The Annals of Probability
1984, Vol. 12, No. 4, 1041–1067

PROBABILITY INEQUALITIES FOR EMPIRICAL PROCESSES AND A LAW OF THE ITERATED LOGARITHM

BY KENNETH S. ALEXANDER[1]

Mathematical Sciences Research Institute[2]

Sharp exponential bounds for the probabilities of deviations of the supremum of a (possibly non-iid) empirical process indexed by a class $\mathscr{F}$ of functions are proved under several kinds of conditions on $\mathscr{F}$. These bounds are used to establish laws of the iterated logarithm for this supremum and to obtain rates of convergence in total variation for empirical processes on the integers.

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of independent random variables taking values in a space $(X, \mathscr{A})$, and $P_{(i)} := \mathscr{L}(X_i)$. We construct the nth empirical measure and process:

$$P_n := n^{-1} \sum_{i=1}^n \delta_{X_i}$$

$$\nu_n := n^{1/2}(P_n - \bar{P}_{(n)})$$

where $\bar{P}_{(n)} := n^{-1} \sum_{i=1}^n P_{(i)}$. Given a measure P and function f on X, let $P(f)$ denote $\int f\, dP$; thus

$$\nu_n(f) = n^{-1/2} \sum_{i=1}^n (f(X_i) - Ef(X_i)).$$

Given a class $\mathscr{F}$ of functions on X, we can view ν_n as a stochastic process indexed by $\mathscr{F}$, and consider limit theorems for this process. To prove such theorems it is often helpful to have bounds on the tail of the r.v. $\sup_{\mathscr{F}} |\nu_n(f)|$ (see for example Dudley, 1978, Kuelbs and Dudley, 1980, Dudley and Phillip, 1983, or Giné and Zinn, 1983.) Our main question of interest is, for what classes $\mathscr{F}$ can "best possible" bounds be obtained?

So first we must ask, what are these best possible bounds? Of course

$$\mathbb{P}[\sup_{\mathscr{F}} |\nu_n(f)| > M] \geq \sup_{\mathscr{F}} \mathbb{P}[|\nu_n(f)| > M] \tag{1.1}$$

so we would like a bound on the left side of (1.1) which is not too far from the best known bounds on the right side of (1.1). Restricting our attention henceforth to uniformly bounded $\mathscr{F}$, specifically $0 \leq f \leq 1$ for each $f \in \mathscr{F}$, we have three

Received July 1983; revised March 1984.

[1] This research is based in part on the author's doctoral dissertation submitted to the Massachusetts Institute of Technology, Cambridge, Massachusetts.

[2] Author now at University of Washington.

AMS 1980 *subject classifications.* Primary 60F10; secondary 60F15, 60G57.

Key words and phrases. Empirical process, exponential bound, law of the iterated logarithm, Vapnik-Červonenkis class, metric entropy.

inequalities from Hoeffding (1963):

(1.2) $$\mathbb{P}[|\nu_n(f)| > M] \le 2\exp(-\psi(M, n, \alpha))$$

where $\alpha = \operatorname{var}(\nu_n(f))$ and ψ is either

$$\psi_1(M, n, \alpha) := Mn^{1/2}h_1(M/n^{1/2}\alpha),$$

$$\psi_2(M, n, \alpha) := Mn^{1/2}h_2\,(M/n^{1/2}\alpha) \quad \text{(Bernstein's inequality)}$$

or

$$\psi_3(M, n, \alpha) := 2M^2.$$

Here

$$h_1(\lambda) := (1 + \lambda^{-1})\log(1 + \lambda) - 1$$

and

$$h_2(\lambda) := \frac{\lambda}{2(1 + \lambda/3)}.$$

We therefore try to bound $\mathbb{P}[\sup_{\mathscr{F}}|\nu_n(f)| > M]$ by something close to the right side of (1.2).

A common example is given by

$$\mathscr{F} = \{1_{(-\infty,x]}: x \in \mathbb{R}^d\}, \text{ with the } X_i \text{ iid } (P).$$

Here P_n is the nth empirical d.f., and Kiefer (1961) proved that for each $\varepsilon > 0$ there exists a constant $R = R(\varepsilon, d)$ such that

(1.3) $$\mathbb{P}[\sup_{\mathscr{F}}|\nu_n(f)| > M] \le 2\exp(-(2 - \varepsilon)M^2) \quad \text{for all} \quad M > R,\ n \ge 1.$$

The exponent in (1.3) is within a factor of ε of the exponent in (1.2) (with $\psi = \psi_3$). If $P((-\infty, x]) = 1/2$ then the bound in (1.2) with $\psi = \psi_3$ is sharp for $f = 1_{(-\infty,x]}$; in this sense (1.3) is best possible.

For general $\mathscr{F}$ we may thus ask whether, given an inequality of form (1.2) for some ψ, we have

(1.4) $$\mathbb{P}[\sup_{\mathscr{F}}|\nu_n(f)| > M] \le K\exp(-(1 - \varepsilon)\psi(M, n, \alpha))$$

for some numerical constant K, and M, n sufficiently large, where

$$\alpha = \sup_{\mathscr{F}}\operatorname{var}(\nu_n(f)) = \sup_{\mathscr{F}} n^{-1}\textstyle\sum_{i=1}^n \operatorname{var} f(X_i).$$

We will show that (1.4) holds for quite general $\mathscr{F}$. When X is a bounded subset of $\mathbb{R}^d$, the functions in $\mathscr{F}$ need only be sufficiently smooth. When $\mathscr{F} = \{1_C: C \in \mathscr{C}\}$ for a class $\mathscr{C}$ of subsets of X, it is enough that $\mathscr{C}$ satisfy a certain combinatorial condition, or that $\mathscr{C}$ consist of convex sets or sets with smooth enough boundaries.

Note that $h_2(\lambda) \ge (1 - \varepsilon)\lambda/2$ if $\lambda \le 3\varepsilon$, so

(1.5) $$\psi_2(M, n, \alpha) \ge ((1 - \varepsilon)M^2/2\alpha) \quad \text{if} \quad M \le 3\varepsilon\alpha n^{1/2}.$$

Thus for moderate M, (1.5) says that $\sup_{\mathscr{F}}|\nu_n(f)|$ has a Gaussian-like tail. This fact can be exploited to prove a law of the iterated logarithm (LIL) for $1\sup_{\mathscr{F}}|\nu_n(f)|$.

When $\alpha = 1/4$, its largest possible value since $0 \le f \le 1$, ψ_3 gives a better bound than ψ_1 or ψ_2. By the lower exponential bounds in Stout (1974, Theorem 5.2.2), ψ_3 is essentially sharp in this case. If $\alpha < 1/4$, ψ_1 or ψ_2 usually gives a better bound than ψ_3: we have

$$h_1(\lambda) \ge \begin{cases} \lambda/4 & \text{if } \lambda \le 4 \\ \tfrac{1}{2} \log \lambda & \text{if } \lambda \ge 4 \end{cases}$$

so

$$\psi_1(M, n, \alpha) \ge \begin{cases} M^2/4\alpha & \text{if } M \le 4n^{1/2}\alpha \\ \tfrac{1}{2} Mn^{1/2}L(M/n^{1/2}\alpha) & \text{if } M \ge 4n^{1/2}\alpha \end{cases} \tag{1.6}$$

where Lx denotes $\log \max(x, e)$. By (1.5) and the above mentioned lower exponential bound, ψ_2 is essentially sharp if $M/n^{1/2}\alpha$ is small. ψ_1 is always better than ψ_2, but ψ_2 is often easier to work with.

But we need not assume that ψ in (1.4) is ψ_1, ψ_2 or ψ_3. We require only that ψ satisfy (1.2) for all choices of a function $0 \le f \le 1$ and r.v.'s $X_1, \cdots, X_n$ with $n^{-1}\sum_{i=1}^n \operatorname{var} f(X_i) \le \alpha$, and all $M > 0$ and $n \ge 1$, and that ψ satisfy

$$\psi(M, n, \alpha) \ge \psi(\theta M, n, \rho\alpha) \ge \theta^2\rho^{-1}\psi(M, n, \alpha) \quad \text{for all } \theta \le 1, \ \rho \ge 1, \ M > 0, \ n \ge 1, \ \alpha > 0. \tag{1.7}$$

Let Ψ be the class of all such ψ; it is easy to verify that $\psi_1, \psi_2, \psi_3 \in \Psi$. The homogeneity condition (1.7) is really only assumed for convenience, but it is natural in view of the Gaussian tail exponent $M^2/2\alpha$ (see (1.5) above.)

When $\mathscr{F} = \{1_C : C \in \mathscr{C}\}$, we will use C and 1_C, and $\mathscr{F}$ and $\mathscr{C}$, interchangeably. In this case one natural condition to impose on $\mathscr{C}$ turns out to be a combinatorial one. For $x_1, \cdots, x_n \in X$, let

$$\Delta^{\mathscr{C}}(x_1, \cdots, x_n) := \operatorname{card}\{C \cap \{x_1, \cdots, x_n\} : C \in \mathscr{C}\} \le 2^n$$

and

$$m^{\mathscr{C}}(n) := \sup\{\Delta^{\mathscr{C}}(x_1, \cdots, x_n) : x_1, \cdots, x_n \in X\}.$$

If $m^{\mathscr{C}}(n) < 2^n$ for some $n \ge 1$, then $\mathscr{C}$ is called a *Vapnik-Červonenkis* (or VC) class. The least $n \ge 1$ for which $m^{\mathscr{C}}(n) < 2^n$ is called the *index* of $\mathscr{C}$ and denoted $V(\mathscr{C})$. Then

$$m^{\mathscr{C}}(n) \le \sum_{j=0}^{V(\mathscr{C})-1} \binom{n}{j} \le \left(\frac{ne}{V(\mathscr{C}) - 1}\right)^{V(\mathscr{C})-1} \tag{1.8}$$

if $n \ge V(\mathscr{C}) - 1$ (Vapnik and Červonenkis, 1971). Many classes of interest in applications are VC classes. Dudley (1978) showed that for $\mathscr{M}$ a finite-dimensional vector space of real-valued functions on X, $\mathscr{C} := \{\{x : f(x) > 0\} : f \in \mathscr{M}\}$ is a VC class. He also showed that if $\mathscr{C}$ is a VC class and $k \ge 1$, then the class of all sets which are in an algebra generated by some k-element subset of $\mathscr{C}$ is a VC class. Since any subset of a VC class is a VC class, it follows in particular that the classes of all rectangles, all ellipsoids, and all half spaces in $\mathbb{R}^d$ are each VC classes. Devroye (1982), refining an earlier result of Vapnik and Červonenkis (1971), showed that for $\mathscr{C}$ a VC class,

$$\mathbb{P}[\sup_{\mathscr{C}} |\nu_n(C)| > M] \le 4e^8\left(\frac{n^2e}{v}\right)^v \exp(-2M^2) \tag{1.9}$$

for all $M > 0$, $n \ge 1$, if $v \ge V(\mathscr{C})$. But this is clearly unsatisfactory if $n \to \infty$ and M remains fixed; we will prove an inequality which eliminates the dependence of the right side of (1.9) on n.

Our results will also hold for those classes $\mathscr{F}$ of functions for which the collection $\mathscr{C} = \{\{(x, t): 0 \le t \le f(x)\}: f \in \mathscr{F}\}$ of regions under the graphs of functions in $\mathscr{F}$ is a VC class of sets. In this case we will call $\mathscr{F}$ a *VC graph class* and $\mathscr{C}$ the *graph region class.* Clearly $\{1_C: C \in \mathscr{C}\}$ is a VC graph class if $\mathscr{C}$ is a VC class. Pollard (1982) studied such classes and used bounds on $\sup_{\mathscr{F}} |\nu_n(f)|$ to obtain results on convergence rates of kernel density estimators.

2. Statement of results.

The key to most of our results is the control of some form of the *metric entropy* of the class $\mathscr{F}$. Given $\varepsilon > 0$, $p > 0$, and a law P on $(X, \mathscr{A})$, set

$$N_p(\varepsilon, \mathscr{F}, P) := \min\{k: \text{there exist } f_1, \cdots, f_K \in \mathscr{F} \text{ such that } \min_{i \le k} \|f - f_i\|_p < \varepsilon \text{ for all } f \in \mathscr{F}\}$$

$$N_p^B(\varepsilon, \mathscr{F}, P) := \min\{k: \text{There exist } f_1^U, f_1^L, \cdots, f_k^U, f_k^L \in \mathscr{F} \text{ such that } f_i^L \le f \le f_i^U \text{ for some } i \text{ for each } f \in \mathscr{F}, \text{ and } \|f_i^U - f_i^L\|_p < \varepsilon\}.$$

The functions

$$H_p := \log N_p, \quad H_p^B := \log N_p^B$$

are called the *metric entropy* and *metric entropy with bracketing* of $\mathscr{F}$ in L^p. When no confusion is possible, we will suppress the $\mathscr{F}$ and P in the entropy notation. When $\mathscr{F}$ consists of sets and $p = 2$, the entropy with bracketing is essentially the entropy with inclusion used by Dudley (1978) and others, though Dudley does not require $f_i^U, f_i^L \in \mathscr{F}$. Let

$$I_p(s, t, \mathscr{F}, P) := \int_s^t H_p(u, \mathscr{F}, P)^{1/2}\, du$$

$$I_p^B(s, t, \mathscr{F}, P) := \int_s^t H_p^B(u, \mathscr{F}, P)^{1/2}\, du.$$

The only facts about the entropy of $\mathscr{F}$ which we will use are upper bounds on the values of I and H^{-1} at specific points. Since H may be replaced by a larger continuous function at the expense of an arbitrarily small increase in these values, we may assume that

(2.1) $H(\cdot, \mathscr{F}, P)$ is continuous and strictly decreasing from ∞ to 0 on $(0, a]$ for some $a > 0$.

This will allow us, given $b > 0$, to define t by $H(t, \mathscr{F}, P) = b$ or $H(t, \mathscr{F}, P) = bt^2$.

For many classes of interest the entropy satisfies a bound of the form

$$H(u, \mathscr{F}, P) \leq Au^{-r}$$

for some $A, r > 0$, and $H = H_p$ or H_p^B. Some examples:

1. Let $\mathscr{F}_{\beta,K}([0, 1]^d)$ be the class of all functions f from $[0, 1]^d$ to $\mathbb{R}$ which are *smooth to order* β with bound K, i.e. have all partial derivatives of order $\leq \beta$ and satisfy

$$\max_{|p|\leq m}\sup\{|D^p f(x)|: x \in [0, 1]^d\}$$
$$+ \max_{|p|=m}\sup\left\{\frac{|D^p f(x) - D^p f(y)|}{|x - y|^\gamma}: x \neq y \in [0, 1]^d\right\} \leq K$$

where $m = [\beta]$ (the integer part), $\gamma = \beta - [\beta]$, and $|p| = p_1 + \cdots + p_d$, $p \in (\mathbb{Z}^+)^d$. Then

$$H_\infty(u, \mathscr{F}_{\beta,K}([0, 1]^d)) \leq Au^{-d/\beta}$$

for some constant $A = A(\beta, K, d)$ (Kolmogorov and Tihomirov, 1959).

2. Let $\mathscr{C}_{\beta,k,d}$ be the class of all regions in $\mathbb{R}^d$ cut out by images of those functions on the sphere S^{d-1} which are smooth to order β with bound K. (More precisely, since "cut out" is a little vague, we'll say $\mathscr{C}_{\beta,k,d}$ is the class $J(d, \beta, k)$ of Dudley, 1978.) Then, by the proof of Theorem 5.12 of Dudley (1978), if $d \geq 2$, $\beta > d - 1$, and P has a bounded density with respect to Lebesgue measure, then

$$H_2^B(u, \mathscr{F}, P) \leq Au^{-2(d-1)/\beta}.$$

3. For the class $\mathscr{C}_d$ of all convex subsets of $[0, 1]^d$, $d \geq 2$, if P has a density bounded by $K < \infty$, then

$$H_2^B(u, \mathscr{C}_d, P) \leq Au^{-(d-1)} \tag{2.2}$$

for some constant $A = A(d, K)$; this follows from a theorem of Bronštein (1976).

4. If $X = \{x_i,\ i \geq 1\}$ (i.e. X is countable) and $(p_m)_{m\geq1}$ is $(P(\{x_i\}))_{i\geq1}$ in nonincreasing order, then

$$I_2^B(0, 1, 2^X, P) < \infty \quad \text{if and only if} \quad \textstyle\sum_{m\geq1} p_m^{1/2} < \infty$$

by Theorem 6.3.1 of Dudley (1983); by modifying his proof it is easy to verify that

$$H_2^B(u, 2^X, P) \le Ar^{-1}(u^2)$$

where $r_m = \sum_{j \ge m} p_j$ and $r^{-1}(u) = \min\{j: r_j \le u\}$. In particular, if $r_m = 0(m^{-\beta})$ for some $\beta > 0$, then

$$H_2^B(u, 2^X, P) \le Au^{2/\beta}. \tag{2.3}$$

The important fact for our purposes is that $I(0, 1, \mathscr{F}, P) < \infty$ if $H(u, \mathscr{F}, P) \le Au^{-r}$ for some $r < 2$.

We begin with results under assumptions on H_∞.

THEOREM 2.1. *Let $\mathscr{F}$ be a class of functions $0 \le f \le 1$ on $(X, \mathscr{A})$. Let $\psi \in \Psi$, $n \ge 1$, $\varepsilon > 0$, and $\alpha \ge \sup_{\mathscr{F}} n^{-1} \sum_{i=1}^n \operatorname{var} f(X_i)$. Define t_0 by $H_\infty(t_0, \mathscr{F}) = \frac{1}{4} \varepsilon \psi(M, n, \alpha)$. If*

$$M > 2^8 \varepsilon^{-3/2} I_\infty \left(\frac{\varepsilon M}{64 n^{1/2}}, t_0 \right) \tag{2.4}$$

then

$$\mathbb{P}^*[\sup_{\mathscr{F}} |\nu_n(f)| > M] \le 5 \exp(-(1 - \varepsilon)\psi(M, n, \alpha)). \tag{2.5}$$

The unwieldy condition (2.4) translates into a natural one when the entropy grows polynomially fast at 0, as is reflected in the following corollary.

COROLLARY 2.2. *Let $\mathscr{F}$, ψ, n, ε, and α be as in Theorem 2.1. There exist constants $K_i = K_i(r, \varepsilon, A)$ such that if $r > 0$,*

$$H_\infty(u, \mathscr{F}) \le Au^{-r} \quad \text{for all} \quad u > 0, \tag{2.6}$$

and

$$M \ge \begin{cases} K_1 \alpha^{(2-r)/4} & \text{if } r < 2 \\ K_2 Ln & \text{if } r = 2 \\ K_3 n^{(r-2)/2(r+2)} & \text{(required for all } r) \end{cases} \tag{2.7}$$

then (2.5) *holds.*

We have not attempted to obtain best numerical constants in the above and following results; techniques which depend on the metric entropy, which is usually known only up to an asymptotic rate, do not lend themselves to this. Our results are intended for asymptotic use.

THEOREM 2.3. *Let $\mathscr{F}$ be a class $\mathscr{C}$ of measurable subsets of X. Let $\psi \in \Psi$, $n \ge 1$, $\varepsilon > 0$, and $\alpha \ge \sup_{\mathscr{C}} n^{-1} \sum_{i=1}^n P_{(i)}(C)(1 - P_{(i)}(C))$. Define t_0 by $H_2^B(t_0, \mathscr{C}, \bar{P}_{(n)}) = \frac{1}{4} \varepsilon \psi(M, n, \alpha)$. If*

$$M \ge 2^9 \varepsilon^{-3/2} I_2^B((\varepsilon M / 64 n^{1/2})^{1/2}, t_0, \mathscr{C}, \bar{P}_{(n)}) \tag{2.8}$$

and

$$H_2^B((\varepsilon M/16n^{1/2})^{1/2}, \mathscr{C}, \bar{P}_{(n)}) \le \varepsilon^2 M n^{1/2}/4 \tag{2.9}$$

then

$$\mathbb{P}^*[\sup_{\mathscr{C}} |\nu_n(C)| > M] \le 5 \exp(-(1-\varepsilon)\psi(M, n, \alpha)). \tag{2.10}$$

Again these conditions become natural when the entropy grows polynomially fast at 0.

COROLLARY 2.4. *Let* $\mathscr{C}$, ψ, n, α, *and* ε *be as in Theorem* 2.3. *There exist constants* $K_i = K_i(r, \varepsilon, A)$ *such that if* $r > 0$,

$$H_2^B(u, \mathscr{C}, \bar{P}_{(n)}) \le Au^{-r} \quad \textit{for all} \quad u > 0 \tag{2.11}$$

and (2.7) *holds, then* (2.10) *holds.*

From Borel-Cantelli and Corollaries 2.2 and 2.4 we get the following improvement of a result of Dudley (1982).

COROLLARY 2.5. *Let* $\mathscr{F}$ (*or* $\mathscr{C}$) *be a class of measurable functions* (*or sets*) *satisfying* (2.6) (*or* (2.11)) *for all n and some* $A > 0$, $r \ge 2$. *Then*

$$\sup_{\mathscr{F}} |\nu_n(f)| = \begin{cases} O(n^{(r-2)/2(r+2)}) & a.s. \quad \textit{if} \quad r > 2 \\ O(Ln) & a.s. \quad \textit{if} \quad r = 2. \end{cases}$$

Corollary 2.5 is sharp for classes of sets, i.e. there exist classes for each $r > 2$ satisfying (2.11) with

$$\lim \sup_n n^{-(r-2)/2(r+2)} \sup_{\mathscr{F}} |\nu_n(f)| > 0 \quad \text{a.s.}$$

and classes with $r = 2$ for which

$$\lim \sup_n (Ln)^{-1} \sup_{\mathscr{F}} |\nu_n(f)| > 0 \quad \text{a.s.}$$

This follows from Remark 2.15 below, or (for $r > 2$) from a result of Dudley (1982).

The following special case of Corollary 2.5 refines a result of Stute (1977) and follows from (2.2). Stute shows it is sharp up to a possible power of Ln.

COROLLARY 2.6 *Let* $X = [0, 1]^d$ *and suppose* $\bar{P}_{(n)}$ *has a density* $\varphi_n(x)$ *uniformly bounded in n and x. Let* $\mathscr{C}$ *be the class of all convex subsets of* $[0, 1]^d$. *Then*

$$\sup_{\mathscr{C}} |\nu_n(C)| = \begin{cases} O(n^{(d-3)/2(d+1)}) & a.s. \quad \textit{if} \quad d > 3 \\ O(Ln) & a.s. \quad \textit{if} \quad d = 3. \end{cases}$$

We turn next to VC graph classes of functions. The key here is the next lemma, which is a refinement of one from Pollard (1982) which is based in turn on one from Dudley (1978).

LEMMA 2.7. *If $\mathscr{F}$ is a* VC *graph class of functions $0 \le f \le 1$ on X with graph region class $\mathscr{G}$, $V(\mathscr{G}) \le v$, and μ is Lebesgue measure on* [0, 1], *then*

$$N_2(u, \mathscr{F}, P) \le N_1(u^2, \mathscr{F}, P) = N_1(u^2, \mathscr{G}, P \times \mu)$$
$$\le (16u^{-2}L(8u^{-2}))^{v-1} \le (8u^{-2})^{2v}$$

for all $0 < u \le 1$ and all laws P on X.

It is the uniformity of this bound in P that makes VC classes tractable.

Vapnik and Červonenkis (1981) showed, under some measurability conditions, that in the iid case with $\mathscr{L}(X_1) = P$, one has for $\mathscr{F}$ uniformly bounded,

$$\lim_n \sup_{\mathscr{F}} | P_n(f) - P(f) | = 0 \quad \text{a.s.}$$

if and only if

$$EH_1(u, \mathscr{F}, P_n) = o(n) \quad \text{for each } u > 0.$$

(Actually, Vapnik and Červonenkis stated this for H_∞ in place of H_1, but their proof shows the condition on H_1 is sufficient. In fact, it doesn't matter what $H_p (p \ge 1)$ is used—see Giné and Zinn, 1983.) Because of this it is natural to seek inequalities like (1.2) under assumptions on the distribution of $H_1(u, \mathscr{F}, P_n)$. We will show that for $\mathscr{F}$ a VC graph class the appropriate assumptions are easily satisfied.

In the course of our proofs we will need to use a special construction of P_n. Let $\{\xi_{ij}, j \ge 1\}$ be independent copies of X_i for each $i \ge 1$. Fix $m < \infty$, to be specified later, and let $\tau(i)$, $i \ge 1$, be independent r.v.'s uniformly distributed on $\{1, \cdots, m\}$ and independent of all the ξ_{ij}. We may then assume

$$X_i = \xi_{i\tau(i)} \tag{2.12}$$

and define

$$\xi := (\xi_{ij})_{i,j \ge 1}$$

$$P_{n\times m} := (nm)^{-1} \textstyle\sum_{i=1}^n \sum_{j=1}^m \delta_{\xi_{ij}}$$

$$\tilde{H}_1(m, n, \varepsilon, \xi, \mathscr{F}) := H_1(\varepsilon, \mathscr{F}, P_{n\times m}).$$

In this context, given $\mathscr{F}$, $\psi \in \Psi$, and n, M, α, ε, $b > 0$ we say (A) *holds* for $(\mathscr{F}, \psi, M, n, \alpha, \varepsilon, b)$ if there exists an integer

$$m \ge \max\left(1 + \frac{2^{11}\alpha}{\varepsilon^2 M^2}, \frac{16}{\varepsilon}, \frac{2^{11}\psi(M, n, \alpha)}{\varepsilon^2 \alpha n}, \frac{2^{12}}{\varepsilon^2 \alpha n}, \frac{8M}{\varepsilon \alpha n^{1/2}}\right) \tag{2.13}$$

such that

$$\mathbb{P}_*\left[\tilde{H}_1\left(m - 1, n, \frac{\varepsilon M}{64 m n^{1/2}}, \xi, \mathscr{F}\right) > \frac{\varepsilon \psi(M, n, \alpha)}{16}\right] \le 1/4 \tag{2.14}$$

and

$$\mathbb{P}^*\left[\tilde{H}_1\left(1, n, \frac{\varepsilon M}{64n^{1/2}}, \xi, \mathscr{F}\right) > \frac{\varepsilon\psi(M, n, \alpha)}{16}\right] \le e^{-b} \tag{2.15}$$

and if $\alpha < 1/4$ then for all $k \ge 1$,

$$\mathbb{P}_*[\tilde{H}_1(2^{k-1}m, n, 2^{k-8}\varepsilon\alpha, \xi, \mathscr{F}) > 2^{2k-15}\varepsilon^2\alpha mn - (k-1)\log 2] \le 1/4. \tag{2.16}$$

The use of inner and outer measures $\mathbb{P}_*$ and $\mathbb{P}^*$ above is required because $\tilde{H}_1$ need not be measurable in general. Because the same holds for $\sup_{\mathscr{F}}|\nu_n(f)|$, some measurability assumptions will be needed for our next theorems. To this end, we assume throughout that the ξ_{ij} are coordinate functions on the product spaces X^∞. We say $\mathscr{F}$ is *n-supremum measurable* (for $(P_{(i)})_{i\le n}$) if $\sup_{\mathscr{F}} Q(f, \xi)$ is measurable for each function $Q(f, \xi)$ which is a linear or quadratic function of finitely many of the $f(\xi_{ij})$, $i \le n$, $j \ge 1$. We say $\mathscr{F}$ is *n-deviation measurable* if both $\mathscr{F}$ and $\{f - g: f, g \in \mathscr{F}, \operatorname{var} \nu_n(f-g) \le \theta\}$ are n-supremum measurable for all $\theta > 0$. We omit the "n-" when either condition is valid for all $n \ge 1$.

We say a class $\mathscr{C}$ in $(X, \mathscr{A})$ is *(v, k)-constructible* if there exists a VC class $\mathscr{D}$ with $V(\mathscr{D}) \le v$ and a function φ: $\mathscr{D}^k \to \mathscr{A}$, constructed from the basic set operations $\cap$, $\cup$, and ^{c}, such that $\mathscr{C} \subset \varphi(\mathscr{D}^k)$. Then $\mathscr{C}$ is a VC class and

$$N_1(u, \mathscr{C}, P) \le N_1(u/k, \mathscr{D}, P)^k$$

for all laws P and $u > 0$. Of course any VC class is $(V(\mathscr{C}), 1)$-constructible. By Lemma 2.7, if $\mathscr{F}$ is a VC graph class with (v, k)-constructible graph region class $\mathscr{C}$, then

$$\begin{aligned} \tilde{H}_1(m, n, u, \mathscr{F}_1, \xi) &\le 2kvL\left(\frac{8k^2}{u^2}\right) \\ H_2(u, \mathscr{F}, \bar{P}_{(n)}) &\le 2kvL\left(\frac{8k^2}{u^2}\right) \end{aligned} \tag{2.17}$$

for all m, n, ξ and all $u > 0$, a fact which will help us establish (A).

EXAMPLE. The class $\mathscr{C}$ of all at-most-k-sided polygons (not necessarily convex) in $\mathbb{R}^2$ is $(4, 3k\text{-}6)$-constructible. Take $\mathscr{D}$ to be all half planes, so $V(\mathscr{D}) = 4$, and note that each set in $\mathscr{C}$ is a union of at most $k - 2$ triangles. Thus we can use

$$\varphi(D_1, \cdots, D_{3k-6}) = (D_1 \cap D_2 \cap D_3) \cup \cdots \cup (D_{3k-8} \cap D_{3k-7} \cap D_{3k-6}).$$

It will suit our purpose to use a slight variant of the entropy N_2. Let $\mathscr{F}_c$ be the class of all constant functions on X and define

$$\begin{aligned} N_2^*(u, \mathscr{F}, (P_{(i)})_{i\le n}) := \min \{k:\ &\text{there exist } f_1, \cdots, f_k \in \mathscr{F} \cup \mathscr{F}_c \text{ such that} \\ &\min_{j\le k} n^{-1} \textstyle\sum_{i=1}^n \operatorname{var}((f - f_j)(X_i)) < u^2 \\ &\text{for all } f \in \mathscr{F}\}. \end{aligned}$$

Then

(2.18) $$\log N_2^*(u, \mathscr{F}, (P_{(i)})_{i\le n}) = 0 \text{ if } u^2 > \sup_{\mathscr{F}} n^{-1} \sum_{i=1}^n \operatorname{var} f(X_i)$$

and

(2.19) $$N_2^*(u, \mathscr{F}, (P_{(i)})_{i\le n}) \le N_2(u, \mathscr{F}, \bar{P}_{(n)}),$$

since $n^{-1} \sum_{i=1}^n \operatorname{var}((f_j - f)(X_i)) \le \|f_j - f\|_2$. Let

$$H_2^* := \log N_2^*$$

$$I_2^*(s, t, \mathscr{F}, (P_{(i)})_{i\le n}) := \int_s^t H_2^*(x, \mathscr{F}, (P_{(i)})_{i\le n})^{1/2}\, dx.$$

THEOREM 2.8. *Let $n \ge 1$ and let $\mathscr{F}$ be an n-deviation measurable* VC *graph class of functions $0 \le f \le 1$ on $(X, \mathscr{A})$ with (v, k)-constructible graph region class $\mathscr{C}$. Let $\psi \in \Psi$, $\varepsilon > 0$, and $\alpha \ge \sup_{\mathscr{F}} n^{-1} \sum_{i=1}^n \operatorname{var} f(X_i)$. If*

(2.20) $$M \ge \alpha^{1/2}$$

and either (i)

$$M > \frac{64kv}{n^{1/2}\varepsilon}$$

and

(2.21) $$\psi(M, n, \alpha) \ge 2^9 kv\varepsilon^{-1} L\left(\frac{n}{v} \vee \frac{k}{\alpha}\right)$$

or (ii)

(2.22) $$M > 2^{14}\varepsilon^{-5/2} kvn^{-1/2} L\left(\frac{\varepsilon n}{v}\right)$$

and

(2.23) $$M > 2^{12}\varepsilon^{-3/2}(kv\alpha L(k/\alpha))^{1/2},$$

then

(2.24) $$\mathbb{P}^*[\sup_{\mathscr{F}} |\nu_n(f)| > M] \le 16 \exp(-(1 - \varepsilon)\psi(M, n, \alpha)).$$

In the examples (2) – (4) above Theorem 2.1, we have to make assumptions on $\bar{P}_{(n)}$ (to control the entropy) to obtain the exponential bound of Corollary 2.4. Note that there are no such assumptions for VC graph classes.

The condition (2.23) is principally important in the way it relates M to the maximum variance α. For fixed k, v, and ε, it says we need $M \ge R(\alpha L\alpha^{-1})^{1/2}$ for some constant $R(k, v, \varepsilon)$. This is sharp in that if $q(\alpha) := (\alpha L\alpha^{-1})^{1/2}$ were replaced with any function $g(\alpha) = o(q(\alpha))$ as $\alpha \to 0$, the theorem would be false. To see this, let $\mathscr{C}(\alpha)$ denote the class of all subintervals of $[0, 1]$ of length less than α, and take $P_{(i)}$ to be the uniform law on $[0, 1]$ for all i. Fix $R > 0$ and $\delta > 0$, then choose $\alpha_0 > 0$ small enough so $\delta q(\alpha) > Rg(\alpha)$ for all $\alpha < \alpha_0$. If (2.23) could be replaced by "$M > Rg(\alpha)$" then for each $\alpha < \alpha_0$ we would have by Theorem 2.8 and (1.6):

$$\lim_n \sup \mathbb{P}[\sup_{\mathscr{C}(\alpha)} |\nu_n(C)| > \delta q(\alpha)] \le 16 \exp\left(-\frac{\delta^2 q^2(\alpha)}{4\alpha}\right). \tag{2.25}$$

Since, by Donsker's well-known invariance principle, the process $\nu_n([0, t])$ converges in distribution as $n \to \infty$ to the Brownian bridge $W_0(t)$, (2.25) implies that

$$\mathbb{P}[\sup_{|s-t|<\alpha} |W_0(s) - W_0(t)| > \delta q(\alpha)] \le 16 \exp(-\tfrac{1}{4}\delta^2 L\alpha^{-1}). \tag{2.26}$$

Set $\alpha_k := k^{-8/\delta^2}$. Then for large enough m, we have been using (2.26):

$$\begin{aligned}
&\mathbb{P}\left[\sup_{|s-t|<\alpha_m} \frac{|W_0(s) - W_0(t)|}{q(|s-t|)} > 2\delta\right] \\
&\quad\le \textstyle\sum_{k\ge m} \mathbb{P}[\sup_{|s-t|<\alpha_k} |W_0(s) - W_0(t)| > 2\delta q(\alpha_{k+1})] \\
&\quad\le \textstyle\sum_{k\ge m} \mathbb{P}[\sup_{|s-t|<\alpha_k} |W_0(s) - W_0(t)| > \delta q(\alpha_k)] \\
&\quad\le \textstyle\sum_{k\ge m} 16 \exp(-\tfrac{1}{4}\delta^2 L\alpha_k^{-1}) \\
&\quad= 16 \textstyle\sum_{k\ge m} k^{-2} \to 0 \quad \text{as} \quad m \to \infty \quad \text{for all} \quad \delta
\end{aligned}$$

contradicting a theorem of Lévy (1954) which tells us that

$$\lim_{\alpha\to 0} \sup_{|s-t|<\alpha} \frac{|W_0(s) - W_0(t)|}{q(|s-t|)} = 2^{1/2} \quad \text{a.s.}$$

A special case of Theorem 2.8 is:

COROLLARY 2.9. *Let $n \ge 1$ and let $\mathscr{C}$ be an n-deviation measurable* VC *class. Let $\varepsilon > 0$, $v \ge V(\mathscr{C})$, and $\alpha \ge \sup_{\mathscr{C}} n^{-1} \sum_{i=1}^n P_{(i)}(C)(1 - P_{(i)}(C))$. If*

$$2^{13}\varepsilon^{-3/2}(v\alpha L(1/\alpha))^{1/2} < M < \varepsilon\alpha n^{1/2}$$

then

$$\mathbb{P}^*[\sup_{\mathscr{C}} |\nu_n(C)| > M] \le 16 \exp\left(-\frac{(1-\varepsilon)M^2}{2\alpha}\right).$$

COROLLARY 2.10. *Let $\mathscr{F}$, v, k, ψ, ε, and α be as in Theorem* 2.8 *and suppose $\mathscr{F}$ is deviation measurable. There exists a constant $K = K(\alpha, v, k, \varepsilon)$ such that*

$$\mathbb{P}^*[\sup_{\mathscr{F}} |\nu_n(f)| > M] \le K \exp(-(1-\varepsilon)\psi(M, n, \alpha)) \quad \textit{for all} \quad M > 0, \quad n \ge 1. \tag{2.27}$$

Corollary 2.10 is immediate from Theorem 2.8. K need only be chosen large enough so the right side of (2.27) is ≥ 1 if (2.20) – (2.23) are not satisfied. We state it to match the original form of Kiefer's (1961) inequality for d.f.'s.

By taking $\psi(M, n, \alpha) = 2M^2$ and $\varepsilon = 2^{11}vM^{-2}LM$ and using a variant of the proof of Theorem 2.8, we obtain the following.

THEOREM 2.11. *Let $n \geq 1$ and let $\mathscr{F}$ be an n-deviation measurable* VC *graph class of functions with graph region class $\mathscr{C}$. Then for all $M \geq 8$,*

$$\mathbb{P}[\sup_{\mathscr{F}} |\nu_n(f)| > M] \leq 16M^{2^{12}V(\mathscr{C})}\exp(-2M^2).$$

Thus ε in the exponent may be replaced by a polynomial in M in front of the exponential.

When (1.4) is valid, with $\psi = \psi_2$, for the proper choices of constants M, α, and ε, using (1.5) and the techniques of Stout (1974, Theorem 5.2.3) it is not hard to show that $\sup_{\mathscr{F}} |\nu_n(f)|$ satisfies an LIL. Specifically, for each $\varepsilon > 0$ we need (1.4) with $\alpha = \alpha_n := \sup_{\mathscr{F}} n^{-1} \sum_{i=1}^{n} \operatorname{var} f(X_i)$ and $M = (1 + 2\varepsilon)(2\alpha_n LLn)^{1/2}$ for all sufficiently large n, and to take advantage of (1.5) we need $n^{-1}LLn = o(\alpha_n)$. Thus from Corollaries 2.2 and 2.4 and Theorem 2.8 we get the following theorems.

THEOREM 2.12. *Let $\mathscr{F}$ be a deviation-measurable class of functions $0 \leq f \leq 1$ on $(X, \mathscr{A})$ such that for some $r < 2$ and $A < \infty$, either*

$$H_\infty(u, \mathscr{F}) \leq Au^{-r} \quad \text{for all} \quad u > 0$$

or

$$H_2^B(u, \mathscr{F}, \bar{P}_{(n)}) \leq Au^{-r} \quad \text{for all} \quad u > 0, \quad n \geq 1, \quad \text{and} \quad \mathscr{F} \text{ consists of sets.}$$

If

$$(LLn)^{-2/r} = o(\alpha_n) \tag{2.28}$$

then

$$\limsup_n \frac{\sup_{\mathscr{F}} |\nu_n(f)|}{(2\alpha_n LLn)^{1/2}} \leq 1 \quad a.s. \tag{2.29}$$

Note that (2.28) ensures that (2.7) holds when needed.

THEOREM 2.13. *Let $\mathscr{F}$ be a deviation-measurable* VC *graph class of functions $0 \leq f \leq 1$ on $(X, \mathscr{A})$. If*

$$L(1/\alpha_n) = o(LLn) \tag{2.30}$$

then the LIL (2.29) *holds.*

For d.f.'s (2.29) is the well-known Chung-Smirnov LIL (Chung, 1949, Smirnov, 1944). In the i.i.d. case for $\mathscr{F}$ a class of sets, it follows from a compact LIL of Kuelbs and Dudley (1980).

The condition (2.30) is sharp in that, given (β_n) with $LLn = o(\beta_n)$, there exist p.m.'s $(P_{(i)})_{i\geq 1}$ on the integers $\mathbb{Z}$ and a VC class $\mathscr{C}$ of subsets of $\mathbb{Z}$ such that

$$L(1/\alpha_n) = o(\beta_n)$$

but

$$\limsup_n \frac{\sup_{\mathscr{C}} |\nu_n(C)|}{(2\alpha_n LLn)^{1/2}} = \infty \quad \text{a.s.}$$

For the proof see Alexander (1982).

From (2.3), Theorem 2.11, Corollary 2.4, and Corollary 2.5, the following is immediate.

COROLLARY 2.14. *Let P be a law on the positive integers and suppose $P([m, \infty)) = O(m^{-\beta})$ for some $\beta > 0$. Then the total variation distance ρ satisfies*

$$\rho(P_n, P) = \begin{cases} O(n^{-1/2}(LLn)^{1/2}) & \text{a.s. if } \beta > 1 \\ O(n^{-1/2}Ln) & \text{a.s. if } \beta = 1 \\ O(n^{-\beta/(\beta+1)}) & \text{a.s. if } 0 < \beta < 1. \end{cases}$$

REMARK 2.15. The bounds in Corollary 2.14 are sharp. For $\beta > 1$, this follows from Theorem 2.12. For $0 < \beta < 1$, consider P with $P(m) = am^{-(\beta+1)}$ for all $m \geq 1$. Fix $\theta > 0$ and let $m_n := [\theta n^{1/(\beta+1)}]$. Define events $B_{nm} := [\nu_n(m) > P(m)^{1/2}]$. Then

$$\rho(P_n, P) \geq T_n := n^{-1/2} \sum_{m=1}^{m_n} P(m)^{1/2} 1_{B_{nm}}.$$

It follows easily from Theorem 5.2.2 of Stout (1974) that if θ is small enough, then there exist n_0 and $\delta > 0$ such that $\mathbb{P}(B_{nm}) \geq \delta$ for all $n \geq n_0$ and $m \leq m_n$. Hence for $n \geq n_0$,

$$ET_n \geq n^{-1/2} \sum_{m=1}^{m_n} \delta P(m)^{1/2} \sim \lambda n^{-\beta/(1+\beta)} \tag{2.31}$$

for some $\lambda > 0$. It is clear that for $k \neq m$ we have $\mathbb{P}(B_{nk} | B_{nm}) \leq \mathbb{P}(B_{nk})$, so $\mathbb{P}(B_{nk} \cap B_{nm}) - \mathbb{P}(B_{nk})\mathbb{P}(B_{nm}) \leq 0$. Thus

$$\begin{aligned} \operatorname{var}(T_n) &= n^{-1} \textstyle\sum_{k,m \leq m_n} P(k)^{1/2} P(m)^{1/2} (\mathbb{P}(B_{nk} \cap B_{nm}) - \mathbb{P}(B_{nk})\mathbb{P}(B_{nm})) \\ &\leq n^{-1} \textstyle\sum_{m \leq m_n} P(m) \leq n^{-1}. \end{aligned}$$

With (2.31) and Chebyshev's inequality this shows

$$\mathbb{P}\left[T_n < \frac{\lambda n^{-\beta/(1+\beta)}}{2}\right] \to 0.$$

It follows that

$$\limsup_n n^{\beta/(\beta+1)} \rho(P_n, P) > 0 \quad \text{a.s.}$$

For $\beta = 1$ the proof is the same except that $ET_n \geq \lambda n^{-1/2} Ln$. □

3. Proofs. The fundamental idea is the familiar "chain argument" first applied to empirical processes by Dudley (1978), then used by LeCam (1983), Pollard (1982), and others. Given a class $\mathscr{F}$ with finite L^p entropy, and $\delta_0 > \delta_1 > \cdots > \delta_N > 0$, there exist $\mathscr{F}_j \subset \mathscr{F}$ $(j \leq N)$ such that $|\mathscr{F}_j| = N_p(\delta_j, \mathscr{F}, \bar{P}_{(n)})$

and for each $f \in \mathscr{F}$ there is an $f_j(f) \in \mathscr{F}_j$ with $\|f - f_j(f)\|_p < \delta_j$. Hence

$$
\begin{aligned}
&\mathbb{P}^*[\sup_{\mathscr{F}} |\nu_n(f)| > M] \\
&\quad \le \mathbb{P}\left[\sup_{\mathscr{F}} |\nu_n(f_0(f))| > \left(1 - \frac{\varepsilon}{4}\right)M\right] \\
&\qquad + \mathbb{P}\left[\sup_{\mathscr{F}} |\nu_n(f_0(f)) - \nu_n(f_N(f))| > \frac{\varepsilon M}{8} - \eta_N\right] \\
&\qquad + \mathbb{P}^*\left[\sup_{\mathscr{F}} |\nu_n(f_N(f)) - \nu_n(f)| > \frac{\varepsilon M}{8} + \eta_N\right] \\
&\quad \le |\mathscr{F}_0| \sup_{\mathscr{F}} \mathbb{P}\left[|\nu_n(f)| > \left(1 - \frac{\varepsilon}{4}\right)M\right] \\
&\qquad + \textstyle\sum_{j=0}^{N-1} |\mathscr{F}_j|\,|\mathscr{F}_{j+1}| \sup_{\mathscr{F}} \mathbb{P}[|\nu_n(f_j(f) - f_{j+1}(f))| > \eta_j] \\
&\qquad + \mathbb{P}^*\left[\sup_{\mathscr{F}} |\nu_n(f_N(f) - f)| > \frac{\varepsilon M}{8} + \eta_N\right] \\
&\quad := \mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3
\end{aligned}
\tag{3.1}
$$

where the η_j are chosen so $\sum_{j=0}^{N} \eta_j < \varepsilon M/8$. If δ_0 is large enough, then $|\mathscr{F}_0|$ is not too large, so $\mathbb{P}_1$ is small. When M is large relative to $\int_{\delta_N}^{\delta_0} H_p(x)^{1/2}\,dx$, the η_j can be chosen large enough so $\mathbb{P}_2$ is small. Control of $\mathbb{P}_3$ depends on δ_N being sufficiently small; the particular method used to bound it varies with the assumptions made on $\mathscr{F}$.

To facilitate the choice of the δ_j's and η_j's we use the following lemma. We omit the straightforward proof.

LEMMA 3.1. *Let H: $(0, t] \to \mathbb{R}^+$ be a decreasing function, and let $0 < s < t$. Set $\delta_0 := t$, $\delta_{j+1} := s \vee \sup\{x \le \delta_j/2\colon H(x) \ge 4H(\delta_j)\}$ for $j \ge 0$, and $N := \min\{j\colon \delta_j = s\}$. Then*

$$\textstyle\sum_{j=0}^{N} \delta_j H(\delta_{j+1})^{1/2} \le 8 \displaystyle\int_{s/4}^{t} H(x)^{1/2}\,dx.$$

PROOF OF THEOREM 2.1. To make $\mathbb{P}_1$ small, we take δ_0 (in (3.1)) $= t_0$. Using (1.7),

$$
\begin{aligned}
\mathbb{P}_1 &\le 2\exp(H(t_0))\exp(-\psi((1 - \varepsilon/4)M, n, \alpha)) \\
&\le 2\exp(-(1 - \varepsilon)\psi(M, n, \alpha)).
\end{aligned}
\tag{3.2}
$$

To handle $\mathbb{P}_3$, take $\delta_N \le \varepsilon M/16n^{1/2}$. Then

$$|\nu_n(f_N(f) - f)| \le 2n^{1/2}\|f_N(f) - f\|_\infty \le \varepsilon M/8$$

so $\mathbb{P}_3 = 0$.

To bound $\mathbb{P}_2$, we may assume $\delta_0 > \varepsilon M/16n^{1/2}$; otherwise we can take $N = 0$ so

$\mathbb{P}_2 = 0$. Let N and δ_j, $0 \le j \le N$, be as in Lemma 3.1, with $H(x) := H_\infty(x, \mathscr{F})$, $t := t_0$, and $s := \varepsilon M/16n^{1/2}$. Let $\eta_j := 4\varepsilon^{-1/2}\delta_j H(\delta_{j+1})^{1/2}$, $j < N$, and $\eta_N := 0$. By (2.4) and Lemma 3.1 we have

$$\sum_{j=0}^{N} \eta_j \le 32\varepsilon^{-1/2} I_\infty(s/4, t) \le \varepsilon M/8. \tag{3.3}$$

Since $(4\delta_j)^{-1} \| f_j(f) - f_{j+1}(f) \|_\infty \le \frac{1}{2}$, if $N \ge 2$ we have

$$\begin{aligned} \mathbb{P}_2 &\le \sum_{j=0}^{N-1} 2 \exp(2H(\delta_{j+1}))\exp(-\psi_3(\eta_j/4\delta_j, n, \tfrac{1}{4})) \\ &= \sum_{j=0}^{N-1} 2 \exp((2 - 2/\varepsilon)H(\delta_{j+1})) \\ &\le \sum_{j=0}^{\infty} 2 \exp((2 - 2/\varepsilon)2^{j+1}H(\delta_0)) \le 3 \exp(-(1 - \varepsilon)\psi(M, n, \alpha)). \end{aligned} \tag{3.4}$$

If $N = 1$, we may redefine η_0 as $8\varepsilon^{-1/2}\delta_0 H(\delta_1)^{1/2}$ without violating (3.3), and the last line of (3.4) again bounds $\mathbb{P}_2$, by a similar calculation. □

To prove Corollary 2.2 we observe that if an entropy function $H(u) \le Au^{-r}$ for all $u > 0$, then for $0 < s < t$,

$$I(s, t) \le \begin{cases} 2A^{1/2}(2 - r)^{-1}t^{(2-r)/2} & \text{if } r < 2 \\ A^{1/2}\log 1/s & \text{if } r = 2 \text{ and } t \le 1 \\ 2A^{1/2}(r - 2)^{-1}s^{(2-r)/2} & \text{if } r > 2. \end{cases} \tag{3.5}$$

If $r < 2$, we may assume equality holds in (2.6) and $\psi_1 := \psi_1(M, n, \alpha) \le \psi(M, n, \alpha)$, so $t_0 \le (4A^2/\varepsilon\psi_1)^{1/r}$. Using (1.6) and (3.5) it is easy to show that

$$M \ge K_1 n^{(r-2)/2(r+2)} \text{ implies (2.4) if } M \ge 4n^{1/2}\alpha$$

$$M \ge K_2 \alpha^{(2-r)/4} \text{ implies (2.4) if } M \le 4n^{1/2}\alpha.$$

If $r = 2$, our assumption (2.1) that H is continuous precludes assuming $t_0 \le 1$, but since $\log N_\infty(u, \mathscr{F}) = 0$ for all $u > 1$, we may assume $t_0 \le 1 + \theta$ for any fixed $\theta > 0$, which has the same effect. Then (3.5) shows that $M \ge K_3 Ln$ implies (2.4). If $r > 2$ then (3.5) shows that $M \ge K_1 n^{(r-2)/2(r+2)}$ implies (2.4).

PROOF OF THEOREM 2.3. As in the proof of Theorem 2.1, to make $\mathbb{P}_1$ small we take $\delta_0 := t_0$; as in (3.2) we have

$$\mathbb{P}_1 \le 2 \exp(-(1 - \varepsilon)\psi(M, n, \alpha)).$$

To make $\mathbb{P}_3$ small, take $\delta_N \le (\varepsilon M/16n^{1/2})^{1/2}$ and $\eta_N := 8\varepsilon^{-1/2}\delta_N H(\delta_N)^{1/2}$. Since our entropy is with bracketing, we may assume that $f_N(f) = f_N^L(f) \le f \le f_N^U(f)$. Then since $\mathscr{F}$ consists of sets,

$$\begin{aligned} |\nu_n(f_N(f) - f)| &\le |\nu_n(f_N^U(f) - f_N^L(f))| + 2n^{1/2}\| f_N^U(f) - f_N^L(f) \|_1 \\ &\le |\nu_n(f_N^U(f) - f_N^U(f))| + 2n^{1/2}\delta_N^2. \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{P}_3 &\le \mathbb{P}[\sup_{\mathscr{F}} |\nu_n(f_N^U(f) - f_N^L(f))| > \eta_N] \\ &\le 2 |\mathscr{F}_N| \exp(-\psi_1(\eta_N, n, \delta_N^2)). \end{aligned} \tag{3.6}$$

To bound $\mathbb{P}_2$, we may assume $\delta_0 > (\varepsilon M/16n^{1/2})^{1/2}$; otherwise we can take

$N = 0$ so $\mathbb{P}_2 = 0$. Let N and δ_j, $0 \le j \le N$ be as in Lemma 3.1, with $H(x) := H_2^B(x, \mathscr{F}, \bar{P}_{(n)})$, $t := t_0$, and $s := (\varepsilon M/16n^{1/2})^{1/2}$. Let $\eta_j := 8\varepsilon^{-1/2}\delta_j H(\delta_{j+1})^{1/2}$, $0 \le j < N$. As in (3.3) we have

$$\textstyle\sum_{j=0}^{N} \eta_j \le \varepsilon M/8.$$

Now by (2.9),

$$\left(\frac{\eta_j}{4\delta_j^2 n^{1/2}}\right)^2 \le \frac{4H(s)}{\varepsilon n s^2} \le 16, \tag{3.7}$$

while $\| f_j(f) - f_{j+1}(f) \|_2 \le 2\delta_j$, so using (1.6),

$$\begin{aligned} \mathbb{P}_2 &\le \textstyle\sum_{j=0}^{N-1} 2 \exp(2H(\delta_{j+1}))\exp(-\psi_1(\eta_j, n, 4\delta_j^2)) \\ &\le \textstyle\sum_{j=0}^{N-1} 2 \exp\left(2H(\delta_{j+1}) - \frac{\eta_j^2}{16\delta_j^2}\right) \qquad (3.8) \\ &\le \textstyle\sum_{j=0}^{N-1} 2 \exp((4 - 4/\varepsilon)4^j H(\delta_0)). \end{aligned}$$

Similarly, (3.6), (3.7), and (1.6) imply

$$\mathbb{P}_3 \le 2 \exp\left(H(\delta_N) - \frac{\eta_N^2}{4\delta_N^2}\right) \le 2 \exp((4 - 4/\varepsilon)4^N H(\delta_0)).$$

With (3.8) and the definition of δ_0, this shows

$$\mathbb{P}_2 + \mathbb{P}_3 \le 3 \exp(-(1 - \varepsilon)\psi(M, n, \alpha)). \ \square$$

The proof of Corollary 2.4 is analogous to that of Corollary 2.2.

Proof of Lemma 2.7. It is sufficient to show that

$$N_1(u, \mathscr{C}, P) \le (16u^{-1}L(8u^{-1}))^{v-1} \tag{3.9}$$

for all $0 < u < 1$ and laws P on $X \times [0, 1]$; the rest of the lemma is easy.

Suppose $C_1, \cdots, C_m \in \mathscr{C}$ with $P(C_i \Delta C_j) \ge u$ for all $i \ne j$. Take n to be the least positive integer such that $\binom{m}{2}(1 - u)^n < 1$, and let $Y_1, \cdots, Y_n$ be i.i.d.(P) and $S := \{Y_1, \cdots, Y_n\}$. Then

$$n \le 1 + (2Lm - \log 2)/u \le 8Lm/eu$$

while

$$\begin{aligned} \mathbb{P}[\Delta^{\mathscr{C}}(Y_1, \cdots, Y_n) < m] &\le \mathbb{P}[S \cap (C_i \Delta C_j) = \phi \text{ for some } i \ne j] \\ &\le \binom{m}{2} (1 - u)^n < 1. \end{aligned}$$

It follows that

$$m \le m^{\mathscr{C}}(n) \le m^{\mathscr{C}}([8Lm/eu]).$$

If $8Lm/eu \geq v - 1$ then by (1.8) this shows $m \leq (8Lm/(v-1)u)^{v-1}$, which implies

$$m \leq (16u^{-1}L(8u^{-1}))^{v-1}. \tag{3.10}$$

If $8Lm/eu < v - 1$ then $Lm < eu(v-1)/8 \leq (v-1)L(16/u)$ so (3.10) again holds. Taking $\{C_1, \cdots, C_m\}$ maximal gives $N_1(u, \mathscr{C}, P) \leq m$, and (3.9) follows. □

In proving Theorem 2.8 the main difficulty is in bounding $\mathbb{P}_3$. For this the key is the following.

PROPOSITION 3.1. *Let $n \geq 1$ and let $\mathscr{F}$ be an n-supremum-measurable class of functions $0 \leq f \leq 1$ on $(X, \mathscr{A})$. Let $\psi \in \Psi$, $\varepsilon > 0$, $\alpha \geq \sup_{\mathscr{F}} n^{-1} \sum_{i=1}^{n} \operatorname{var} f(X_i)$, $M > 0$, and $b := \psi(M, n, \alpha)$. Suppose that* (A) *holds for $(\mathscr{F}, \psi, M, n, \alpha, \varepsilon, b)$. Then*

$$\mathbb{P}^*[\sup_{\mathscr{F}} |\nu_n(f)| > M] \leq 11 \exp(-(1-\varepsilon)\psi(M, n, \alpha)).$$

To avoid measurability difficulties in the proof of Proposition 3.1, we use the following two lemmas, whose proofs are straightforward modifications of the well-known measurable cases.

LEMMA 3.2. *Let $(\Omega_1, \mathscr{A}_1, Q_1)$, $(\Omega_2, \mathscr{A}_2, Q_2)$ be probability spaces, and let $A \subset \Omega_1$, $B \subset \Omega_1 \times \Omega_2$, and $\beta > 0$. Suppose for each $\omega_1 \in A$ there is a subset $C_{\omega_1} \subset \Omega_2$ with $Q_2^*(C_{\omega_1}) \geq \beta$ and $\{\omega_1\} \times C_{\omega_1} \subset B$. Then*

$$Q_1^*(A) \leq \beta^{-1}(Q_1 \times Q_2)^*(B).$$

LEMMA 3.3. *Let $(\Omega_1, \mathscr{A}_1, Q_1)$, $(\Omega_2, \mathscr{A}_2, Q_2)$ be probability spaces and let $A \in \mathscr{A}_1 \times \mathscr{A}_2$, $B \subset \Omega_1$. Then*

$$(Q_1 \times Q_2)^*(A \cap (B \times \Omega_2)) \leq \sup_{\omega_1 \in B} Q_2(\{\omega_2 \colon (\omega_1, \omega_2) \in A\}). \ \square$$

Let us outline the proof of Proposition 3.1 heuristically before giving the full story. P_n is normally constructed by sampling one point under each law $P_{(i)}$ $(i \leq n)$. In our alternate construction (2.12) we sample m points under each $P_{(i)}$, then resample one from each group of m to use in constructing P_n. This resampling may be viewed as the construction of a "secondary" empirical process, call it ν_n^0, in which the laws for the sampling are the empirical measures constructed from each group of m points from the first sampling. That is, we construct $\nu_n^0 = n^{1/2}(P_n - P_{n \times m})$. We show that little is lost if we replace ν_n with ν_n^0 in the probability we wish to bound. The advantage of this replacement is that $|\nu_n^0(f-g)|$ is small whenever $P_{n\times m}|f-g|$ is small, so we need only consider a finite subcollection, call it $\mathscr{G}$, of $\mathscr{F}$ which is nearly dense in $\mathscr{F}$ with respect to the $L^1(P_{n\times m})$ metric. Let $p := \max_{f \in \mathscr{G}} \mathbb{P}[|\nu_n^0(f)| > M]$; then $|\mathscr{G}|p$ provides an upper bound for $\mathbb{P}^*[\sup_{\mathscr{F}} |\nu_n^0(f)| > M]$. Now $|\mathscr{G}|$ and p are random, so the upper bound is conditional on $P_{n\times m}$. For it to be a good bound, $P_{n\times m}$ must satisfy two conditions with high probability: $|\mathscr{G}|$ must be much less than p^{-1}, and the (random) variance $\operatorname{var} \nu_n^0(f)$ for the resampling process must be close to the variance var $\nu_n(f)$ for the original process for all f. The former is guaranteed by condition (A) (cf. (2.15).) The latter can be proved, if m is large enough (cf.(2.13)), by an iteration scheme. In this scheme, symmetrization is used to make the resampling-variance process var $\nu_n^0(\cdot)$ look somewhat like a new (symmetrized) empirical process on which this entire procedure is repeated, this time with a larger m. After sufficient repetitions, the random variance is an adequate approximation to var $\nu_n(f)$.

A useful observation for the proof is that

$$\tilde{H}_1\left(m_1 + m_2, \cdot, \frac{m_1\varepsilon_1 + m_2\varepsilon_2}{m_1 + m_2}, \cdot, \cdot\right) \leq \tilde{H}_1\left(m_1, \cdot, \frac{\varepsilon_1}{2}, \cdot, \cdot\right) + \tilde{H}_1\left(m_2, \cdot, \frac{\varepsilon_2}{2}, \cdot, \cdot\right). \tag{3.11}$$

PROOF OF PROPOSITION 3.1. The key will be the randomization defined in (2.12) of the choice of the X_i. To take advantage of it we must first modify $\nu_n(f)$ so it depends only on $\{f(\xi_{ij}): i \leq n, j \leq m\}$ and not on $\bar{P}_{(n)}(f)$, where m is the integer in (2.13). We do so by approximating $\bar{P}_{(n)}$ by $P_{n\times m}$. Define ξ' to be ξ with the X_i's deleted, that is, $\xi'_{ij} := \xi_{ij}$ if $j < \tau(i)$, $\xi'_{ij} := \xi_{i,j+1}$ if $j \geq \tau(i)$, and let

$$P'_{n\times(m-1)} := (m-1)^{-1}n^{-1} \sum_{i=1}^{n} \sum_{j\leq m, j\neq\tau(i)} \delta_{\xi_{ij}}$$

$$\psi_0 := \psi(M, n, \alpha)$$

$$A_0 := \left[\tilde{H}_1\left(1, n, \frac{\varepsilon M}{64n^{1/2}}, \xi, \mathscr{F}\right) \leq \frac{\varepsilon\psi_0}{16}\right]$$

$$A'_0 := \left[\tilde{H}_1\left(m-1, n, \frac{\varepsilon m}{64mn^{1/2}}, \xi', \mathscr{F}\right) \leq \frac{\varepsilon\psi_0}{16}\right]$$

$$A''_0 := \left[\tilde{H}_1\left(m, n, \frac{\varepsilon M}{16mn^{1/2}}, \xi, \mathscr{F}\right) \leq \frac{\varepsilon\psi_0}{8}\right].$$

Note that $P'_{n\times(m-1)}$ is independent of P_n. If $\xi \in [\sup_{\mathscr{F}}|\nu_n(f)| > M]$, there exists $f_\xi \in \mathscr{F}$ with $|\nu_n(f_\xi)| > M$, so define events

$$C'_\xi := [|n^{1/2}(P'_{n\times(m-1)}(f_\xi) - \bar{P}_{(n)}(f_\xi))| \leq (4\alpha/(m-1))^{1/2}]$$

$$C_\xi := C'_\xi \cap A'_0.$$

By Chebyshev's inequality we have $\mathbb{P}(C'_\xi) \geq 3/4$. By (2.14) it follows that $\mathbb{P}^*(C_\xi) \geq 1/2$, so by Lemma 3.2,

$$\begin{aligned} &\mathbb{P}^*[\sup_{\mathscr{F}}|\nu_n(f)| > M] - \mathbb{P}^*(A_0^c) \\ &\leq \mathbb{P}^*[\sup_{\mathscr{F}}|\nu_n(f)| > M; A_0] \\ &\leq 2\mathbb{P}^*\left[\sup_{\mathscr{F}}|n^{1/2}(P_n(f) - P'_{n\times(m-1)}(f))| > M - \left(\frac{4\alpha}{m-1}\right)^{1/2}; A_0; A'_0\right] \\ &:= (I). \end{aligned} \tag{3.12}$$

Let

$$\nu_n^0 := n^{1/2}(P_n - P_{n\times m}) = (1 - (1/m))n^{1/2}(P_n - P'_{n\times(m-1)}).$$

By (3.11) we have $A_0 \cap A_0' \subset A_0''$, so by (2.13)

$$(I) \le 2\mathbb{P}^*\left[\sup_{\mathscr{F}} |\nu_n^0(f)| > \left(1 - \frac{\varepsilon}{8}\right)M;\ A_0''\right]. \tag{3.13}$$

We would like to be able to condition on ξ and bound the probability in (3.13) with only the choice of X_i from among the ξ_{ij} remaining random. We have for each $f \in \mathscr{F}$

$$\mathbb{P}[|\nu_n^0(f)| > (1 - \varepsilon/8)M \mid \xi] \le 2\exp(-\psi((1 - \varepsilon/8)M, n, \operatorname{var}(\nu_n(f) \mid \xi)))$$

which gives a good bound only if $\operatorname{var}(\nu_n(f) \mid \xi)$ is not much larger than α, so we must condition on this latter event. Set

$$\bar{f}_i^{(r)} := r^{-1} \textstyle\sum_{j=1}^r f(\xi_{ij}),\ r \ge 1$$

$$V(r, n, f, \xi) := (rn)^{-1} \textstyle\sum_{i=1}^n \sum_{j=1}^r (f(\xi_{ij}) - \bar{f}_i^{(r)})^2$$

$$= \operatorname{var}(\nu_n(f) \mid \xi) \qquad \text{if } r = m$$

$$B_k := [\sup_{\mathscr{F}} V(2^{k-1}m, n, f, \xi) \le (1 + 2^{k-2}\varepsilon)\alpha], \quad k \ge 1$$

$$s_0 := \frac{\varepsilon M}{16mn^{1/2}}$$

and for each r, ξ, s let $\mathscr{F}_{r\xi s} \subset \mathscr{F}$ be such that $|\mathscr{F}_{r\xi s}| = \tilde{N}_1(r, n, s, \xi, \mathscr{F})$ and for each $f \in \mathscr{F}$ there exists $f_{r\xi s}(f) \in \mathscr{F}_{r\xi s}$ with $P_{n\times r}|f - f_{r\xi s}(f)| < s$. Then

$$|\nu_n^0(f) - \nu_n^0(g)| \le mn^{1/2}P_{n\times m}|f - g|$$

and from n-supremum measurability we know $\sup_{\mathscr{F}}|\nu_n^0(f)|$ is measurable, so using (3.12), (3.13), Lemmas 3.2 and 3.3, and (1.7), we get

$$\begin{aligned}
&\mathbb{P}^*[\sup_{\mathscr{F}}|\nu_n(f)| > M] - \mathbb{P}^*(A_0^c) - 2\mathbb{P}^*(A_0'' \cap B_1^c)\\
&\le 2\mathbb{P}^*[\sup_{\mathscr{F}} \nu_n^0(f)| > \left(1 - \frac{\varepsilon}{8}\right) M;\ A_0'', B_1]\\
&\le 2\mathbb{P}^*\left[\sup_{\mathscr{F}} |\nu_n^0(f_{r\xi s_0}(f))| > \left(1 - \frac{3\varepsilon}{16}\right) M;\ A_0'', B_1\right]\\
&\le 2\sup_{\xi\in A_0''\cap B_1} |\mathscr{F}_{r\xi s_0}| \sup_{\mathscr{F}} \mathbb{P}\left[|\nu_n^0(f)| > \left(1 - \frac{3\varepsilon}{16}\right) M \mid \xi\right]\\
&\le 4\exp(\varepsilon\psi_0/8)\exp\left(-\psi\left(\left(1 - \frac{3\varepsilon}{16}\right) M, n, (1 + \varepsilon/2)\alpha\right)\right)\\
&\le 4\exp(-(1 - \varepsilon)\psi(M, n, \alpha)).
\end{aligned} \tag{3.14}$$

By (2.15) we have

$$\mathbb{P}^*(A_0^c) \le \exp(-\psi(M, n, \alpha)) \tag{3.15}$$

so it remains to bound $\mathbb{P}^*(A_0'' \cap B_1^c)$. If $\alpha \ge 1/4$ then $B_1^c = \phi$ and we are done. Hence we may assume (2.16) holds. We will proceed by an iteration. Define $\xi^{(r)}$ by $\xi_{ij}^{(r)} = \xi_{i,r+j}$ and set

$$s_k := 2^{k-8}\varepsilon\alpha$$

$$m_k := 2^{k-1}m$$

$$A_k := [\tilde{H}_1(m_k, n, s_k, \xi, \mathscr{F}) \le 2^{2k-15}\varepsilon^2\alpha mn - (k-1)\log 2]$$

$$A_k' := [\tilde{H}_1(m_k, n, s_k, \xi^{(m_k)}, \mathscr{F}) \le 2^{2k-15}\varepsilon^2\alpha mn - (k-1)\log 2]$$

for all $k \ge 1$. Suppose we can show that

$$\begin{aligned} \mathbb{P}^*(A_k \cap B_k^c) &\le 2\exp(-2^{2k-13}mn\varepsilon^2\alpha - (k-1)\log 2) \\ &\quad + 2\mathbb{P}^*(A_{k+1} \cap B_{k+1}^c), \quad k \ge 1. \end{aligned} \tag{3.16}$$

By (2.13) we have $s_1 \ge \varepsilon M/16mn^{1/2}$ and $\varepsilon\psi_0/8 \le 2^{-13}\varepsilon^2\alpha mn$, so $A_0'' \subset A_1$. For all sufficiently large k we have $(1 + 2^{k-2}\varepsilon)\alpha \ge 1/4$ so $B_k^c = \phi$. Hence from (3.16) we obtain by induction

$$\begin{aligned} \mathbb{P}^*(A_0'' \cap B_1^c) &\le \mathbb{P}^*(A_1 \cap B_1^c) \\ &\le \textstyle\sum_{k=1}^\infty 2^k \exp(-2^{2k-13}mn\varepsilon^2\alpha - (k-1)\log 2) \\ &= 2\textstyle\sum_{k=1}^\infty \exp(-2^{2k-13}mn\varepsilon^2\alpha) \\ &\le 3\exp(-\psi(M, n, \alpha)) \end{aligned}$$

using (2.13). In combination with (3.14) and (3.15), this proves the proposition.

It remains to establish (3.16). Fix $k \ge 1$ and set

$$\bar{f}_i := \bar{f}_i^{(m_k)} = m_k^{-1} \textstyle\sum_{j=1}^{m_k} f(\xi_{ij})$$

$$\bar{f}_i' := m_k^{-1} \textstyle\sum_{j=m_k+1}^{2m_k} f(\xi_{ij})$$

$$V(f) := V(m_k, n, f, \xi)$$

$$V'(f) := V(m_k, n, f, \xi^{(m_k)})$$

$$V_0(f) := V(f) - V'(f).$$

Then $V(f)$ and $V'(f)$ are independent copies of the "sample variance" for the choice of X_i from among the ξ_{ij}, but with the "sample size" m_k instead of m for each i. We will take advantage of the fact that as k increases, $V(f)$ becomes increasingly likely (conditional on B_{k+1}) to be not much larger (uniformly in f) than α. We have

$$EV(f) = (m_k n)^{-1}(m_k - 1) \textstyle\sum_{i=1}^n \operatorname{var} f(X_i) < \alpha$$

$$\operatorname{var} V(f) = n^{-2}m_k^{-3}(m_k - 1)^2 \textstyle\sum_{i=1}^n E(f(X_i) - Ef(X_i))^4 < \alpha/nm_k.$$

By Chebyshev's inequality, (2.13), (2.16), (3.11), and Lemma 3.2, we have analogously to (3.12):

$$\begin{aligned}\mathbb{P}^*(A_k \cap B_k^c) &= \mathbb{P}^*[\sup_{\mathcal{F}} V(f) > (1 + 2^{k-2}\varepsilon)\alpha;\ A_k] \\ &\le 2\mathbb{P}^*\left[\sup_{\mathcal{F}} V_0(f) > 2^{k-2}\varepsilon\alpha - \left(\frac{4\alpha}{m_k n}\right)^{1/2};\ A_k;\ A_k'\right] \\ &\le 2\mathbb{P}^*[\sup_{\mathcal{F}} V_0(f) > 2^{k-3}\varepsilon\alpha;\ A_{k+1}].\end{aligned} \tag{3.17}$$

Now consider the symmetrized "sample variance" $V_0(f)$:

$$\begin{aligned}V_0(f) &= (m_k n)^{-1} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m_k} ((f(\xi_{ij}) - \bar{f}_i)^2 - (f(\xi_{i,m_k+j}) - \bar{f}_i')^2) \\ &= (m_k n)^{-1} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m_k} [f(\xi_{ij}) + f(\xi_{i,m_k+j}) - \bar{f}_i - \bar{f}_i'] \\ &\qquad \cdot [f(\xi_{ij}) - f(\xi_{i,m_k+j})].\end{aligned} \tag{3.18}$$

Let (θ_{ij}) be iid with $\mathbb{P}(\theta_{ij} = 1) = \mathbb{P}(\theta_{ij} = -1) = 1/2$, and define

$$Y_{ij}(f) := f(\xi_{ij}) - f(\xi_{i,m_k+j})$$

$$Z_{ij}(f) := f(\xi_{ij}) + f(\xi_{i,m_k+j}) - \bar{f}_i - \bar{f}_i'$$

$$W(f) := (m_k n)^{-1} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m_k} \theta_{ij} Y_{ij}(f) Z_{ij}(f).$$

Then $W(f)$ has the same law as $V_0(f)$ by (3.18), and in fact the last probability in (3.17) is unchanged if V_0 is replaced by W. By Theorem 2 of Hoeffding (1963),

$$\mathbb{P}[W(f) > t \mid \xi] \le \exp(-m_k^2 n^2 t^2 / 2 \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m_k} Y_{ij}^2(f) Z_{ij}^2(f)). \tag{3.19}$$

Since

$$\begin{aligned}(m_k n)^{-1} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m_k} Y_{ij}^2(f) Z_{ij}^2(f) &\le (m_k n)^{-1} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m_k} Z_{ij}^2(f) \\ &\le 2(m_k n)^{-1} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m_{k+1}} (f(\xi_{ij}) - \bar{f}_i^{(m_{k+1})})^2 \\ &= 4V(m_{k+1}, n, f, \xi) \le 4(1 + 2^{k-1}\varepsilon)\alpha\end{aligned}$$

for $\xi \in B_{k+1}$, and

$$\frac{m_k n (2^{k-4}\varepsilon\alpha)^2}{8(1 + 2^{k-1}\varepsilon)\alpha} \ge 2^{2k-12}\varepsilon^2 \alpha m n,$$

it follows from (3.19) that

$$\mathbb{P}[W(f) > 2^{k-4}\varepsilon\alpha \mid \xi \in B_{k+1}] \le \exp(-2^{2k-12}\varepsilon^2 \alpha m n). \tag{3.20}$$

The other observation needed to bound the right side of (3.17) is that

$$\begin{aligned}| V(f) - V(g) | &\le (m_k n)^{-1} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m_k} |(f(\xi_{ij}) - \bar{f}_i)^2 - (g(\xi_{ij}) - \bar{g}_i)^2| \\ &\le 2(m_k n)^{-1} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m_k} ||f(\xi_{ij}) - \bar{f}_i| - |g(\xi_{ij}) - \bar{g}_i|| \\ &\le 4P_{n \times m_k} |f - g|\end{aligned}$$

so $|V_0(f) - V_0(g)| \le 8P_{n\times m_{k+1}}|f-g|$, and therefore

$$|W(f) - W(g)| \le 8P_{n\times m_{k+1}}|f-g|. \tag{3.21}$$

Hence by (3.17), (3.20), (3.21), and Lemmas 3.2 and 3.3, analogously to (3.14) we get

$$\begin{aligned}P^*(A_k \cap B_k^c) &- 2\mathbb{P}^*(A_{k+1} \cap B_{k+1}^c)\\ &\le 2\mathbb{P}^*[\sup_{\mathscr{F}} W(f) > 2^{k-3}\varepsilon\alpha;\ A_{k+1};\ B_{k+1}]\\ &\le 2\mathbb{P}^*[\sup_{\mathscr{F}} W(f_{m_{k+1}\xi s_{k+1}}(f)) > 2^{k-4}\varepsilon\alpha;\ A_{k+1};\ B_{k+1}]\\ &\le 2\sup_{\xi\in A_{k+1}\cap B_{k+1}} |\mathscr{F}_{m_{k+1}\xi s_{k+1}}| \sup_{\mathscr{F}} \mathbb{P}[W(f) > 2^{k-4}\varepsilon\alpha \mid \xi]\\ &\le 2\exp(-2^{2k-13}mn\varepsilon^2\alpha - (k-1)\log 2)\end{aligned}$$

so (3.16) holds and we are done. □

PROPOSITION 3.4. *Let n, $\mathscr{F}$, ψ, ε, α, and M be as Proposition 3.1, with $\mathscr{F}$ n-deviation measurable. Define t_0, t_1, and b by*

$$H_2^*(t_0, \mathscr{F}, (P_{(i)})_{i\le n}) = \tfrac{1}{4}\,\varepsilon\psi(M, n, \alpha)$$

$$H_2^*(t_1, \mathscr{F}, (P_{(i)})_{i\le n}) = 4\varepsilon n t_1^2$$

$$b := \psi(M, n, \alpha).$$

Suppose that one of the following holds:

(i) (A) *holds for* $(\mathscr{F}, \psi, M, n, \alpha, \varepsilon, b)$;

(ii) (A) *holds for* $(\mathscr{F}, \varphi, \varepsilon M/16, n, s^2/4, \tfrac{1}{4}, b)$ *for some* $s \ge t_0$ *and* $\varphi \in \psi$ *such that*

$$\varphi(\varepsilon M/16, n, s^2/4) \ge 2\psi(M, n, \alpha); \tag{3.22}$$

(iii) (A) *holds for* $(\mathscr{F}, \varphi, \varepsilon M/16, n, s^2/4, \tfrac{1}{4}, b)$ *for some* $s \ge t_1$ *and* $\varphi \in \psi$ *such that* (3.22) *holds, and*

$$M > 2^9\varepsilon^{-3/2} I_2^*(s/4, t_0, \mathscr{F}, (P_{(i)})_{i\le n}). \tag{3.23}$$

Then

$$\mathbb{P}^*[\sup_{\mathscr{F}} |\nu_n(f)| > M] \le 16\exp(-(1-\varepsilon)\psi(M, n, \alpha)).$$

PROOF. (i) is just Proposition 3.1, where $\psi(M, n, \alpha)$ is large enough that the "chaining" of (3.1) is not needed. Under (iii) we construct the chain of (3.1) from $\delta_0 = t_0$ to $\delta_N = s$; having $s \ge t_1$ ensures that we can use the upper inequality in (1.6) in bounding $\mathbb{P}_2$, as in (3.7) and (3.8). Under (ii) we can omit the middle of the chain, taking $N = 0$ so $\mathbb{P}_2 = 0$.

For the specifics, we begin with (iii). We may assume $s < t_0$, otherwise (ii) holds. Let N and δ_j, $0 \le j \le N$, be as in Lemma 3.1, with $H(x) := H_2^*(x, \mathscr{F}, (P_{(i)})_{i\le n})$, $t := t_0$, and s from (iii). Let $\eta_j := 8\varepsilon^{-1/2}\delta_j H(\delta_{j+1})^{1/2}$, $0 \le j \le N$. As in

(3.2) and (3.3) we have by (3.23):

$$\sum_{j=0}^{N} \eta_j \le \varepsilon M/8$$

$$\mathbb{P}_1 \le 2 \exp(-(1-\varepsilon)\psi(M, n, \alpha)),$$

while as in (3.7) and (3.8), we get from the assumption that $s \ge t_1$ that

$$\mathbb{P}_2 \le 3 \exp(-(1-\varepsilon)\psi(M, n, \alpha)).$$

Let $\mathscr{F}_0 := \{1 + (f-g)/2: f, g \in \mathscr{F}, \operatorname{var} \nu_n(f-g) \le s^2\}$. Then $\tilde{H}_1(r, n, u, \mathscr{F}_0, \xi) \le 2\tilde{H}_1(r, n, u, \mathscr{F}, \xi)$, so we have (A) holding for $(\mathscr{F}_0, \varphi, \varepsilon M/16, n, s^2/4, 1/2, b)$. Since $\mathscr{F}_0$ is n-supremum measurable, from Proposition 3.1 and (3.22) we get that

$$\mathbb{P}_3 \le \mathbb{P}^*\left[\sup_{\mathscr{F}_0} |\nu_n(f)| > \frac{\varepsilon M}{16}\right]$$
$$\le 11 \exp(-\psi(M, n, \alpha)).$$

If (ii) holds, then $N = 0$ so the above proof works without (3.23) or the assumption $s \ge t_1$. □

REMARK 3.5. At times (when proving a central limit theorem, for example), we do not need the sharpest possible bound on $\mathbb{P}^*[\sup_{\mathscr{F}} |\nu_n(f)| > M]$. The following modification of Proposition 3.1 may then be useful. Let (A′) be the same as (A) except that (2.13) is weakened to

$$m \ge \max\left(1 + \frac{2^{11}\alpha}{\varepsilon^2 M^2}, \frac{16}{\varepsilon}, \frac{2^{12}}{\varepsilon^2 \alpha n}\right)$$

and the right side of (2.16) is replaced with e^{-b} for $k = 1$. Let n, $\mathscr{F}$, ψ, ε, α, and M be as in Proposition 3.1, and let $b > 0$. If (A′) holds for $(\mathscr{F}, \psi, M, n, \alpha, \varepsilon, b)$, then

$$\mathbb{P}^*[\sup_{\mathscr{F}} |\nu_n(f)| > M]$$
$$\le 4 \exp(-(1-\varepsilon)\psi(M, n, \alpha)) + 6 \exp(-2^{-11} mn\varepsilon^2\alpha) + 3e^{-b}.$$

The only changes in the proof are in the bounding of $\mathbb{P}^*(A_0^c)$ and $\mathbb{P}^*(A_0'' \cap B_1^c)$ (we no longer know $A_0'' \subset A_1$). This time $\mathbb{P}^*(A_0^c) < e^{-b}$, and we must use the following:

$$\mathbb{P}^*(A_0'' \cap B_1^c) \le \mathbb{P}^*(A_1^c) + \mathbb{P}^*(A_1 \cap B_1^c)$$
$$\le e^{-b} + 1 \wedge 2 \sum_{k=1}^{\infty} \exp(-2^{2k-13} mn\varepsilon^2\alpha)$$
$$\le e^{-b} + 3 \exp(-2^{-11} mn\varepsilon^2\alpha).$$

This leads to a modified version of Proposition 3.4 (iii). Let n, $\mathscr{F}$, ψ, ε, α, M, t_0, and t_1 be as in Proposition 3.4, and let $b > 0$ and

$$\mathscr{F}_0(s) := \{1 + (f-g)/2: f, g \in \mathscr{F}, \operatorname{var} \nu_n(f-g) \le s^2\}.$$

If (A′) holds for $(\mathscr{F}_0(s), \varphi, \varepsilon M/16, n, s^2/4, 1/2, b)$ for some $s \ge t_1$ and $\varphi \in \Psi$, and if (3.23) holds, then

$$\mathbb{P}^*[\sup_{\mathscr{F}}|\nu_n(f)| > M]$$
$$\le 5\exp(-(1-\varepsilon)\psi(M,n,\alpha)) + 4\exp(-\varphi(\varepsilon M/16, n, s^2/4)/2)$$
$$+ 6\exp(-2^{-15}mns^2) + 3e^{-b}. \quad \square$$

To apply Proposition 3.4 to VC graph classes, we need the next lemma.

LEMMA 3.6. *Let $\mathscr{F}$ be a* VC *graph class of functions $0 \le f \le 1$, with (v, k)-constructible graph region class. Let $\psi \in \Psi$, $M > 0$, $0 < \varepsilon < 1$, $\alpha > 0$, and $n \ge 1$. If*

$$\alpha^{1/2} \le M \le n^{1/2} \tag{3.24}$$

$$M \ge \frac{64kv}{n^{1/2}\varepsilon} \tag{3.25}$$

$$\psi(M, n, \alpha) \ge 2^9 k\, v\, \varepsilon^{-1} L\left(\frac{n}{v} \vee \frac{k}{\alpha}\right) \tag{3.26}$$

then (A) *holds for $(\mathscr{F}, \psi, M, n, \alpha, \varepsilon, u)$ for all $u > 0$.*

PROOF. Let $\psi_0 := \psi(M, n, \alpha)$, and let m be the integer part of $2m_0$, where $m_0 := (8n)^{-1}\exp(\varepsilon\psi_0/64kv)$. Suppose we can show that m satisfies (2.13). Then $2^{k-8}\varepsilon\alpha \ge \varepsilon M/64mn^{1/2}$ and $\varepsilon\psi_0/16 \le 2^{2k-15}\varepsilon^2\alpha mn - (k-1)\log 2$ for all $k \ge 1$, so by (2.17), (A) will follow if

$$4kvL(4k/u) \le \varepsilon\psi_0/16 \quad \text{for} \quad u = \varepsilon M/64mn^{1/2}. \tag{3.27}$$

But $u \ge kv/mn$ by (3.25), so (3.27) is immediate from the definition of m, and the lemma follows.

To prove (2.13), set

$$a := \frac{n}{v} \vee \frac{k}{\alpha}.$$

Then by (3.24), (3.25), and (3.26) we have a $\ge 64/\varepsilon$ and

$$m_0 \ge \left(\frac{v}{8n}\right)a^4 \ge \frac{a^3}{8} \ge \frac{2^{15}}{\varepsilon^2} \ge 1 + \frac{2^{10}\alpha}{\varepsilon^2 M^2}$$

so the first two terms in (2.13) are smaller than m_0. Since $e^{8x} \ge a^8x/La \ge 2^{39}kx/\varepsilon^6\alpha$ for all $x \ge La$, by (3.26) we have

$$m_0 \ge \frac{v}{8n}\,\frac{2^{39}k}{\varepsilon^6\alpha}\,\frac{\varepsilon\psi_0}{2^9kv} \ge \frac{2^{11}\psi_0}{\varepsilon^2\alpha n}$$

which takes care of the third term in (2.13). For the last two terms we have from (3.24) and (3.27)

$$m_0 \ge \left(\frac{v}{8n}\right)e^{8La} \ge a^3 \ge \frac{2^{12}}{\varepsilon^2\alpha} \ge \max\left(\frac{2^{12}}{\varepsilon^2 n\alpha}, \frac{8M}{\varepsilon\alpha n^{1/2}}\right). \quad \square$$

PROOF OF THEOREM 2.8. We may assume $M \le n^{1/2}$ and $\alpha \le 1/4$. Under (i), the theorem then follows from Lemma 3.6 and Proposition 3.4(i).

Under (ii), we may assume (2.21) is false (otherwise (i) applies.) We will use Proposition 3.4(iii). Let t_0, t_1 be as in Proposition 3.4. By Lemma 2.7, (2.19), (2.17), and (2.23) we have

$$I_2^*(s/4, \alpha^{1/2}) \le \int_{s/4}^{\alpha^{1/2}} \left(2kvL\left(\frac{8k^2}{x^2}\right)\right)^{1/2} dx$$
$$\le 8\left(kv\alpha L\left(\frac{k}{\alpha}\right)\right)^{1/2} \le 2^{-9}\varepsilon^{3/2}M \tag{3.28}$$

for all $s > 0$. Hence (3.23) holds for $\alpha^{1/2}$ in place of t_0, but by (2.18) we may then assume (3.23) holds for t_0.

To establish (3.22), take $\varphi = \psi_1$ and $s := (vkL(\varepsilon n/v)/\varepsilon n)^{1/2}$; by (2.22) we have $\varepsilon M/16 \ge 2^{10}n^{1/2}s^2$, so by (1.6), (2.22), and the falsity of (2.21),

$$\psi_1(\varepsilon M/16, n, s^2/4) \ge \varepsilon M n^{1/2}/4 \ge 2^{11}\varepsilon^{-3/2}kvL(\varepsilon n/v)$$
$$\ge 2^{11}\varepsilon^{-1}kvL\left(\frac{n}{v} \vee \frac{4k}{s^2}\right) \ge 2\psi(M, n, \alpha). \tag{3.29}$$

Now (2.22) and $M \le n^{1/2}$ give $8k \le \varepsilon n/v$, so

$$H_2^*(s, \mathscr{F}, (P_{(i)})_{i\le n}) \le 2kvL(8k^2/s^2)$$
$$\le 4kvL(\varepsilon n/v) = 4\varepsilon n s^2 \tag{3.30}$$

so $s \ge t_1$. Finally, (2.20), (2.22), and (3.29) give us (A) for

$$\left(\mathscr{F}, \psi_1, \frac{\varepsilon M}{16}, n, \frac{s^2}{4}, 1/4, b\right)$$

by Lemma 3.6, so Proposition 3.4(iii) finishes the proof. □

PROOF OF THEOREM 2.11. Take $k := 1$, $\alpha := 1/4$, $\psi(M, n, \alpha) := 2M^2$, and $\varepsilon := 2^{11}vLM/M^2$, where $v := V(\mathscr{C})$. We may assume $M < n^{1/2}$ and $\varepsilon < 1$.

If $M \ge (n/v)^{1/8}$, Theorem 2.8(i) is easily shown to apply, proving the theorem.

If $M < (n/v)^{1/8}$, we use Proposition 3.4(iii), with $\varphi = \psi_1$ and $s := (vL(\varepsilon n/v)/\varepsilon n)^{1/2}$. By (2.17) we have $t_0^2 \le 8M^{-512}$, so as in (3.28),

$$I_2^*(s/4, t_0) \le 8(vt_0^2Lt_0^{-2})^{1/2} \le 2^{-9}\varepsilon^{3/2}M$$

so (3.23) holds. Since $8k < 2^{11} \le \varepsilon M^2/v \le \varepsilon n/v$, (3.30) shows $s \ge t_1$.

To obtain (A) we use Lemma 3.6. From

$$\frac{\varepsilon M}{16} = \frac{(2^{11}vLM)^2}{16\varepsilon M^3} \ge \frac{(2^{11}v)^2}{16\varepsilon}\frac{(L(n/v)/8)^2}{(n/v)^{3/8}} \ge 2^{14}n^{5/8}s^2 \tag{3.31}$$

we get using (1.6) that

$$\psi_1\left(\frac{\varepsilon M}{16}, n, \frac{s^2}{4}\right) \geq \frac{1}{32}\,\varepsilon M n^{1/2} L\left(\frac{\varepsilon M}{4n^{1/2}s^2}\right) \geq 8M^3 v LMLn \geq 2^{12} vL\left(\frac{n}{v} \vee \frac{4}{s^2}\right) \tag{3.32}$$

and (3.31) also gives

$$\frac{\varepsilon M}{16} \geq 2^{14} n^{1/2} s^2 \geq \max\left(s, \frac{256v}{n^{1/2}}\right),$$

so (A) follows for $(\mathscr{F}, \psi_1, \varepsilon M/16, n, s^2/4, 1/4, b)$. The second line of (3.32) proves (3.22), so we are done. □

Acknowledgements. I am grateful to my thesis advisor Professor Richard M. Dudley for his devotion of much time and effort to guidance of my work.

REFERENCES

ALEXANDER, K. S. (1982). Some limit theorems and inequalities for weighted and non-identically distributed empirical processes. Ph.D. dissertation, Massachusetts Institute of Technology.

BRONŠTEIN, E. M. (1976). ε-entropy of convex sets and functions. *Siberian Math. J.* **17** 393–398. (*Sibirsk. Mat. Ž.* **17** 508–514, in Russian.)

CHUNG, K. L. (1949). An estimate concerning the Kolmogoroff limit distribution. *Trans. Amer. Math. Soc.* **67** 36–50.

DEVROYE, L. (1982). Bounds for the uniform deviation of empirical measures. *J. Multivariate Anal.* **12** 72–79.

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929.

DUDLEY, R. M. (1982). Empirical and Poisson processes on classes of sets or functions too large for central limit theorems. *Z. Wahrsch. verw. Gebiete* **61** 355–368.

DUDLEY, R. M. (1983). A course on empirical processes. Ecole d'eté de probabilitiés de St.-Flour, 1982. To appear in *Lecture Notes in Math.* (Springer).

DUDLEY, R. M., and PHILIPP, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. verw. Gebiete* **62** 509–552.

GINÉ, E., and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–989.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

KIEFER, J. (1961). On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm. *Pacific J. Math.* **11** 649–660.

KOLMOGOROV, A. N., and TIHOMIROV, V. M. (1959). ε-entropy and ε-capacity of sets in functional spaces. *Amer. Math. Soc. Transl.* **17** 277–364 (*Uspekhi Mat. Nauk.* **14** vyp. 2 (86) 3–86, in Russian.)

LECAM, L. (1983). A remark on empirical measures. *A Festschrift for Erich L. Lehmann in Honor of his 65th Birthday* (P. Bickel, K. Doksum, and J. Hodges, editors). Wadsworth, Belmont, California.

LÉVY, P. (1954). *Théorie de L'addition des Variables Aléatoires.* Gauthier-Villars, Paris.

POLLARD, D. (1982). A central limit theorem for empirical processes. *J. Austral. Math. Soc. Ser.* A **33** 235–248.

SMIRNOV, N. V. (1944). Approximation of the laws of distributions of random variables by empirical data (in Russian). *Uspekhi Mat. Nauk.* (Old series) **10** 179–206.

STOUT, W. (1974). *Almost Sure Convergence.* Academic, New York.

STUTE, W. (1977). Convergence rates for the isotrope discrepancy. *Ann. Probab.* **5** 707–723.

VAPNIK, V. N. and ČERVONENKIS, A. YA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264–280 (*Teor. Verojatnost. i Primenen.* **16** 264–297, in Russian).

VAPNIK, V. N., and ČERVONENKIS, A. YA. (1981). Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theor. Probab. Appl.* **26** 532–553.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195