A central limit theorem for ℓ^∞-valued martingale difference arrays and its application *

Yoichi Nishiyama

The Institute of Statistical Mathematics and Utrecht University

August 20, 1996 †

Abstract

A central limit theorem for $\ell^\infty(T)$-valued martingale difference arrays is given, where T is a non-empty set. As its application, the asymptotic behavior of log-likelihood ratio random fields in general statistical models is derived.

1 Introduction

Let T be a non-empty set. We denote by $\ell^\infty(T)$ the space of bounded, $\mathbb{R}$-valued functions defined on T, and equip it with the sup-norm $||\cdot||_T$. For every $n \in \mathbb{N}$, let $\mathfrak{B}^n = (\Omega^n, \mathcal{F}^n, \mathbf{F}^n, P^n)$ be a stochastic base, where $(\Omega^n, \mathcal{F}^n, P^n)$ is a probability space and $\mathbf{F}^n = \{\mathcal{F}_i^n\}_{i\in\mathbb{N}_0}$ is a non-decreasing sequence of sub-σ-fields of $\mathcal{F}^n$ indexed by $\mathbb{N}_0 = \{0\}\cup\mathbb{N}$. Here we make a definition.

Definition 1 *$\{M_i^n\}_{i\in\mathbb{N}} = \{(M_i^n(t)|t \in T)\}_{i\in\mathbb{N}}$ is an $\ell^\infty(T)$-valued martingale difference array on $\mathfrak{B}^n$ if*
(i) *M_i^n is a mapping from Ω^n to $\ell^\infty(T)$ for every $i \in \mathbb{N}$;*
(ii) *$\{M_i^n(t)\}_{i\in\mathbb{N}}$ is a $\mathbb{R}$-valued martingale difference array on $\mathfrak{B}^n$ for every $t \in T$.*

It is required in (ii) above that $M_i^n(t)$ is $\mathcal{F}_i^n$-measurable and $E_{i-1}^n M_i^n(t) = 0$ almost surely, where E_{i-1}^n denotes the $\mathcal{F}_{i-1}^n$-conditional expectation. Notice that we do not require any measurability of the $\ell^\infty(T)$-valued random element M_i^n.

The purpose of this paper is to give some sufficient conditions to ensure the weak convergence of sequences of $\ell^\infty(T)$-valued random elements

$$\sum_{i=1}^{\sigma^n} M_i^n = \left(\sum_{i=1}^{\sigma^n} M_i^n(t) \middle| \, t \in T \right),$$

where σ^n is a finite stopping time on $\mathfrak{B}^n$. Such problems have been investigated mainly for i.i.d. and row-independent cases, in the theory of empirical processes: see Dudley

*This research was performed in the Department of Mathematics, Utrecht University, while the author held a JSPS Fellowship for Research Abroad from the Japan Society for the Promotion of Science.

AMS 1991 subject classifications: 60F05, 60F17, 62E20.

Keywords and phrases. Central limith theorem, uniform metric, martingale, likelihood ratio, Markov chain.

†This postscript file was produced on November 3, 1997. But there is no difference between the original version and the present, except for font size.

(1978), Pollard (1982), Giné and Zinn (1984), Ossiander (1987), Andersen et al. (1988), among others. The recent book by Van der Vaart and Wellner (1996) gives a nice exposition about the weak convergence theory and broad results up to row-independent cases, with some new developments and applications to statistics. In the last few years, some authors have considered the problem to remove the assumption of independence: see Arcones and Yu (1994) and Doukhan et al. (1995) for stationary sequences; Levental (1989) and Bae and Levental (1995) for stationary martingale difference arrays; Nishiyama (1996) for semimartingales with continuous time parameters. The result given in Section 2 of the present paper generalizes Theorem 3.1 of Ossiander (1987), Theorem 2.11.9 of Van der Vaart and Wellner (1996), and Proposition 1.1 of Bae and Levental (1995), although the line of proof is exactly the same as the preceding ones. A "Jain-Marcus central limit theorem" for martingale difference arrays is also obtained as an application of our result.

Another purpose of this paper, exhibited in Section 3, is to make an extension of the *local asymptotic normality (LAN) theorem* which has the origin in the work by Le Cam (1960). For every $n \in \mathbb{N}$, let $\mathfrak{B}^n = (\Omega^n, \mathcal{F}^n, \mathbf{F}^n = \{\mathcal{F}_i^n\}_{i \in \mathbb{N}_0}, P^n)$ be a stochastic base. Let T be a non-empty set, and $\mathbf{P}^n = \{P^{n,t}; t \in T\}$ be a family of probability measures on $(\Omega^n, \mathcal{F}^n)$ such that $P^{n,t} \ll P^n$ for every $t \in T$. Here we define $Z_i^n(t) = dP_i^{n,t}/dP_i^n$, where $P_i^{n,t}$ (resp. P_i^n) denotes the restriction of $P^{n,t}$ (resp. P^n) on the σ-field $\mathcal{F}_i^n$. By using the result in Section 2, we derive the asymptotic behavior of log-likelihood ratio random fields

$$\log Z_{\sigma^n}^n = (\log Z_{\sigma^n}^n(t) | t \in T)$$

under some appropriate conditions. Such results, namely *functional LAN theorems*, have been investigated in finite-dimensional parametric models (i.e. the index set T is Euclidean): e.g., Le Cam (1970), Inagaki and Ogata (1975), Ogata and Inagaki (1977), Ibragimov and Has'minskii (1981), Kutoyants (1984), and Vostrikova (1987). On the other hand, the result in Section 3 seems the first one which concerns the general situation in which the index set T is arbitrary. This progress makes it possible to treat infinite-dimensional parametric (so-called, non- and semi-parametric) models. As an illustration, we discuss a non-parametric model of Markov chains by setting T to be a subset of a certain function space.

The proofs are given in Sections 4 and 5. To close the present section we make some more notations and definitions. Let a probability space $(\Omega^n, \mathcal{F}^n, P^n)$ be given. We define the *outer integral* of an arbitrary mapping $X^n : \Omega^n \to \mathbb{R}$, and the *outer probability* of an arbitrary subset B of Ω^n, exactly in the same way as on p.6 of Van der Vaart and Wellner (1996), and we denote them by $E^{n*}X^n$ and $P^{n*}(B)$ respectively. Next let a stochastic base $\mathfrak{B}^n$ be given.

Definition 2 *The* $\mathcal{F}_i^n$**-measurable covering** *of an arbitrary mapping* $X^n : \Omega^n \to \mathbb{R} \cup \{\infty\}$ *is defined as any* $\mathcal{F}_i^n$*-measurable mapping* $[X^n]_{\mathcal{F}_i^n} : \Omega^n \to \mathbb{R} \cup \{\infty\}$ *such that*
(i) $[X^n]_{\mathcal{F}_i^n} \geq X^n$ *identically;*
(ii) $[X^n]_{\mathcal{F}_i^n} \leq U^n$ *almost surely for every* $\mathcal{F}_i^n$*-measurable mapping* $U^n : \Omega^n \to \mathbb{R} \cup \{\infty\}$ *such that* $U^n \geq X^n$ *almost surely.*

It follows from Lemma 1.2.1 of Van der Vaart and Wellner (1996) that $[X^n]_{\mathcal{F}_i^n}$ always exists uniquely up to a P^n-null set, hence is well-defined in that sense. We denote by $\stackrel{P^n}{\longrightarrow}$ (resp. $\stackrel{P^{n*}}{\longrightarrow}$) the convergence in P^n-probability (resp. P^{n*}-probability). A sequence of mappings $X^n : \Omega^n \to \ell^\infty(T)$ converges weakly to a Borel law L on $(\ell^\infty(T), ||\cdot||_T)$ if

$$\lim_{n\to\infty} E^{n*} f(X^n) = \int_{\ell^\infty(T)} f(x)L(dx) \quad \forall f \in C_b(\ell^\infty(T)), \tag{1}$$

where $C_b(\ell^\infty(T))$ denotes the space of bounded, $||\cdot||_T$-continuous $\mathbb{R}$-valued functions defined on $\ell^\infty(T)$. We define $X^n \stackrel{P^n}{\Longrightarrow} X$ to mean that there exists a Borel measurable mapping X from a probability space to $\ell^\infty(T)$ such that its Borel law L_X is *tight* and satisfies (1) with $L = L_X$. Notice that the latter definition includes the *tightness* of the Borel law L_X, because we are concerned with such cases only.

2 Central limit theorem

Let $\mathfrak{B}^n$ be a stochastic base, on which an $\ell^\infty(T)$-valued martingale difference array $\{M_i^n\}_{i\in\mathbb{N}}$ and a finite stopping time σ^n are defined. For a given finite partition $\mathfrak{P} = \{S_k; 1 \le k \le N\}$ of T, we define

$$\hat{M}^n(\mathfrak{P}) = \max_{1\le k\le N} \sum_{i=1}^{\sigma^n} E_{i-1}^n |\Delta_i^n(S_k)|^2,$$

where

$$\Delta_i^n(S) = \left[\sup_{t,s\in S} |M_i^n(t) - M_i^n(s)|\right]_{\mathcal{F}_i^n} \quad \forall S \subset T. \tag{2}$$

Also we denote $|||M_i^n||| = [\sup_{t\in T} |M_i^n(t)|]_{\mathcal{F}_i^n}$. Here is the main theorem.

Theorem 1 *Let T be a non-empty set. For every $n \in \mathbb{N}$, let $\{M_i^n\}_{i\in\mathbb{N}}$ be an $\ell^\infty(T)$-valued martingale difference array and σ^n be a finite stopping time on a stochastic base $\mathfrak{B}^n$. Suppose the following conditions hold:*
(A) *every finite-dimensional marginal of $\sum_{i=1}^{\sigma^n} M_i^n$ converges weakly to a (tight,) Borel law;*
(B) $\sum_{i=1}^{\sigma^n} E_{i-1}^n |||M_i^n||| 1\{|||M_i^n||| > \varepsilon\} \stackrel{P^n}{\longrightarrow} 0$ *for every $\varepsilon > 0$;*
(C) *for every $\varepsilon \in (0,1)$ there exists a partition $\mathfrak{P}_\varepsilon$ of T with finite size N_ε such that*

$$\sup_{\varepsilon\in(0,1)\cap\mathbb{Q}} \frac{\hat{M}^n(\mathfrak{P}_\varepsilon)}{\varepsilon^2} = O_{P^n}(1) \quad \textit{and} \quad \int_0^1 \sqrt{\log N_\varepsilon}\, d\varepsilon < \infty.$$

Then $\sum_{i=1}^{\sigma^n} M_i^n$ converges weakly to a tight, Borel law.

We can use the usual central limit theorems for martingales to establish the weak convergence of the marginals. See e.g. Jacod and Shiryaev (1987). Here let us state a "Lindeberg-Feller" type theorem.

Corollary 2 *In the same situation as Theorem 1, suppose the following conditions* **(A′)**, **(B′)** *and* **(C)** *above hold:*
(A′) $\sum_{i=1}^{\sigma^n} E_{i-1}^n M_i^n(t) M_i^n(s) \stackrel{P^n}{\longrightarrow} C(t,s)$ *[some constant] for every $t, s \in T$;*
(B′) $\sum_{i=1}^{\sigma^n} E_{i-1}^n |||M_i^n|||^2 1\{|||M_i^n||| > \varepsilon\} \stackrel{P^n}{\longrightarrow} 0$ *for every $\varepsilon > 0$.*
Then it holds that $\sum_{i=1}^{\sigma^n} M_i^n \stackrel{P^n}{\Longrightarrow} X$, where each marginal $(X(t_1), ..., X(t_d))$ has the normal distribution $N(0, \Sigma)$ with $\Sigma = \{C(t_i, t_j)\}_{ij}$. Furthermore, if we set

$$\rho(t,s) = \sqrt{C(t,t) + C(s,s) - 2C(t,s)} \quad \forall t, s \in T, \tag{3}$$

then ρ defines a pseudo-metric on T such that (T,ρ) is totally bounded and almost all paths of X are uniformly ρ-continuous.

The condition **(C)**, namely the *partitioning entropy condition*, is a modification of the one in Theorem 2.11.9 of Van der Vaart and Wellner (1996), and close to the one in Proposition 1.1 of Bae and Levental (1995). It plays the same role as the usual *metric entropy condition (with bracketing)*, however, we do not have to equip T with any pseudo-metric ρ a-priori. This progress is due to the new tightness criterion, namely Theorem 1.5.6 of Van der Vaart and Wellner (1996) [Lemma 3 of this paper], in which no pseudo-metric is required. On the other hand, it is also an important problem to discuss the continuity of sample paths of limits. Corollary 2 asserts that the pseudo-metric ρ defined by (3) is suitable to do it. This fact comes from the Gaussian property of the limits. See Example 1.5.10 of Van der Vaart and Wellner (1996) for the details.

The proof of Theorem 1 is given in Section 4, in which we follow exactly the same line as the one for Theorem 2.11.9 of Van der Vaart and Wellner (1996) with some modifications for the present situation. The key point is just the usage of Bernstein-Freedman's inequality based on the condition **(C)**. However, our framework and results are more general than the previous ones, and in fact there are some new applications. First, let us briefly mention the generalization of Jain-Marcus's theorem to the case of martingale difference arrays at the end of this section. Another application, concerning the functional LAN theorem, is exhibited in the next section. Hopefully, they will be the prototypes of various applications.

Example: Jain-Marcus central limit theorem. Let (T,ρ) be a totally bounded pseudo-metric space, and for every $\varepsilon > 0$ we denote by $N(T,\rho;\varepsilon)$ the ε-covering number of T with respect to ρ. Consider the $\ell^\infty(T)$-valued martingale difference array $\{M_i^n\}_{i\in\mathbb{N}_0}$ on $\mathfrak{B}^n$ such that

$$|M_i^n(t) - M_i^n(s)| \le K_i^n \rho(t,s) \quad \forall t,s \in T,$$

where $\{K_i^n\}_{i\in\mathbb{N}_0}$ is an $\mathbb{R}$-valued adapted process. For given finite stopping time σ^n, a sufficient condition for **(C)** is

$$\sum_{i=1}^{\sigma^n} E_{i-1}^n |K_i^n|^2 = O_{P^n}(1) \quad \text{and} \quad \int_0^1 \sqrt{\log N(T,\rho;\varepsilon)}\,d\varepsilon < \infty. \tag{4}$$

Therefore, if the conditions **(A)** and **(B)** [resp. **(A′)** and **(B′)**] are satisfied in addition to the above (4), then we obtain the weak convergence of $\sum_{i=1}^{\sigma^n} M_i^n$ by Theorem 1 [resp. the central limit theorem of $\sum_{i=1}^{\sigma^n} M_i^n$ by Corollary 2].

3 Log-likelihood ratio random fields

For every $n \in \mathbb{N}$, let $\mathfrak{B}^n = (\Omega^n, \mathcal{F}^n, \mathbf{F}^n = \{\mathcal{F}_i^n\}_{i\in\mathbb{N}_0}, P^n)$ be a stochastic base. Let T be a non-empty set, and $\mathbf{P}^n = \{P^{n,t}; t \in T\}$ be a family of probability measures on $(\Omega^n, \mathcal{F}^n)$ such that $P^{n,t} \ll P^n$ for every $t \in T$. Here we define

$$Z_i^n(t) = \frac{dP_i^{n,t}}{dP_i^n},$$

where $P_i^{n,t}$ (resp. P_i^n) denotes the restriction of $P^{n,t}$ (resp. P^n) on the σ-field $\mathcal{F}_i^n$. We suppose $P_0^{n,t} = P_0^n$ for every $t \in T$, hence we can set $Z_0^n(t) = 1$. Let σ^n be a finite stopping time on $\mathfrak{B}^n$, and we also suppose the random element $\log Z^n_{\sigma^n} = (\log Z^n_{\sigma^n}(t)|t \in T)$ takes values in $\ell^\infty(T)$. Here we set

$$\begin{aligned}\tilde{Z}^n_i(t) &= \frac{Z^n_{i\wedge\sigma^n}(t)}{Z^n_{(i-1)\wedge\sigma^n}(t)},\\ \zeta^n_i(t) &= \sqrt{\frac{Z^n_{i\wedge\sigma^n}(t)}{Z^n_{(i-1)\wedge\sigma^n}(t)}} - 1.\end{aligned}$$

Further, let $g : (0,\infty) \to \mathbb{R}$ be a function which satisfies the following condition:

(g) there exists a constant $\delta \in (0,1)$ such that

$$|\log x - \log y| \leq |g(x) - g(y)| \quad \text{whenever} \quad x, y \in [1-\delta, 1+\delta].$$

Notice that the functions of the form $g(x) = Kx^\gamma$, with $\gamma > 0$ and large $K > 0$, satisfy the condition **(g)**, and also that the values of g on the complement of $[1-\delta, 1+\delta]$ can be set to zero. Now we define the following notation: for given function g which satisfies **(g)** and given finite partition $\mathfrak{P} = \{S_k; 1 \leq k \leq N\}$ of T

$$\hat{Z}^n_g(\mathfrak{P}) = \max_{1\leq k\leq N} \sum_{i=1}^{\sigma^n} E^n_{i-1}\left[\sup_{t,s\in S_k} |g(\tilde{Z}^n_i(t)) - g(\tilde{Z}^n_i(s))|\right]^2_{\mathcal{F}^n_i}.$$

Also we set $|||\zeta^n_i||| = [\sup_{t\in T}|\zeta^n_i(t)|]_{\mathcal{F}^n_i}$.

Theorem 3 *In the above situation, we suppose the following conditions hold:*
(a$_1$) $\sum_{i=1}^{\sigma^n} 4E^n_{i-1}\zeta^n_i(t)\zeta^n_i(s) \xrightarrow{P^n} C(t,s)$ *[some constant] for every $t, s \in T$;*
(a$_2$) $\left|\left|\sum_{i=1}^{\sigma^n} 4E^n_{i-1}|\zeta^n_i(t)|^2 - C(t,t)\right|\right|_T \xrightarrow{P^{n*}} 0$*;*
(b) $\sum_{i=1}^{\sigma^n} E^n_{i-1}|||\zeta^n_i|||^2 1\{|||\zeta^n_i||| > \varepsilon\} \xrightarrow{P^n} 0$ *for every $\varepsilon > 0$;*
(c) *for some $g : (0,\infty) \to \mathbb{R}$ satisfying* **(g)**, *it holds that for every $\varepsilon \in (0,1)$ there exists a partition $\mathfrak{P}_\varepsilon$ of T with finite size N_ε such that*

$$\sup_{\varepsilon\in(0,1)\cap\mathbb{Q}} \frac{\hat{Z}^n_g(\mathfrak{P}_\varepsilon)}{\varepsilon^2} = O_{P^n}(1) \quad \text{and} \quad \int_0^1 \sqrt{\log N_\varepsilon}\, d\varepsilon < \infty.$$

Then it holds that $\log Z^n_{\sigma^n} \overset{P^n}{\Longrightarrow} X$, where each marginal $(X(t_1), ..., X(t_d))$ has the normal distribution $N(-\frac{1}{2}\mathrm{diag}\Sigma, \Sigma)$ with $\Sigma = \{C(t_i, t_j)\}_{ij}$. Furthermore, the formula (3) defines a pseudo-metric ρ on T such that (T, ρ) is totally bounded and almost all paths of X are uniformly ρ-continuous.

The proof is given in Section 5. Below let us apply the above result to ergodic Markov chains.

Example: Ergodic Markov chains. Let $\{X_i\}_{i\in\mathbb{N}_0}$ be an ergodic Markov chain, defined on a probability space $(\Omega, \mathcal{F}, P)$, with values in an arbitrary state space $(E, \mathcal{E})$. Let $\mu(dx)$ denote the initial distribution, $p(x, dy)$ the transition distribution, and $\pi(dx)$ the invariant distribution. Let us equip the space $\mathbb{H} = L^2(E \times E, \pi(dx)p(x,dy))$ with the inner product given by

$$(h_1, h_2)_{\mathbb{H}} = \int_{E\times E} h_1(x,y)h_2(x,y)\pi(dx)p(x,dy) \quad \forall h_1, h_2 \in \mathbb{H}.$$

Next we define the subset $\mathbb{H}_0$ of $\mathbb{H}$ by

$$\mathbb{H}_0 = \left\{ h \in \mathbb{H}; \quad \int_E h(x,y)p(x,dy) = 0 \quad \forall x \in E \quad \text{and} \quad h > -1 \right\}.$$

Fix a subset $T \subset \mathbb{H}_0$, and for every $n \in \mathbb{N}$ let us consider a family of probability measures $\mathbf{P}^n = \{P^{n,h}; h \in T\}$ on $(\Omega, \mathcal{F})$ such that: under $P^{n,h}$, the process $\{X_i\}_{i \in \mathbb{N}_0}$ is the Markov chain with initial distribution μ and transition distribution $p^{n,h}$ given by

$$p^{n,h}(x,dy) = \left(1 + \frac{h(x,y)}{\sqrt{n}}\right) p(x,dy).$$

Here we set $\mathcal{F}_i = \sigma\{X_0, ..., X_i\}$. Then it holds that

$$Z_i^n(h) = \frac{dP_i^{n,h}}{dP_i} = \prod_{j=1}^{i} \left(1 + \frac{h(X_{j-1}, X_j)}{\sqrt{n}}\right).$$

We need some more notations to state the following result, which concerns the asymptotic behavior of $\log Z_n^n = (\log Z_n^n(h) | h \in T)$. For given $K \in L^2(E, \pi(dx))$ we define the pseudo-metric ρ_K on $\mathbb{H}$ by

$$\rho_K(h_1, h_2) = \sup_{x \in E} \frac{\rho_x(h_1, h_2)}{|K(x)| \vee 1} \quad \forall h_1, h_2 \in \mathbb{H},$$

where

$$\rho_x(h_1, h_2) = \sqrt{\int_E |h_1(x,y) - h_2(x,y)|^2 p(x,dy)} \quad \forall x \in E.$$

For every $\varepsilon > 0$ the ε-bracketing number $N_{[]}(T, \rho_K; \varepsilon)$ is defined as the smallest N such that: there exist N-pairs $l_k, u_k \in \mathbb{H}$ with $\rho_K(l_k, u_k) < \varepsilon$ $(k = 1, ..., N)$ such that for every $h \in T$ it holds $l_k \le h \le u_k$ for some k.

Proposition 4 *Let $(\Omega, \mathcal{F}, \mathbf{F} = \{\mathcal{F}_i\}_{i \in \mathbb{N}_0}, P)$ and $\mathbf{P}^n = \{P^{n,h}; h \in T\}$ as above be given. Suppose that there exists $h^* \in L^4(E \times E, \pi(dx)p(x,dy))$ such that $\sup_{h \in T} |h| \le h^*$, and also that there exists $K \in L^2(E, \pi(dx))$ such that*

$$\int_0^1 \sqrt{\log N_{[]}(T, \rho_K; \varepsilon)} d\varepsilon < \infty.$$

Then it holds that $\log Z_n^n \overset{P}{\Longrightarrow} X$, where each marginal $(X(h_1), ..., X(h_d))$ has the normal distribution $N(-\frac{1}{2}\mathrm{diag}\Sigma, \Sigma)$ with $\Sigma = \{(h_i, h_j)_{\mathbb{H}}\}_{ij}$. Furthermore, almost all paths of X are uniformly ρ-continuous, where ρ is the pseudo-metric on T defined by

$$\rho(h_1, h_2) = \sqrt{(h_1 - h_2, h_1 - h_2)_{\mathbb{H}}} \quad \forall h_1, h_2 \in T.$$

This result is easily derived from the ergodic theorem and Theorem 3 if we set $g(x) = x$ in the condition **(g)**, hence the proof is omitted.

Here we give a statistical application. Fix a subset $T \subset \mathbb{H}_0$ such that $(h,h)_{\mathbb{H}} = 1$ for every $h \in T$, and let us consider the testing problem:

$$\begin{array}{lll} \text{hypothesis} & H_0: & p \\ \text{against} & H_1^n: & p^{n,h} \quad \text{for some } h \in T. \end{array}$$

We propose the test statistics

$$S^n = \frac{1}{2} + \sup_{h \in T} \log Z_n^n(h).$$

Assume the same conditions as in Proposition 4. Then, under the null hypothesis H_0, it holds that $S^n \xrightarrow{P} S = \sup_{h \in T} Y(h)$ where $(Y(h_1), ..., Y(h_d))$ has the normal distribution $N(0, \Sigma)$ with $\Sigma = \{(h_i, h_j)_{\mathbb{H}}\}_{ij}$. This fact follows easily from Proposition 4 and the continuous mapping theorem.

4 Proof of Theorem 1

First of all, we prepare two kinds of inequalities. The first one is the Bernstein-Freedman inequality for martingale difference array. See Freedman (1975) or Section 4.13 of Liptser and Shiryaev (1989) for the proof.

Lemma 1 *Let $\{M_i\}_{i \in \mathbb{N}}$ be an $\mathbb{R}$-valued martingale difference array such that $|M_i| \leq a$ for all $i \in \mathbb{N}$. Let σ be a bounded stopping time. Then for any $b > 0$*

$$P\left(\max_{1 \leq l \leq \sigma} \left|\sum_{i=1}^{l} M_i\right| > \varepsilon,\ \sum_{i=1}^{\sigma} E_{i-1}|M_i|^2 \leq b\right) \leq 2 \exp\left(-\frac{\varepsilon^2}{2[a\varepsilon + b]}\right) \quad \forall \varepsilon > 0.$$

The second inequality, a minor modification of Lemma 2.11.17 of Van der Vaart and Wellner (1996), is used in connection with the preceding one.

Lemma 2 *Let X be a random variable and B be a measurable set such that for some $a, b > 0$*

$$P(|X| > \varepsilon,\ B) \leq 2 \exp\left(-\frac{\varepsilon^2}{2[a\varepsilon + b]}\right) \quad \forall \varepsilon > 0.$$

Then there exists a universal constant D such that for every measurable set A

$$E|X| 1_{A \cap B} \leq D\left(a \log \mu + \sqrt{b}\sqrt{\log \mu}\right)\left(P(A) + \frac{1}{\mu}\right) \quad \forall \mu > e.$$

Our result is based on the following criterion for weak convergence of $\ell^\infty(T)$-valued random elements, in which a *stochastic process* $X = (X(t)|t \in T)$ means a family of $\mathbb{R}$-valued Borel measurable random variables $X(t)$ defined on a common probability space. See Theorem 1.5.4 and 1.5.6 of Van der Vaart and Wellner (1996) for the proof.

Lemma 3 *Let T be a non-empty set. For every $n \in \mathbb{N}$ let $(\Omega^n, \mathcal{F}^n, P^n)$ be a probability space, and X^n be a mapping from Ω^n to $\ell^\infty(T)$. Consider the following statements:*
(i) *X^n converges weakly to a tight, Borel law;*
(ii) *every finite-dimensional marginal of X^n converges weakly to a (tight,) Borel law;*
(iii) *for every $\varepsilon, \eta > 0$ there exists a finite partition $\mathfrak{P} = \{S_k; 1 \leq k \leq N\}$ of T such that*

$$\limsup_{n \to \infty} P^{n*}\left(\max_{1 \leq k \leq N} \sup_{t,s \in S_k} |X^n(t) - X^n(s)| > \varepsilon\right) \leq \eta;$$

Then there is the equivalence (i) $\Longleftrightarrow$ (ii) + (iii). *Furthermore, if the marginals of a stochastic process $X = (X(t)|t \in T)$ have the same laws as the limits in* (ii), *then there exists a version $\tilde{X}$ of X such that $X^n \xrightarrow{P^n} \tilde{X}$.*

Proof of Theorem 1. By the assumption **(C)** we can choose a *nested* sequence of finite partitions $\tilde{\mathfrak{P}}_q = \{\tilde{S}_{qk}; 1 \leq k \leq \tilde{N}_q\}$ $(q \in \mathbb{N})$ of T such that

$$\sup_{q\in\mathbb{N}} \frac{\hat{M}^n(\tilde{\mathfrak{P}}_q)}{2^{-2q}} = O_{P^n}(1) \quad \text{and} \quad \sum_{q=1}^{\infty} 2^{-q}\sqrt{\log \tilde{N}_q} < \infty.$$

To see it, first choose a sequence of finite partitions $\mathfrak{P}_{2^{-q}} = \{S_{2^{-q},k}; 1 \leq k \leq N_{2^{-q}}\}$ $(q \in \mathbb{N})$ of T according to **(C)**, and next define $\tilde{\mathfrak{P}}_q$ as the finite partition generated by $\mathfrak{P}_1, ..., \mathfrak{P}_q$, whose size $\tilde{N}_q$ is at most $\prod_{j=1}^{q} N_{2^{-j}}$. Then the constructed sequence $\tilde{\mathfrak{P}}_q$ $(q \in \mathbb{N})$ satisfies all of the requirements. In fact, for instance, we can easily see

$$\begin{aligned}\sum_{q=1}^{\infty} 2^{-q}\sqrt{\log \tilde{N}_q} &\leq \sum_{q=1}^{\infty} 2^{-q}\sum_{j=1}^{q}\sqrt{\log N_{2^{-j}}} \\ &= \sum_{j=1}^{\infty}\sqrt{\log N_{2^{-j}}}\sum_{q=j}^{\infty} 2^{-q} = 2\sum_{j=1}^{\infty} 2^{-j}\sqrt{\log N_{2^{-j}}} < \infty.\end{aligned}$$

For simplicity, we denote $\tilde{\mathfrak{P}}_q = \{\tilde{S}_{qk}; 1 \leq k \leq \tilde{N}_q\}$ by $\mathfrak{P}_q = \{S_{qk}; 1 \leq k \leq N_q\}$. We can assume $N_q > e$ without loss of generality.

Now, choose an element t_{qk} from each partitioning set S_{qk} and define for every $t \in T$

$$\left\{\begin{array}{rcl}\tau_q t &=& t_{qk} \\ S_q t &=& S_{qk}\end{array}\right. \qquad \text{if } t \in S_{qk}.$$

Also we set $a_q = 2^{-q}/\sqrt{\log N_{q+1}}$.

Due to Lemma 3, it is enough to show that for every $\varepsilon, \eta > 0$ there exists $q_0 \in \mathbb{N}$ such that

$$\limsup_{n\to\infty} P^{n*}\left(\left\|\sum_{i=1}^{\sigma^n} M_i^n(t) - \sum_{i=1}^{\sigma^n} M_i^n(\tau_{q_0}t)\right\|_T > \varepsilon\right) \leq \eta.$$

In order to do it, fix any q_0 for a while, and we define $\mathcal{F}_i^n$-measurable indicator functions

$$\begin{aligned}A_{i,q-1}^n(t) &= 1\{\Delta_i^n(S_{q_0}t) \leq a_{q_0}, ..., \Delta_i^n(S_{q-1}t) \leq a_{q-1}\}, \\ B_{i,q}^n(t) &= 1\{\Delta_i^n(S_{q_0}t) \leq a_{q_0}, ..., \Delta_i^n(S_{q-1}t) \leq a_{q-1}, \Delta_i^n(S_q t) > a_q\}, \\ B_{i,q_0}^n(t) &= 1\{\Delta_i^n(S_{q_0}t) > a_{q_0}\},\end{aligned}$$

where recall the notation (2). Here we decompose

$$\begin{aligned}M_i^n(t) - M_i^n(\tau_{q_0}t) &= (M_i^n(t) - M_i^n(\tau_{q_0}t))\, B_{i,q_0}^n(t) \\ &\quad + \sum_{q=q_0+1}^{\infty} (M_i^n(t) - M_i^n(\tau_q t))\, B_{i,q}^n(t) \\ &\quad + \sum_{q=q_0+1}^{\infty} \Big(M_i^n(\tau_q^n t) - M_i^n(\tau_{q-1}t)\Big)\, A_{i,q-1}^n(t).\end{aligned} \tag{5}$$

Now notice that

$$\begin{aligned}0 &= E_{i-1}^n\,(M_i^n(t) - M_i^n(\tau_{q_0}t))\, B_{i,q_0}^n(t) \\ &\quad + \sum_{q=q_0+1}^{\infty} E_{i-1}^n\,(M_i^n(t) - M_i^n(\tau_q t))\, B_{i,q}^n(t) \\ &\quad + \sum_{q=q_0+1}^{\infty} E_{i-1}^n\,(M_i^n(\tau_q t) - M_i^n(\tau_{q-1}t))\, A_{i,q-1}^n(t)\end{aligned} \tag{6}$$

almost surely,

$$(7) \qquad \left|E^n_{i-1}\left(M^n_i(t) - M^n_i(\tau_q t)\right) B^n_{i,q}(t)\right| \leq E^n_{i-1}\Delta^n_i(S_q t)B^n_{i,q}(t) \leq a_{q-1}$$

almost surely, and that

$$(8) \qquad \left|E^n_{i-1}\left(M^n_i(\tau_q t) - M^n_i(\tau_{q-1} t)\right) A^n_{i,q-1}(t)\right| \leq a_{q-1}$$

almost surely. Here we choose the versions of conditional expectations as follows: first choose a version of $E^n_{i-1}\Delta^n_i(S_q t)B^n_{i,q}(t)$ which is non-negative and not larger than a_{q-1} identically; next, on the exceptional sets of (6), (7) and (8), we define the values of other conditional expectations as zero. Then, the values of $E^n_{i-1}\Delta^n_i(S_q t)B^n_{i,q}(t)$ and $E^n_{i-1}\left(M^n_i(\tau_q t) - M^n_i(\tau_{q-1} t)\right) A^n_{i,q-1}(t)$ depend on t only through $S_{q_0}t, ..., S_q t$ and $\tau_{q-1}t, \tau_q t$, while (6), (7) and (8) hold *identically* for *all* $t \in T$.

From (5), (6) and (7), we obtain

$$\left|\left|\sum_{i=1}^{\sigma^n} M^n_i(t) - \sum_{i=1}^{\sigma^n} M^n_i(\tau_{q_0} t)\right|\right|_T \leq (I_1) + (I_2) + (II_1) + (II_2) + (III)$$

identically, where

$$\begin{aligned}
(I_1) &= \left|\left|\sum_{i=1}^{\sigma^n} \Delta^n_i(S_{q_0}t)B^n_{i,q_0}(t)\right|\right|_T, \\
(I_2) &= \left|\left|\sum_{i=1}^{\sigma^n} E^n_{i-1}\Delta^n_i(S_{q_0}t)B^n_{i,q_0}(t)\right|\right|_T, \\
(II_1) &= \left|\left|\sum_{i=1}^{\sigma^n}\sum_{q=q_0+1}^{\infty} \Delta^n_i(S_q t)B^n_{i,q}(t)\right|\right|_T, \\
(II_2) &= \left|\left|\sum_{i=1}^{\sigma^n}\sum_{q=q_0+1}^{\infty} E^n_{i-1}\Delta^n_i(S_q t)B^n_{i,q}(t)\right|\right|_T, \\
(III) &= \left|\left|\sum_{i=1}^{\sigma^n}\sum_{q=q_0+1}^{\infty} \left(M^n_i(\tau_q t) - M^n_i(\tau_{q-1} t)\right) A^n_{i,q-1}(t)\right.\right. \\
&\qquad \left.\left. - \sum_{i=1}^{\sigma^n}\sum_{q=q_0+1}^{\infty} E^n_{i-1}\left(M^n_i(\tau_q t) - M^n_i(\tau_{q-1} t)\right) A^n_{i,q-1}(t)\right|\right|_T.
\end{aligned}$$

Further, it holds that $(II_1) \leq (II_1') + (II_2)$ identically, where

$$(II_1') = \left|\left|\sum_{i=1}^{\sigma^n}\sum_{q=q_0+1}^{\infty} \Delta^n_i(S_q t)B^n_{i,q}(t) - \sum_{i=1}^{\sigma^n}\sum_{q=q_0+1}^{\infty} E^n_{i-1}\Delta^n_i(S_q t)B^n_{i,q}(t)\right|\right|_T.$$

Hereafter we perform estimations for terms $(I_1), (I_2), (II_1'), (II_2)$ and (III).

Estimation of (I_1) *and* (I_2). It is clear that $\Delta^n_i(S_{q_0}t) \leq 2|||M^n_i|||$ almost surely, hence we have

$$\Delta^n_i(S_{q_0}t)B^n_{i,q_0}(t) \leq 2|||M^n_i|||1\{2|||M^n_i||| > a_{q_0}\}$$

almost surely. So the assumption **(B)** implies that the term (I_2) converges to zero in probability as $n \to \infty$ for any fixed q_0. Also, it follows from the Lenglart inequality (see e.g. Lemma I.3.30 of Jacod and Shiryaev (1987) for the continuous time case) that for any $\varepsilon > 0$

$$\begin{aligned} & P^n\left(\sum_{i=1}^{\sigma^n} 2|||M_i^n|||1\{2|||M_i^n||| > a_{q_0}\} \geq \varepsilon\right) \\ \leq \quad & \varepsilon + P^n\left(\sum_{i=1}^{\sigma^n} E_{i-1}^n 2|||M_i^n|||1\{2|||M_i^n||| > a_{q_0}\} \geq \varepsilon^2\right). \end{aligned}$$

Hence the term (I_1) also converges to zero in probability as $n \to \infty$ for any fixed q_0.

In order to estimate the terms (II_1'), (II_2) and (III), let us notice that: by the construction of $\mathfrak{P}_q$ $(q \in \mathbb{N})$ based on the condition **(C)**, for any $\eta > 0$ there exists a constant $K = K_\eta > 0$, not depending on q_0, such that $\limsup_{n\to\infty} P^n(\Omega^n \setminus \Omega_K^n) \leq \eta$ where

$$\Omega_K^n = \left\{\sup_{q\in\mathbb{N}} \frac{\hat{M}^n(\mathfrak{P}_q)}{2^{-2q}} \leq K\right\}.$$

Estimation of (II_2). We can easily see

$$\begin{aligned} (II_2) \quad \leq \quad & \left\|\sum_{i=1}^{\sigma^n} \sum_{q=q_0+1}^{\infty} \frac{1}{a_q} E_{i-1}^n |\Delta_i^n(S_q t)|^2 B_{i,q}^n(t)\right\|_T \\ \leq \quad & \sup_{q\geq q_0+1} \left\|\sum_{i=1}^{\sigma^n} \frac{E_{i-1}^n |\Delta_i^n(S_q t)|^2 B_{i,q}^n(t)}{2^{-2q}}\right\|_T \cdot \sum_{q=q_0+1}^{\infty} \frac{2^{-2q}}{a_q} \\ \leq \quad & K \sum_{q=q_0+1}^{\infty} 2^{-q}\sqrt{\log N_{q+1}} \quad \text{almost surely on the set } \Omega_K^n. \end{aligned}$$

Estimation of (II_1'). Since

$$|\Delta_i^n(S_q t)B_{i,q}^n(t) - E_{i-1}^n \Delta_i^n(S_q t)B_{i,q}^n(t)| \leq 2a_{q-1} \quad \text{identically}$$

and for every $m \in \mathbb{N}$

$$\sum_{i=1}^{m\wedge\sigma^n} E_{i-1}^n |\Delta_i^n(S_q t)|^2 B_{i,q}^n(t) \leq K 2^{-2q} \quad \text{almost surely on the set } \Omega_K^n,$$

it follows from Lemma 1 that

$$\begin{aligned} & P^n\left(\max_{1\leq l\leq (m\wedge\sigma^n)} \left|\sum_{i=1}^{l} \left(\Delta_i^n(S_q t)B_{i,q}^n(t) - E_{i-1}^n \Delta_i^n(S_q t)B_{i,q}^n(t)\right)\right| > \varepsilon, \Omega_K^n\right) \\ & \qquad \leq 2\exp\left(-\frac{\varepsilon^2}{2[2a_{q-1}\varepsilon + K2^{-2q}]}\right) \quad \forall \varepsilon > 0. \end{aligned}$$

After letting $m \to \infty$, apply Lemma 2. Then we have for every $A \in \mathcal{F}^n$

$$\begin{aligned} & E^n\left|\sum_{i=1}^{\sigma^n} \left(\Delta_i^n(S_q t)B_{i,q}^n(t) - E_{i-1}^n \Delta_i^n(S_q t)B_{i,q}^n(t)\right)\right| 1_{A\cap\Omega_K^n} \\ \leq \quad & D\left(2a_{q-1}\log N_q + \sqrt{K}2^{-q}\sqrt{\log N_q}\right)\left(P^n(A) + \frac{1}{N_q}\right) \\ = \quad & D(2+\sqrt{K})2^{-q}\sqrt{\log N_q}\left(P^n(A) + \frac{1}{N_q}\right). \end{aligned}$$

For every q, let $\{\Omega^n_{qk}; 1 \le k \le N_q\}$ be a partition of Ω^n such that the maximum

$$\left|\left|\sum_{i=1}^{\sigma^n}\left(\Delta^n_i(S_q t)B^n_{i,q}(t) - E^n_{i-1}\Delta^n_i(S_q t)B^n_{i,q}(t)\right)\right|\right|_T$$

is achieved at S_{qk} on the set Ω^n_{qk}. Then we have

$$\begin{aligned}
&E^n\left|\left|\sum_{i=1}^{\sigma^n}\sum_{q=q_0+1}^{\infty}\left(\Delta^n_i(S_q t)B^n_{i,q}(t) - E^n_{i-1}\Delta^n_i(S_q t)B^n_{i,q}(t)\right)\right|\right|_T 1_{\Omega^n_K} \\
&\le \sum_{q=q_0+1}^{\infty} E^n\left|\left|\sum_{i=1}^{\sigma^n}\left(\Delta^n_i(S_q t)B^n_{i,q}(t) - E^n_{i-1}\Delta^n_i(S_q t)B^n_{i,q}(t)\right)\right|\right|_T 1_{\Omega^n_K} \\
&\le \sum_{q=q_0+1}^{\infty}\sum_{k=1}^{N_q} E^n\left|\left|\sum_{i=1}^{\sigma^n}\left(\Delta^n_i(S_q t)B^n_{i,q}(t) - E^n_{i-1}\Delta^n_i(S_q t)B^n_{i,q}(t)\right)\right|\right|_T 1_{\Omega^n_{qk}\cap\Omega^n_K} \\
&\le D(2+\sqrt{K})\sum_{q=q_0+1}^{\infty}\sum_{k=1}^{N_q} 2^{-q}\sqrt{\log N_q}\left(P^n(\Omega^n_{qk}) + \frac{1}{N_q}\right) \\
&\le D(4+2\sqrt{K})\sum_{q=q_0+1}^{\infty} 2^{-q}\sqrt{\log N_q}.
\end{aligned}$$

Estimation of (III). Exactly the same calculation as (II_1') yields that

$$E^n(III)1_{\Omega^n_K} \le D(4+4\sqrt{K})\sum_{q=q_0+1}^{\infty} 2^{-q}\sqrt{\log N_q}.$$

Those estimations imply the assertion: choose large K, next choose large q_0, and then, take $\limsup_{n\to\infty}$. □

5 Proof of Theorem 3

For every $\varepsilon \in (0,1)$ we set

$$\begin{aligned}
\lambda^{n,\varepsilon}_i(t) &= \log \tilde{Z}^n_i(t)1\{|||\zeta^n_i||| \le \varepsilon\}, \\
B^{n,\varepsilon}(t) &= \sum_{i=1}^{\sigma^n} E^n_{i-1}\lambda^{n,\varepsilon}_i(t), \\
C^{n,\varepsilon}(t,s) &= \sum_{i=1}^{\sigma^n} E^n_{i-1}\lambda^{n,\varepsilon}_i(t)\lambda^{n,\varepsilon}_i(s).
\end{aligned}$$

Here we perform some computations.

Lemma 4 *Suppose* $(\mathbf{a}_1)$, $(\mathbf{a}_2)$ *and* $(\mathbf{b})$. *For given* $a \in (0,1)$ *it holds that:*
(i) *if* $\sup_{t\in T} C(t,t) < \infty$ *then* $||B^{n,a}(t) + \frac{1}{2}C(t,t)||_T \xrightarrow{P^{n*}} 0$;
(ii) $C^{n,a}(t,s) \xrightarrow{P^n} C(t,s) \quad \forall t,s \in T$;
(iii) $\sum_{i=1}^{\sigma^n} |E^n_{i-1}\lambda^{n,a}_i(t)|^2 \xrightarrow{P^n} 0 \quad \forall t \in T$.

Remark. In the proof of Theorem 3, we will see that $(\mathbf{a}_1)$, **(b)** and **(c)** imply $\sup_{t\in T} C(t,t) < \infty$.

Proof. [STEP 1] First we prove the following facts: for given $a \in (0,1)$ there exist constants $K_1, K_2, K_3 > 0$ such that for every $\varepsilon \in (0,a]$

$$\left\|B^{n,\varepsilon}(t) + \frac{1}{2}C(t,t)\right\|_T \le \varepsilon K_1 + o_{P^{n*}}(1), \tag{9}$$

$$|C^{n,\varepsilon}(t,s) - C(t,s)| \le \varepsilon K_2 + o_{P^n}(1) \quad \forall t,s \in T, \tag{10}$$

$$\sum_{i=1}^{\sigma^n} |E_{i-1}^n \lambda_i^{n,\varepsilon}(t)| \le \frac{1}{2}C(t,t) + \varepsilon K_3 + o_{P^n}(1) \quad \forall t \in T. \tag{11}$$

In order to show (9), first notice that there exists a constant $K > 0$ such that $|\log x - (x-1) + 2(\sqrt{x}-1)^2| \le K|\sqrt{x}-1|^3$ whenever $|\sqrt{x}-1| \le a$. Hence, for fixed $\varepsilon \in (0,a]$ we obtain

$$\begin{aligned}
&\left|B^{n,\varepsilon}(t) + 2\sum_{i=1}^{\sigma^n} E_{i-1}^n \left|\zeta_i^n(t)\right|^2 1\left\{|||\zeta_i^n||| \le \varepsilon\right\}\right| \\
&\le \quad \varepsilon K \sum_{i=1}^{\sigma^n} E_{i-1}^n \left|\zeta_i^n(t)\right|^2 1\left\{|||\zeta_i^n||| \le \varepsilon\right\} \\
&\qquad + \left|\sum_{i=1}^{\sigma^n} E_{i-1}^n (\tilde{Z}_i^n(t) - 1) 1\left\{|||\zeta_i^n||| \le \varepsilon\right\}\right|
\end{aligned}$$

almost surely. Since $E_{i-1}^n(\tilde{Z}_i^n(t)-1) = 0$ almost surely, the last term of the right hand side is not larger than

$$\begin{aligned}
&\left|\sum_{i=1}^{\sigma^n} E_{i-1}^n (\tilde{Z}_i^n(t) - 1) 1\left\{|||\zeta_i^n||| > \varepsilon\right\}\right| \\
&\le \quad \sum_{i=1}^{\sigma^n} E_{i-1}^n |\tilde{Z}_i^n(t) - 1| 1\left\{|||\zeta_i^n||| > \varepsilon\right\} \\
&\le \quad \frac{\varepsilon+2}{\varepsilon} \sum_{i=1}^{\sigma^n} E_{i-1}^n \left|\zeta_i^n(t)\right|^2 1\left\{|||\zeta_i^n||| > \varepsilon\right\}
\end{aligned}$$

almost surely. Thus we obtain

$$\begin{aligned}
&\left|B^{n,\varepsilon}(t) + \frac{1}{2}C(t,t)\right| \\
&\le \quad (2+\varepsilon K)\left|\sum_{i=1}^{\sigma^n} E_{i-1}^n \left|\zeta_i^n(t)\right|^2 - \frac{1}{4}C(t,t)\right| + \varepsilon K \cdot \frac{1}{4}C(t,t) \\
&\qquad + \left(2 + \frac{\varepsilon+2}{\varepsilon}\right) \sum_{i=1}^{\sigma^n} E_{i-1}^n |||\zeta_i^n|||^2 1\left\{|||\zeta_i^n||| > \varepsilon\right\}
\end{aligned} \tag{12}$$

almost surely. In order to get the estimate for all $\omega \in \Omega^n$, we can choose the versions of conditional expectations as follows: first choose a version of $E_{i-1}^n |||\zeta_i^n|||^2 1\{|||\zeta_i^n||| > \varepsilon\}$ which is non-negative identically; next, on the union of all exceptional sets for the estimates appeared above, we define the values of all other conditional expectations as zero. Then, the inequality (12) holds *identically* for *all* $t \in T$. By taking the supremum of (12) with respect to $t \in T$, and letting $n \to \infty$, we obtain the assertion (9).

A similar argument yields (11). In fact, it is much easier than (9), because the assertion of (11) is t-wise, for which we do not need any argument about versions of conditional expectations. Also, it is easy to show (10) if we notice the following fact: for given $a \in (0,1)$ there exists a constant $K > 0$ such that $|\log x \cdot \log y + 4(\sqrt{x}-1)(\sqrt{y}-1)| \le K \max\{|\sqrt{x}-1|^3, |\sqrt{y}-1|^3\}$ whenever $\max\{|\sqrt{x}-1|, |\sqrt{y}-1|\} \le a$.

[STEP 2] Next we prove the following facts:

$$||B^{n,a}(t) - B^{n,\varepsilon}(t)||_T \xrightarrow{P^{n*}} 0 \quad \forall \varepsilon \in (0,a); \tag{13}$$

$$|C^{n,a}(t,s) - C^{n,\varepsilon}(t,s)| \xrightarrow{P^n} 0 \quad \forall t,s \in T \ \forall \varepsilon \in (0,a). \tag{14}$$

In order to show (13), notice that for given $a \in (0,1)$ there exists a constant $K > 0$ such that $|\log x| \le K|\sqrt{x}-1|^2$ whenever $|\sqrt{x}-1| \le a$. For every $\varepsilon \in (0,a)$ it holds that

$$\begin{aligned} & \left| \sum_{i=1}^{\sigma^n} E^n_{i-1} \lambda^{n,a}_i(t) - \sum_{i=1}^{\sigma^n} E^n_{i-1} \lambda^{n,\varepsilon}_i(t) \right| \\ = & \left| \sum_{i=1}^{\sigma^n} E^n_{i-1} \log \tilde{Z}^n_i(t) 1\{\varepsilon < |||\zeta^n_i||| \le a\} \right| \\ \le & \ K \sum_{i=1}^{\sigma^n} E^n_{i-1} |\zeta^n_i(t)|^2 1\{\varepsilon < |||\zeta^n_i||| \le a\} \\ \le & \ K \sum_{i=1}^{\sigma^n} E^n_{i-1} |||\zeta^n_i|||^2 1\{|||\zeta^n_i||| > \varepsilon\}, \end{aligned} \tag{15}$$

almost surely. We can choose some versions of conditional expectations such that the estimate above holds identically for all $t \in T$, by the same way as in the proof of (9). Take the supremum of (15) with respect to $t \in T$, and let $n \to \infty$, then we get (13). A similar computation yields (14).

[STEP 3] Now it is easy to see that (9) and (13) yield the assertion (i), and that (10) and (14) yield the assertion (ii); first choose $\varepsilon > 0$ small enough, and then, let $n \to \infty$. In order to show the assertion (iii), notice that for any $\varepsilon \in (0,a)$

$$\begin{aligned} & \left| \sum_{i=1}^{\sigma^n} |E^n_{i-1} \lambda^{n,a}_i(t)|^2 - \sum_{i=1}^{\sigma^n} |E^n_{i-1} \lambda^{n,\varepsilon}_i(t)|^2 \right| \\ = & \left| \sum_{i=1}^{\sigma^n} E^n_{i-1}(\lambda^{n,a}_i(t) + \lambda^{n,\varepsilon}_i(t)) E^n_{i-1}(\lambda^{n,a}_i(t) - \lambda^{n,\varepsilon}_i(t)) \right| \\ \le & \ 2|\log(1-a^2)| \sum_{i=1}^{\sigma^n} \left| E^n_{i-1} \log \tilde{Z}^n_i(t) 1\{\varepsilon < |||\zeta^n_i||| \le a\} \right| \ = \ o_{P^n}(1), \end{aligned}$$

hence

$$\begin{aligned} \sum_{i=1}^{\sigma^n} |E^n_{i-1} \lambda^{n,a}_i(t)|^2 &= \sum_{i=1}^{\sigma^n} |E^n_{i-1} \lambda^{n,\varepsilon}_i(t)|^2 + o_{P^n}(1) \\ &\le |\log(1-\varepsilon^2)| \sum_{i=1}^{\sigma^n} |E^n_{i-1} \lambda^{n,\varepsilon}_i(t)| + o_{P^n}(1). \end{aligned}$$

Therefore we obtain (iii) by virtue of (11); first choose $\varepsilon > 0$ small enough, and then let $n \to \infty$. $\square$

Proof of Theorem 3. Let $\delta \in (0,1)$ be the constant appeared in the condition **(g)**. Set $a = \sqrt{1+\delta} - 1$, then $\{x; |\sqrt{x} - 1| \le a\} \subset \{x; |x-1| \le \delta\}$. Since

$$\log Z^n_{\sigma^n}(t) = \sum_{i=1}^{\sigma^n} \log \tilde{Z}^n_i(t),$$

we can approximate $\log Z^n_{\sigma^n} = (\log Z^n_{\sigma^n}(t)|t \in T)$ by $\Lambda^n = (\Lambda^n(t)|t \in T)$ which is defined by

$$\Lambda^n(t) = \sum_{i=1}^{\sigma^n} \lambda^{n,a}_i(t) = \sum_{i=1}^{\sigma^n} \log \tilde{Z}^n_i(t) 1\{|||\zeta^n_i||| \le a\}.$$

In fact, it holds that

$$\begin{aligned} \sup_{t \in T} 1\{\log Z^n_{\sigma^n}(t) \neq \Lambda^n(t)\} &\le \sum_{i=1}^{\sigma^n} 1\{|||\zeta^n_i||| > a\} \\ &\le \frac{1}{a^2} \sum_{i=1}^{\sigma^n} |||\zeta^n_i|||^2 1\{|||\zeta^n_i||| > a\}, \end{aligned}$$

hence using also the Lenglart inequality we obtain $|| \log Z^n_{\sigma^n}(t) - \Lambda^n(t)||_T \xrightarrow{P^{n*}} 0$.

Next we consider the decomposition

$$\Lambda^n = B^{n,a} + \sum_{i=1}^{\sigma^n} M^n_i$$

where $\{M^n_i\}_{i \in \mathbb{N}_0} = \{(M^n_i(t)|t \in T)\}_{i \in \mathbb{N}_0}$ is the martingale difference array given by $M^n_i(t) = \lambda^{n,a}_i(t) - E^n_{i-1}\lambda^{n,a}_i(t)$. First we show the weak convergence of $\sum_{i=1}^{\sigma^n} M^n_i$, to which we apply Corollary 2. By virtue of (ii) and (iii) of Lemma 4 we can easily see that the condition **(A′)** is satisfied. It is also easy to see that the assumption **(b)** implies the Lindeberg condition **(B′)**. Finally, in view of the condition **(g)** and the relation between a and δ, we have for any subset $S \subset T$

$$\begin{aligned} & E^n_{i-1}\left[\sup_{t,s \in S} |\lambda^{n,a}_i(t) - \lambda^{n,a}_i(s)|\right]^2_{\mathcal{F}^n_i} \\ &= E^n_{i-1}\left[\sup_{t,s \in S} \left|\log \tilde{Z}^n_i(t) - \log \tilde{Z}^n_i(t)\right|\right]^2_{\mathcal{F}^n_i} 1\{|||\zeta^n_i||| \le a\} \\ &\le E^n_{i-1}\left[\sup_{t,s \in S} \left|g(\tilde{Z}^n_i(t)) - g(\tilde{Z}^n_i(s))\right|\right]^2_{\mathcal{F}^n_i}. \end{aligned}$$

Thus the assumption **(c)** implies the condition **(C)**. Therefore Corollary 2 yields $\sum_i^n M^n_i \stackrel{P^n}{\Longrightarrow} Y$, where $(Y(t_1), ..., Y(t_d))$ has the normal distribution $N(0, \Sigma)$ with $\Sigma = \{C(t_i, t_j)\}_{ij}$.

It remains to show that $||B^{n,a}(t) + \frac{1}{2}C(t,t)||_T \xrightarrow{P^{n*}} 0$ and that $t \rightsquigarrow C(t,t)$ is uniformly ρ-continuous. To do it, observe that

$$\begin{aligned} \sqrt{C(t,t)} &= \sqrt{E|Y(t)|^2} \\ &\le \sqrt{E|Y(t) - Y(s)|^2} + \sqrt{E|Y(s)|^2} = \rho(t,s) + \sqrt{C(s,s)}. \end{aligned}$$

The above inequality and the totally boundedness of (T,ρ), a consequence of Corollary 2, imply that $\sup_{t\in T} C(t,t) < \infty$. Hence the first assertion follows from (i) of Lemma 4. Also, it is trivial from the above inequality that $t \rightsquigarrow \sqrt{C(t,t)}$ is uniformly ρ-continuous, thus $t \rightsquigarrow C(t,t)$ is so. $\square$

Acknowledgements. I thank Prof. R.D. Gill for some comments.

References

[1] Andersen, N.T., Giné, E., Ossiander, M. and Zinn, J. (1988). The central limit theorem and the law of iterated logarithm for empirical processes under local conditions. *Probab. Theory Relat. Fields* **77** 271-305.

[2] Arcones, M.A. and Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.* **7** 47-71.

[3] Bae, J. and Levental, S. (1995). Uniform CLT for Markov chains and its invariance principle: a martingale approach. *J. Theoret. Probab.* **8** 549-570.

[4] Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. Henri Poincaré - Probab. Statist.* **31** 393-427.

[5] Dudley, R.M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899-929. Correction, **7** 909-911.

[6] Freedman, D.A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100-118.

[7] Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929-989.

[8] Ibragimov, I.A. and Has'minskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

[9] Inagaki, N. and Ogata, Y. (1975). The weak convergence of likelihood ratio random fields and its applications. *Ann. Inst. Statist. Math.* **27** 391-419.

[10] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes.* Springer, Berlin.

[11] Jain, N. and Marcus, M. (1975). Central limit theorems for $C(S)$-valued random variables. *J. Funct. Anal.* **19** 216-231.

[12] Kutoyants, Yu.A. (1984). *Parameter Estimation for Stochastic Processes.* Heldermann Verlag, Berlin.

[13] Le Cam, L. (1960). Locally asymptotically normal families of distributions. *Univ. of Calif. Publ. Stat.* **3** 27-98.

[14] Le Cam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statist.* **41** 802-828.

[15] Levental, S. (1989). A uniform CLT for uniformly bounded families of martingale differences. *J. Theoret. Probab.* **2** 271-287.

[16] Liptser, R.S. and Shiryayev A.N. (1989). *Theory of Martingales.* Kluwer Academic Publishers, Dordrecht.

[17] Nishiyama, Y. (1996). Some central limit theorems for ℓ^∞-valued semimartingales and their applications. In: *Stochastic Calculus and Statistics* 103-141. Inst. Statist. Math., Tokyo.

[18] Ogata, Y. and Inagaki, N. (1977). The weak convergence of the likelihood ratio random fields for Markov observations. *Ann. Inst. Statist. Math.* **29** 165-187.

[19] Ossiander, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897-919.

[20] Pollard, D. (1982). A central limit theorem for empirical processes. *J. Australian Math. Soc. Ser. A* **33** 235-248.

[21] Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer, New York.

[22] Vostrikova, L. (1987). On the weak convergence of likelihood ratio processes of general statistical parametric models. *Stochastics* **23** 277-298.

The Institute of Statistical Mathematics
4-6-7 Minami-Azabu
Minato-ku
Tokyo 106
Japan

Department of Mathematics
Utrecht University
Budapestlaan 6
3584 CD Utrecht
The Netherlands