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

BY
J.E. YUKICH

1. Introduction

Let \((S, \mathcal{S}, P)\) 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 \((S^N, \mathcal{S}^N, P^N)\) of copies of \((S, \mathcal{S}, P).\) Let the \(n\)th empirical measure for \(P\) be defined by

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

where \(\delta_x\) is the unit mass at \(x \in S.\)

Recent research has yielded new limit theorems for the empirical process

\[
\left\{ \int f(dP_n - dP) : f \in \mathcal{F} \right\},
\]

where \(\mathcal{F}\) 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 \(\mathcal{F}.\) Related research has concentrated on the empirical process indexed by a sequence of classes of functions, say \(\mathcal{G}_n, n \geq 1.\) 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

\[
(1.1) \quad \left\{ \int g(dP_n - dP) : g \in \mathcal{G}_n \right\}
\]

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

Received July 18, 1986.
there exist non-zero finite constants $C_1$ and $C_2$ such that

\begin{equation}
C_1 \leq \limsup_{n \to \infty} \sup_{g \in \mathcal{G}_n} \alpha(n) \int g(dP_n - dP) \leq C_2 \quad \text{a.s.} \quad (P^N),
\end{equation}

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

Taking $\mathcal{G}_1 = \mathcal{G}_2 = \cdots = \mathcal{G}_n$, $\alpha(n) = 1$, and $C_1 = C_2 = 0$, and by considering (1.1) as a stochastic process with values in $(L^\infty(\mathcal{G}), \| \cdot \|_\mathcal{G})$ (the Banach space of bounded real functions on $\mathcal{G}$, 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

$$(S^N, \mathcal{F}^N, P^N) \times (\Omega, \Sigma, Q),$$

where $(\Omega, \Sigma, Q)$ is a probability space independent of $(S, \mathcal{F}, P)$. This enlarged space will accommodate certain randomization techniques. In particular we will consider a Rademacher sequence $\epsilon_i, i \geq 1$, on $(\Omega, \Sigma, Q)$, i.e., the $\epsilon_i, i \geq 1$, are i.i.d. with $Q(\epsilon_i = 1) = Q(\epsilon_i = -1) = 1/2$. We will also consider an orthogonal-gaussian sequence $g_i, i \geq 1$, on $(\Omega, \Sigma, Q)$, i.e., the $g_i$ are i.i.d. $N(0,1)$. In this context $E_\epsilon$ and $E_g$ denote integration with respect to $Q$ and $E_X$ denotes integration with respect to $P^N$. 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 $Pr := P^N \times Q$ and let

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

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

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

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

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

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

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

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

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

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

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

Let $(T, \rho)$ be a metric or pseudo-metric space. The covering number $N(\epsilon, T, \rho), \epsilon \in [0, |T|_p], \rho$, where $|T|_\rho$ denotes the diameter of $(T, \rho)$, is defined
$N(\varepsilon, T, \rho) = \min \left\{ n : \exists t_1, \ldots, t_n \in T \text{ such that} \right.$

\[ \min_i \rho(t_i, t) \leq \varepsilon \text{ for all } t \in T \left. \right\}.$

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 $\mathcal{L}^2(S, \mathcal{F}, P)$, let

\[ e_p(f, g) = \left( \int (f - g)^2 \, dP \right)^{1/2} \]

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

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

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

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

\[ N_{n,p}(\varepsilon, \mathcal{F}) = N(\varepsilon, \mathcal{F}, d_{n,p}). \]

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

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

\[ [f, g] := \{ h : S \to \mathbb{R} \text{ such that } f(x) \leq h(x) \leq g(x) \text{ for all } x \in S \text{ and } h \text{ is measurable} \}. \]

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

\[ \mathcal{F} \subset \bigcup_{[f^-, f^+]} \mathcal{N}(\varepsilon, \mathcal{F}, P) \]

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

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

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

\[ N_\mathcal{F}(\varepsilon, \mathcal{F}) := \text{card } \mathcal{N}(\varepsilon, \mathcal{F}_0, P) \]

where $\mathcal{F}_0$ is the collection $\{ \text{Re}(f) : f \in \mathcal{F} \} \cup \{ \text{Im}(f) : f \in \mathcal{F} \}$. 
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 \( \mathcal{H} \) denote the collection of functions \( h: \mathbb{R}^+ \to \mathbb{R}^+ \) such that

(i) as \( x \downarrow 0 \), \( f(x)/h(f(x)) \) is monotonically increasing whenever \( f: \mathbb{R} \to \mathbb{R} \) 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 \( \mathcal{G}_n, n \geq 1 \), be a sequence of classes of functions on \((S, \mathcal{S}, P)\). Assume that \( \| S_n \|_j \leq \| S_n \|_n \) for all \( j \leq n \). Suppose that there is an \( h \in \mathcal{H} \) and constants \( C_1 > 0 \) and \( \varepsilon_0 \) such that for all \( 0 < \varepsilon \leq \varepsilon_0 \) and all \( n > 1 \),

\[
C_1 2^{h(n)} \leq N_{\{e, \mathcal{G}_n\}} \leq 2^{h(n)} N(\varepsilon),
\]

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

\[
\int_0^1 \left( \frac{\log N(e^2)}{h(\log N(e^2))} \right)^{1/2} de < \infty.
\]

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

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

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

\[
C_1 2^{h(n)} \leq N_{\{e, \mathcal{G}_n\}} \leq 2^{h(n)} N_{\{e, G_1\}},
\]
and also

\[ \int_0^1 \left( \frac{\log N_{\mathcal{F}}(e^2, \mathcal{F}_1)}{h(\log N_{\mathcal{F}}(e^2, \mathcal{F}_1))} \right)^{1/2} \, \, d\varepsilon < \infty. \]

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 \( \mathcal{F}_1 \) is a \( P \)-Donsker class of functions [7], [20]; also, if \( \mathcal{F}_1 \subset \mathcal{L}^2(S, \mathcal{F}, P) \) satisfies the total boundedness condition \( N_{\mathcal{F}_1}(e, \mathcal{F}_1) < \infty \) for all \( e > 0 \) then

\[ \limsup_{n \to \infty} \frac{1}{n} \| S_n \|_{\mathcal{F}_1} = 0 \quad \text{a.s.,} \]

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 \( N^{(2)}_{\mathcal{F}}(e, \mathcal{F}) \) denotes metric entropy with bracketing in the \( \mathcal{L}^2(P) \) norm, then the condition

\[ \int_0^1 \left( \frac{\log N^{(2)}_{\mathcal{F}}(e, \mathcal{F})}{\log \log \log N^{(2)}_{\mathcal{F}}(e, \mathcal{F})} \right)^{1/2} \, \, d\varepsilon < \infty \]

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

\[ \limsup_{n \to \infty} \left( \frac{1}{n \log \log n} \right)^{1/2} \| S_n \|_{\mathcal{F}} < \infty \quad \text{a.s.} \]

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

\[ \mathcal{F} := \{ x \to e^{itx} : t \in [-1/2, 1/2] \} \]

together with the function

\[ H_1: x \to \left( x \left( \log \frac{1}{x} \right)^2 \log \log \frac{1}{x} \right)^{-1}, \quad 0 < x \leq e^{-1}. \]
Also, a probability measure $P$ on $\mathbb{R}$ is said to satisfy (\(\ast\)) if

(\(\ast\)) $P$ 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 (\(\ast\)) and assume that the entropy function

\[
H_1(e) = H_1(e, \mathcal{F})
\]

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

\[
H_1(e) = O(H_1(e))
\]

and therefore for all $\tau > 0$,

\[
(2.8) \quad \int_0^\infty \left( \frac{\log N_1(e^2, \mathcal{F})}{(\log \log \log N_1(e^2, \mathcal{F}))^{1+\tau}} \right)^{1/2} \, de < \infty.
\]

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 $N_1(e, \mathcal{F})$ 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 $\mathbb{R}^d$, $d \geq 1$. 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

\[
(2.9) \quad M_p(e) = \inf \{M: M \geq 1 \text{ and } P(\|X\| > M) < e\}, \quad 0 < e \leq 1.
\]
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 \( I \subset \mathbb{R} \) 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 \( M_P(e) \). 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 \( \mathbb{R}^d, 1 \leq d < \infty \).

**Corollary 2.4.** Let \( P \) be a probability measure on \( \mathbb{R} \) satisfying \( \limsup_{t \to \infty} \Re c(t) < 1 \). Suppose that there is an \( h \in \mathcal{H} \) such that

\[
\int_0^\infty \left( \frac{\log M_P(e^2)}{h(\log M_P(e^2))} \right)^{1/2} \, de < \infty. \tag{2.11}
\]

Then there is a finite constant \( U_1 \) such that

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

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 (\( \ast \)) the integral condition of Theorem 2.3 implies the tail condition [29]

\[
\int_0^\infty \left( \frac{\log M_P(e^2)}{(\log \log M_P(e^2))^{1+\tau}} \right)^{1/2} \, de < \infty \quad \text{for all } \tau > 0. \tag{2.13}
\]
Thus the integral condition (2.11) cannot, in general, be substantially weakened. The reader should also recall [27], [18] that if

\[(2.14) \quad \int_0^\infty (\log M_p(\epsilon^2))^{1/2} \, d\epsilon < \infty, \]

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 \(C_n(t), \quad t \in [-1/2, 1/2],\) satisfies a compact law of the iterated logarithm [29]; this follows from [17] and the relation

\[M_p(\epsilon) \geq \epsilon^2 N(3\epsilon^2, \mathcal{F}, \| \cdot \|_1)\]

of [27].

When taken together, these results clearly show that the various forms of the limiting behavior of \(C_n(t), \quad t \in [-1/2, 1/2],\) can essentially be characterized through integral conditions on the tail function \(M_p(\epsilon)\); 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 \(\mathbb{R}^d, \quad d \geq 1.\)

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 \(\mathcal{G}_n, \quad n \geq 1,\) be a sequence of classes of functions; assume that

\[\| \mathcal{S}_n \|_j \leq \| \mathcal{S}_n \|_n \quad \text{for all } j \leq n.\]

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

\[(3.1) \quad \sum_{k=2}^{\infty} \sum_{j=\log(1/\epsilon)}^{k/2} N^2_{\epsilon^2} \left( 2^{-j-1}, \mathcal{G}_{2^j} \right) \left( N_{\epsilon^2} \left( \mathcal{G}_{2^j} \right) \right)^{-\gamma s(j)^{2j}} < \infty.\]
Letting $H := H(n) := H[1](\varepsilon, G_n)$ suppose that

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

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

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

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

$$(3.2) \quad \sum_{k=1}^{\infty} \left( N[1](\varepsilon, G_n) \right)^{-K} < \infty;$$

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

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

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 $s: \mathbb{R}^+ \to \mathbb{R}^+$ with $\sum_j s(j))^{1/2} < \infty$, and an integer $j_0$ such that for all $j \geq j_0$ and $n \in \mathbb{N}^+$,

$$C_12^{h(n)} \leq N[1](2^{-j}, G_n) \leq 2^{h(n)}2^{s(j)h(2^{2j})/2^j}.$$ Using the function $s(\cdot)$, condition (iii) and the above inequality, it is also easily verified that given $\varepsilon_0$ 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 $\varepsilon$, $0 < \varepsilon < \varepsilon_0$, that $H(n) \leq 2h(n)$ 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 $\sum_{j} (j^{-4})^{1/2} < \infty$. In fact, a close reading of the proof of Theorem 2.2 [31] reveals that the function $j \rightarrow j^{-4}$ can be replaced by any function $s: j \rightarrow s(j)$ satisfying $\sum_{j} (s(j))^{1/2} < \infty$; the proof remains unchanged, save for the appearance of new constants. Let us briefly show how this may be done.

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

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

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

$$\Pr^*\left\{ \sup_{g \in \mathcal{G}_n} |S_n(g)| > (A + \beta)(HnV)^{1/2} \right\}$$

$$\leq \Pr^*\left\{ \max_{\tilde{g} \in \mathcal{G}_n} |S_n(\tilde{g})| > A(HnV)^{1/2} \right\}$$

$$+ \Pr^*\left\{ |S_n(g)| - \sup_{g \in \mathcal{G}_n} |S_n(g)|| > \beta(HnV)^{1/2} \right\}.$$ 

(3.4)

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

$$2N^{-K}_1(e, \mathcal{G}_n) + 2 \sum_{j = \ln(1/\epsilon)}^{m} N_{1}^{2}(2^{-j-1}, \mathcal{G}_n)(N_{1}(e, \mathcal{G}_n))^{-\gamma_1(j)/2},$$

(3.5)

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

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

Finally, the case $H(n)V(n)/n$ 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 $\mathcal{F}$ satisfies the BLIL, the first step is to a.s. approximate the random distances $d_{n,2}(f, g)$, $f, g \in \mathcal{F}$, 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 $N_{1}$. 


Now by considering real parts, it is clear that if $F$ satisfies the BLIL, then so does $\{\cos tx: t \in [-\frac{1}{2}, \frac{1}{2}]\}$. By stationarity and the relation
\[ \sin^2(tx) = \frac{1 - \cos 2tx}{2}, \]
the class $\mathcal{F} := \{(f - g)^2: f, g \in \mathcal{F}\}$ also satisfies the BLIL, i.e.,
\[ \limsup_{n \to \infty} \sup_{h \in \mathcal{F}} \left( \frac{n}{\log \log n} \right)^{1/2} \left| n^{-1} \sum_{i=1}^{n} (h(X_i) - Eh(X_i)) \right| \leq C \text{ a.s.}, \]
where here and elsewhere, $C$ denotes a positive, finite constant possibly changing from line to line. Thus for $n \geq n_0$ there are measurable sets $\Omega_n$ such that $\Pr(\Omega_n) \to 1$ as $n \to \infty$ and
\[ \sup_{f, g \in \mathcal{F}} \left| d_{n,2}^2(f, g) - e_p^2(f, g) \right| \leq \left( \frac{\log n}{n} \right)^{1/2} \text{ for all } \omega \in \Omega_n. \]
Thus, for all $f, g \in \mathcal{F}$ and for all $n \geq n_0$,
\[ e_p^2(f, g) \leq d_{n,2}^2(f, g) + \left( \frac{\log n}{n} \right)^{1/2} \text{ for all } \omega \in \Omega_n. \]
Therefore, for all $\varepsilon > 0$ and for all $n \geq n_0$ this implies
\[ (4.1) \quad N\left( \varepsilon^2 + \left( \frac{\log n}{n} \right)^{1/2} \right)^{1/2}, \mathcal{F}, e_p \leq N_n(\varepsilon, \mathcal{F}) \text{ for all } \omega \in \Omega_n, \]
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 $u \leq u_0$ that
\[ N\left( \varepsilon^2 + u^2 \right)^{1/2}, \mathcal{F}, e_p \leq N_{u^{-1}}(\varepsilon, \mathcal{F}) \text{ for all } \omega \in \Omega_{u^{-1}}. \]
Taking logarithms and expectations for $u_0$ small enough we obtain for all $u \leq u_0$,
\[ (4.2) \quad \log N\left( \varepsilon^2 + u^2 \right)^{1/2} \leq 2E_X^* \log N_{u^{-1}}(\varepsilon, \mathcal{F}), \]
where we have suppressed the arguments $\mathcal{F}$ and $e_p$ in $N$. Now if $F$ satisfies the BLIL, then a straightforward application of Lemma 2.9 of [10] shows that
\[ (4.3) \quad \sup_{n \geq n_0} E_X^* E_g \sup_{\mathcal{F}} \left| \frac{1}{n \log \log n} \sum_{i=1}^{n} g_i f(X_i) \right|^{1/2} \leq C. \]
Combining this with Fernique's minorization for stationary Gaussian processes [9] gives
\[ \sup_{n \geq n_0} E_X^* \int_0^1 \left( \frac{\log N_n(e, \mathcal{F})}{\log \log n} \right)^{1/2} de \leq C. \]

Letting \( u = n^{-1/5} \) and using Fubini's theorem and (4.2) in that order, (4.3) yields
\[ \sup_{0 < u \leq u_0} \int_0^1 \left( \frac{\log N((e^2 + u^2)^{1/2})}{\log \log(u^{-5})} \right)^{1/2} de \leq C. \]

Under condition (\( \ast \)), Lemma 1 of [27] shows that
\[ x^2 N_{[1]}(3x^2, \mathcal{F}) = O(N(x, \mathcal{F}, e_\rho)) \]
and thus the above inequality remains valid with \( \log N(\cdot)^{1/2} \) replaced by \( \log N_{[1]}(\cdot) \), i.e.,
\[ (4.4) \quad \sup_{0 < u \leq u_0} \int_0^1 \left( \frac{H_{[1]}(e^2 + u^2)}{\log \log(u^{-1})} \right)^{1/2} de \leq C. \]

Under the conditions of Theorem 2.3 we may actually verify that (4.4) holds if and only if
\[ (4.5) \quad H_{[1]}(x) = O(H_1(x)) \quad \text{as } x \downarrow 0. \]

To see this, let
\[ I(t, u, f) := \int_0^t (f(e^2 + u^2))^{1/2} de \]
for all functions \( f: \mathbb{R}^+ \rightarrow \mathbb{R}^+ \) and for all \( t \geq 0, u > 0 \). If (4.5) holds, then the decomposition
\[ I(t, u, H_{[1]}) = \int_0^u + \int_u^t \]
shows that
\[ I(t, u, H_{[1]}) = O\left( \left( \log \log \frac{1}{u} \right)^{1/2} \right) \]
for \( t \) small enough.
Conversely, suppose that \( H_{t_1} \) 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 \( u_i := u_i(t) \) such that for all \( u \leq u_i(t) \),

\[
\left( \log \log \frac{1}{\sqrt{2}u} \right)^{1/2} - \left( \log \log \frac{1}{\sqrt{2}t} \right)^{1/2} \geq \frac{1}{2} \left( \log \log \frac{1}{\sqrt{2}u} \right)^{1/2}.
\]

For all \( 0 < u \leq u_1 \) and all \( 0 \leq x \leq t \) set

\[
\lambda_u(x) := \left( H_{t_1}(x^2 + u^2)/H_1(x^2 + u^2) \right)^{1/2},
\]

we may assume without loss of generality that \( \lambda_u \) 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 \( 0 < u \leq u_1 \),

\[
(4.6) \quad \frac{I(t, u, H_{t_1})}{I(t, u, H_1)} = \frac{\int_0^t \lambda_u(e)(H_1(e^2 + u^2))^{1/2} de}{\int_0^t (H_1(e^2 + u^2))^{1/2} de} = \lambda_u(\xi)
\]

for some \( 0 \leq \xi \leq t \). 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 \( 0 < u \leq u_1 \), we have by definition of \( u_1 \),

\[
I(t, u, H_1) \geq \int_u^t (H_1(e^2 + u^2))^{1/2} de \\
\geq 2 \left[ \left( \log \log \frac{1}{\sqrt{2}u} \right)^{1/2} - \left( \log \log \frac{1}{\sqrt{2}t} \right)^{1/2} \right] \\
\geq \left( \log \log \frac{1}{\sqrt{2}u} \right)^{1/2},
\]

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

\[
\sup_{0 < u \leq u_1} \frac{I(t, u, H_{t_1})}{I(t, u, H_1)} \leq C.
\]

This contradiction shows that \( H_{t_1} \) 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

$\mathcal{G}_n := \{ x \to e^{itx} : |t| \leq 2^{h(n)} \}, \quad n \geq 1$

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

$$\int_0^\infty \left( \frac{\log(M_p(e^2)/e^2)}{h(\log(M_p(e^2)/e^2))} \right)^{1/2} \, de < \infty,$$

we conclude that it is enough to verify condition (2.1) with $N(e) = 16M_p(e)/e$; in fact it suffices to show that there is an $\epsilon_0$ such that for all $0 < \epsilon \leq \epsilon_0$ and all $n \geq 1$,

$$C_1 2^{h(n)} \leq N_1(\epsilon, \mathcal{G}_n) \leq 16 \cdot 2^{h(n)} M_p \left( \frac{\epsilon}{3} \right)^{\frac{\epsilon}{3}}. \quad (5.1)$$

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

By hypothesis, there is an $\epsilon_0$ and a constant $M := M(\epsilon_0) > 0$ such that for all $m > M$ we have $\int (1 - \cos mx) \, dP > 2\epsilon_0$. Let $j$ and $k$ be any reals belonging to $[-2^{h(n)}, 2^{h(n)}]$ such that $|j - k| \geq M$. Then

$$\int |e^{ikx} - e^{ijx}| \, dP \geq \int |1 - e^{ix(j-k)}| \, dP$$

$$\geq \int (1 - \cos(|j-k|x)) \, dP$$

$$\geq 2\epsilon_0.$$

Thus there are at least $2 \cdot 2^{h(n)}/M$ functions of the form $f_k(x) = e^{ikx}$ such that for all $j \neq k$, $\int |f_j - f_k| \, dP \geq 2\epsilon_0$, showing that

$$N_1(\epsilon, \mathcal{G}_n) \geq N(2\epsilon, \mathcal{G}_n, \| \cdot \|_1) \geq 2 \cdot 2^{h(n)}/M \quad \text{for all} \quad 0 < \epsilon \leq \epsilon_0.$$

To establish the second inequality in (5.1) we observe that for all $0 < \epsilon \leq \epsilon_0$,

$$N_1(\epsilon, \mathcal{G}_n) \leq 2^{h(n)} N_1(\epsilon, \mathcal{F}) \leq 16 \cdot 2^{h(n)} M_p \left( \frac{\epsilon}{3} \right)^{\frac{\epsilon}{3}},$$

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

Q.E.D.
References


Lehigh University
Bethlehem, Pennsylvania