Comm. Korean Math. Soc. **14** (1999), No. 3, pp. 581–595

THE UNIFORM CLT FOR MARTINGALE DIFFERENCE OF FUNCTION-INDEXED PROCESS UNDER UNIFORMLY INTEGRABLE ENTROPY

JONGSIG BAE AND MOON JOO CHOI

ABSTRACT. In the present paper we provide a short proof of the uniform CLT for the function-indexed martingale difference process under the uniformly integrable entropy by establishing a maximal inequality.

1. Introduction

In the recent theory of weak convergence and empirical process, authors often attempt to relax the structure of independent and identically distributed (IID) random variables or, in statistical terms, a random sample from a fixed population. One way of relaxing the structure is to remove the assumption of identical distribution by considering a sequence of independent random variables or even a triangular array of random variables that is pioneered by the Lindeberg and Feller CLT. For a recent result on a triangular array of rowwise independent but not necessarily IID, see [17] among others. Another way of relaxing the structure is to remove the assumption of independence by considering problems of martingale differences, of Markov chains, or of various types of mixing. See for example, [9], [3] for the problem of stationary Markov chains, [11], [1] for the problem of stationary martingale differences, and [4] for the problem under β-mixing condition. The other way of relaxing the IID structure based on

Received November 4, 1998. Revised April 23, 1999.

1991 Mathematics Subject Classification: Primary 60F17, 60G42; Secondary 62G20.

Key words and phrases: UCLT; martingale differences; function-indexed process; uniformly integrable entropy; Freedman inequality; sub-Gaussian inequality; eventual uniform equicontinuity; sequential empirical process.

This work was supported by Faculty Research Fund, Sungkyunkwan University, 1997.

the complete data of random sampling is to consider the so-called random censoring model. See [10], [15], and [2].

In the study of a uniform central limit theorem (UCLT), two main assumptions on the size of the index set that is usually up to generality of a metric space or a function space are turned out to be suitable. The first one is bracketing entropy which is applied by [12] and [1] among others. The second one is uniformly integrable entropy that is used, for example, in the recent paper of [17]. There are trade off relations between two assumptions. The former does not require an extra assumption on the envelope of the indexed class but there is a restriction of including so-called VC graph class, a wide class of function spaces that have many applications. The later includes VC graph classes, however, needs a moment condition on the envelope. See [16] and [17].

In this paper we deal with the UCLT for a function-indexed process constructed from a sequence of martingale differences under the assumption of uniformly integrable entropy, relaxing both the assumptions on the independence and the identical distribution of the IID structure. The main method of getting the result is to study a maximal inequality for martingale differences based on the Freedman inequality as [17] did in the problem of independent process based on the sub-Gaussian inequality for Rademacher averages.

The result partially improves that of [17] in the sense that we are dealing with the dependent process. The result improves that of [1] in the sense that we are dealing with a martingale difference that is not necessarily stationary and the proof is shorter.

In section 2, we introduce some preliminaries on a setup of martingale differences, function-indexed stochastic process and the size of the index class of functions. In section 3, we state the main result of the UCLT for martingale differences. In section 4, we complete the proof of an eventual uniform equicontinuity result that gives the UCLT by using a truncation argument and a stopping time argument. Finally, in section 5, we consider a weak convergence result for a sequential empirical process for martingale differences. See section 2.12 of [16] for an analogous result on IID setting.

2. Preliminaries

We begin with a given probability space $(\Omega, \mathcal{E}, P)$. Let $\mathcal{E}_0 \subseteq \mathcal{E}_1 \subseteq \mathcal{E}_2 \cdots$ be an increasing sequence of sub-σ-fields of $\mathcal{E}$. Let (T, ρ) be a pseudometric space. Consider a martingale difference process $\{d_j : 1 \leq j \leq n,\ n \in \mathbf{N}\}$ indexed by T with respect to the increasing σ-fields $\{\mathcal{E}_j : 0 \leq j \leq n,\ n \in \mathbf{N}\}$. By that $\{d_j : 1 \leq j \leq n,\ n \in \mathbf{N}\}$ is a martingale difference process indexed by T we mean, for each $t \in T$, $\{d_j(t) : 1 \leq j \leq n,\ n \in \mathbf{N}\}$[1] is a sequence of random variables satisfying

$$(a)\quad E(d_j(t)|\mathcal{E}_{j-1}) = 0 \text{ for } t \in T;$$
$$(b)\quad d_j(t) \text{ is } \mathcal{E}_j\text{-measurable for } t \in T.$$

We simply denote $E_{j-1}d$ to mean $E(d|\mathcal{E}_{j-1})$, the conditional expectation of the random element d given the σ-field $\mathcal{E}_{j-1}$. Define the conditional variance process

$$v_j(t) = E_{j-1}d_j^2(t) \text{ for } t \in T.$$

Notice that, for each $t \in T$, $v_j(t)$ is an $\mathcal{E}_{j-1}$-measurable random variable.

In the context of the one dimensional central limit theorem, the partial sum of a martingale difference sequence is known to be asymptotically normal. The following CLT for martingale differences appears in the literature. See, for example, chapter 8 of [13].

PROPOSITION 1. *Let $t \in T$ be fixed and let $\{d_j(t) : 1 \leq j \leq n,\ n \in \mathbf{N}\}$ be a sequence of martingale differences. If as $n \to \infty$,*

$$(1)\qquad \frac{1}{n}\sum_{j=1}^{n} v_j(t) \to^P \sigma^2(t), \text{ where } \sigma^2(t) \text{ is a positive constant;}$$

$$(2)\qquad \frac{1}{n}\sum_{j=1}^{n} E_{j-1}(d_j^2(t)\{|d_j(t)| > \epsilon\}) \to^P 0 \text{ for every } \epsilon > 0;$$

then $\frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_j(t)$ converges in distribution to $N(0, \sigma^2(t))$.

Throughout the paper events are identified with their indicator functions when there is no risk of ambiguity. So, for example, the expression $\{|d_j(t)| > \epsilon\}$ in the summand of (2) means the indicator functions of the events $\{|d_j(t)| > \epsilon\}$.

[1]Throughout this paper we use the notation $\{a_j : 1 \leq j \leq n,\ n \in \mathbf{N}\}$ to denote a sequence $\{a_n\}_{n=1}^{\infty}$.

The goal of this paper is to establish the UCLT for the process $\{S_n(t) : t \in T\}$ defined by

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} d_j(t) \text{ for } t \in T, \tag{3}$$

where $\{d_j : 1 \leq j \leq n,\ n \in \mathbf{N}\}$ is a martingale difference process indexed by T with respect to the increasing σ-fields $\{\mathcal{E}_j : 0 \leq j \leq n, n \in \mathbf{N}\}$ on $(\Omega, \mathcal{E}, P)$.

Establishing a UCLT essentially means showing that $\mathcal{L}(S_n(t) : t \in T) \to \mathcal{L}(Z(t) : t \in T)$, where the processes involved here are indexed by T and are considered as random elements of the Banach space

$$B(T) := \{z : T \to R : ||z||_T := \sup_{t \in T} |z(t)| < \infty\},$$

the space of the bounded real-valued functions on T, taken with the sup norm. The limiting process $Z = (Z(t) : t \in T)$ is a Gaussian process whose sample paths are contained in

$$U_B(T, \rho) := \{z \in B(T) : z \text{ is uniformly continuous with respect to } \rho\ \}.$$

Notice that $(B(T), || \cdot ||_T)$ is a Banach space and $U_B(T, \rho)$ is a closed subspace of $(B(T), || \cdot ||_T)$ and hence is a Banach space. In particular $U_B(T, \rho)$ is separable if and only if (T, ρ) is totally bounded.

Until recent years, it is common to introduce a class of functions as an index set of a stochastic process in the theory of weak convergence and empirical process. See the recent text of [16] among others. In order to introduce this concept, let $(X, \mathcal{X})$ be a measurable space. Let $(\mathcal{F}, || \cdot ||)$ be a subset of a normed vector space of real functions from X to R. Here, of course, $|| \cdot ||$ is the norm inherited by the normed vector space R^X, the space of all functions from X to R. In this case by that $\{d_j : 1 \leq j \leq n,\ n \in \mathbf{N}\}$ is a martingale difference process indexed by $\mathcal{F}$ with respect to the increasing σ-fields $\{\mathcal{E}_j : 0 \leq j \leq n,\ n \in \mathbf{N}\}$ we mean, for each $f \in \mathcal{F}$, $\{f(d_j) : 1 \leq j \leq n,\ n \in \mathbf{N}\}$ is a sequence of random variables satisfying

$$(a)\quad E_{j-1} f(d_j) = 0 \text{ for } f \in \mathcal{F};$$
$$(b)\quad f(d_j) \text{ is } \mathcal{E}_j\text{-measurable for } f \in \mathcal{F}.$$

Define, similarly as before, the conditional variance process

$$v_j(f) = E_{j-1} f^2(d_j) \text{ for } f \in \mathcal{F}.$$

In measuring the size of the class $\mathcal{F}$ we are going to use the concepts of covering numbers and packing numbers. See [12] and [1] where the assumption of bracketing entropy is used. We introduce the definitions in [16]. See also [5].

DEFINITION 1. (Covering number) The covering number $N(\epsilon, \mathcal{F}, ||\cdot||)$ is the minimum number of balls $\{g : ||g - h|| < \epsilon\}$ of radius ϵ needed to cover $\mathcal{F}$.

DEFINITION 2. (Packing number) Call a collection of points ϵ-separated if the distance between each pair of points is strictly larger than ϵ. The packing number $D(\epsilon, \mathcal{F}, ||\cdot||)$ is the maximum number of ϵ-separated points in $\mathcal{F}$.

The concepts of covering number and packing number are equivalent in the sense that $N(2\epsilon, \mathcal{F}, ||\cdot||) \leq D(2\epsilon, \mathcal{F}, ||\cdot||) \leq N(\epsilon, \mathcal{F}, ||\cdot||)$. By definition, the space $(\mathcal{F}, ||\cdot||)$ is totally bounded if and only if the covering numbers and/or the packing numbers are finite for every $\epsilon > 0$.

Finally we introduce the concept of uniformly integrable entropy. Let F be an envelope of $\mathcal{F}$. That is, F is a measurable function from X to $[0, \infty)$ such that $\sup_{f \in \mathcal{F}} |f(x)| \leq F(x)$ for all $x \in X$. Let $M(X, F)$ be the set of all measures γ on $(X, \mathcal{X})$ with $\gamma(F^2) := \int_X F^2 d\gamma < \infty$.

DEFINITION 3. (Uniformly integrable entropy) Say that $\mathcal{F}$ has *uniformly integrable entropy* with respect to L_2-norm if

$$\int_0^\infty \sup_{\gamma \in M(X,F)} [\ln N(\epsilon[\gamma(F^2)]^{1/2}, \mathcal{F}, d_\gamma^{(2)})]^{1/2} d\epsilon < \infty,$$

where $d_\gamma^{(2)}(f, g) := [\int_X (f - g)^2 d\gamma]^{1/2}$.

When the class $\mathcal{F}$ has uniformly integrable entropy, $(\mathcal{F}, d_\gamma^{(2)})$ is totally bounded for any measure γ. Many important classes of functions, such as VC graph classes, have uniformly integrable entropy. See section 2.6 of [16].

3. The Uniform CLT for Martingale Differences

Consider a class $\mathcal{F}$ of measurable functions defined on a measurable space $(X, \mathcal{X})$ with envelope F. Given a martingale difference process $\{d_j : 1 \leq j \leq n, \ n \in \mathbf{N}\}$ indexed by $\mathcal{F}$ with respect to an increasing

σ-fields $\{\mathcal{E}_j : 0 \le j \le n,\ n \in \mathrm{N}\}$ in $(X, \mathcal{X})$, we consider the process $\{S_n(f) : f \in \mathcal{F}\}$ defined by

$$S_n(f) := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} f(d_j), \text{ for } f \in \mathcal{F}. \tag{4}$$

Suppose that $\mathcal{F}$ has uniformly integrable entropy with respect to L_2-norm:

$$\int_0^\infty \sup_{\gamma \in M(X,F)} [\ln N(\epsilon[\gamma(F^2)]^{1/2}, \mathcal{F}, d_\gamma^{(2)})]^{1/2} d\epsilon < \infty,$$

where $M(X, F)$ is the set of all measures γ on $(X, \mathcal{X})$ with $\gamma(F^2) := \int_X F^2 d\gamma < \infty$ and $d_\gamma^{(2)}(f,g) := [\int_X (f-g)^2 d\gamma]^{1/2}$. Write

$$d(f,g) := d_P^{(2)}(f,g) := [E(f-g)^2]^{1/2}. \tag{5}$$

We equip the space $\mathcal{F}$ with the pseudometric d so that $(\mathcal{F}, d)$ is totally bounded.

Let

$$\sigma_n^2(f,g) := \frac{1}{n} \sum_{j=1}^{n} E_{j-1}(f(d_j) - g(d_j))^2 \text{ for } f, g \in \mathcal{F}.$$

We use the following definition of weak convergence which is originally due to [8]. Recall that $B(\mathcal{F})$ is the space of the bounded real-valued functions on $\mathcal{F}$.

DEFINITION 4. A sequence of $B(\mathcal{F})$-valued random functions $\{Y_n : n \ge 1\}$ converges in law to a $B(\mathcal{F})$-valued Borel measurable random function Y whose law concentrates on a separable subset of $B(\mathcal{F})$, denoted $Y_n \Rightarrow Y$, if

$$Eg(Y) = \lim_{n\to\infty} E^* g(Y_n), \forall g \in C(B(\mathcal{F}), ||\cdot||_{\mathcal{F}}),$$

where $C(B(\mathcal{F}), ||\cdot||_{\mathcal{F}})$ is the set of all bounded, continuous functions from $(B(\mathcal{F}), ||\cdot||_{\mathcal{F}})$ into R. Here E^* denotes upper expectation with respect to the outer probability P^*.

We are ready to state the UCLT for martingale differences.

THEOREM 1. *Let $\{d_j : 1 \le j \le n,\ n \in \mathrm{N}\}$ be a martingale difference process indexed by a class $\mathcal{F}$ of measurable functions with envelop F on* *a measurable space* $(X, \mathcal{X})$. *Suppose that* $\mathcal{F}$ *has uniformly integrable entropy with* $EF^2 < \infty$. *Assume that there exists a constant* D *such that*

$$P^*\left(\sup_{f,g\in\mathcal{F}} \frac{\sigma_n^2(f,g)}{d^2(f,g)} \geq D\right) \to 0, \text{ as } n \to \infty. \tag{6}$$

Suppose that, as $n \to \infty$,

$$\frac{1}{n}\sum_{j=1}^{n} E_{j-1}F^2(d_j) \to^P \sigma^2(F), \tag{7}$$

where $\sigma^2(F)$ *is a positive constant; and,*

$$\frac{1}{n}\sum_{j=1}^{n} E_{j-1}(F^2(d_j)\{F(d_j) > \epsilon\sqrt{n}\}) \to^P 0, \quad \text{for every } \epsilon > 0. \tag{8}$$

Suppose there exists a Gaussian process Z *such that the finite dimensional distributions of* S_n *converge to those of* Z. *Then*

$$S_n \Rightarrow Z \text{ as random elements of } B(\mathcal{F}).$$

The limiting process $Z = \{Z(f) : f \in \mathcal{F}\}$ *is mean zero Gaussian with covariance structure* $EZ(f)Z(g)$ *and the sample paths of* Z *are bounded and uniformly continuous with respect to the metric* d.

REMARK.

1. We need the moment condition on the envelope, $EF^2 < \infty$, for the class $\mathcal{F}$ of functions when we use uniformly integrable entropy. On the other hand, one can derive this condition from the assumption of integrability of $\mathcal{L}_2$-bracketing entropy. See, for example, [1].
2. A sufficient condition to the Lipschitz condition (6) is that

$$E^* \sup_{f,g\in\mathcal{F}} \sum_{j=1}^{n} \frac{E_{j-1}(f(d_j) - g(d_j))^2}{nd^2(f,g)} \text{ converges.}$$

3. The condition (7) on the conditional variances and the Lindeberg condition (8) are essential in the sense that the parallel conditions (1) and (2) are required in Proposition 1.

PROOF. The result is a consequence of Theorem 2 and the assumption on the finite dimensional distribution convergence of S_n to those of Z by applying Theorem 10. 2 of [14] to the process $\{S_n(f) : f \in \mathcal{F}\}$ indexed by the totally bounded pseudometric space $(\mathcal{F}, d)$. □

Our program in establishing the UCLT for the martingale difference process is to prove the following eventual uniform equicontinuity of the process $\{S_n(f) : f \in \mathcal{F}\}$ by the results on the Freedman inequality and a chaining argument, a truncation argument and a stopping time argument.

THEOREM 2. *Under the assumptions* (6), (7), *and* (8) *of* Theorem 1, *we have the eventual uniform equicontinuity of the process* $\{S_n(f) : f \in \mathcal{F}\}$. *That is, for every* $\epsilon > 0$ *there exists* $\eta > 0$ *such that*

$$\limsup_{n\to\infty} P^* \left(\sup_{d(f,g)<\eta} |\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (f(d_j) - g(d_j))| > \epsilon \right) < \epsilon.$$

In the following section we are going to make our efforts to complete the proof of Theorem 2.

4. Proof of the Theorem 2

We begin this section with a modified version of Freedman inequality and a maximal inequality for the process that can be applied when the process is a sub-Gaussian. Notice that for any bounded random variable ξ on Ω, we use the notation $||\xi||_\infty$ to denote $\sup_{\omega\in\Omega} |\xi(\omega)|$.

LEMMA 1. *Let* $(d_j)_{1\le j\le n}$ *be a martingale difference with respect to increasing* σ*-fields* $(\mathcal{E}_j)_{0\le j\le n}$. *That is,* $E(d_j|\mathcal{E}_{j-1}) = 0, j = 1, \cdots, n$. *Suppose that* $||d_j||_\infty \le M$ *for a constant* $M < \infty, j = 1, \cdots, n$. *Let* $\tau \le n$ *be a stopping time relative to* $(\mathcal{E}_i)$ *that satisfies* $|| \sum_{j=1}^{\tau} E(d_j^2|\mathcal{E}_{j-1})||_\infty \le L$ *for a constant* L. *If* $0 \le \eta \le \frac{L}{2M}$, *then*

$$P\left(|\sum_{j=1}^{\tau} d_j| > \eta \right) \le 2 \cdot \exp\left\{ -\frac{\eta^2}{3L} \right\}.$$

PROOF. Since the bound for the tail probability $P\left(\left|\sum_{j=1}^{\tau} d_j\right| > \eta\right)$ in Freedman inequality, see Proposition 2.1 of [7], is given by

$$2 \cdot \exp\left\{ -\frac{\eta^2}{2(L + M\eta)} \right\} \text{ for } \eta \ge 0,$$

the result directly follows from the restriction $0 \le \eta \le \frac{L}{2M}$. □

The following Lemma whose proof is based on the chaining argument appears in [16]. For an analogous result on a probability version, see the chaining lemma in [13].

LEMMA 2. *Let (T, ρ) be a pseudometric space. Let $\{X(t) : t \in T\}$ be a separable sub-Gaussian process in the sense that*

$$P(|X(s) - X(t)| > \eta) \leq 2\exp\left\{-\frac{\eta^2}{\rho^2(s,t)}\right\} \text{ for every } s, t \in T,\ \eta > 0.$$

Then for $\delta > 0$ there exists a universal constant K such that

$$E \sup_{\rho(s,t)\leq\delta} |X(s) - X(t)| \leq K \int_0^\delta \sqrt{\log D(\epsilon, T, \rho)}\, d\epsilon,$$

where $D(\epsilon, T, \rho)$ is the packing number for (T, ρ).

REMARK.

1. One can work with an upper bound $K\exp\{-\frac{\eta^2}{C\rho^2(s,t)}\}$ for certain constants K and C instead of $2\exp\{-\frac{\eta^2}{\rho^2(s,t)}\}$ in the Lemma 2 without affecting the problem. See Problem 2.2.14 of [16].
2. The separability of the process $\{X(t) : t \in T\}$ in the Lemma 2 means that $\sup_{\rho(s,t)\leq\delta} |X(s) - X(t)|$ remains almost surely the same if the index set T is replaced by a suitable countable subset.

From now on, we make the technical assumption that the process $\{S_n(f) : f \in \mathcal{F}\}$ is separable where d is the metric given in (5). Compare with the statement of the chaining lemma in [13] where the separability is assumed in a different way.

We introduce the following truncation argument. For $\theta > 0$, $n \geq 1$, $f \in \mathcal{F}$, let

$$f^{(\sqrt{n}\theta)}(\cdot) = \begin{cases} \sqrt{n}\theta & \text{if } f(\cdot) > \sqrt{n}\theta \\ f(\cdot) & \text{if } |f(\cdot)| \leq \sqrt{n}\theta \\ -\sqrt{n}\theta & \text{if } f(\cdot) < -\sqrt{n}\theta. \end{cases}$$

so that $f^{(\sqrt{n}\theta)}(\cdot)$ is a truncation of $f(\cdot)$ at the level $\sqrt{n}\theta$. For fixed $\theta > 0$, we simplify the notation by writing

$$\bar{f}(d_j) := f^{(\sqrt{n}\theta)}(d_j) - E_{j-1} f^{(\sqrt{n}\theta)}(d_j).$$

Define

$$S_n^{(\theta)}(f) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \bar{f}(d_j) \text{ for } f \in \mathcal{F}\ .$$

The following maximal inequality for the truncated and stopped martingale differences plays a key role in the proof of the main result. See Theorem 3.1 of [17] for the parallel result in the case of independent process. See also Proposition 1 of [1] in the case of stationary martingale difference where a delicate chaining argument with stratification is employed under the bracketing entropy.

THEOREM 3. *Let $\{d_j : 1 \le j \le n,\ n \in \mathbf{N}\}$ be a sequence of martingale differences of $\mathcal{L}_2$-process indexed by a class $\mathcal{F}$ of measurable functions with envelop F on a measurable space $(X, \mathcal{X})$. Let τ be a finite stopping time relative to the increasing σ-fields $\{\mathcal{E}_j : 0 \le j \le n,\ n \in \mathbf{N}\}$ that satisfies*

$$\sigma_\tau^2(f,g) \le Dd^2(f,g) \text{ for } f,g \in \mathcal{F} \text{ and for a constant } D.$$

Then, for every $\delta > 0$, and for every $\theta \le (\frac{D\delta^2}{12})^{1/2}$, there exist universal constants K and C such that

$$E \sup_{d(f,g)\le\delta} |S_\tau^{(\theta)}(f) - S_\tau^{(\theta)}(g)| \le KJ(C\delta)$$

where

$$J(\delta) := \int_0^\delta \sup_{\gamma\in M(X,F)} (\log N(\epsilon[\gamma(F^2)]^{1/2}, \mathcal{F}, d_\gamma^{(2)}))^{1/2} d\epsilon.$$

PROOF. Let $\delta > 0$, let $0 < \theta \le (\frac{D\delta^2}{12})^{1/2}$. Let $f, g \in \mathcal{F}$ fixed. Then the martingale difference $d_j := \frac{\bar{f}(d_j)-\bar{g}(d_j)}{\sqrt{n}}$ has an upper bound $M := 4\theta$. Take $L := 5Dd^2(f,g)$. Then we have

$$\frac{1}{n}\sum_{j=1}^{\tau} E_{j-1}(\bar{f}(d_j) - \bar{g}(d_j))^2 \le 12\tau\theta^2 + 4\sigma_\tau^2 \le 12\tau\theta^2 + 4Dd^2(f,g) \le L.$$

Now, apply Lemma 1 to obtain for $0 < \eta < \frac{L}{2M}$

$$P\left(|S_\tau^{(\theta)}(f) - S_\tau^{(\theta)}(g)| > \eta\right) \le 2\cdot\exp\left\{-\frac{\eta^2}{15Dd^2(f,g)}\right\}.$$

Next, apply Lemma 2 to the process $\{S_\tau^{(\theta)}(f) : f \in \mathcal{F}\}$ to conclude that

$$E \sup_{d(f,g)\le\delta} |S_\tau^{(\theta)}(f) - S_\tau^{(\theta)}(g)| \le K \int_0^\delta (\log D(\epsilon, \mathcal{F}, d))^{1/2} d\epsilon \tag{9}$$

for a universal constant K. Finally an observation of the equivalence between the packing number and the covering number, and a change of variables gives the bound $KJ(C\delta)$ where K and C are universal constants. This completes the proof of Theorem 3. □

We are now ready to finish the proof of Theorem 2.

PROOF. Let $\epsilon > 0$ be fixed. Since $\{f(d_j), 1 \le j \le n, n \in \mathbf{N}\}$ is a sequence of martingale differences with respect to $\{\mathcal{E}_j, 0 \le j \le n, n \in \mathbf{N}\}$, we have

$$|E_{j-1}(f(d_j)\{|f(d_j)| > \sqrt{n}\theta\})| = |E_{j-1}(f(d_j)\{|f(d_j)| \le \sqrt{n}\theta\})|. \tag{10}$$

Note that for any $\theta > 0$, $f \in \mathcal{F}$, we have

$$\begin{aligned}
&\sup_{f\in\mathcal{F}} |S_n(f) - S_n^{(\theta)}(f)| \\
&= \sup_{f\in\mathcal{F}} \frac{1}{\sqrt{n}} \left| \sum_{j=1}^{n} \{f(d_j) - f^{(\sqrt{n}\theta)}(d_j) + E_{j-1}(f^{(\sqrt{n}\theta)}(d_j))\} \right| \\
&\le \sup_{f\in\mathcal{F}} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \{|f(d_j) - f^{(\sqrt{n}\theta)}(d_j)| + |E_{j-1}(f^{(\sqrt{n}\theta)}(d_j))|\} \\
&\le \frac{1}{\sqrt{n}} \sum_{j=1}^{n} F(d_j)\{F(d_j) > \sqrt{n}\theta\} \\
&\quad + \sup_{f\in\mathcal{F}} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} |E_{j-1}(f(d_j)\{|f(d_j)| \le \sqrt{n}\theta\})| \\
&\quad + \sup_{f\in\mathcal{F}} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} |E_{j-1}(\sqrt{n}\theta)\{|f(d_j)| > \sqrt{n}\theta\})| \\
&\le \frac{1}{\sqrt{n}} \sum_{j=1}^{n} F(d_j)\{F(d_j) > \sqrt{n}\theta\} \\
&\quad + \sup_{f\in\mathcal{F}} \frac{2}{\sqrt{n}} \sum_{j=1}^{n} |E_{j-1}(f(d_j))\{|f(d_j)| > \sqrt{n}\theta\})| \text{ by (10)} \\
&\le \frac{1}{\sqrt{n}} \sum_{j=1}^{n} F(d_j)\{F(d_j) > \sqrt{n}\theta\} + \frac{2}{\sqrt{n}} \sum_{j=1}^{n} E_{j-1}(F(d_j)\{F(d_j) > \sqrt{n}\theta\}) \\
&\le \frac{1}{\theta n} \sum_{j=1}^{n} F^2(d_j)\{F(d_j) > \sqrt{n}\theta\} + \frac{2}{\theta n} \sum_{j=1}^{n} E_{j-1}F^2(d_j)\{F(d_j) > \sqrt{n}\theta\}\ .
\end{aligned}$$

The last two bounds converge to zero in probability by the condition (7) on the conditional variances and the Lindeberg condition (8). Therefore we have $P\{\sup_{f\in\mathcal{F}}|S_n(f)-S_n^{(\theta)}(f)|>\frac{\epsilon}{4}\}=o(1)$. Since for every $\eta>0$

$$\sup_{d(f,g)<\eta}|S_n(f)-S_n(g)|\leq \sup_{d(f,g)<\eta}|S_n^{(\theta)}(f)-S_n^{(\theta)}(g)|+2\sup_{f\in\mathcal{F}}|S_n(f)-S_n^{(\theta)}(f)|,$$

it remains to show for n large enough, there exists $\eta>0$ such that

$$P\left(\sup_{d(f,g)<\eta}|S_n^{(\theta)}(f)-S_n^{(\theta)}(g)|>\frac{\epsilon}{2}\right)\leq\frac{\epsilon}{2}. \tag{11}$$

We define a sequence of random variables τ_n for $n\geq 1$

$$\tau_n:=n\wedge\max\left\{k\geq 0:\sup_{f,g\in\mathcal{F}}\frac{\sigma_k^2(f,g)}{d^2(f,g)}<D\right\}.$$

Being the random variables $\sigma_k^2(f,g)$ predictable, we see that τ_n is a stopping time. Notice that

$$P\left(\sup_{f,g\in\mathcal{F}}\frac{\sigma_{\tau_n}^2(f,g)}{d^2(f,g)}\geq D\right)=0. \tag{12}$$

Since $P(\tau_n<n)\to 0$ as $n\to\infty$, it is enough to prove that for every $\eta>0$, and for every $\theta\leq(\frac{D\eta^2}{12})^{1/2}$

$$P^*\left(\sup_{d(f,g)<\eta}|S_{\tau_n}^{(\theta)}(f)-S_{\tau_n}^{(\theta)}(g)|>\frac{\epsilon}{2}\right)<\frac{\epsilon}{2}, \tag{13}$$

where

$$S_{\tau_n}^{(\theta)}(f)=\frac{1}{\sqrt{n}}\sum_{j=1}^{\tau_n}\bar{f}(d_j).$$

Notice that from (12) we have

$$\sigma_{\tau_n}^2(f,g)\leq Dd^2(f,g).$$

From Theorem 3, we conclude that for every $\eta>0$, and for every $\theta\leq(\frac{D\eta^2}{12})^{1/2}$, there exist universal constants K and C such that

$$E^*\sup_{d(f,g)\leq\eta}|S_\tau^{(\theta)}(f)-S_\tau^{(\theta)}(g)|\leq KJ(C\eta)$$

where

$$J(C\eta) := \int_0^{C\eta} \sup_{\gamma \in M(X,F)} (\log N(\epsilon[\gamma(F^2)]^{1/2}, \mathcal{F}, d_\gamma^{(2)}))^{1/2} d\epsilon.$$

Since $\mathcal{F}$ has uniformly integrable entropy, we can choose $\eta > 0$ small enough to have $\frac{2}{\epsilon} K J(C\eta) < \frac{\epsilon}{2}$. Now notice that

$$\begin{aligned} &P^* \left(\sup_{d(f,g)<\eta} |S_{\tau_n}^{(\theta)}(f) - S_{\tau_n}^{(\theta)}(g)| > \frac{\epsilon}{2} \right) \\ &\leq \frac{2}{\epsilon} E^* \sup_{d(f,g)<\eta} |S_{\tau_n}^{(\theta)}(f) - S_{\tau_n}^{(\theta)}(g)| \\ &\leq \frac{2KJ(C\eta)}{\epsilon}, \end{aligned}$$

which is less than $\frac{\epsilon}{2}$ by the choice of η. The proof of Theorem 2 is completed. □

5. The Sequential Empirical Process for Martingale Differences

Let $(X, \mathcal{X})$ be a measurable space. We consider $(\Omega = X^{\mathbf{Z}}, \mathcal{E} = \mathcal{X}^{\mathbf{Z}}, P)$ as the basic probability space. We denote by S the left shift on Ω. We assume that P is invariant under S, i.e., $PS^{-1} = P$, and that S is ergodic. We denote by $\xi = \cdots, \xi_{-1}, \xi_0, \xi_1, \cdots$ the coordinate maps on Ω. From our assumptions it follows that $\{\xi_j\}_{j \in \mathbf{Z}}$ is a stationary and ergodic process. Next we define for $j \in \mathbf{Z}$ a σ-field $\mathcal{E}_j := \sigma(\xi_i : i \leq j)$ and $H_j := \{f : \Omega \to R : f$ is $\mathcal{E}_j$-measurable and $f \in L^2(\Omega)\}$. We denote for $f \in L^2(\Omega)$, $E_{j-1}(f) := E(f|\mathcal{E}_{j-1})$, and $H_0 \ominus H_{-1} := \{f \in H_0 : E(fg) = 0$ for $g \in H_{-1}\}$. Finally for every f, $g \in L^2(\Omega)$ we put $d(f,g) := [E(f-g)^2]^{1/2}$. Consider $\mathcal{F} \subseteq H_0 \ominus H_{-1}$ with envelope F satisfying $EF^2 < \infty$. From our setup it follows that for every $f \in \mathcal{F}$, $\{f(S^j(\xi)), \mathcal{E}_j\}$ is a stationary martingale difference sequence.

Consider now the sequential empirical process $\{Z_n(s,f) : s \in [0,1], f \in \mathcal{F}\}$ for the stationary martingale differences defined by

$$Z_n(s,f) = \frac{1}{\sqrt{n}} \sum_{j=1}^{[ns]} f(d_j) \text{ for } s \in [0,1] \text{ and } f \in \mathcal{F}, \tag{14}$$

where $d_j := S^j(\xi)$, $d := S^0(\xi)(=\xi)$ and $[x]$ denotes the integral part of x.

The following CLT for the sequential empirical process Z_n generalizes that of IID result to the stationary and ergodic martingale differences.

THEOREM 4. *Assume that $\mathcal{F}$ has uniformly integrable entropy with $EF^2 < \infty$. Suppose that*

$$E^* \sup_{f,g\in\mathcal{F}} \frac{E_{-1}[f(d) - g(d)]^2}{d^2(f,g)} < \infty. \tag{15}$$

Then

$$Z_n \Rightarrow Z \text{ as random elements of } B([0,1] \times \mathcal{F}).$$

The limiting process $Z = \{Z(s,f) : s \in [0,1], f \in \mathcal{F}\}$, known as a Kiefer-Muller process, is mean zero Gaussian with covariance function

$$EZ(s,f)Z(t,g) = (s \wedge t)Ef(\xi)g(\xi)$$

and the sample paths of Z is bounded and uniformly continuous with respect to the metric $|s - t| + d(f,g)$.

In the proof of Theorem 4, we are going to use some well known facts such as a multivariate CLT for stationary martingale differences, a permanence of the uniform entropy bound, and Lebesgue Dominated Convergence Theorem.

PROOF. Since the covariance function of Z_n is given by

$$cov(Z_n(s,f), Z_n(t,g)) = \frac{[ns] \wedge [nt]}{n} Ef(\xi)g(\xi),$$

we have $EZ(s,f)Z(t,g) = (s \wedge t)Ef(\xi)g(\xi)$. By the multivariate CLT for stationary and ergodic martingale differences, see for example Theorem 7.7.5 of [6], the finite dimensional distributions of the process $\{Z_n(s,f) : s \in [0,1], f \in \mathcal{F}\}$ converge to those of Z. By Theorem 2.10.20 of [16], the uniformly integrable entropy condition of the class $[0,1] \times \mathcal{F}$ is inherited from that of $\mathcal{F}$. Next, use the stationarity to observe that the condition (6) is reduced to the condition (15) and the condition (7) follows from the assumption $EF^2 < \infty$. Finally apply Lebesgue Dominated Convergence Theorem to verify the Lindeberg condition (8). Now, apply Theorem 1 to complete the proof. □

References

[1] J. Bae, *An Empirical CLT for Stationary Martingale Differences*, J. Korean Math. Soc. **32** (1995), 427-446.

[2] ———, *The Uniform CLT for the Function-Indexed Kaplan-Meier Integral Process*, Preprint, 1998.

[3] J. Bae and S. Levental, *Uniform CLT for Markov Chains and Its Invariance Principle: A Martingale Approach*, J. Theoret. Probab. **8** (1995), 549-570.

[4] P. Doukhan, P. Massart, and E. Rio, *Invariance principles for absolute regular empirical process*, Ann. Inst. H. Poincare Probab. Statist. **31** (1995), 393-427.

[5] R. M. Dudley, *A Course on Empirical Processes,* Lecture notes in Math. 1097 Springer-Verlag, New York, 1984.

[6] R. Durrett, *Probability: Theory and Examples,* Wadsworth, Belmont, CA, 1991.

[7] D. Freedman, *On Tail Probabilities for Martingales*, Ann. Probab. **3** (1975), 100-118.

[8] J. Hoffmann-Jϕrgensen, *Stochastic processes on Polish spaces,* Aarhus Universitet., Matematisk Institut, 1991.

[9] M. I. Gordin and B. A. Lifsic, B. A, *The central limit theorem for stationary Markov Processes*, Sov. Math. Dokl. **19** (1978), 392-394.

[10] E. L. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations*, J. Amer. Statist. Assoc. **53** (1958), 457-481.

[11] S. Levental, *A Uniform CLT for Uniformly Bounded Families of Martingale Differences*, J. Theoret. Probab. **2** (1989), 271-287.

[12] M. Ossiander, *A Central Limit Theorem under Metric Entropy with L_2 Bracketing*, Ann. Probab. **15** (1987), 897-919.

[13] D. Pollard, *Convergence of Stochastic Processes,* Springer series in Statistics, Springer-Verlag, New York, 1984.

[14] ———, *Empirical Processes: Theory and Applications,* Regional conference series in Probability and Statistics 2, Inst. Math. Statist., Hayward, CA, 1990.

[15] W. Stute, *The central limit theorem under random censorship*, Ann. Statist. **23** (1995), 422-439.

[16] A. W. Van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes with Applications to Statistics,* Springer series in Statistics, Springer-Verlag, New York, 1996.

[17] K. Ziegler, *Functional Central Limit Theorem for Triangular Arrays of Function-Indexed Processes under Uniformly Integrable Entropy Conditions,* J. Multivariate. Anal. **62** (1997), 233-272.

Department of Mathematics
Sungkyunkwan University
Suwon 440-746, Korea
E-mail: jsbae@yurim.skku.ac.kr
E-mail: mjc@math.skku.ac.kr