum is uniformly integrable and hence

$$E \sum_{j=1}^{\nu_n} Z_{n,j} = E \sum_{j=1}^{\nu_n} E_{j-1}(Z_{n,j}) \to 0 \quad , \quad \text{as } n \to \infty \quad .$$

Thus, for any $\varepsilon>0$,

$$P(\sum_{j=1}^{\tau_n} Z_{n,j}>\varepsilon) \le P(\tau_n \ne \nu_n) + P(\sum_{j=1}^{\nu_n} Z_{n,j}>\varepsilon) \to 0 \quad , \quad \text{as } n \to \infty,$$

using Chebycheff's inequality for the last term.

(ii) Define in this case

$$\nu_n' = \inf\{k\ge 1; \sum_{j=1}^{k} Z_{n,j}>1\} \ ,$$

$$\nu_n = \nu_n' \wedge \tau_n \ ,$$

so that ν_n again is a stopping time, and note that $0 \le \sum_{j=1}^{\nu_n} Z_{n,j} \le 1 + \max\{Z_{n,j}; 1\le j\le \tau_n\}$. By the assumption this shows that $\sum_{j=1}^{\nu_n} Z_{n,j}$ is uniformly integrable, and the proof can then be completed in the same way as part (i). □

The proof of the following frequently used result is left to the reader.

Lemma 2.6. Let $\{X_n\}$ be real random variables. Then

(i) $X_n \overset{P}{\to} 0$ if and only if there exists constants $\varepsilon_n \to 0$, such that $P(|X_n| >\varepsilon_n)\to 0$, as $n \to \infty$, and

(ii) $\{X_n\}_{n=1}^{\infty}$ is tight if and only if $X_n/a_n \overset{P}{\to} 0$ for any sequence $\{a_n\}$ of constants such that $a_n \to \infty$ as $n \to \infty$.

Combining Lemma 2.6(ii) and Lemma 2.5, and noting that $\{\max\{Z_{n,j}; 1\le j\le\tau_n\}\}_{n=1}^{\infty}$ is uniformly integrable implies that $\{\max\{Z_{n,j}/a_n; 1\le j\le\tau_n\}\}_{n=1}^{\infty}$ is uniformly integrable if $a_n \to \infty$, leads to the next lemma.

Lemma 2.7. Let $\{Z_{n,j},\mathcal{B}_{n,j}\}$ and $\{\tau_n\}$ be as in Lemma 2.5. Then

(i) if $\{\sum_{j=1}^{\tau_n} E_{j-1}(Z_{n,j})\}_{n=1}^{\infty}$ is tight then $\{\sum_{j=1}^{\tau_n} Z_{n,j}\}_{n=1}^{\infty}$ is tight, and

(ii) if $\{\max\{Z_{n,j}; 1\le j\le\tau_n\}\}_{n=1}^{\infty}$ is uniformly integrable, and $\{\sum_{j=1}^{\tau_n} Z_{n,j}\}_{n=1}^{\infty}$ is tight, then $\{\sum_{j=1}^{\tau_n} E_{j-1} Z_{n,j}\}_{n=1}^{\infty}$ is tight.

3. Functional Central Limit Theorems for Martingales

Perhaps the most intuitively appealing explanation of the martingale central limit theorem is the Lévy-Doob-Dubins-Schwarz characterization of the Brownian motion--a martingale which has continuous sample paths and squared variation equal to the identity is necessarily a Brownian motion (see [6]). (However, this of course goes both ways: The martingale central limit theorem on the other hand throws light on the characterization, and it can be used to provide a simple proof of it). We start by proving an approximation version of the characterization, informally that if the jumps are uniformly small and the squared variation is uniformly close to the identity, then a martingale is approximately a Brownian motion. In the proof of finite dimensional convergence, we use ideas borrowed from Kunita and Watanabe's proof of the characterization of Brownian motion but could as well have used the customary proof of finite dimensional convergence, as e.g. in [8], which simplifies considerably in the present situation. However, the present proof seems to tie in better with our point of view. Once this result has been proved, the most general central limit theorems for martingales follow simply by random change of time and truncation.

Lemma 3.1. Suppose $\{X_{n,j}, \mathcal{B}_{n,j};\ j\geq 1,\ n\geq 1\}$ is a m.d.a. and $\{\tau_n(t);\ t\geq 0\}$ adapted timescales, with $\tau_n(0)=0$ such that there exist constants $\varepsilon_n \to 0$ satisfying

$$(3.1)\qquad |X_{n,j}| \leq \varepsilon_n \quad , \quad \Big|\sum_{j=1}^{\tau_n(t)} X_{n,j}^2 - t\Big| \leq \varepsilon_n \ ,$$

for all $j,n\geq 1$ and $t\geq 0$. Then

$$S\circ\tau_n \stackrel{d}{\to} B \quad , \quad \text{as } n\to\infty, \text{ in } D[0,\infty).$$

Proof: Using the notation of (1.1) and of Proposition 2.3 we will verify the hypothesis of Proposition 2.1.

(i) Tightness: By applying in turn Proposition 2.4(iii) for p = 4 and (3.1) we obtain that

$$\sum_{\ell=1}^{k} P(vS\circ\tau_n(\ell)\geq\varepsilon) \leq \sum_{\ell=1}^{k} \frac{C}{\varepsilon^4} E\Big(\sum_{j=\tau_n(t_{\ell-1})+1}^{\tau_n(t_\ell)} X_{n,j}^2\Big)^2$$

$$\leq \frac{C}{\varepsilon^4}\sum_{\ell=1}^{k}(t_\ell - t_{\ell-1}+2\varepsilon_n)^2 .$$

Since

$$\limsup_{n\to\infty}\sum_{\ell=1}^{k}(t_\ell - t_{\ell-1}+2\varepsilon_n)^2 = \sum_{\ell=1}^{k}(t_\ell - t_{\ell-1})^2$$

$$\leq \Delta T ,$$

this proves tightness.

(ii) Finite dimensional convergence: We have to prove that if $S\circ\tau_{k_n} \xrightarrow{d} X$ as $n\to\infty$, where X is continuous, then X has the same finite dimensional distributions as B. For simplicity of notation, we will assume that $k_n=n$, so that $S\circ\tau_n \xrightarrow{d} X$. The general case is then obtained simply by changing n to k_n in the computations below. By the continuity of X, $S\circ\tau_n(t) \xrightarrow{d} X(t)$, for $t\geq 0$, so that

$$Ee^{iuS\circ\tau_n(t)} \to Ee^{iuX(t)} = \phi_t(u) \quad , \text{ say,}$$

for each u and each $t>0$ and we only have to show that $\phi_t(u)$ is the characteristic function of B(t) and the corresponding fact for the k-dimensional characteristic functions.

Again with the notation of (1.1),

$$(3.2) \quad e^{iuS\circ\tau_n(T)} - 1 = \sum_{\ell=1}^{k} e^{iuS\circ\tau_n(t_{\ell-1})}(e^{iu\Delta S\circ\tau_n(\ell)} - 1)$$

$$= \sum_{\ell=1}^{k} e^{iuS\circ\tau_n(t_{\ell-1})}\{iu\Delta S\circ\tau_n(\ell) - \frac{u^2}{2}\ \Delta S\circ\tau_n(\ell)^{\,2}\} + r_n$$

where, by Taylor's formula

$$|r_n| \le \frac{u^3}{3!} \sum_{\ell=1}^{k} |\Delta S\circ\tau_n(\ell)|^3 .$$

Thus, using first Proposition 2.4 (ii) with p = 3 and then (3.1)

$$E|r_n| \le \frac{Cu^3}{3!} \sum_{\ell=1}^{k} E\Big(\sum_{j=\tau_n(t_{\ell-1})+1}^{\tau_n(t_\ell)} X_{n,j}^2\Big)^{3/2}$$

$$\le \frac{Cu^3}{3!} \sum_{\ell=1}^{k} (t_\ell - t_{\ell-1} + 2\varepsilon_n)^{3/2}$$

and hence, for $K = C\,u^3/3!$,

(3.3) $$\limsup_{n\to\infty} E|r_n| \le K\Delta^{1/2}T .$$

Now, with the obvious identifications, the hypothesis of Proposition 2.3 is satisfied, and hence

(3.4) $$E\sum_{\ell=1}^{k} e^{iuS\circ\tau_n(t_{\ell-1})}\Delta S\circ\tau_n(\ell) = \sum_{\ell=1}^{k} E\{e^{iuS\circ\tau_n(t_{\ell-1})}E(\Delta S\circ\tau_n(\ell)\,\|\,\mathcal{B}_{n,\tau_n(t_{\ell-1})})\}$$

$$= 0$$

and

(3.5) $$E\sum_{\ell=1}^{k} e^{iuS\circ\tau_n(t_{\ell-1})}\Delta S\circ\tau_n(\ell)^2 = \sum_{\ell=1}^{k} E\{e^{iuS\circ\tau_n(t_{\ell-1})}E(\Delta S\circ\tau_n(\ell)^2\,\|\,\mathcal{B}_{n,\tau_n(t_{\ell-1})})\}$$

$$= \sum_{\ell=1}^{k} E\{e^{iuS\circ\tau_n(t_{\ell-1})}E\Big(\sum_{j=\tau_n(t_{\ell-1})+1}^{\tau_n(t_\ell)} X_{n,j}^2\,\|\,\mathcal{B}_{n,\tau_n(t_{\ell-1})}\Big)\}$$

$$= \sum_{\ell=1}^{k} Ee^{iuS\circ\tau_n(t_{\ell-1})}\Delta(\ell) + R_n ,$$

where

(3.6) $$|R_n| \le \sum_{\ell=1}^{k} E\Big|\sum_{j=\tau_n(t_{\ell-1})+1}^{\tau_n(t_\ell)} X_{n,j}^2 - \Delta(\ell)\Big|$$

$$\leq k \quad \varepsilon_n \to 0 \quad , \quad \text{as } n \to \infty$$

by (3.1).

Taking expectations of both sides of (3.2), inserting (3.3)-(3.6), and letting $n \to \infty$ now proves that

$$|\phi_T(u) - 1 + \frac{u^2}{2} \sum_{\ell=1}^{k} \phi_{t_{\ell-1}}(u)| \leq K\Delta^{1/2} T \quad .$$

Since the partition $0=t_0<t_1<\ldots<t_k = T$ is arbitrary, this shows that $\phi_t(u)$ is Rieman integrable in t and that

$$\phi_T(u) - 1 = -\frac{u^2}{2} \int_0^T \phi_t(u)dt \quad .$$

Since $\phi_0(u) = 1$, the only solution to this equation is

$$\phi_t(u) = e^{-\frac{u^2}{2} t} = Ee^{iuB(t)} .$$

To conclude the proof it only remains to prove the corresponding result for the multidimensional characteristic functions. However, if $uSo\tau_n(T)$ is replaced by $\int_0^T u(t)dSo\tau_n(t)$, where the function u is assumed to be piecewise constant, with only finitely many jumps, then the same calculations show that

$$\begin{aligned} E \exp(i\int_0^T u(t)dX(t)) &= \exp(-1/2\int_0^T u(t)^2 dt) \\ &= E \exp(i\int_0^T u(t)dB(t)) \quad . \end{aligned}$$

□

The first step in weakening the hypothesis of Lemma 3.1 is concerned with the second part of (3.1).

Lemma 3.2. Suppose $\{X_{n,j}, \mathcal{B}_{n,j}\}$ is a m.d.a. and $\{\tau_n(t);\ t\in[0,1]\}$ are timescales such that

(3.7)
$$|X_{n,j}| \leq \varepsilon_n \to 0 \quad , \quad \text{as } n \to \infty, \text{ and}$$

$$\sum_{j=1}^{\tau_n(t)} X_{n,j}^2 \xrightarrow{P} \tau(t) \quad , \quad \text{as } n \to \infty \quad ,$$

for all $n,j\geq 1$ and $t\in[0,1]$, where τ is non-random and continuous, then

(3.8) $S\circ\tau_n \overset{d}{\rightarrow} B$, as $n\rightarrow\infty$, in $D[0,\infty)$.

Proof: Define, for $t\geq 0$

$$\eta_n(t) = \sum_{j=1}^{t} X^2_{n,j}$$

$$\eta_n^{-1}(t) = \inf\{s\geq 0;\ \eta_n(s)>t\} ,$$

(where we, without loss of generality, assume that $\eta_n(t) \overset{a.s.}{\rightarrow} \infty$, as $t\rightarrow\infty$) so that $\{\eta_n\}$ and $\{\eta_n^{-1}\}$ are timescales and η_n^{-1} in addition is adapted. Clearly

$$t < \sum_{j=1}^{\eta_n^{-1}(t)} X^2_{n,j} \leq t + \varepsilon_n^2 ,$$

and hence $S_n\circ\eta_n^{-1} \overset{d}{\rightarrow} B$ in $D[0,\infty)$, by Lemma 3.1. Since furthermore $\{\eta_n\circ\tau_n\}$ are timescales, and

$$\eta_n\circ\tau_n(t) = \sum_{j=1}^{\tau_n(t)} X^2_{n,j} \overset{P}{\rightarrow} t \quad , \quad \text{as } n\rightarrow\infty ,$$

for $t\in[0,1]$, by assumption, Proposition 2.2 (i) implies that

$$S_n\circ\eta_n^{-1}\circ\eta_n\circ\tau_n \overset{d}{\rightarrow} B\circ\tau \quad , \quad \text{in } D[0,1].$$

It is easily seen that moreover

$$\sup_{0\leq t\leq 1} |S\circ\tau_n(t) - S_n\circ\eta_n^{-1}\circ\eta_n\circ\tau_n(t)| \leq \sup_{k\geq 1} |S_n(k)-S_n\circ\eta_n^{-1}\circ\eta_n(k)|$$

$$\leq \varepsilon_n \rightarrow \infty \quad , \quad \text{as } n\rightarrow\infty ,$$

and, by Proposition 2.2(ii), this proves (3.8). □

In a sense, Lemma 3.2 says all there is to say about the central limit theorem for martingales, since if $S\circ\tau_n$ converges to Brownian motion, then the maximum of the

$X_{n,j}$'s has to tend to zero, which (more or less) leads to the first part of (3.7), and then the second part, with $\tau(t) = t$ is minimal. However, we will derive further conditions, which may be easier to check. The first result applies not only to m.d.a.'s, but to arrays which are asymptotically close to m.d.a.'s in an appropriate way.

Theorem 3.3. Suppose that $\{X_{n,j}, \mathcal{B}_{n,j};\ j\ge 1,\ n\ge 1\}$ is an adapted array and $\{\tau_n(t);\ t\in[0,1]\}$ are adapted timescales such that, for some $a>0$,

$$(3.9)\qquad \begin{aligned} &\sum_{j=1}^{\tau_n(t)} |E_{j-1}(X_{n,j}1\{|X_{n,j}|\le a\})| \xrightarrow{P} 0 \\ &\sum_{j=1}^{\tau_n(t)} X_{n,j}^2 \xrightarrow{P} \tau(t) \quad , \quad \text{as } n\to\infty , \end{aligned}$$

for $t\in[0,1]$, where τ is non-random and continuous. Then

$$(3.10)\qquad S\circ\tau_n \xrightarrow{d} B\circ\tau \quad , \quad \text{as } n\to\infty \text{ , in } D[0,1].$$

Further, if (3.9) holds for one $a>0$, then it holds for all $a>0$.

Proof: It is easily seen that the second part of (3.9) implies that $\max\{|X_{n,j}|;\ 1\le j\le\tau_n(1)\} \xrightarrow{P} 0$, and thus, by Lemma 2.6(i), there exist constants $\varepsilon_n \to 0$ such that

$$(3.11)\qquad P(\max_{1\le j\le\tau_n(1)} |X_{n,j}| > \varepsilon_n) \to 0 \quad , \quad \text{as } n\to\infty .$$

Hence

$$\sum_{j=1}^{\tau_n(1)} |X_{n,j}| 1\{a\ge |X_{n,j}|>\varepsilon_n\} \xrightarrow{P} 0 \quad , \quad \text{as } n\to\infty ,$$

since all the summands are zero on the set $\{\max_{1\le j\le\tau_n(1)} |X_{n,j}|\le\varepsilon_n\}$, and then, according to Lemma 2.5(ii),

$$(3.12)\qquad \sum_{j=1}^{\tau_n(1)} E_{j-1}(|X_{n,j}| 1\{a\ge |X_{n,j}|>\varepsilon_n\}) \xrightarrow{P} 0 \quad , \quad \text{as } n\to\infty .$$

Thus, using the first part of (3.9),

$$(3.13)\quad \sum_{j=1}^{\tau_n(1)} |E_{j-1}(X_{n,j}1\{|X_{n,j}|\le\varepsilon_n\})| \le \sum_{j=1}^{\tau_n(1)} |E_{j-1}(X_{n,j}1\{|X_{n,j}|\le a\})|$$

$$+ \sum_{j=1}^{\tau_n(1)} E_{j-1}(|X_{n,j}|1\{a\ge|X_{n,j}|>\varepsilon_n\})$$

$$\overset{P}{\to} 0 \quad , \quad \text{as } n\to\infty .$$

Now, put

$$X'_{n,j} = X_{n,j}1\{|X_{n,j}|\le\varepsilon_n\} \quad , \quad X''_{n,j} = X_{n,j} - X'_{n,j} \quad ,$$

$$Y_{n,j} = X'_{n,j} - E_{j-1}(X'_{n,j}) \quad , \quad S'\circ\tau_n(t) = \sum_{j=1}^{\tau_n(t)} Y_{n,j} \quad .$$

Then by (3.11) and (3.13)

$$\sup_{0\le t\le 1} |S\circ\tau_n(t) - S'\circ\tau_n(t)| \le \sum_{j=1}^{\tau_n(1)} |X''_{n,j}| + \sum_{j=1}^{\tau_n(1)} |E_{j-1}(X'_{n,j})|$$

$$\overset{P}{\to} 0 \quad , \quad \text{as } n\to\infty ,$$

and thus, according to Proposition 2.2(ii), the conclusion (3.10) will follow if we prove that $S'\circ\tau_n$ satisfies the conditions of Lemma 3.2. Clearly $\{Y_{n,j}, \mathcal{B}_{n,j}\}$ is a m.d.a., and $|Y_{n,j}| \le 2\varepsilon_n$, so it only remains to be shown that

$$(3.14)\quad \sum_{j=1}^{\tau_n(t)} Y_{n,j}^2 \overset{P}{\to} \tau(t) \quad , \quad \text{as } n\to\infty$$

for all $t\in[0,1]$. Here

$$\sum_{j=1}^{\tau_n(t)} Y_{n,j}^2 = \sum_{j=1}^{\tau_n(t)} (X'_{n,j})^2 - 2\sum_{j=1}^{\tau_n(t)} X'_{n,j}E_{j-1}(X'_{n,j}) + \sum_{j=1}^{\tau_n(t)} E_{j-1}(X'_{n,j})^2 ,$$

where

$$\sum_{j=1}^{\tau_n(t)} (X'_{n,j})^2 = \sum_{j=1}^{\tau_n(t)} X_{n,j}^2 - \sum_{j=1}^{\tau_n(t)} (X''_{n,j})^2$$

$$\overset{P}{\to} \tau(t) - 0 = \tau(t) \ , \quad \text{as } n \to \infty ,$$

since the $X''_{n,j}$'s are zero on $\{\max_{1\le j\le \tau_n(1)} |X_{n,j}| \le \varepsilon_n\}$. Further, by (3.13),

$$\left|\sum_{j=1}^{\tau_n(t)} X'_{n,j}E_{j-1}(X'_{n,j})\right| \le \varepsilon_n \sum_{j=1}^{\tau_n(1)} |E_{j-1}(X'_{n,j})|$$

$$\overset{P}{\to} 0 \ , \quad \text{as } n \to \infty ,$$

and similarly

$$\sum_{j=1}^{\tau_n(t)} E_{j-1}(X'_{n,j})^2 \le \varepsilon_n \sum_{j=1}^{\tau_n(1)} |E_{j-1}(X'_{n,j})|$$

$$\overset{P}{\to} 0 \ , \quad \text{as } n \to \infty ,$$

by (3.13) and thus (3.14) holds.

Finally for the last assertion of the theorem, if $a<a'$, say, using that $\max\{|X_{n,j}|;\ 1\le j\le \tau_n(1)\} \overset{P}{\to} 0$, we have

$$\left|\sum_{j=1}^{\tau_n(1)} E_{j-1}(X_{n,j}1\{|X_{n,j}|\le a\}) - \sum_{j=1}^{\tau_n(1)} E_{j-1}(X_{n,j}1\{|X_{n,j}|\le a'\})\right|$$

$$\le \sum_{j=1}^{\tau_n(1)} E_{j-1}(|X_{n,j}|1\{a<|X_{n,j}|\le a'\})$$

$$\overset{P}{\to} 0 \ , \quad \text{as } n \to \infty ,$$

by the same argument as for (3.12). □

As an easy corollary, we will now obtain conditions which insure that norming with sums of squares and with conditional variances is asymptotically equivalent.

Corollary 3.4. Let $\{X_{n,j}, \mathcal{B}_{n,j}\}$ be an adapted array, and $\{\tau_n(t);\ t\in[0,1]\}$ adapted timescales.

(i) If

$$\sum_{j=1}^{\tau_n(1)} E_{j-1}(X_{n,j}^2 1\{|X_{n,j}|>\varepsilon\}) \overset{P}{\to} 0 \quad , \quad \text{as } n\to\infty \quad , \quad \forall\varepsilon>0 \ , \tag{3.15}$$

and if either $\{\sum_{j=1}^{\tau_n(1)} E_{j-1}(X_{n,j}^2)\}_{n=1}^{\infty}$ or $\{\sum_{j=1}^{\tau_n(1)} X_{n,j}^2\}_{n=1}^{\infty}$ is tight (which in particular holds if either sum converges in probability) then

$$\sup_{0\le t\le 1} \left| \sum_{j=1}^{\tau_n(t)} X_{n,j}^2 - \sum_{j=1}^{\tau_n(t)} E_{j-1}(X_{n,j}^2) \right| \overset{P}{\to} 0 \quad , \quad \text{as } n\to\infty \ .$$

(ii) The hypothesis (3.15) is equivalent to the assumption that (3.15) holds for one fixed $\varepsilon>0$, and that $\max\{|X_{n,j}|;\ 1\le j\le\tau_n(1)\} \overset{P}{\to} 0$.

Proof: (i) Write $\tilde{X}_{n,j} = X_{n,j}1\{|X_{n,j}|>1\}$. Then by (3.15), $\sum_{j=1}^{\tau_n(1)} E_{j-1}(\tilde{X}_{n,j}^2) \overset{P}{\to} 0$, and it follows from Lemma 2.5(i) that $\sum_{j=1}^{\tau_n(1)} \tilde{X}_{n,j}^2 \overset{P}{\to} 0$, so we may in the proof assume that $|X_{n,j}| \le 1$, for $n,j\ge 1$.

Then both $\{\sum_{j=1}^{\tau_n(1)} E_{j-1}(X_{n,j}^2)\}_{n=1}^{\infty}$ and $\{\sum_{j=1}^{\tau_n(1)} X_{n,j}^2\}_{n=1}^{\infty}$ are tight, by the assumption combined with Lemma 2.7. Further, (3.15) implies that

$$\overline{M}_n = \max_{1\le j\le\tau_n(1)} E_{j-1}(X_{n,j}^2) \overset{P}{\to} 0 \quad , \quad \text{as } n\to\infty \ , \tag{3.16}$$

and by Lemma 2.5(i), that

$$\sum_{j=1}^{\tau_n(1)} X_{n,j}^2 1\{|X_{n,j}|>\varepsilon\} \overset{P}{\to} 0 \quad , \quad \text{as } n\to\infty \quad , \quad \forall\varepsilon>0 \ ,$$

which in the same way gives that

(3.17) $$M_n = \max_{1\le j\le \tau_n(1)} X_{n,j}^2 \xrightarrow{P} 0 \quad , \quad \text{as } n\to\infty .$$

Clearly, $\{Y_{n,j} = X_{n,j}^2 - E_{j-1}X_{n,j}^2\}$ is a m.d.a. with $|Y_{n,j}| \le 2$, and since

$$\sum_{j=1}^{\tau_n(t)} Y_{n,j}^2 \le 2\{\sum_{j=1}^{\tau_n(1)} X_{n,j}^4 + \sum_{j=1}^{\tau_n(1)} E_{j-1}(X_{n,j}^2)^2\}$$

$$\le 2M_n \sum_{j=1}^{\tau_n(1)} X_{n,j}^2 + 2\tilde{M}_n \sum_{j=1}^{\tau_n(1)} E_{j-1}(X_{n,j}^2)$$

$$\xrightarrow{P} 0 \quad , \quad \text{as } n\to\infty ,$$

by tightness and (3.16), (3.17), and the Corollary then follows from the theorem (with $\tau(t) \equiv 0$).

(ii) From (3.17) follows that (3.15) implies that $\max\{|X_{n,j}|;\ 1\le j\le\tau_n(1)\} \xrightarrow{P} 0$. The other implication follows in the same way as the last assertion of the theorem. □

We can now prove the general functional central limit theorem for martingales. Of course, the most important special case of it is when $\tau(t) \equiv t$, and the limiting process is an ordinary Brownian motion.

<u>Theorem 3.5</u>. Suppose $\{X_{n,j},\mathcal{B}_{n,j}\}$ is a m.d.a., $\{\tau_n(t);\ t\in[0,1]\}$ are adapted time-scales and $\tau(t);\ t\in[0,1]$ is a continuous, non-random function and suppose that one of the following three sets of conditions holds:

(3.18)
$$E \max_{1\le j\le\tau_n(1)} |X_{n,j}| \to 0 \quad , \text{ as } n\to\infty ,$$

$$\sum_{j=1}^{\tau_n(t)} X_{n,j}^2 \xrightarrow{P} \tau(t) \quad , \quad \text{as } n\to\infty \quad , \quad \text{for } t\in[0,1] ,$$

or

$$\sum_{j=1}^{\tau_n(1)} E_{j-1}(X_{n,j}^2 1\{|X_{n,j}|>\varepsilon\}) \xrightarrow{P} 0 \quad , \quad \text{as } n \to \infty \quad , \quad \forall \varepsilon>0$$

(3.19)

$$\sum_{j=1}^{\tau_n(t)} E_{j-1}(X_{n,j}^2) \xrightarrow{P} \tau(t) \quad , \quad \text{as } n \to \infty \quad , \quad \text{for } t\in[0,1] \; ,$$

or

(3.20) (3.19) holds for one $\varepsilon>0$, and $\max_{1\le j\le\tau_n(1)} |X_{n,j}| \xrightarrow{P} 0$, as $n \to \infty$.

Then

$$S\circ\tau_n \xrightarrow{d} B\circ\tau \quad \text{as } n \to \infty \quad , \quad \text{in } D[0,1] \; .$$

<u>Proof</u>: Assume first (3.18) is satisfied. Then $\{\max_{1\le j\le\tau_n(1)} |X_{n,j}|\}_{n=1}^{\infty}$ is uniformly integrable and

$$\sum_{j=1}^{\tau_n(1)} |X_{n,j}| 1\{|X_{n,j}|>1\} \xrightarrow{P} 0 \quad , \quad \text{as } n \to \infty \; .$$

Since $\{X_{n,j}\}$ is a m.d.a., it follows, using Lemma 2.5(ii) that

$$\sum_{j=1}^{\tau_n(1)} |E_{j-1}(X_{n,j} 1\{|X_{n,j}|\le 1\})| = \sum_{j=1}^{\tau_n(1)} |E_{j-1}(X_{n,j} 1\{|X_{n,j}|>1\})|$$

$$\le \sum_{j=1}^{\tau_n(1)} E_{j-1}(|X_{n,j}| 1\{|X_{n,j}|>1\})$$

$$\xrightarrow{P} 0 \quad , \quad \text{as } n \to \infty \; ,$$

and thus the conditions of Theorem 3.3 are satisfied, and $S\circ\tau_n \to B\circ\tau$, as required.

Next, assume (3.19) holds. By Corollary 3.4, it follows that

$$\sum_{j=1}^{\tau_n(t)} X_{n,j}^2 \xrightarrow{P} \tau(t) \quad \text{as } n \to \infty \quad , \quad \text{for } t \in [0,1] \ ,$$

and, again using that $\{X_{n,j}\}$ is a m.d.a.,

$$\sum_{j=1}^{\tau_n(1)} |E_{j-1}(X_{n,j} 1\{|X_{n,j}| \le 1\})| \le \sum_{j=1}^{\tau_n(1)} |E_{j-1}(X_{n,j} 1\{|X_{n,j}| > 1\})|$$

$$\le \sum_{j=1}^{\tau_n(1)} E_{j-1}(X_{n,j}^2 1\{|X_{n,j}| > 1\})$$

$$\xrightarrow{P} 0 \ ,$$

by assumption, so again the conditions of Theorem 3.3 are satisfied, and $S \circ \tau_n \xrightarrow{d} B \circ \tau$.

Finally, by Corollary 3.4(ii), the conditions (3.19) and (3.20) are equivalent, so the result holds also under (3.20). □

Corollary 3.6. Suppose $\{X_{n,j}, \mathcal{B}_{n,j}\}$ is a m.d.a. and for each n, τ_n is a stopping time with respect to $\mathcal{B}_{n,1}, \mathcal{B}_{n,2}, \ldots$ and suppose one of the following three sets of conditions holds:

$$E \max_{1 \le j \le \tau_n} |X_{n,j}| \to 0 \ , \qquad \text{as } n \to \infty \ ,$$

(3.21)

$$\sum_{j=1}^{\tau_n} X_{n,j}^2 \xrightarrow{P} 1 \ , \qquad \text{as } n \to \infty \ ,$$

or

$$\sum_{j=1}^{\tau_n} E_{j-1}(X_{n,j}^2 1\{|X_{n,j}| > \varepsilon\}) \xrightarrow{P} 0 \quad , \quad \text{as } n \to \infty \quad , \quad \forall \varepsilon > 0 \ ,$$

(3.22)

$$\sum_{j=1}^{\tau_n} E_{j-1}(X_{n,j}^2) \xrightarrow{P} 1 \quad , \qquad \text{as } n \to \infty \ ,$$

or

(3.23) (3.22) holds for one $\varepsilon>0$, and

$$\max_{1\le j\le \tau_n} |X_{n,j}| \xrightarrow{P} 0 \ .$$

Then

$$(3.24) \qquad \sum_{j=1}^{\tau_n} X_{n,j} \xrightarrow{d} N(0,1) \quad , \quad \text{as } n\to\infty \ .$$

Proof: To prove the result assuming (3.21), define adapted timescales $\{\tau_n(t);\ t\in[0,1]\}_{n=1}^{\infty}$ (similar to $\eta_n^{-1}(t)$ in the proof of Lemma 3.2) by

$$\tau_n(t) = \inf\{k;\ \sum_{j=1}^{k} X_{n,j}^2 > t\}\wedge\tau_n \quad , \quad \text{for } 0\le t<1 \quad ,$$

$$\tau_n(1) = \tau_n \ .$$

It is easily seen that $\{\tau_n(t)\}$ satisfies the condition (3.18) of the theorem (cf. the proof of Lemma 3.2), so in particular $S\circ\tau_n(1) \xrightarrow{d} B(1)$, which is just another way of writing (3.24).

The proof under (3.22), or the equivalent condition (3.23) is similar; one just has to replace $X_{n,j}^2$ by $E_{j-1}(X_{n,j}^2)$ in the definition of $\tau_n(t)$. □

We conclude this section with several comments on the results.

(i) In reasonable circumstances the conditions are also necessary for the functional martingale central limit theorem. In fact, if $\{\max\{|X_{n,j}|;\ 1\le j\le\tau_n(1)\}\}$ is uniformly integrable, and if "τ_n takes all relevant values" (see [19] for a definition--this holds in particular if $\tau_n(t)$ only has jumps of size one, as e.g. when $\tau_n(t) = [nt]$), then $S\circ\tau_n \xrightarrow{d} B\circ\tau$ implies (3.18), see [19]. Easy examples show that neither uniform integrability nor "τ_n takes all relevant values," can be entirely dispensed with in this statement.

Similarly, if $\limsup_{n\to\infty} E \sum_{j=1}^{\tau_n(1)} X_{n,j}^2 \leq 1$, then $SO\tau_n$ implies (3.19) and the equivalent condition (3.20), see [7], [19]. Furthermore, $\{\max\{|X_{n,j}|;\ 1\leq j\leq \tau_n(1)\}\}_{n=1}^{\infty}$ then is uniformly integrable, and (3.18) follows from (3.19) and (3.20) by Corollary 3.4.

(ii) One important special case of the theorem is the degenerate one, when $\tau(t) \equiv 0$. From the theorem, if $\{\max\{|X_{n,j}|;\ 1\leq j\leq \tau_n\}_{n=1}^{\infty}$ is uniformly integrable, and $\sum_{j=1}^{\tau_n} X_{n,j}^2 \overset{P}{\to} 0$, for some sequence $\{\tau_n\}_{n=1}^{\infty}$ of stopping times, then

$$(3.25) \qquad \sup_{1\leq k\leq \tau_n} \left| \sum_{j=1}^{k} X_{n,j} \right| \overset{P}{\to} 0, \quad \text{as } n\to\infty ,$$

and conversely, if (3.25) holds, and $\{\max\{|X_{n,j}|;\ 1\leq j\leq \tau_n\}_{n=1}^{\infty}$ is uniformly integrable, then $\sum_1^{\tau_n} X_{n,j}^2 \overset{P}{\to} 0$.

Similarly, if $\sum_{j=1}^{\tau_n} E_{j-1}(X_{n,j}^2) \overset{P}{\to} 0$, then (3.25) holds, and conversely, (3.25) and $\limsup_{n\to\infty} E \sum_{j=1}^{\tau_n} X_{n,j}^2 < \infty$ implies that $\sum_{j=1}^{\tau_n} E_{i-1}(X_{n,i}^2) \overset{P}{\to} 0$.

(iii) If the $X_{n,j}$'s in each row are independent, and the $\tau_n(t)$'s are non-random, then for $\mathcal{B}_{n,j} = \sigma(X_{n,1},\ldots,X_{n,j})$ the second part of (3.19) just says that the that the normalization is such that $V(\sum_{j=1}^{\tau_n(t)} X_{n,i}^2) \to \tau(t)$, and the first part is Lindeberg's condition.

(iv) There is another important special case in which the conditions of the theorem are particularly easy to check: if the $X_{n,j}$'s are obtained by normalizing a single stationary ergodic sequence $\ldots X_{-1}, X_0, X_1,\ldots,$

$$X_{n,j} = \frac{X_j - E(X_j \| \mathcal{B}_{j-1})}{\sigma\sqrt{n}} ,$$

with $\mathcal{B}_j = \mathcal{B}_{n,j} = \sigma(\ldots, X_{j-1}, X_j)$ and $\sigma^2 = E(X_j - E(X_j\|\mathcal{B}_{j-1}))^2$ assumed to be strictly positive and finite, then, for $\tau_n(t) = [nt]$, it follows at once from the ergodic theorem that (3.19) holds and hence

$$\sum_{j=1}^{[nt]} (X_j - E(X_j \| \mathcal{B}_{j-1}))/\sqrt{n} \xrightarrow{d} \sigma B(t) .$$

In nonstationary cases, the conditions of the theorem often have to be checked by computing higher moments, e.g. the first parts of (3.18), (3.19) follow if

$$E \sum_{j=1}^{\tau_n(1)} X_{n,j}^{\alpha} \to 0 \quad , \quad \text{as } n \to \infty \quad , \quad \text{for some } \alpha > 2 \ ,$$

and the second parts of (3.18), (3.19) may be obtained by computing means and variances of the sums on the left hand sides.

(v) Throughout, we have (implicitly) assumed that $\tau_n(1)$, and hence $\tau_n(t)$ for $t<1$, is finite a.s. However, a small further argument, using (ii) above shows that this can be dispensed with, and that the theorem (and the corollary) holds also if $\tau_n(1)$ (or τ_n) are extended stopping times which may be infinite with positive probabilities provided $\sum_{j=1}^{\tau_n(1)} X_{n,j}$ converges with a probability which tends to one (this is automatic under (3.19), (3.20)).

(vii) It is obvious from the proof of Proposition 2.2(i) that the second part of (3.18) can be weakened to requiring that $\sum_{j=1}^{\tau_n(\cdot)} X_{n,j}^2 \xrightarrow{d} \tau(\cdot)$ in $D_0[0,1]$, with $\tau(\cdot)$ non-random, but possibly discontinuous, and similar remarks apply to (3.19) and (3.20).

4. Mixing and Multidimensional Processes

In this section, the convergence in the previous results will be strengthened to Rényi-mixing--which in turn will make it possible to remove the condition that $\tau(t)$ is non-random--and multivariate versions of the results will be obtained.

While the purpose of the previous sections is to provide complete proofs, starting from scratch, of the basic results of martingale central limit theory, the intention of the present section is only to indicate one possible direction for developing the results further--examples of other directions being provided by limit theory for continuous parameter martingales and diffusion approximations--and we will accordingly give a more sketchy development, sometimes leaving details of arguments to the reader, and referring to results from other areas as they are needed, rather than explicitly collecting them at the beginning.

Some further notions are needed for the results. As in Section 2, let X, $\{X_n\}_{n=1}^{\infty}$ be random variables in a complete separable metric space (S,ρ), and in addition assume that all the X_n's are defined on the same probability space $(\Omega,\mathcal{B},P)$. Then $\{X_n\}$ is *Rényi-mixing* (or just *mixing*) *with limit* X, $X_n \overset{d}{\to} X$ (mixing) if $X_n \overset{d}{\to} X$ in (S,ρ), with respect to the conditional probability $P(\cdot|B)$, for any $B\in\mathcal{B}$ with $P(B)>0$. Further, $\{X_n\}$ is *Rényi-stable* (or just *stable*) if X_n converges in distribution to some limit, with respect to $P(\cdot|B)$, for any $B\in\mathcal{B}$ with $P(B)>0$. Thus a mixing sequence is stable, and conversely if a stable sequence has the same limit with respect to $P(\cdot|B)$, for all B, then it is mixing. Loosely speaking, a sequence is mixing if it converges in distribution and is "asymptotically independent" of any fixed events. Prominent examples of sequences which are stable but not mixing is given by sequences which converge in probability, or almost surely, to a non-degenerate limit. Some indication of the use and interest of mixing is given by the following result.

Proposition 4.1. The following three assertions are equivalent:

(i) $X_n \to X$ (mixing),

(ii) $X_n \overset{d}{\to} X$, as $n\to\infty$, with respect to $P(\cdot|B)$ for all $B\in\mathcal{B}_0$ with $P(B)>0$ for some algebra $\mathcal{B}_0$ which generates $\sigma(X_1, X_2,\ldots)$, and

(iii) if Y, $\{Y_n\}_{n=1}^{\infty}$ are random variables in another complete separable metric space, with $Y_n \overset{d}{\to} Y$, as $n \to \infty$, then $(X_n, Y_n) \overset{d}{\to} (X,Y)$.

<u>Sketch of proof</u>. Clearly (i) is equivalent to

(4.1) $Eh(X_n)1_B \to Eh(X)P(B) \quad , \quad \text{as } n \to \infty ,$

for any bounded continuous function h: $S \to \mathbb{R}$ and event $B \in \mathcal{B}$, and similarly for (ii), with $\mathcal{B}$ replaced by $\mathcal{B}_0$. Clearly (i) implies (ii). Further, to any $\varepsilon>0$ and $B\in\mathcal{B}$ there exists a $B^{\varepsilon}\in\mathcal{B}_0$ with $P(B\Delta B^{\varepsilon}) < \varepsilon$, cf. [5], p. 606, and hence if $|h| \le C$, then

$$|Eh(X_n)1_B - Eh(X_n)1_{B^{\varepsilon}}| \le C\varepsilon .$$

This is easily seen to imply (4.1), if $Eh(X_n)1_{B^{\varepsilon}} \to Eh(X)P(B^{\varepsilon})$, for all $\varepsilon>0$, and hence (ii) implies that (4.1) holds for $B \in \sigma(X_1, X_2, \ldots)$. The case of general $B\in\mathcal{B}$ then follows by a further small argument, as in [1].

It is straightforward to see that (iii) holds if and only if $(X_n, Y) \overset{d}{\to} (X,Y)$ for any fixed random variable Y in (S', ρ'), and this in turn is equivalent to

(4.2) $Eh(X_n)1\{Y\in\tilde{B}\} \overset{d}{\to} Eh(X)P(\tilde{B}) \quad , \quad \text{as } n \to \infty \quad ,$

for any continuous bounded h; $S \to \mathbb{R}$ and any Y-continuity set $\tilde{B}$, cf. [3], p. 20. Obviously (i) implies (4.2), and conversely, for an arbitrary event $B\in\mathcal{B}$, taking $Y = 1_B$ and $\tilde{B} = (1/2, 3/2)$, say, (4.1) follows from (4.2). Hence (i) and (iii) are equivalent. □

Additional interesting properties of Rényi-mixing is that it is preserved under absolutely continuous change of measure, and that it implies that the sample paths fluctuate strongly, see [1,17,19]. Furthermore it should be noted that, with obvious changes only, Lemma 4.1 holds also if mixing convergence is replaced by stability. Using Proposition 4.1(ii) it is easy to see that the limits of the previous section are mixing.

Theorem 4.2. If the σ-algebras $\{\mathcal{B}_{n,j}\}$ are nested, i.e. if $\mathcal{B}_{n,j} \subset \mathcal{B}_{n+1,j}$, for all n,j, then the conclusions of Theorems 3.3 and 3.5 can be strengthened to mixing convergence, i.e., under the same hypotheses, $S\circ\tau_n \xrightarrow{d} B\circ\tau$ (mixing), and similarly for Corollary 3.6.

Proof: the proofs under the different hypotheses are all similar, so we will only give one as an example, say the first part of Theorem 3.5.

Thus, we will assume (3.18) holds. According to Proposition 4.1 it is sufficient to show that if B is a fixed event in the algebra $\cup_{n\geq 1, j\geq 1} \mathcal{B}_{n,j}$ with $P(B)>0$, then

(4.3) $\quad S\circ\tau_n \xrightarrow{d} B\circ\tau \quad , \quad \text{as } n \to \infty \quad , \quad \text{in } D[0,1] \, ,$

with respect to $P(\cdot|B)$. Since $B \in \cup_{n\geq 1, j\geq 1} \mathcal{B}_{n,j}$, there are n_0, j_0 with $B \in \mathcal{B}_{n_0,j_0}$, and since the $\mathcal{B}_{n,j}$'s are nested and increasing, it follows that

$$B \in \mathcal{B}_{n,j} \quad , \quad \text{for } n \geq n_0 \; , \quad j \geq j_0 \; .$$

Let $X'_{n,j} = X_{n,j} 1\{j \geq j_0\}$ and write $S'\circ\tau_n(t) = \sum_{j=1}^{\tau_n(t)} X'_{n,j}$. Then

$$\sup_{0\leq t\leq 1} |S\circ\tau_n(t) - S'\circ\tau_n(t)| \leq \sum_{j=1}^{j_0} |X_{n,j}| \leq j_0 \max_{1\leq j\leq \tau_n(1)} |X_{n,j}|$$

$$\to 0 \; ,$$

in P-prcbability, and thus, as is immediately seen, in $P(\cdot|B)$-probability. Hence by Proposition 2.2(ii), (4.3) follows if

(4.4) $\quad S'\circ\tau_n \xrightarrow{d} B\circ\tau \quad , \quad \text{as } n \to \infty \quad , \quad \text{in } D[0,1] \, ,$

under $P(\cdot|B)$. Clearly, $\{X'_{n,j}, \mathcal{B}_{n,j}; n\geq n_0, j\geq 1\}$ is a $P(\cdot|B)$-m.d.a.,

$$E(\max_{1\leq j\leq \tau_n(1)} |X'_{n,j}| \,|B) \leq \frac{1}{P(B)} E \max_{1\leq j\leq \tau_n(1)} |X_{n,j}|$$

$$\to 0 \quad , \quad \text{as } n \to \infty \; ,$$

and

$$\sum_{j=1}^{\tau_n(t)} (X'_{n,j})^2 = \sum_{j=1}^{\tau_n(t)} X^2_{n,j} - \sum_{j=1}^{j_0 \wedge \tau_n(t)} X^2_{n,j}$$

$$\to \tau(t) - 0 = \tau(t) \ ,$$

in P--and hence in $P(\cdot|B)$-probability, and thus (4.4) follows from the first part of Theorem (3.5), applied to $\{X'_{n,j}, \mathcal{B}_{n,j}\}$. □

For the extension to several dimensions it will be useful to have a slightly different description of the limit process $B \circ \tau$, and to emphasize this we will change notation and will in the sequel write B_τ instead of $B \circ \tau$. Clearly, B_τ can be characterized as the normal process which has mean zero, variance function $\tau(t)$ and independent increments. Further, of course, $B_\tau = B \circ \tau$ makes sense also if τ is a stochastic process, and, in particular, if τ is independent of B, which we will assume throughout, then B_τ can be described by saying that conditional on τ, B_τ is normal, with zero mean, variance function $\tau(t)$, and with independent increments. Similarly, given a nonnegative definite, nondecreasing matrix valued, possibly random, function $\tau(t) = (\tau_{j,k}(t);\ 1 \le j,k \le d)$, we define a d-dimensional process $B_\tau = (B_\tau^{(1)}, \ldots, B_\tau^{(d)})$ by requiring that conditional on $\tau(t)$ it is a normal process with mean zero, variance matrix $V(B_\tau(t) \| \tau) = \tau(t)$, and with independent increments. We then have the following extension of Theorem 3.5. (Theorem 3.3 has the analogous extension--this is, however, left to the reader).

<u>Theorem 4.3</u>. Suppose $\{X_{n,j}, \mathcal{B}_{n,j}\}$ is a d-dimensional m.d.a., i.e., that $X_{n,j} = (X_{n,j}^{(1)}, \ldots, X_{n,j}^{(d)})^T$ are d-dimensional random vectors, such that $\{X_{n,j}^{(k)}, \mathcal{B}_{n,j}\}$ is a m.d.a., for k = 1,2,...,d, that $\{\tau_n(t);\ t \in [0,1]\}$ are adapted timescales, and that $\tau(t) = (\tau_{j,k}(t);\ 1 \le j,k \le d)$ is a continuous, possibly random, matrix function.

(i) If one of the following three sets of conditions holds,

$$E \max_{1 \le j \le \tau_n(1)} |X_{n,j}^{(k)}| \to 0 \quad , \quad \text{as } n \to \infty \quad , \quad \text{for } k=1,\ldots,d,$$

(4.5)

$$\sum_{j=1}^{\tau_n(t)} X_{n,j}X_{n,j}^T \xrightarrow{P} \tau(t) \quad , \quad \text{as } n \to \infty \quad , \quad \text{for } t\in[0,1]$$

or

$$\sum_{j=1}^{\tau_n(1)} E_{j-1}((X_{n,j}^{(k)})^2 1\{|X_{n,j}^{(k)}|>\varepsilon\}) \xrightarrow{P} 0 \quad , \quad \text{as } n \to \infty, \ \forall\, \varepsilon>0, \text{ and for } k=1,\ldots,d,$$

(4.6)

$$\sum_{j=1}^{\tau_n(t)} E_{j-1}(X_{n,j}X_{n,j}^T) \xrightarrow{P} \tau(t) \quad , \quad \text{as } n \to \infty \quad , \quad \text{for } t\in[0,1] \ ,$$

or

(4.7) (4.6) holds for one $\varepsilon>0$, and $\max\limits_{1\le j\le \tau_n(1)} |X_{n,j}^{(k)}| \xrightarrow{P} 0$, as $n\to\infty$, for $k=1,\ldots,d$,

and if in addition the σ-algebras $\{\mathcal{B}_{n,j}\}$ are nested, then

$$S\circ\tau_n \xrightarrow{d} B_\tau \quad , \quad \text{as } n \to \infty \quad , \quad \text{in } D[0,1]^d \ ,$$

and the convergence is stable: If $Y_n \xrightarrow{P} Y$ in (S',ρ'), then $(S\circ\tau_n, Y_n) \xrightarrow{d} (B_\tau, Y)$ in $D[0,1]^d \times S'$, where the distribution of (B_τ,Y) is determined by the requirement that B and Y are independent, and that the distribution of (τ,Y) is the limit of the distributions of $(\sum_{j=1}^{\tau_n(\cdot)} X_{n,j}X_{n,j}^T, Y_n)$ or of $(\sum_{j=1}^{\tau_n(\cdot)} E_{j-1}(X_{n,j}X_{n,j}^T), Y_n)$, respectively.

(ii) If in (i) the limit $\tau(t)$ is non-random, then the hypothesis that the $\mathcal{B}_{n,j}$'s are nested can be deleted, and it still follows that $S\circ\tau_n \xrightarrow{d} B_\tau$ in $D[0,1]^k$, but the convergence is not necessarily stable.

Proof: Suppose first that d=1 and that (4.5) holds. Without loss of generality we may as in Lemma 3.2 assume that the m.d.a. $\{X_{n,j},\mathcal{B}_{n,j}\}$ satisfies $E \max\limits_{1\le j}|X_{n,j}|\to 0$ as $n \to \infty$, and that $\sum_{j=1}^{k} X_{n,j}^2 \xrightarrow{a.s.} \infty$, as $k \to \infty$ for each n. We will now proceed as in Lemma 3.2, defining

$$\eta_n(t) = \sum_{j=1}^{t} X_{n,j}^2$$

(4.8)

$$\eta_n^{-1}(t) = \inf\{s\geq 0;\ \eta_n(s)>t\} .$$

It follows at once from Theorem 4.2 that $S\circ\eta_n^{-1} \stackrel{d}{\to} B$ (mixing) in $D[0,1]$. Clearly Theorem 4.2 can be immediately translated to $D[0,T]$, for any $T>0$, and it follows that $S\circ\eta_n^{-1} \stackrel{d}{\to} B$ (mixing) in $D[0,T]$, for any $T>0$, which in turn implies that $S\circ\eta_n^{-1} \stackrel{d}{\to} B$ (mixing) in $D[0,\infty)$, cf. [21]. By assumption,

$$\eta_n\circ\tau_n = \sum_{j=1}^{\tau_n(t)} X_{n,j}^2 \stackrel{P}{\to} \tau(t) \quad ,$$

and if in addition $Y_n \stackrel{P}{\to} Y$, then by Proposition 4.1,

$$(S\circ\eta_n^{-1},\ \eta_n\circ\tau_n,\ Y_n) \stackrel{d}{\to} (B,\tau,Y) \ ,$$

where B is independent of (τ,Y). By a minor extension of Proposition 2.2(i), it follows that

$$(S\circ\eta_n^{-1}\circ\eta_n\circ\tau_n,\ Y_n) \stackrel{d}{\to} (B_\tau,\ Y) \ ,$$

and then since

$$\sup_{0\leq t\leq 1} |S\circ\tau_n(t) - S\circ\eta_n^{-1}\circ\eta_n\circ\tau_n(t)| \leq \max_{1\leq j}|X_{n,j}|$$

$$\stackrel{P}{\to} 0 \quad , \quad \text{as } n\to\infty \quad ,$$

the desired result follows, that

$$(S\circ\tau_n,\ Y_n) \stackrel{d}{\to} (B_\tau,\ Y) \quad , \quad \text{as } n\to\infty \ ,$$

this time by a small extension of the second part of Proposition 2.2 .

Still assuming (4.5), $\{S\circ\tau_n\}$ is tight also for $d>1$, since by what just has been proved its components converge in distribution, and hence are tight. Thus, similarly for convergence in $D[0,1]^d$ as in $D[0,1]$, it only has to be shown that the finite dimensional distributions converge. Equivalently, as is easily seen by considering multidimensional characteristic functions, it is sufficient to prove that

(4.9) $$\int u(t)\, dS\circ\tau_n(t) \xrightarrow{d} \int u(t)\, dB_\tau(t) ,$$

as $n \to \infty$, if $u(t)$ is non-random, piecewise constant and with only finitely many jumps (this is just the Cramér-Wold theorem, [3], p. 49). However, clearly the left hand side of (4.9) is a sum of martingale differences and the convergence follows easily--though with some notational qualms--from what has been proven for the case d=1.

The proof of part (i) under the hypotheses (4.6) or (4.7) differs from the above in only one place--instead of defining η_n, η_n^{-1} by (4.8) it is convenient to use

$$\eta_n(t) = \sum_{j=1}^{t} E_{j-1}(X_{n,j}^2) .$$

Finally, the one-dimensional version of part (ii) is just Theorem 3.5, and the multidimensional version is then obtained in precisely the same way as above. □

Corollary 4.4. Suppose $\{X_{n,j}, \mathcal{B}_{n,j}\}$ is a d-dimensional m.d.a., $\{\tau_n\}$ is a sequence of stopping times, and $\tau = (\tau_{j,k};\ 1 \le j,k \le d)$ a, possibly random, matrix.

(i) If one of the conditions (4.5)-(4.7) holds, with $\tau_n(1)$, $\tau_n(t)$ replaced by τ_n, and $\tau(t)$ replaced by τ, and if the σ-algebras $\{\mathcal{B}_{n,j}\}$ are nested, then

$$\sum_{j=1}^{\tau_n} X_{n,j} \xrightarrow{d} \tilde{B}_\tau , \quad \text{as } n \to \infty ,$$

where conditional on τ, $\tilde{B}_\tau$ is normal with mean zero and variance matrix τ. If furthermore τ is strictly positive definite a.s. and the modified version of (4.5) holds, then

$$\Big(\sum_{j=1}^{\tau_n} X_{n,j} X_{n,j}^T\Big)^{-1/2} \sum_{j=1}^{\tau_n} X_{n,j} \xrightarrow{d} N_d(0,I) , \quad \text{as } n \to \infty ,$$

and if the modified version of (4.6) or (4.7) holds, then

$$(\sum_{j=1}^{\tau_n} E_{j-1}(X_{n,j}X_{n,j}^T))^{-1/2} \sum_{j=1}^{\tau_n} X_{n,j} \overset{d}{\to} N_d(0,I) \quad , \quad \text{as } n \to \infty .$$

(ii) If τ is non-random, then the hypothesis that the σ-algebras are nested can be deleted.

Proof: The first part of (i) follows from the theorem in a similar way as Corollary 3.6 follows from Theorem 3.5, after reducing the problem to the case d=1 by considering linear functionals, $\sum_{k=1}^{d} u_k \sum_{j=1}^{\tau_n} X_{n,j}^{(i)}$ (the Cramér-Wold theorem). The second part is then immediate after noting that stability implies that, e.g. under the modified version of (4.5), there is joint convergence,

$$(\sum_{j=1}^{\tau_n} X_{n,j}X_{n,j}^T, \sum_{j=1}^{\tau_n} X_{n,j}) \overset{d}{\to} (\tau, B_\tau) \quad ,$$

and hence

$$(\sum_{j=1}^{\tau_n} X_{n,j}X_{n,j}^T)^{-1/2} \sum_{j=1}^{\tau_n} X_{n,j} \overset{d}{\to} \tau^{-1/2}B_\tau$$
$$\overset{d}{=} N_d(0,I) .$$

The proof of part (ii) is similar. □

Clearly the remarks (iii)-(vii) after Corollary 3.6 apply, also to the present situation. Further, as a final remark, considering e.g. d=1 and Condition 4.5, the requirement that the sums of squares converge in probability, and not only in distribution is only used to insure asymptotic independence, so that the marginal convergences $\sum_{j=1}^{\tau_n(\cdot)} X_{n,j}^2 \to \tau$ and $S\circ\eta_n^{-1} \to B$ imply joint convergence

(4.10) $$(\sum_{j=1}^{\tau_n(\cdot)} X_{n,j}^2, S\circ\eta_n^{-1}) \overset{d}{\to} (\tau, B) \quad , \quad \text{as } n \to \infty ,$$

where τ and B are independent, and hence convergence in probability can be replaced by any weaker condition which still insures (4.10). One set of such conditions is given in Theorem 3.4 of Hall and Heyde [8].

APPENDIX

Here we will prove the simple special case p=4 of the Burkholder inequality stated in Proposition 2.4(ii), which is the only case needed for this paper. In fact the proof of finite dimensional convergence, p. 12, ℓ 4 also uses p=3, but that might as well have been reduced to p=4 by first using the Liapunov inequality.

$$E|\Delta S\circ\tau_n(\ell)|^3 \le \{E\Delta S\circ\tau_n(\ell)^4\}^{3/4} .$$

Thus let $\{X_j, \mathcal{B}_j\}$ be a martingale difference sequence. We will start by assuming that Doob's inequality

$$\text{(A.1)} \quad E \max_{1\le k\le n} |\sum_{j=1}^{k} X_j|^4 \le C'E(\sum_{j=1}^{n} X_j)^4$$

is known, and prove that then

$$\text{(A.2)} \quad E(\sum_{j=1}^{n} X_j)^4 \le CE(\sum_{j=1}^{n} X_j^2)^2 ,$$

where C' and C are universal constants.

Let $S(n) = \sum_{j=1}^{n} X_j$ and $M(n) = \max\{|\sum_{j=1}^{k} X_j|;\ 1\le k\le n\}$. By the martingale difference property, $ES(j)^3X_{j+1} = E\{S(j)^3E(X_{j+1}\|\mathcal{B}_j)\} = 0$, so that, expanding $S(j+1)^4 = (S(j)+X_{j+1})^4$ we have

$$ES(j+1)^4 - ES(j)^4 = EX_{j+1}^4 + 4ES(j)X_{j+1}^3 + 6ES(j)^2X_{j+1}^2 ,$$

nd thus, summing over j,

$$\text{(A.3)} \quad ES(n)^4 = \sum_{j=1}^{n} EX_j^4 + 4\sum_{j=1}^{n} ES(j-1)X_j^3 + 6\sum_{j=1}^{n} ES(j-1)^2X_j^2 .$$

Inserting the obvious inequalities $|S(j-1)| \le M(n)$ and $|X_j| \le 2M(n)$ into (A.3) we obtain

$$\text{(A.4)} \quad ES(n)^4 \le (4+8+6)E(M(n)^2 \sum_{j=1}^{n} X_j^2) .$$

Now, by the Cauchy-Schwarz inequality and by (A.1)

$$EM(n)^2 \sum_{j=1}^{n} X_j^2 \le \{EM(n)^4\}^{1/2}\{E(\sum_{j=1}^{n} X_j^2)^2\}^{1/2}$$

$$\le \{C'ES(n)^4\}^{1/2}\{E(\sum_{j=1}^{n} X_j^2)^2\}^{1/2} ,$$

and inserting this into (A.4) and dividing through by $\{ES(n)^4\}^{1/2}$ we obtain (A.2) with $C = 18^2C'$ (which is not the best possible value of C, cf. [4]).

The general assertion of Proposition 2.4(ii), for p=4, now follows easily by replacing X_j by $X_j 1\{\tau<X_j\le\tau'\}$ in (A.2) and letting n tend to infinity, using Fatou's lemma on the left hand side. (In fact the proof works not only for p=4, but for any even integer p.)

REFERENCES

[1] Aldous, D.J. and Eagleson, G.K. (1978): On mixing and stability of limit theorems. Ann. Probab. 6, 325-331.

[2] Beśka, M., Kłopotowski, A. and Słomiński, L. (1982): Limit theorems for random sums of dependent d-dimensional vectors. Z. Wahrsch. verw. Geb. 61, 43-57.

[3] Billingsley, P. (1968): *Convergence of Probability Measures*. Wiley, New York.

[4] Burkholder, D.L. (1973): Distribution function inequalities for martingales. Ann. Probab. 1, 19-42.

[5] Doob, J.L. (1953): *Stochastic Processes*. Wiley, New York.

[6] Dubins, L.E. and Schwarz, G. (1965): On continuous martingales. Proc. Nat. Acad. Sci. 53, 913-916.

[7] Gänssler, P. and Häusler, E. (1980): Remarks on the functional central limit theorem for martingales. Z. Wahrsch. Verw. Geb. 50, 237-243.

[8] Hall, P. and Heyde, C.C. (1980): *Martingale Limit Theory*. Academic Press, New York.

[9] Helland, I. (1982): Central limit theorems for martingales with discrete or continuous time. Scand. J. Statist. 9, 79-94.

[10] Jeganathan, P. (1981): A solution to the martingale central limit problem, parts I, II, and III, plus addendum to part II. Techn. reports, Stat.-Math. division, Indian Statist. Inst., 203 BT Road, Calcutta 700035.

[11] Jeganathan, P. (1981): Martingale invariance principles using Lévy's martingale characterization of the Brownian motion. Techn. report, Stat.-Math. division, Indian Statist. Inst., 203 BT Road, Calcutta 700035.

[12] Kłopotowski, A. (1980): Mixtures of infinitely divisible distributions as limit laws for sums of dependent random variables. Z. Wahrsch. verw. Geb. 51, 101-113.

[13] Lipster, R. Sh. and Shirayev, A.N. (1980): A functional central limit theorem for semimartingales. Theory Probab. Appl. 25, 667-688.

[14] Lipster, R. Sh. and Shirayev, A.N. (1981): On necessary and sufficient conditions in the functional central limit theorem for semimartingales. Theory Probab. Appl. 26, 130-135.

[15] Rebolledo, R. (1977): Remarques sur la convergence en loi de martingales vers des martingales continues. C.R. Acad. Sc. Paris 285, 517-520.

[16] Rebolledo, R. (1980): Central limit theorems for local martingales. Z. Wahrsch. verw. Geb. 51, 269-286.

[17] Rényi, A. (1958): On mixing sequences of sets. Acta Math. Acad. Sci. Hungar. 9, 215-228.

[18] Rootzén, H. (1976): Fluctuations of sequences which converge in distribution. Ann. Probab. 4, 456-463.

[19] Rootzén, H. (1980): On the functional central limit theorem for martingales, II. Z. Wahrsch. verw. Geb. 51, 79-93.

[20] Rosén, B. (1967): On the central limit theorem for sums of dependent random variables. Z. Wahrsch. verw. Geb. 7, 48-82.

[21] Whitt, W. (1980): Some useful functions for functional limit theorems. Math. Operations Res. 5, 67-85.

PREPRINTS 1982

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Holmgaard, Simon and Yu, Song Yu: Gaussian Markov Random Fields Applied to Image Segmentation.

No. 2 Andersson, Steen A., Brøns, Hans K. and Jensen, Søren Tolver: Distribution of Eigenvalues in Multivariate Statistical Analysis.

No. 3 Tjur, Tue: Variance Component Models in Orthogonal Designs.

No. 4 Jacobsen, Martin: Maximum-Likelihood Estimation in the Multiplicative Intensity Model.

No. 5 Leadbetter, M.R.: Extremes and Local Dependence in Stationary Sequences.

No. 6 Henningsen, Inge and Liestøl, Knut: A Model of Neurons with Pacemaker Behaviour Recieving Strong Synaptic Input.

No. 7 Asmussen, Søren and Edwards, David: Collapsibility and Response Variables in Contingency Tables.

No. 8 Hald, A. and Johansen, S.: On de Moivre's Recursion Formulae for Duration of Play.

No. 9 Asmussen, Søren: Approximations for the Probability of Ruin Within Finite Time.

PREPRINTS 1983

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Jacobsen, Martin: Two Operational Characterizations of Cooptional Times.

No. 2 Hald, Anders: Nicholas Bernoulli's Theorem.

No. 3 Jensen, Ernst Lykke and Rootzén, Holger: A Note on De Moivre's Limit Theorems: Easy Proofs.

No. 4 Asmussen, Søren: Conjugate Distributions and Variance Reduction in Ruin Probability Simulation.

No. 5 Rootzén, Holger: Central Limit Theory for Martingales via Random Change of Time.