ILLINOIS JOURNAL OF MATHEMATICS
Volume 32, Number 1, Spring 1988

CONVERGENCE RATES FOR FUNCTION CLASSES WITH APPLICATIONS TO THE EMPIRICAL CHARACTERISTIC FUNCTION

BY

J.E. YUKICH

1. Introduction

Let $(S, \mathscr{S}, P)$ be a probability space and let X_i, $i \geq 1$, be independent, identically distributed (i.i.d.) S-valued random variables with common law P. We shall consider the X_i, $i \geq 1$, to be the coordinates for a countable product $(S^{\mathbf{N}}, \mathscr{S}^{\mathbf{N}}, P^{\mathbf{N}})$ of copies of $(S, \mathscr{S}, P)$. Let the nth empirical measure for P be defined by

$$P_n := n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n}),$$

where δ_x is the unit mass at $x \in S$.

Recent research has yielded new limit theorems for the empirical process

$$\left\{\int f(dP_n - dP) : f \in \mathscr{F}\right\},$$

where $\mathscr{F}$ is a class of measurable functions on S. We refer the reader to [5], [7], [10], [11], [22], [25] where attention is focused on the empirical process indexed by a single class of functions $\mathscr{F}$. Related research has concentrated on the empirical process indexed by a sequence of classes of functions, say $\mathscr{G}_n$, $n \geq 1$. For example, see [14], [21], [28], [31].

In recent work [31], the author has used randomization techniques and metric entropy methods to study the limit behavior of

$$\left\{\int g(dP_n - dP) : g \in \mathscr{G}_n\right\} \tag{1.1}$$

where $\mathscr{G}_n$, $n \geq 1$, is a sequence of function classes on $(S, \mathscr{S}, P)$. Under weak metric entropy and growth conditions on $\mathscr{G}_n$, $n \geq 1$, it is shown in [31], that

Received July 18, 1986.

there exist non-zero finite constants C_1 and C_2 such that

$$(1.2) \qquad C_1 \leq \limsup_{n\to\infty} \sup_{g\in\mathscr{G}_n} \left|\alpha(n)\int g(dP_n - dP)\right| \leq C_2 \quad \text{a.s.}\,(P^{\mathbf{N}}),$$

where $\alpha(n)$, the rate of convergence, is closely connected to the metric entropy of the classes $\mathscr{G}_n$, $n \geq 1$.

Taking $\mathscr{G}_1 = \mathscr{G}_2 = \cdots = \mathscr{G}_n$, $\alpha(n) \equiv 1$, and $C_1 = C_2 = 0$, and by considering (1.1) as a stochastic process with values in $(l^\infty(\mathscr{G}), \|\cdot\|_{\mathscr{G}})$ (the Banach space of bounded real functions on $\mathscr{G}$, equipped with the sup norm), we see that (1.2) becomes an infinite dimensional law of large numbers. In this context, (1.2) has been characterized by Hoffmann-Jørgensen [12], Giné and Zinn [10], and also by Talagrand [23].

In this article we shall only concern ourselves with the majorization part of (1.2); we shall first improve and reformulate an earlier result of the author (cf. Theorem 2.2 [31]) guaranteeing the existence of the upper bound C_2 in (1.2). Our improved version (see Theorem 2.1 below), which constitutes one of the main results of this article, shows that the author's previous sufficient conditions involving a double summation [29] may essentially be replaced by an elegant metric entropy integral condition. In this way we obtain a rate of convergence theorem which fits in nicely with existing limit theorems for the function indexed empirical process; it should be noted that we are, in effect, determining rates of convergence for the infinite dimensional strong law of large numbers. Finally, as in [31], we will reveal the close connection between the rate of convergence $\alpha(n)$ and the metric entropy of the underlying classes of functions.

Additionally, we will show that our sufficient metric entropy integral condition is essentially the weakest possible. This is done by considering the log log behavior of the empirical characteristic function. See Theorem 2.3 below.

By way of important application, we show that our main result yields new and improved rates of uniform a.s. convergence of the empirical characteristic function over intervals expanding with n. Perhaps more significantly, by relating metric entropy to the tail behavior of the underlying distribution, we are able to essentially characterize rates of uniform a.s. convergence in terms of elegant integral conditions involving the tail behavior (see Corollary 2.4 below). The results considerably extend and generalize upon previous work in this area; for example, see [2], [3], [4], [13], [17], [19], [28], [31].

This paper is organized as follows. The remainder of this section is devoted to terminology and notation. Section 2, provides statements of the main results, as announced in Theorem 2.1, Corollary 2.2, Theorem 2.3 and Corollary 2.4. The proof of Theorem 2.1 occupies Section 2 where in fact a slightly stronger result is obtained. Theorem 2.3 is proved in Section 3 and Corollary 2.4 is proved in Section 5.

Basic preliminaries. For definiteness we take as our underlying probability space

$$(S^{\mathbf{N}}, \mathscr{S}^{\mathbf{N}}, P^{\mathbf{N}}) \times (\Omega, \Sigma, Q),$$

where (Ω, Σ, Q) is a probability space independent of $(S, \mathscr{S}, P)$. This enlarged space will accommodate certain randomization techniques. In particular we will consider a Rademacher sequence ε_i, $i \geq 1$, on (Ω, Σ, Q), i.e., the ε_i, $i \geq 1$, are i.i.d. with $Q(\varepsilon_i = 1) = Q(\varepsilon_i = -1) = 1/2$. We will also consider an orthogaussian sequence g_i, $i \geq 1$, on (Ω, Σ, Q), i.e., the g_i are i.i.d. $N(0,1)$. In this context E_ε and E_g denote integration with respect to Q and E_X denotes integration with respect to $P^{\mathbf{N}}$. Also, since the supremum of the empirical process over an uncountable class of functions may not be measurable, this will necessitate use of outer probability measure. Let $Pr := P^{\mathbf{N}} \times Q$ and let

$$Pr^*(B) = \inf\{Pr(C) : C \supset B\}, \quad B \subset S^{\mathbf{N}} \times \Omega.$$

For every real-valued function f on S define the upper integral

$$E^*f := \inf\left\{\int g\,dP : g \geq f,\ g \textit{ is } \mathscr{S}\text{-measurable}\right\}.$$

Throughout, let $\mathscr{G}_n$, $n \geq 1$, be a sequence of classes of functions on $(S, \mathscr{S}, P)$. As in [10], we assume that the quantities

$$\sup_{g \in \mathscr{G}_n} \left\{ \sum_{i=1}^{n} (a_i g(X_i) - bEg(X_i)) \right\}, \quad a_i, b \in \mathbf{R},\ n \in \mathbf{N}^+,$$

are $P^{\mathbf{N}}$-completion measurable. For all $j, k \in \mathbf{N}^+$, let

$$\|S_k\|_j := \sup_{g \in \mathscr{G}_j} \left| \sum_{i=1}^{k} (g(X_i) - Eg(X_i)) \right|.$$

As previously indicated, this article studies growth rates for the normed sums $\|S_n\|_n$.

Given $g \in \mathscr{G}_n$, let $\|g\|_\infty$ denote the essential supremum of g and $\|g\|_p$ the $\mathscr{L}^p(P)$ norm of g. Define the respective maximal sup and $\mathscr{L}^2$ norms by

$$B(n) := \sup_{g \in \mathscr{G}_n} \|g\|_\infty \quad \text{and} \quad V(n) := \sup_{g \in \mathscr{G}_n} \|g\|_2^2. \tag{1.3}$$

When understood we write B and V for $B(n)$ and $V(n)$, respectively.

Let (T, ρ) be a metric or pseudo-metric space. The covering number $N(\varepsilon, T, \rho)$, $\varepsilon \in [0, |T|_\rho]$, where $|T|_\rho$ denotes the diameter of (T, ρ), is defined as

$$N(\varepsilon, T, \rho) = \min\Big\{n\colon \exists t_1, \ldots, t_n \in T \quad \text{such that} \quad \min_i \rho(t, t_i) \le \varepsilon \quad \text{for all} \quad t \in T\Big\}.$$

The metric entropy $H(\varepsilon, T, \rho)$ is defined as $H(\varepsilon, T, \rho) = \log N(\varepsilon, T, \rho)$.

If f and g are functions in $\mathscr{L}^2(S, \mathscr{S}, P)$, let

$$e_P(f, g) = \left(\int (f-g)^2 \, dP\right)^{1/2}$$

denote their $\mathscr{L}^2(P)$ distance.

If f and g are functions in $\mathscr{L}^0(S, \mathscr{S}, P)$ and if $\{X_i\}_{i=1}^n$ is a sample from P, we may consider the random distances

$$d_{n,p}(f, g) := \left(n^{-1} \sum_{i=1}^n |f(X_i) - g(X_i)|^p\right)^{1/(p \vee 1)}, \quad 0 < p < \infty.$$

The associated covering numbers for a class $\mathscr{F} \subset \mathscr{L}^0(S, \mathscr{S}, P)$ are defined by

$$N_{n,p}(\varepsilon, \mathscr{F}) = N(\varepsilon, \mathscr{F}, d_{n,p}).$$

We note that the $N_{n,p}$ are random covering numbers and are not necessarily measurable.

Given a pair of functions $f, g\colon S \to \mathbf{R}$, define their bracket by

$$[f, g] := \{h\colon S \to \mathbf{R} \quad \text{such that } f(x) \le h(x) \le g(x) \quad \text{for all} \quad x \in S \quad \text{and } h \text{ is measurable}\}.$$

As in [5], [7], given $\mathscr{F} \subset \mathscr{L}^1(S, \mathscr{S}, P)$ and $\varepsilon > 0$, let $\mathscr{N}(\varepsilon, \mathscr{F}, P)$ be a collection of minimum cardinality of brackets $[f^-, f^+]$ such that

$$\mathscr{F} \subset \bigcup_{[f^-, f^+]} [f^-, f^+],$$

where $\|f^+ - f^-\|_1 < \varepsilon$. Let

$$N_{[\,]}(\varepsilon, \mathscr{F}) = \operatorname{card} \mathscr{N}(\varepsilon, \mathscr{F}, P) \quad \text{and} \quad H_{[\,]}(\varepsilon, \mathscr{F}) = \log N_{[\,]}(\varepsilon, \mathscr{F}).$$

In the case that $\mathscr{F}$ consists of complex valued functions, let

$$N_{[\,]}(\varepsilon, \mathscr{F}) := \operatorname{card} \mathscr{N}(\varepsilon, \mathscr{F}_0, P)$$

where $\mathscr{F}_0$ is the collection $\{\mathrm{Re}(f)\colon f \in \mathscr{F}\} \cup \{\mathrm{Im}(f)\colon f \in \mathscr{F}\}$.

Finally, we note that all a.s. statements are meant to be a.s. statements with respect to Pr^*.

2. The main results

As indicated, we shall be concerned with optimal a.s. growth rates for the normed partial sums $\|S_n\|_n$; in order that our results may be stated in their fullest generality, let $\mathscr{H}$ denote the collection of functions $h: \mathbf{R}^+ \to \mathbf{R}^+$ such that

(i) as $x \downarrow 0$, $f(x)/h(f(x))$ is monotonically increasing whenever $f: \mathbf{R}^+ \to \mathbf{R}^+$ is monotonically increasing,
(ii) $h(x) = O(x^{1/2})$ as $x \uparrow \infty$ and
(iii) $\log\log x = O(h(x))$.

Throughout, $B(n)$and $V(n)$ are as in (1.3).

THEOREM 2.1. *Let $\mathscr{G}_n$, $n \geq 1$, be a sequence of classes of functions on $(S, \mathscr{S}, P)$. Assume that $\|S_n\|_j \leq \|S_n\|_n$ for all $j \leq n$. Suppose that there is an $h \in \mathscr{H}$ and constants $C_1 > 0$ and ε_0 such that for all $0 < \varepsilon \leq \varepsilon_0$ and all $n \geq 1$,*

$$C_1 2^{h(n)} \leq N_{[\,]}(\varepsilon, \mathscr{G}_n) \leq 2^{h(n)} N(\varepsilon), \tag{2.1}$$

where $N(\varepsilon)$ is some continuous, monotonically increasing function satisfying

$$\int_0 \left(\frac{\log N(\varepsilon^2)}{h(\log N(\varepsilon^2))} \right)^{1/2} d\varepsilon < \infty. \tag{2.2}$$

If $h(n)V(n)/n = O(1)$ and if $\sup_{n \geq 1} B(n)/(V(n)C_2) \leq 1$ for some finite constant C_2, then there is a finite constant U_1 such that

$$\limsup_{n \to \infty} \left(\frac{1}{h \cdot V \cdot n} \right)^{1/2} \|S_n\|_n \leq U_1 \quad \text{a.s.} \tag{2.3}$$

In general, as shown by Theorem 2.1 of [31], the growth rate provided by (2.3) is essentially the best possible; more precisely, under weak additional hypotheses (2.3) may be bounded below by a strictly positive constant.

We immediately deduce the following result.

COROLLARY 2.2. *Let $\mathscr{G}_n$, $n \geq 1$, be a sequence of classes of functions satisfying all the conditions of Theorem 2.1 with $N(\varepsilon)$ equal to $N_{[\,]}(\varepsilon, G_1)$; i.e., for all $0 < \varepsilon \leq \varepsilon_0$ and all $n \geq 1$,*

$$C_1 2^{h(n)} \leq N_{[\,]}(\varepsilon, \mathscr{G}_n) \leq 2^{h(n)} N_{[\,]}(\varepsilon, \mathscr{G}_1), \tag{2.4}$$

and also

$$\int_0 \left(\frac{\log N_{[\,]}(\varepsilon^2, \mathscr{G}_1)}{h\left(\log N_{[\,]}(\varepsilon^2, \mathscr{G}_1)\right)} \right)^{1/2} d\varepsilon < \infty. \tag{2.5}$$

Then (2.3) *holds.*

Let us see how Theorem 2.1 and Corollary 2.2 fit in with existing limit theorems for the function indexed empirical process. Recall, for example, that if h is a constant function in (2.5) then $\mathscr{G}_1$ is a P-Donsker class of functions [7], [20]; also, if $\mathscr{G}_1 \subset \mathscr{L}^1(S, \mathscr{S}, P)$ satisfies the total boundedness condition $N_{[\,]}(\varepsilon, \mathscr{G}_1) < \infty$ for all $\varepsilon > 0$ then

$$\limsup_{n \to \infty} \frac{1}{n} \|S_n\|_{\mathscr{G}_1} = 0 \quad \text{a.s.},$$

i.e., (2.3) is satisfied with $h(n) = n$, V constant and $U_1 = 0$. See [7]. It has also been shown [30] that if $N_{[\,]}^{(2)}(\varepsilon, \mathscr{G})$ denotes metric entropy with bracketing in the $\mathscr{L}^2(P)$ norm, then the condition

$$\int_0 \left(\frac{\log N_{[\,]}^{(2)}(\varepsilon, \mathscr{G})}{\log\log\log N_{[\,]}^{(2)}(\varepsilon, \mathscr{G})} \right)^{1/2} d\varepsilon < \infty$$

implies that $\mathscr{G}$ satisfies a bounded law of the iterated logarithm (abbreviated BLIL), i.e.,

$$\limsup_{n \to \infty} \left(\frac{1}{n \log\log n} \right)^{1/2} \|S_n\|_{\mathscr{G}} < \infty \quad \text{a.s.} \tag{2.6}$$

This latter result extends and generalizes a similar result of Kuelbs and Dudley [16]. Seen in the context of the above remarks, the integral conditions (2.2) and (2.5) appear as very natural ones.

We may actually show that the entropy condition (2.5) cannot be substantially weakened; this will be done (see Theorem 2.3 below) by considering the class of functions

$$\mathscr{F} := \left\{ x \to e^{itx} : t \in [-1/2, 1/2] \right\} \tag{2.7}$$

together with the function

$$H_1 \colon x \to \left(x \left(\log \frac{1}{x} \right)^2 \log\log \frac{1}{x} \right)^{-1}, \quad 0 < x \le e^{-1}.$$

Also, a probability measure P on $\mathbf{R}$ is said to satisfy $(*)$ if

$(*)$ P has a density $f(x)$ and $\{f(x)+f(-x)\}$ is decreasing for x large.

As we will see later, the following theorem also has importance in the study of the empirical characteristic function.

THEOREM 2.3. *Let P satisfy $(*)$ and assume that the entropy function*

$$H_{[\,]}(\varepsilon) = H_{[\,]}(\varepsilon, \mathscr{F})$$

satisfies at least one of the two regularity conditions: either $H_{[\,]}(\varepsilon) = O(H_1(\varepsilon))$ or $H_1(\varepsilon) = O(H_{[\,]}(\varepsilon))$ as $\varepsilon \downarrow 0$. If $\mathscr{F}$ satisfies the BLIL, then

$$H_{[\,]}(\varepsilon) = O(H_1(\varepsilon))$$

and therefore for all $\tau > 0$,

$$\int_0 \left(\frac{\log N_{[\,]}(\varepsilon^2, \mathscr{F})}{\left(\log\log\log N_{[\,]}(\varepsilon^2, \mathscr{F})\right)^{1+\tau}} \right)^{1/2} d\varepsilon < \infty. \tag{2.8}$$

Returning to Theorem 2.1, we see that its significance lies with its generality and relatively easy applications. While it is often possible to determine rates of convergence for particular function classes, each class usually requires different techniques. Theorem 2.1, however, requires only estimations of $N_{[\,]}(\varepsilon, \mathscr{G}_n)$ and there are many function classes for which such estimates are readily available; see [6], [7], [15]. In this way, the above provides a unified approach to the rate of convergence problem. As general as it is, Theorem 2.1 cannot possibly furnish exact rates of convergence; it does, however, yield improved rates of convergence in many instances. They are, for example, straightforward applications of (2.3) to non-parametric density estimation; we postpone a discussion of this and discuss only a.s. rates of uniform convergence of the empirical characteristic function.

We will see that Theorem 2.1 enables us to significantly extend and generalize known results [3], [4], [28] concerning limit theorems for the empirical characteristic function. In fact, Theorem 2.3, along with results in [17], [27], [30] and the forthcoming Corollary 2.4, essentially characterize most forms of the limiting behavior of the empirical characteristic function on $\mathbf{R}^d$, $d \geq 1$. If P is the underlying distribution, we will see that the limiting behavior can be closely connected to integral conditions on the tail function [27], [31] defined by

$$M_P(\varepsilon) = \inf\{M\colon M \geq 1 \text{ and } P(\|X\| > M) < \varepsilon\}, \quad 0 < \varepsilon \leq 1. \tag{2.9}$$

Before proceeding to our main result in this area, we first recall some basic preliminaries.

Given a probability measure P on the real line, let $c(t)$ denotes its characteristic function $\int e^{itx}\,dP(x)$ and $c_n(t)$ the empirical characteristic function $\int e^{itx}\,dP_n(x)$. Recently, considerable attention has been given to determining rates of convergence for $\sup_{t\in I}|c_n(t) - c(t)|$, where $I \subset \mathbf{R}$ is a fixed interval; more generally, I may depend upon n [3], [4], [28], [29]. It is known [4], [28], for example, that if $I(n) = [-\exp(n/\alpha(n)), \exp(n/\alpha(n))]$ with $\alpha(n) \to \infty$, then

$$\limsup_{n\to\infty} \sup_{t\in I(n)} |c_n(t) - c(t)| = 0 \quad \text{a.s.} \tag{2.10}$$

for all P; however, in general, this fails when $I(n)$ is replaced by the entire real line.

Our next result shows how to deduce rates of convergence for

$$\sup_t |c_n(t) - c(t)|$$

in terms of integral conditions on the tail function $M_P(\varepsilon)$. Although we confine our attention to the real line for the sake of simplicity, it should be clear that Corollary 2.4 can be stated in the context of probability measures on general Euclidean spaces $\mathbf{R}^d$, $1 \le d < \infty$.

COROLLARY 2.4. *Let P be a probability measure on* $\mathbf{R}$ *satisfying* $\limsup_{t\to\infty}\mathrm{Re}\, c(t) < 1$. *Suppose that there is an $h \in \mathscr{H}$ such that*

$$\int_0 \left(\frac{\log M_P(\varepsilon^2)}{h\left(\log M_P(\varepsilon^2)\right)} \right)^{1/2} d\varepsilon < \infty. \tag{2.11}$$

Then there is a finite constant U_1 such that

$$\limsup_{n\to\infty} \sup_{|t| \le 2^{h(n)}} \left(\frac{n}{h(n)} \right)^{1/2} |c_n(t) - c(t)| \le U_1 \quad \text{a.s.} \tag{2.12}$$

At this point, a few remarks are in order. When working with the integral condition (2.11) in the context of Theorem 2.1, we require the bound $h(x) = O(x^{1/2})$. We conjecture that Corollary 2.4 remains true without this asymptotic condition.

It should be noted that under $(*)$ the integral condition of Theorem 2.3 implies the tail condition [29]

$$\int_0 \left(\frac{\log M_P(\varepsilon^2)}{\left(\log\log\log M_P(\varepsilon^2)\right)^{1+\tau}} \right)^{1/2} d\varepsilon < \infty \quad \text{for all } \tau > 0. \tag{2.13}$$

Thus the integral condition (2.11) cannot, in general, be substantially weakened. The reader should also recall [27], [18] that if

$$\int_0 \left(\log M_P(\varepsilon^2)\right)^{1/2} d\varepsilon < \infty, \tag{2.14}$$

then the normalized empirical process

$$C_n(t) := n^{1/2}(c_n(t) - c(t)), \quad t \in [-1/2, 1/2],$$

converges weakly to a Gaussian process on the space of continuous, complex valued functions on $[-1/2, 1/2]$; moreover, under $(*)$, (2.14) is actually necessary for weak convergence. Finally, the integral condition (2.13), with $\tau = 0$ there, implies that $C_n(t)$, $t \in [-1/2, 1/2]$, satisfies a compact law of the iterated logarithm [29]; this follows from [17] and the relation

$$M_P(\varepsilon) \geq \varepsilon^2 N\left(3\varepsilon^2, \mathscr{F}, \|\cdot\|_1\right)$$

of [27].

When taken together, these results clearly show that the various forms of the limiting behavior of $C_n(t)$, $t \in [-1/2, 1/2]$, can essentially be characterized through integral conditions on the tail function $M_P(\varepsilon)$; this general conclusion is worth noting and serves as a focal point around which existing theorems for $C_n(t)$ may be studied. Finally, as indicated earlier, only trivial modifications are needed in order that the above results apply equally well to empirical characteristic functions on general Euclidean spaces $\mathbf{R}^d$, $d \geq 1$.

3. Proof of Theorem 2.1

The proof of Theorem 2.1 will follow from the following stronger, albeit more cumbersome result.

THEOREM 3.1. *Let $\mathscr{G}_n$, $n \geq 1$, be a sequence of classes of functions; assume that*

$$\|S_n\|_j \leq \|S_n\|_n \quad \textit{for all } j \leq n.$$

Assume that there are constants $\varepsilon > 0$ and $\gamma := \gamma(\varepsilon) > 0$ and a decreasing function s: $\mathbf{R}^+ \to \mathbf{R}^+$ with $\Sigma_j (s(j))^{1/2} = M < \infty$ such that

$$\sum_{k=2}^{\infty} \sum_{j=\log(1/\varepsilon)}^{k/2} N_{[\,]}^2\left(2^{-j-1}, \mathscr{G}_{2^k}\right)\left(N_{[\,]}(\varepsilon, \mathscr{G}_{2^k})\right)^{-\gamma s(j) 2^j} < \infty. \tag{3.1}$$

Letting $H := H(n) := H_{[\,]}(\varepsilon, \mathscr{G}_n)$ *suppose that*

$$H(n)V(n)/n = O(1), \quad \log\log n = O(H(n))$$

and $\sup_{n\ge 1} B(n)/(V(n)C_2) \le 1$ *for some finite constant* C_2. *Then there is a finite constant* U_2 *such that*

$$\limsup_{n\to\infty} \sup_{g\in\mathscr{G}_n} \left(\frac{1}{H\cdot V\cdot n}\right)^{1/2} \|S_n\|_n \le U_2 \quad \text{a.s.}$$

Remarks. This result represents a slight improvement over Theorem 2.2 of [31], in which the function $s(j)$ equals j^{-4}.

Without loss of generality we may assume that $\log(1/\varepsilon)$ is integral.

We note that the minimal growth rate condition on H implies the existence of $K := K(\varepsilon)$ such that

$$\sum_{k=1}^{\infty} \left(N_{[\,]}(\varepsilon, \mathscr{G}_{2^k})\right)^{-K} < \infty; \tag{3.2}$$

using this K the proof of Theorem 3.1 will show that U_2 can be as small as

$$8\left(2(K+1)^{1/2} + 12M(\gamma C_2)^{1/2}\right). \tag{3.3}$$

Equipped with Theorem 3.1, the proof of Theorem 2.1 now becomes quite easy.

Proof of Theorem 2.1. Using conditions (i) and (ii) it is easily verified that (2.1) and (2.2) imply the existence of a function s: $\mathbf{R}^+ \to \mathbf{R}^+$ with $\Sigma_j(s(j))^{1/2} < \infty$, and an integer j_0 such that for all $j \ge j_0$ and $n \in \mathbf{N}^+$,

$$C_1 2^{h(n)} \le N_{[\,]}\left(2^{-j}, \mathscr{G}_n\right) \le 2^{h(n)} 2^{s(j)h(2^{2^j})2^j}.$$

Using the function $s(\cdot)$, condition (iii) and the above inequality, it is also easily verified that given ε_0 there are constants $\varepsilon > 0$, $0 < \varepsilon < \varepsilon_0$, and $\gamma := \gamma(\varepsilon) > 0$ satisfying (3.1). The proof is completed by noting for fixed ε, $0 < \varepsilon < \varepsilon_0$, that $H(n) \le 2h(n)$ for n large. Q.E.D.

We now turn to the proof of Theorem 3.1.

Proof of Theorem 3.1. First assume that $H(n)V(n)/n = o(1)$. The proof in this case is a straightforward modification of the proof of Theorem 2.2 of [31]. There, we note that the function $j \to j^{-4}$ used in the chaining argument was chosen only because of its convenient summability $\Sigma_j(j^{-4})^{1/2} < \infty$. In fact, a close reading of the proof of Theorem 2.2 [31] reveals that the function $j \to j^{-4}$ can be replaced by any function s: $j \to s(j)$ satisfying $\Sigma_j(s(j))^{1/2} < \infty$; the proof remains unchanged, save for the appearance of new constants. Let us briefly show how this may be done.

Find fixed values for $\varepsilon > 0$, $\gamma := \gamma(\varepsilon) > 0$, and $K := K(\varepsilon)$ such that conditions (3.1) and (3.2) are satisfied. Throughout we will write, for fixed $g \in \mathscr{G}_n$, $S_n(g) := \Sigma_{i=1}^n(g(X_i) - Eg(X_i))$. For each bracket $[g^-, g^+]$ in $\mathscr{N}(\varepsilon, \mathscr{G}_n, P)$ find a $\tilde{g} \in [g^-, g^+] \cap \mathscr{G}_n$. Denote the finite collection $\{\tilde{g}\}$ by $\tilde{\mathscr{G}}_n$. Let $\beta := 12(C_2\gamma)^{1/2}M$. Let $0 < \mu < 1$ and find $n_0 := (\varepsilon, \mu)$ such that for all $n \geq n_0$

$$B(3H(K+1)/nV)^{1/2} \leq \mu \quad \text{and} \quad \beta(HV/n)^{1/2} < 4\varepsilon.$$

Let $A := ((2 + \mu)(K + 1))^{1/2}$. As in [31], the proof is centered around the following basic inequality:

$$\begin{aligned}
&Pr^*\left\{\sup_{g\in\mathscr{G}_n}|S_n(g)| > (A+\beta)(HnV)^{1/2}\right\} \\
(3.4)\qquad &\leq Pr^*\left\{\max_{\tilde{g}\in\tilde{\mathscr{G}}_n}|S_n(\tilde{g})| > A(HnV)^{1/2}\right\} \\
&\quad + Pr^*\left\{\left|\max_{\tilde{g}\in\tilde{G}_n}|S_n(g)| - \sup_{g\in\mathscr{G}_n}|S_n(g)|\right| > \beta(HnV)^{1/2}\right\}.
\end{aligned}$$

The proof consists of two main steps. The first consists of fixing n, $n \geq n_0$, and bounding the right hand side of (3.4) by

$$(3.5)\quad 2N_{[\,]}^{-K}(\varepsilon, \mathscr{G}_n) + 2\sum_{j=\ln(1/\varepsilon)}^{m} N_{[\,]}^2(2^{-j-1}, \mathscr{G}_n)\left(N_{[\,]}(\varepsilon, \mathscr{G}_n)\right)^{-\gamma s(j)2^j},$$

where $m := [1 + \log_2\{(n/VH)^{1/2}\}]$. The second step consists of applying a maximal inequality with $n = 2^k$, the Borel-Cantelli Lemma, and letting μ tend to zero.

Now carry out the proof exactly as on pp. 82–85 of [31].

Finally, the case $H(n)V(n)/n \simeq$ constant may be handled in an analogous way via trivial modifications of Theorem 2.3 in [31]. Q.E.D.

4. Proof of Theorem 2.3

The proof of Theorem 2.3 consists of two steps. Assuming that $\mathscr{F}$ satisfies the BLIL, the first step is to a.s. approximate the random distances $d_{n,2}(f, g)$, $f, g \in \mathscr{F}$, by the non-random distance $e_P(f, g)$. The second step consists of finding necessary random entropy conditions (in terms of $N_{n,2}$) and then, using the first step, replacing the random entropy $N_{n,2}$ by $N_{[\,]}$.

Now by considering real parts, it is clear that if $\mathscr{F}$ satisfies the BLIL, then so does $\{\cos tx\colon t \in [-\frac{1}{2}, \frac{1}{2}]\}$. By stationarity and the relation

$$\sin^2(tx) = (1 - \cos 2tx)/2,$$

the class $\tilde{\mathscr{F}} := \{(f-g)^2 : f, g \in \mathscr{F}\}$ also satisfies the BLIL, i.e.,

$$\limsup_{n\to\infty} \sup_{h\in\tilde{\mathscr{F}}} \left(\frac{n}{\log\log n}\right)^{1/2} \left| n^{-1} \sum_{i=1}^{n} (h(X_i) - Eh(X_i)) \right| \le C \quad \text{a.s.,}$$

where here and elsewhere, C denotes a positive, finite constant possibly changing from line to line. Thus for $n \ge n_0$ there are measurable sets Ω_n such that $Pr(\Omega_n) \to 1$ as $n \to \infty$ and

$$\sup_{f,g\in\mathscr{F}} |d_{n,2}^2(f,g) - e_P^2(f,g)| \le \left(\frac{\log n}{n}\right)^{1/2} \quad \text{for all } \omega \in \Omega_n.$$

Thus, for all $f, g \in \mathscr{F}$ and for all $n \ge n_0$,

$$e_P^2(f,g) \le d_{n,2}^2(f,g) + \left(\frac{\log n}{n}\right)^{1/2} \quad \text{for all } \omega \in \Omega_n.$$

Therefore, for all $\varepsilon > 0$ and for all $n \ge n_0$ this implies

$$N\left(\left(\varepsilon^2 + \left(\frac{\log n}{n}\right)^{1/2}\right)^{1/2}, \mathscr{F}, e_P\right) \le N_n(\varepsilon, \mathscr{F}) \quad \text{for all } \omega \in \Omega_n, \tag{4.1}$$

where we suppress the index 2 on $N_{n,2}$ for notational convenience.

Letting $u = n^{-1/5}$ and $u_0 = n_0^{-1/5}$, (4.1) implies for all $\varepsilon > 0$ and for all $u \le u_0$ that

$$N\left((\varepsilon^2 + u^2)^{1/2}, \mathscr{F}, e_P\right) \le N_{u^{-5}}(\varepsilon, \mathscr{F}) \quad \text{for all } \omega \in \Omega_{u^{-5}}.$$

Taking logarithms and expectations for u_0 small enough we obtain for all $u \le u_0$,

$$\log N\left((\varepsilon^2 + u^2)^{1/2}\right) \le 2E_X^* \log N_{u^{-5}}(\varepsilon, \mathscr{F}), \tag{4.2}$$

where we have suppressed the arguments $\mathscr{F}$ and e_p in N.

Now if $\mathscr{F}$ satisfies the BLIL, then a straightforward application of Lemma 2.9 of [10] shows that

$$\sup_{n\ge n_0} E_X^* E_g \sup_{\mathscr{F}} \left| \left(\frac{1}{n\log\log n}\right)^{1/2} \sum_{i=1}^{n} g_i f(X_i) \right| \le C. \tag{4.3}$$

Combining this with Fernique's minorization for stationary Gaussian processes [9] gives

$$\sup_{n \geq n_0} E_X^* \int_0^1 \left(\frac{\log N_n(\varepsilon, \mathscr{F})}{\log \log n} \right)^{1/2} d\varepsilon \leq C.$$

Letting $u = n^{-1/5}$ and using Fubini's theorem and (4.2) in that order, (4.3) yields

$$\sup_{0 < u \leq u_0} \int_0^1 \left(\frac{\log N\big((\varepsilon^2 + u^2)^{1/2}\big)}{\log \log(u^{-5})} \right)^{1/2} d\varepsilon \leq C.$$

Under condition $(*)$, Lemma 1 of [27] shows that

$$x^2 N_{[\,]}(3x^2, \mathscr{F}) = O(N(x, \mathscr{F}, e_P))$$

and thus the above inequality remains valid with $\log N(\cdot)^{1/2}$ replaced by $\log N_{[\,]}(\cdot)$, i.e.,

$$\sup_{0 < u \leq u_0} \int_0^1 \left(\frac{H_{[\,]}(\varepsilon^2 + u^2)}{\log \log(u^{-1})} \right)^{1/2} d\varepsilon \leq C. \tag{4.4}$$

Under the conditions of Theorem 2.3 we may actually verify that (4.4) holds if and only if

$$H_{[\,]}(x) = O(H_1(x)) \quad \text{as } x \downarrow 0. \tag{4.5}$$

To see this, let

$$I(t, u, f) := \int_0^t \left(f(\varepsilon^2 + u^2) \right)^{1/2} d\varepsilon$$

for all functions f: $\mathbf{R}^+ \to \mathbf{R}^+$ and for all $t \geq 0$, $u > 0$. If (4.5) holds, then the decomposition

$$I(t, u, H_{[\,]}) = \int_0^u + \int_u^t$$

shows that

$$I(t, u, H_{[\,]}) = o\left(\left(\log \log \frac{1}{u} \right)^{1/2} \right)$$

for t small enough.

Conversely, suppose that $H_{[\,]}$ satisfies (4.4) but not (4.5). From now on "all $t > 0$" means all $t > 0$ such that $\log\log(1/\sqrt{2}t)$ is defined. For all $t > 0$ define $u_1 := u_1(t)$ such that for all $u \le u_1(t)$,

$$\left(\log\log\frac{1}{\sqrt{2}u}\right)^{1/2} - \left(\log\log\frac{1}{\sqrt{2}t}\right)^{1/2} \ge \frac{1}{2}\left(\log\log\frac{1}{\sqrt{2}u}\right)^{1/2}.$$

For all $0 < u \le u_1$ and all $0 \le x \le t$ set

$$\lambda_u(x) := \left(H_{[\,]}(x^2+u^2)/H_1(x^2+u^2)\right)^{1/2};$$

we may assume without loss of generality that λ_u is continuous on $[0, t]$. Note that $\lambda_u(x)$ becomes arbitrarily large when $x^2 + u^2$ becomes small. By the Mean Value Theorem (e.g., see [23], pp. 123–124) we have for all $t > 0$ and $0 < u \le u_1$,

$$\text{(4.6)} \qquad \frac{I(t,u,H_{[\,]})}{I(t,u,H_1)} = \frac{\int_0^t \lambda_u(\varepsilon)\left(H_1(\varepsilon^2+u^2)\right)^{1/2} d\varepsilon}{\int_0^t \left(H_1(\varepsilon^2+u^2)\right)^{1/2} d\varepsilon} = \lambda_u(\xi)$$

for some $0 \le \xi \le t$. By taking t and u arbitrarily small, it is clear that the ratio (4.6) can be made arbitrarily large.

But this leads to a contradiction, since for all $t > 0$ and all $0 < u \le u_1$, we have by definition of u_1,

$$\begin{aligned} I(t,u,H_1) &\ge \int_u^t \left(H_1(\varepsilon^2+u^2)\right)^{1/2} d\varepsilon \\ &\ge 2\left[\left(\log\log\frac{1}{\sqrt{2}u}\right)^{1/2} - \left(\log\log\frac{1}{\sqrt{2}t}\right)^{1/2}\right] \\ &\ge \left(\log\log\frac{1}{\sqrt{2}u}\right)^{1/2}, \end{aligned}$$

and thus, by (4.4) we obtain for all $t > 0$,

$$\sup_{0<u\le u_1} \frac{I(t,u,H_{[\,]})}{I(t,u,H_1)} \le C.$$

This contradiction shows that $H_{[\,]}$ satisfies (4.5), completing the proof of Theorem 2.3. Q.E.D.

5. Proof of Corollary 2.4

The proof of Corollary 2.4 simply reduces to verifying that the function classes

$$\mathscr{G}_n := \{x \to e^{itx} : |t| \le 2^{h(n)}\}, \quad n \ge 1$$

satisfy the hypotheses of Theorem 2.1. Noting that condition (2.11) implies

$$\int_0 \left(\frac{\log(M_P(\varepsilon^2)/\varepsilon^2)}{h(\log(M_P(\varepsilon^2)/\varepsilon^2))} \right)^{1/2} d\varepsilon < \infty,$$

we conclude that it is enough to verify condition (2.1) with $N(\varepsilon) = 16 M_P(\varepsilon)/\varepsilon$; in fact it suffices to show that there is an ε_0 such that for all $0 < \varepsilon \le \varepsilon_0$ and all $n \ge 1$,

$$C_1 2^{h(n)} \le N_{[\,]}(\varepsilon, \mathscr{G}_n) \le 16 \cdot 2^{h(n)} M_P\left(\frac{\varepsilon}{3}\right) \Big/ \frac{\varepsilon}{3}. \tag{5.1}$$

The first inequality in (5.1) may be seen as follows [31].

By hypothesis, there is an ε_0 and a constant $M := M(\varepsilon_0) > 0$ such that for all $m \ge M$ we have $\int (1 - \cos mx)\, dP > 2\varepsilon_0$. Let j and k be any reals belonging to $[-2^{h(n)}, 2^{h(n)}]$ such that $|j - k| \ge M$. Then

$$\begin{aligned} \int |e^{ikx} - e^{ijx}|\, dP &\ge \int |1 - e^{ix|j-k|}|\, dP \\ &\ge \int (1 - \cos(|j-k|x))\, dP \\ &\ge 2\varepsilon_0. \end{aligned}$$

Thus there are at least $2 \cdot 2^{h(n)}/M$ functions of the form $f_k(x) = e^{ikx}$ such that for all $j \ne k$, $\int |f_j - f_k|\, dP \ge 2\varepsilon_0$, showing that

$$N_{[\,]}(\varepsilon, \mathscr{G}_n) \ge N(2\varepsilon, \mathscr{G}_n, \|\cdot\|_1) \ge 2 \cdot 2^{h(n)}/M \quad \text{for all } 0 < \varepsilon \le \varepsilon_0.$$

To establish the second inequality in (5.1) we observe that for all $0 < \varepsilon \le \varepsilon_0$,

$$N_{[\,]}(\varepsilon, \mathscr{G}_n) \le 2^{h(n)} N_{[\,]}(\varepsilon, \mathscr{F}) \le 16 \cdot 2^{h(n)} M_P\left(\frac{\varepsilon}{3}\right) \Big/ \frac{\varepsilon}{3},$$

where $\mathscr{F}$ is as in (2.7) and where the second inequality follows from [27]. Q.E.D.

REFERENCES

1. K. ALEXANDER, *Rates of growth and sample moduli for weighted empirical processes indexed by sets*, Probab. Th. Rel. Fields, vol. 75 (1987), pp. 379–423.
2. S. CSÖRGŐ, *Limit behavior of the empirical characteristic function*, Ann. Probab., vol. 9 (1981), pp. 130–144.
3. ______, *Rates of uniform convergence for the empirical characteristic function*, Acta Sci. Math (Szeged), vol. 48 (1985), pp. 97–102.
4. S. CSÖRGŐ and V. TOTIK, *On how long interval is the empirical characteristic function uniformly consistent*?, Acta Sci. Math. (Szeged), vol. 45 (1981), pp. 141–149.
5. R.M. DUDLEY, *Central limit theorems for empirical measures*, Ann. Probab., vol. 6 (1978), pp. 899–929; Correction in vol. 7 (1979), pp. 909–911.
6. ______, *Empirical and Poisson processes on classes of sets or functions too large for central limit theorems*, Z. Wahrsch. Verw. Gebiete, vol. 61 (1982), pp. 355–368.
7. ______, *A course on empirical processes*, Lecture Notes in Mathematics, no. 1097, Springer, N.Y., 1984, pp. 1–142.
8. R.M. DUDLEY AND WALTER PHILLIP, *Invariance principles for sums of Banach space valued random elements and empirical processes*, Z. Wahrsch. Verw. Gebiete, vol. 62 (1983), pp. 509–552.
9. X. FERNIQUE, *Regularité des trajectoires des fonctions aléatoires gaussiennes*, Lecture Notes in Mathematics, no. 480, Springer, N.Y., 1974, pp. 1–96.
10. E. GINÉ and J. ZINN, *Some limit theorems for empirical processes*, Ann. Probab., vol. 12 (1984), pp. 929–989.
11. ______, *Lectures on the central limit theorem for empirical processes*, Lecture Notes in Mathematics, no. 1221, Springer, N.Y., 1987, pp. 50–113.
12. J. HOFFMANN-JØRGENSEN, *The law of large numbers for nonmeasurable and non-separable random elements*, Astérisque, no. 1, vol. 1 (1985), pp. 299–356.
13. N. JAIN and M. MARCUS, *Continuity of subgaussian processes*, *Probability on Banach Spaces IV* (ed. J. Kuelbs), Marcel Dekker, New York, 1978.
14. V.I. KOLCINSKI, *Some limit theorems for empirical measures*, Theory Probab. Math. Statist., vol. 21 (1980), pp. 79–86.
15. A.N. KOLMOGOROV AND V.M. TIKHOMIROV, *ε-entropy and ε-capacity of sets in function spaces*, Amer. Math. Soc. Transl. Ser. 2, vol. 17 (1961), pp. 277–364. Uspekhi Mat. Nauk., vol. 14, pp. 3–86.
16. J. KUELBS AND R.M. DUDLEY, *Log log laws for empirical measures*, Ann. Probab., vol. 8 (1980), pp. 405–418.
17. M. LEDOUX, (1982) *Loi du logarithme itéré dans $C(S)$ et fonction caracteristique empirique*, Z. Wahrsch. Verw. Gebiete, vol. 60 (1982), pp. 425–435.
18. M. MARCUS, *Weak convergence of the empirical characteristic function*, Ann. Probab., vol. 9 (1981), pp. 194–201.
19. M. MARCUS and WALTER PHILIPP, *Almost sure invariance principles for sums of B-valued random variables with applications to random Fourier series and the empirical characteristic process*, Tran. Amer. Math. Soc., vol. 269 (1982), pp. 67–90.
20. M. OSSIANDER, *A central limit theorem under metric entropy with L^2 bracketing*, Ann. Probab., vol. 15 (1987), pp. 897–919.
21. D. POLLARD, *Rates of strong uniform convergence*, preprint.
22. ______, *Convergence of stochastic processes*, Springer, Berlin, 1984.
23. W. RUDIN, *Principles of mathematical analysis*, second ed., McGraw-Hill, N.Y., 1964.
24. M. TALAGRAND, *The Glivenko-Cantelli problem*, Ann. Probab., vol. 15 (1987), pp. 837–870.
25. ______, *Measurability properties for empirical processes*, Ann. Prob., vol. 15 (1987), pp. 204–212.
26. ______, *Donsker classes and random geometry*, to appear.

27. J.E. YUKICH, *Weak convergence of the empirical characteristic function*, Proc. Amer. Math. Soc., vol. 95 (1985), 3, pp. 470–473.
28. ______, *Laws of large numbers for classes of functions*, J. Multivariate Analysis, vol. 17 (1985), pp. 245–260.
29. ______, *Log log laws for classes of functions*, Proc. of IVth Vilnius Conference on Prob. Math. Statist., VNU Sci. Press. Utrecht, 1985, pp. 346–348.
30. ______, *The law of the iterated logarithm for empirical processes*, Lecture Notes in Mathematics, Springer, N.Y., to appear.
31. ______, (1987) *Some limit theorems for the empirical process indexed by functions*, Probab. Theory Rel. Fields, vol. 74 (1987), pp. 71–90.

LEHIGH UNIVERSITY
BETHLEHEM, PENNSYLVANIA