The Annals of Probability
1986, Vol. 14, No. 2, 598–611

SAMPLE MODULI FOR SET-INDEXED GAUSSIAN PROCESSES[1]

BY KENNETH S. ALEXANDER

University of Washington

Sample path behavior is studied for Gaussian processes W_P indexed by classes $\mathscr{C}$ of subsets of a probability space $(X, \mathscr{A}, P)$ with covariance $EW_P(A)W_P(B) = P(A \cap B)$. A function ψ is found in some cases such that $\limsup_{t \to 0} \sup\{|W_P(C)|/\psi(P(C)): C \in \mathscr{C}, P(C) \le t\} = 1$ a.s. This unifies and generalizes the LIL and Lévy's Hölder condition for Brownian motion, and some results of Orey and Pruitt for the Brownian sheet.

1. Introduction and statement of results. Let $\mathscr{C}$ be a class of measurable subsets of a probability space $(X, \mathscr{A}, P)$. There exists a Gaussian process W_P indexed by $\mathscr{C}$, defined on, say, $(\Omega, \mathscr{F}, \mathbb{P})$, with covariance

$$EW_P(A)W_P(B) = P(A \cap B).$$

For example, if $\mathscr{C}$ is $\{[0, x]: x \in [0,1]^d\}$ and P is uniform, then W_P is Brownian motion ($d = 1$) or a Brownian sheet ($d > 1$). Processes W_P or modifications thereof arise as limits of set-indexed empirical processes [Dudley (1978) and Giné and Zinn (1986)] or of partial sum processes [Pyke (1982), Bass and Pyke (1984), and Alexander and Pyke (1986)]. The canonical examples, due to Donsker (1951, 1952), are the convergence of the normalized empirical distribution function to a Brownian bridge, and the convergence of the partial-sum process $S_n(t) = n^{-1/2}\sum_{j \le nt} X_j$ (X_j, $j \ge 1$, iid) to Brownian motion.

There is of course a vast literature on the sample path behavior of Brownian motion. For the Brownian sheet, many results were obtained by Orey and Pruitt (1973). A study of the sample-path behavior of general processes W_P is therefore of interest for two reasons: first, to unify some known results for these special cases, and second, because more general index classes do arise naturally as limits, as mentioned above. Our study must begin with some definitions. For $\varepsilon > 0$, set

$$N(\varepsilon, \mathscr{C}, P) := \min\Big\{k: \text{there exist } C_1, \dots, C_k \in \mathscr{C} \text{ with } \min_{i \le k} P(C \,\triangle\, C_i) < \varepsilon \text{ for all } C \in \mathscr{C}\Big\};$$

$$H(\varepsilon, \mathscr{C}, P) := \log N(\varepsilon, \mathscr{C}, P).$$

H is called the *metric entropy* of $\mathscr{C}$ for the metric $d_P(A, B) := P(A \,\triangle\, B) = E(W_P(A) - W_P(B))^2$. A nondecreasing function η on $[0, \infty)$ with

$$\eta(0) = 0, \qquad \eta(x + y) \le \eta(x) + \eta(y) \tag{1.1}$$

is called a *sample modulus* for W_P if there exists almost surely a (random)

Received April 1984; revised April 1985.
[1]Research supported under NSF Grant No. MCS-83-11686.
AMS 1980 *subject classifications*. Primary 60G15; secondary 60G17.
Key words and phrases. Gaussian process, set-indexed process, sample modulus, Vapnik–Červonenkis class.

constant K such that

$$(1.2)\qquad |W_P(A) - W_P(B)| \le K\eta(P(A \vartriangle B)) \quad \text{for all } A, B \in \mathscr{C}$$

and is called a *local sample modulus at* ϕ if (1.2) is replaced by

$$|W_P(A)| \le K\eta(P(A)).$$

If $\phi \in \mathscr{C}$, then a sample modulus is also a local sample modulus at ϕ. The sample path behavior of W_P is intimately tied to the metric entropy of $\mathscr{C}$ by the following result of Dudley (1973):

$$(1.3)\qquad \eta(t) := \int_0^t H(\varepsilon^2, \mathscr{C}, P)^{1/2}\, d\varepsilon \text{ is a sample modulus for } W_P \text{ on } \mathscr{C},$$

and, under mild conditions on H, this η is optimal in the sense that $\eta(t) = O(\varphi(t))$ for any other sample modulus φ.

The class $\mathscr{C}$ is called a Vapnik and Červonenkis (or VC) class [see Vapnik and Červonenkis (1971)] if

$$(1.4)\qquad m^{\mathscr{C}}(n) := \sup\{\text{card}\{F \cap C\colon C \in \mathscr{C}\}\colon F \subset X, \text{card}(F) = n\} < 2^n$$

for some $n \ge 0$, that is if no n-point subset of X can be "cut up" by $\mathscr{C}$ into all 2^n of its subsets. The least $n \ge 0$ such that (1.4) holds is called the *index* of $\mathscr{C}$ and denoted $V(\mathscr{C})$. To avoid trivialities we always assume $V(\mathscr{C}) \ge 1$. Examples of VC classes in $\mathbb{R}^d$ include all ellipsoids, all rectangles, or all polyhedra with at most k sides (k fixed). If $\mathscr{C}$ is a VC class then so are $\{A \setminus B\colon A, B \in \mathscr{C}\}$, $\{A^c\colon A \in \mathscr{C}\}$, and $\{A \cup B\colon A, B \in \mathscr{C}\}$. A finite union of VC classes, or a subset of a VC class, is a VC class. These and further facts about VC classes can be found in Dudley (1978).

For our purposes, the key fact about VC classes is the following result of Dudley (1978) [as refined by Alexander (1984a)]: if $\mathscr{C}$ is a VC class, then

$$(1.5)\qquad N(\varepsilon, \mathscr{C}, P) \le \left(16\varepsilon^{-1}L(8\varepsilon^{-1})\right)^{V(\mathscr{C})-1}, \qquad 0 < \varepsilon \le 1,$$

where $Lx := \log\max(x, e)$. The fact that this bound is independent of P leads to a scaling property of the metric entropy which will enable us to improve on (1.3) in many cases, especially when we only want a local sample modulus at ϕ. Define for $0 < t \le 1$:

$$\mathscr{C}_t := \{C \in \mathscr{C}\colon P(C) \le t\},$$

$$E_t := \bigcup_{C \in \mathscr{C}_t} C,$$

$$a(t) := P(E_t) \vee t,$$

$$g(t) := a(t)/t.$$

[If E_t is not measurable, replace it throughout by a set $F_t \supset E_t$ with $P(F_t) = P^*(E_t)$.] g is called the *capacity function* of $\mathscr{C}$ (for P). The motivation for the name and definition of g will be given in Remark 1.1, following some further definitions. A slightly different definition of the capacity function was used in Alexander (1984b), where the processes under consideration were "tied down" like a Brownian bridge, but in those examples to follow which refer to results from that paper, this will not matter. Define probability measures

$$P_t(\cdot) := P(\cdot|E_t)$$

and observe that $P_t(C) = P(C)/P(E_t)$ for $C \in \mathscr{C}_t$. It follows using (1.5) that for any $\mathscr{D}_t \subset \mathscr{C}_t$, $\delta > 0$, and $0 < u \leq 1$,

$$\begin{aligned} N(ut, \mathscr{D}_t, P) &= N(ut/P(E_t), \mathscr{D}_t, P_t) \\ &\leq \left(16g(t)u^{-1}L(8g(t)u^{-1})\right)^{V(\mathscr{C})-1} \qquad (1.6) \\ &= O\left(g(t)^{V(\mathscr{C})-1+\delta}\right) \quad \text{as } t \to 0. \end{aligned}$$

Defining

$$\rho_1(\mathscr{C}) := \inf\{\rho \geq 0: N(ut, \mathscr{C}_t, P) = O(g(t)^{\rho}) \text{ as } t \to 0 \text{ for all sufficiently small } u > 0\}$$

it follows that $\rho_1(\mathscr{C}) \leq V(\mathscr{C}) - 1$.

For technical reasons (specifically, the possibility of large jumps in g) we must also consider a slight variant of the capacity function. Define

$$\mathscr{C}_t^* := \{C \in \mathscr{C}: P(C) < t\}$$

and define g^* similarly to g, but using $\mathscr{C}_t^*$ in place of $\mathscr{C}_t$. Then set

$$\rho_2(\mathscr{C}) := \inf\{\rho \geq 0: N(ut, \mathscr{C}_t^*, P) = O(g^*(t)^{\rho}) \text{ as } t \to 0 \text{ for all sufficiently small } u > 0\},$$

and note that, analogously to (1.6), $\rho_2(\mathscr{C}) \leq V(\mathscr{C}) - 1$. Therefore

$$\rho(\mathscr{C}) := \max(\rho_1(\mathscr{C}), \rho_2(\mathscr{C})) \leq V(\mathscr{C}) - 1. \qquad (1.7)$$

We say a positive function φ on $(0,1]$ is *approximately nonincreasing* if $\varphi(t) \sim \theta(t)$ as $t \to 0$ for some nonincreasing function θ. Since $g(t-) \leq g^*(t) \leq g(t)$, if $Lg(t)$ is approximately nonincreasing then $Lg(t) \sim Lg^*(t)$ and $\rho_1(\mathscr{C}) = \rho_2(\mathscr{C}) = \rho(\mathscr{C})$.

For $r \geq 0$ we say C is *r-full* (for P and g) if for each $\lambda > 0$ there is an $\varepsilon_\lambda > 0$ such that for all sufficiently small t, there exist $k \geq \varepsilon_\lambda g(t)^r$ sets $C_1, \dots, C_k$ in $\mathscr{C}$ such that

$$P(C_i) = t \quad \text{for all } i, \qquad P(C_i \cap C_j) \leq \lambda t \quad \text{for all } i \neq j.$$

Observe that $\mathscr{C}$ is 0-full if for all small t there is a $C \in \mathscr{C}$ with $P(C) = t$, for we can then take $k = \varepsilon_\lambda = 1$. Define

$$r(\mathscr{C}) := \begin{cases} \sup\{r \geq 0: \mathscr{C} \text{ is } r\text{-full}\} & \text{if } g(t) \text{ is unbounded as } t \to 0, \\ 0 & \text{if } g \text{ is bounded.} \end{cases}$$

$$\tilde{\rho}(\mathscr{C}) := \begin{cases} \rho(\mathscr{C}) & \text{if } \rho(\mathscr{C}) > 0 \text{ or } Lg(t) = O(LLt^{-1}), \\ 1 & \text{otherwise} \end{cases}$$

taking $\sup \phi$ to be 0. Clearly $r(\mathscr{C}) \leq \rho(\mathscr{C})$. The reason for using $\tilde{\rho}(\mathscr{C})$ instead of

$\rho(\mathscr{C})$ is to ensure that $(tLg(t))^{1/2} = O(\psi(t))$ for the ψ of Theorem 1.2 below. The proof of that theorem [cf. (2.5) below] will show why this is necessary. In all our examples below for which values are calculated, ρ and $\tilde{\rho}$ are equal. Any example to the contrary would necessarily have $g(t)$ unbounded but $N(ut, \mathscr{C}_t, P) = o(g(t))$, which is somewhat pathological, as the following remark shows.

REMARK 1.1. Heuristically, $g(t) \approx P(E_t)/t$ is the maximum number of disjoint sets of size t which will "fit" in $\mathscr{C}$. In "regular" cases, then, $g(t)$ should be a lower bound on the number $N(ut, \mathscr{C}_t, P)$ of sets needed to approximate all sets of size at most t in $\mathscr{C}$ to within a fraction of t. If $\rho(\mathscr{C}) = 1$, then $g(t)$ is nearly also an upper bound.

The only intrinsic property of the capacity function g which we will actually use is that $a(t)/t \le g(t) \le 1/t$. Therefore our results remain valid if g is replaced throughout [including in the definitions of $\rho(\mathscr{C})$ and "r-full"] by a function $g_0 \ge g$ with these properties. The same applies to g^*.

Here is our main result.

THEOREM 1.2. *Let $\mathscr{C}$ be a VC class with capacity function g, and*

$$\psi(t) := \left(2t\left(\tilde{\rho}(\mathscr{C})Lg(t) + LLt^{-1}\right)\right)^{1/2},$$
$$\varphi(t) := \left(2t\left(r(\mathscr{C})Lg(t) + LLt^{-1}\right)\right)^{1/2}. \tag{1.8}$$

Then

$$\limsup_{t\to 0} \sup\left\{|W_P(C)|/\psi(P(C)) \colon C \in \mathscr{C}, P(C) \le t\right\} \le 1 \quad a.s. \tag{1.9}$$

and if $\mathscr{C}$ is 0-*full,*

$$\limsup_{t\to 0} \sup\left\{W_P(C)/\varphi(P(C)) \colon C \in \mathscr{C}, P(C) \le t\right\} \ge 1 \quad a.s. \tag{1.10}$$

If

$$LLt^{-1} = o(Lg(t)) \quad and \quad r(\mathscr{C}) > 0 \tag{1.11}$$

then "lim sup" *may be replaced by* "lim inf" *in* (1.10) *along any subsequence* (t_n) *approaching* 0 *at least geometrically fast* [*i.e.*, $n = O(Lt_n^{-1})$].

Define

$$\bar{g}(t) := \sup\{g(s) \colon s \ge t\}$$
$$\bar{\psi}(t) := \left(2t\left(\tilde{\rho}(\mathscr{C})L\bar{g}(t) + LLt^{-1}\right)\right)^{1/2}.$$

If $Lg(t)$ is unbounded and approximately nonincreasing then $Lg(t) \sim L\bar{g}(t)$. If $Lg(t)$ is approximately nonincreasing or $Lg(t) = o(LLt^{-1})$, then $\psi(t) \sim \bar{\psi}(t)$.

COROLLARY 1.3. *Let $\psi(t)$ be as in* (1.8) *and $\bar{\psi}$ as above. Then $\bar{\psi}$ is a local sample modulus at ϕ for W_P. If $r(\mathscr{C}) = \rho(\mathscr{C}) > 0$ or $Lg(t) = o(LLt^{-1})$, then*

$$\limsup_{t\to 0} \sup\left\{|W_P(C)|/\psi(P(C)) \colon P(C) = t\right\} = 1 \quad a.s.$$

If also (1.11) *holds, then*

$$\limsup_{n\to\infty}\left\{|W_P(C)|/(2P(C)Lg(P(C)))^{1/2}\colon P(C)=t_n\right\}=\rho(\mathscr{C})^{1/2}\quad a.s.$$

whenever $n = O(Lt_n^{-1})$.

Using only (1.5), the best that (1.3) can tell us is that $(tLt^{-1})^{1/2}$ is a sample modulus for W_P on $\mathscr{C}$. Locally at ϕ, Corollary 1.3 is an improvement on this whenever $Lg(t) = o(Lt^{-1})$.

The values of g, $r(\mathscr{C})$, and $\rho(\mathscr{C})$ in the first three of the following examples were essentially established in Alexander (1984b).

EXAMPLE 1.4. $X = [0,1]^d$, $s \in [0,1]^d$, $\mathscr{C} = \{[s,x]\colon x \in [0,1]^d\}$, and P uniform. (Here $[s,x]$ is the k-dimensional rectangle with opposite corners at s and x, even if s is not "below" x.) If $s = 0$, then $W(x) := W_P([0,x])$ is a Brownian sheet. We have $g(t) \sim 2^{n(s)}(Lt^{-1})^{d-1}/(d-1)!$ where $n(s) \le d$ is the number of noninteger coordinates s has, and $\rho(\mathscr{C}) = r(\mathscr{C}) = 1$. By Corollary 1.3,

$$\limsup_{t\to 0}\sup\left\{|W_P(C)|/\left(2dP(C)LLP(C)^{-1}\right)^{1/2}\colon P(C)=t\right\}=1\quad\text{a.s.}\tag{1.12}$$

For $d = 1$ (1.12) is the standard LIL for Brownian motion; for $d > 1$ it is due to Orey and Pruitt (1973).

EXAMPLE 1.5. $X = [0,1]^d$, $\mathscr{C} = \{[s,x]\colon s, x \in [0,1]^d\}$, and P uniform. Here W_P gives the increments over rectangles of the Brownian sheet. We have $g(t) = t^{-1}$ and $r(\mathscr{C}) = \rho(\mathscr{C}) = 1$. This time Corollary 1.3 says

$$\limsup_{t\to 0}\sup\left\{|W_P(C)|/\left(2P(C)LP(C)^{-1}\right)^{1/2}\colon P(C)=t\right\}=1\quad\text{a.s.}$$

For $d = 1$ this is Lévy's familiar Hölder condition or Brownian motion; for $d > 1$ it is a statement about the Brownian sheet essentially due to Orey and Pruitt (1973).

EXAMPLE 1.6. $X = \mathbb{R}^d$, P is a nondegenerate normal law, and $\mathscr{C}$ consists of all closed half spaces. Then $g(t) \sim K_d(Lt^{-1})^{(d-1)/2}$ for some constant K_d, and $r(\mathscr{C}) = \rho(\mathscr{C}) = 1$ if $d > 1$, 0 if $d = 1$. [Note here $V(\mathscr{C}) = d + 2$ (see Dudley (1978)), so (1.7) is not sharp in general.] By Corollary 1.3,

$$\limsup_{t\to 0}\sup\left\{|W_P(C)|/\left((d+1)P(C)LLP(C)^{-1}\right)^{1/2}\colon P(C)=t\right\}=1\quad\text{a.s.}$$

EXAMPLE 1.7. Let $X := \{x \in \mathbb{R}^d\colon |x| \le R\}$ $(d > 1)$, let P have density $f(x) = c_d/|x|^{d-1}$ on X (where c_d is a constant), and let $\mathscr{C} = \{C_x\colon x \in X\}$ where C_x is the ball with diameter endpoints at 0 and x. Here c_d is chosen so $P(C_x) = |x|$ for all x, and R is then chosen so $P(X) = 1$. Ossiander and Pyke (1984) showed that $W(x) := W_P(C_x)$ is Lévy's Brownian motion, a Gaussian process on X with $W(0) = 0$ and $E(W(x) - W(y))^2 = |x - y|$. Since $P(C_x)$ and $P(\{y\colon |y| \le |x|\})$ are both proportional to $|x|$, g is bounded. Hence according to

Corollary 1.3,

$$\limsup_{t\to 0}\sup\left\{|W(x)|/\left(2|x|LL|x|^{-1}\right)^{1/2}: |x| = t\right\} = 1 \quad \text{a.s.}$$

Since the increments of W have a translation-invariant law, it follows that for any fixed y,

$$\limsup_{t\to 0}\sup\left\{|W(x) - W(y)|/\left(2|x-y|LL|x-y|^{-1}\right)^{1/2}: |x-y| = t\right\} = 1 \quad \text{a.s.} \tag{1.13}$$

Equation (1.13) is due to Lévy (1948).

EXAMPLE 1.8. $X = \mathbb{R}^d$, $\mathscr{C} = \{(-\infty, x]: x \in \mathbb{R}^d\}$, and P arbitrary. Then W_F given by $W_F(x) := W_P((-\infty, x]) - F(x)W_P(\mathbb{R}^d)$ is the limit in law of the normalized empirical d.f. $n^{1/2}(F_n - F)$, where F is the d.f. of P. Here $V(\mathscr{C}) = d + 1$ [Wenocur and Dudley (1981)]. Hence by (1.7) and Theorem 1.2,

$$\limsup_{t\to 0}\sup\left\{|W_F(x)|/\psi(F(x)): F(x) = t\right\} \le 1 \quad \text{a.s.,}$$

where

$$\psi(t) := \left(2t\left(dL\left(t^{-1}P[F < t]\right) + LLt^{-1}\right)\right)^{1/2}.$$

Unfortunately, Example 1.4 (with $s = 0$) shows this is not sharp in general.

EXAMPLE 1.9. For Brownian motion $B(t)$ on $[0,1]$, intermediate between the LIL at 0 and Lévy's global Hölder condition, we can consider the sample modulus of B uniformly over some subset T of $[0,1]$. Thus we take $X = [0,1]$, P uniform, and $\mathscr{C} = \{[s,t]: s,t \in [0,1],\ s \text{ or } t \text{ in } T\}$, so $B(t) - B(s) = W_P([s,t])$. Then E_ε is $T^\varepsilon := \{s \in [0,1]: |s - t| < \varepsilon \text{ for some } t \in T\}$, the ε-neighborhood of T. It is not hard to see that $g(\varepsilon) = P(T^\varepsilon)/\varepsilon$ is decreasing, and $\rho(\mathscr{C}) = r(\mathscr{C}) = 1$, whatever T may be. Hence by Corollary 1.3,

$$\lim_{\varepsilon\to 0}\sup_{t\in T}\sup\left\{|B(s) - B(t)|/\psi(|s-t|): |s-t| \le \varepsilon\right\} = 1 \quad \text{a.s.,}$$

where

$$\psi(\varepsilon) := \left(2\varepsilon\left(L\left(\varepsilon^{-1}P(T^\varepsilon)\right) + LL\varepsilon^{-1}\right)\right)^{1/2}.$$

This result can also be obtained as a corollary to Theorem 1 of Mueller (1981). When T is the Cantor set, it can be shown that if $\varepsilon = 3^{-(n+1)}/2$, then $P(T^\varepsilon) = 2(2/3)^{n+1} - 2\varepsilon$. It follows that $L(P(T^\varepsilon)/\varepsilon) \sim (\log 2/\log 3)L\varepsilon^{-1}$, so $\psi(\varepsilon) \sim (2(\log 2/\log 3)\varepsilon L\varepsilon^{-1})^{1/2}$. This suggests any infinite T must be very "thin" if $L(\varepsilon^{-1}P(T^\varepsilon))$ is not the dominant of the two terms added in the above definition of ψ. But such T do exist. Let $T_0 := \{1/2\}$, $T_{n+1} := \{x \pm \exp(-\exp(e^n)): x \in T_n\}$, and $T := \bigcup_{n\ge 0}T_n$. Then without much difficulty we obtain $\varepsilon^{-1}P(T^\varepsilon) = O(LL\varepsilon^{-1})$, so $\psi(\varepsilon) \sim (2\varepsilon LL\varepsilon^{-1})^{1/2}$.

These examples show Theorem 1.2 unifies a number of results on sample moduli for particular Gaussian processes. In special cases, considerably more precise results are possible. For the increments of the Brownian sheet (Examples

1.4 and 1.5) Orey and Pruitt (1973) gave an integral test for upper and lower class functions. For the increments of Brownian motion over a variety of collections of intervals, Mueller (1981) gave integral tests and proved functional limit theorems. In both these cases, however, the proofs depend critically on the special ordered structure of the index family of sets, through the Markov properties of the Brownian sheet or of Brownian motion. The price we pay to free ourselves from dependence on the Markov property is the inability to obtain a result as precise as an integral test.

When we seek a global sample modulus for W_P on $\mathscr{C}$ the problem can sometimes be reduced to finding a local sample modulus at ϕ for a different class $\mathscr{C}'$. For example, to study Brownian motion we can take $\mathscr{C} = \{[0, x]: x \in [0,1]\}$ and $\mathscr{C}' = \{[x, y]: x < y \in [0,1]\}$. In general, since $W_P(A) - W_P(B) = W_P(A \setminus B) - W_P(B \setminus A)$, if η is a local sample modulus at ϕ for W_P on $\mathscr{C}' = \{A \setminus B: A, B \in \mathscr{C}\}$, then 2η is a (global) sample modulus on $\mathscr{C}$. When we seek a sharp upper bound as in Theorem 1.2, however, the added factor of 2 may be unsatisfactory, so we will give a modification of Theorem 1.2 which may be better.

Set

$$\tilde{N}(u, t, \mathscr{C}, P) = \min\Big\{k: \text{There exist } A_1, \dots, A_k, B_1, \dots, B_k \in \mathscr{C} \text{ such that}$$
$$\min_{i \le k}\big(P[(A \setminus B) \triangle (A_i \setminus B_i)] + P[(B \setminus A) \triangle (B_i \setminus A_i)]\big)$$
$$\le ut \text{ for all } A, B \in \mathscr{C} \text{ with } P(A \triangle B) \le t\Big\},$$

$$\rho_0(\mathscr{C}) := \inf\{\rho \ge 0: \tilde{N}(u, t, \mathscr{C}, P) = O(t^{-\rho}) \text{ as } t \to 0 \text{ for all sufficiently small } u > 0\},$$

$$\mathscr{C}' := \{A \triangle B: A, B \in \mathscr{C}\},$$

$$r_0(\mathscr{C}) := \sup\{r \ge 0: \mathscr{C}' \text{ is } r\text{-full (for } P \text{ and for } g(t) = t^{-1})\}.$$

Note $\tilde{N}(u, t, \mathscr{C}, P)$ is the number of pairs A_i, B_i needed to simultaneously approximate $A \setminus B$ and $B \setminus A$ to within ut for all A, B with $P(A \triangle B) \le t$. Since $P[(A \setminus B) \triangle (A_i \setminus B_i)] \le P(A \triangle A_i) + P(B \triangle B_i)$, we have

$$\tilde{N}(u, t, \mathscr{C}, P) \le N(ut/4, \mathscr{C}, P)^2,$$

so by (1.5),

$$\rho_0(\mathscr{C}) \le 2(V(\mathscr{C}) - 1).$$

Theorem 1.10. *Let $\mathscr{C}$ be a VC class and $\psi(t) := (2tLt^{-1})^{1/2}$. Then*

$$r_0(\mathscr{C})^{1/2} \le \liminf_{t \to 0} \sup\{|W_P(A) - W_P(B)|/\psi(P(A \triangle B)): P(A \triangle B) \le t\}$$
$$\le \limsup_{t \to 0} \sup\{|W_P(A) - W_P(B)|/\psi(P(A \triangle B)): P(A \triangle B) \le t\}$$
$$\le \rho_0(\mathscr{C})^{1/2} \quad a.s.$$

Example 1.11. The Brownian sheet: $X = [0,1]^d$, $\mathscr{C} = \{[0, x]: x \in [0,1]^d\}$, P uniform. We will show that $\rho_0(\mathscr{C}) \le d$. Write C_x for $[0, x]$ and C_{xy} for $C_x \setminus C_y$. Fix $0 < t \le 1$, $0 < u \le 1$, and $0 < \mu < 1$ with $\mu^{-d} - 1 < u/8$, and let m be the least positive integer with $m^{-1} < ut/16d$. Fix x, y with $P(C_x \triangle C_y) \le t$.

Let $I := \{i: x_i \geq y_i\}$ and

$$r_i := x_i \wedge y_i, \qquad s_i := |x_i - y_i|,$$
$$u_i := \max\{k/m: k/m \leq r_i,\ k \text{ an integer}\},$$
$$J := \min\{j: \mu^j \leq ut/16d\},$$
$$v_i := \min\{\mu^j: j \leq J,\ u_i + \mu^j \geq x_i \vee y_i\} \wedge (1 - u_i),$$
$$a_i := \begin{cases} u_i + v_i & \text{if } i \in I, \\ u_i & \text{if } i \notin I, \end{cases}$$
$$b_i := \begin{cases} u_i + v_i & \text{if } i \notin I, \\ u_i & \text{if } i \in I. \end{cases}$$

Observe that the rectangle whose corners include u, $u + v$, a, and b contains the rectangle with corresponding corners r, $r + s$, x, and y. We wish to show that

$$(1.14) \qquad P(C_{xy} \vartriangle C_{ab}) + P(C_{yx} \vartriangle C_{ba}) \leq ut.$$

Now $C_{ab} = C_{au}$ and $C_{xy} = C_{xr}$, so

$$(1.15) \quad P(C_{xy} \vartriangle C_{ab}) = P(C_{xr} \vartriangle C_{au}) = P(C_{au}) - P(C_{xr}) + 2P(C_{xr} \setminus C_{au}).$$

We have $\mu v_i \leq s_i + ut/8d$, so

$$\begin{aligned} P(C_{au}) &= \Big(\prod_{i \in I} a_i - \prod_{i \in I} u_i\Big) \prod_{i \notin I} u_i \\ &\leq \Big(\prod_{i \in I} \big((u_i + \mu^{-1}s_i + ut/4d) \wedge 1\big) - \prod_{i \in I} u_i\Big) \prod_{i \notin I} u_i \\ &\leq ut/4 + \Big(\prod_{i \in I} (u_i + \mu^{-1}s_i) - \prod_{i \in I} u_i\Big) \prod_{i \notin I} u_i \\ &\leq ut/4 + \mu^{-d}\Big(\prod_{i \in I} (u_i + s_i) - \prod_{i \in I} u_i\Big) \prod_{i \notin I} u_i \\ &\leq ut/4 + \mu^{-d}\Big(\prod_{i \in I} (r_i + s_i) - \prod_{i \in I} r_i\Big) \prod_{i \notin I} r_i \\ &\leq ut/4 + \mu^{-d} P(C_{xr}), \end{aligned}$$

while $P(C_{xr} \setminus C_{au}) \leq d/m \leq ut/16$. Hence by (1.15),

$$\begin{aligned} P(C_{xy} \vartriangle C_{ab}) &\leq ut/4 + (\mu^{-d} - 1)P(C_{xr}) + ut/8 \\ &\leq ut/2. \end{aligned}$$

Similarly, $P(C_{yx} \vartriangle C_{ba}) \leq ut/2$, and (1.14) follows.

It is clear that for a given I, the number of choices of (a, b) is at most $(m + 1)^d (J + 1)^d$ which is $O((t^{-1}Lt^{-1})^d) = O(t^{-(d+\delta)})$ as $t \to 0$. Since I takes only 2^d values, it follows that $\rho_0(\mathscr{C}) \leq d$.

Orey and Pruitt (1973) showed that for the Brownian sheet W and $\psi(t) := (2tLt^{-1})^{1/2}$,

$$(1.16) \quad \limsup_{t \to 0} \{|W(x) - W(y)|/\psi(P(C_x \vartriangle C_y)): P(C_x \vartriangle C_y) \leq t\} = d^{1/2} \quad \text{a.s.}$$

This would follow from Theorem 1.10 if we would show $r_0(\mathscr{C}) \geq d$.

Unfortunately, the best we have been able to do is $r_0(\mathscr{C}) \geq d - 1$, so Theorem 1.10 only proves the " $\leq$ " half of (1.16).

2. Proofs. We will only prove Theorem 1.2; the proof of Theorem 1.10 is similar but simpler, and Corollary 1.3 is immediate when we observe that (1.1) for $\bar{\psi}$ follows from monotonicity of $\bar{g}$.

To prove (1.9) it suffices to show

$$\mathbb{P}\left[\sup\{|W_P(C)|/\psi(P(C)) \colon P(C) \leq t\} > 1 + 4\varepsilon\right] \to 0 \tag{2.1}$$

for all $\varepsilon > 0$. Fix ε, $\beta \in (0, 1/12)$, choose η_0 small enough so

$$\eta_0 \sum_{j=0}^{\infty} (j+1)^{1/2} \beta^{j/2} \leq 1, \tag{2.2}$$

then $u < \varepsilon$ small enough so

$$\eta_0^2 \varepsilon^2 \geq 32 u \mu^{-3} V(\mathscr{C})(b + L(64/\mu\beta)), \tag{2.3}$$

where $\mu := 1 - u/2$ and $b := \sup_{t>0} 2tLg(t)/\psi^2(t) < \infty$. Fix $0 < t \leq 1$ and for each $j \geq 0$ let $I_j := (\mu^{j+1}t, \mu^j t]$ and $J_j := \{s \in [0,1] \colon \mu^{j+1} < a(s) \leq \mu^j\}$. Set

$$\mathscr{C}(j,k) := \{C \in \mathscr{C} \colon P(C) \in I_j \cap J_k\},$$
$$t_{jk} := \sup(I_j \cap J_k),$$
$$\psi_{jk} := \inf\{\psi(s) \colon s \in I_j \cap J_k\}$$

for those j, k for which $I_j \cap J_k \neq \phi$. Let $\mathscr{F}(j,k) \subset \mathscr{C}(j,k)$ such that

$$|\mathscr{F}(j,k)| = N(ut_{jk}, \mathscr{C}(j,k), P)$$

and such that for each $C \in \mathscr{C}(j,k)$ there is an $F_{jk}(C) \in \mathscr{F}(j,k)$ with $P(C \,\triangle\, F_{jk}(C)) < ut_{jk}$. Define

$$\mathscr{E}(j,k) := \{A \setminus B \colon A, B \in \mathscr{C}(j,k), P(A \setminus B) < u\mu^j t\}.$$

Define $k(j)$ to be the largest $k \geq 0$ for which $\mu^k \geq \mu^{j+1}t$. Since $a(s) \geq s$, $\mathscr{C}(j,k)$ is empty whenever $k > k(j)$.

Since W_P is finitely additive a.s., we have $|W_P(C)| \leq |W_P(F_{jk}(C))| + |W_P(C \setminus F_{jk}(C))| + |W_P(F_{jk}(C) \setminus C)|$, so

$$\begin{aligned}
&\mathbb{P}\left[\sup\{|W_P(C)|/\psi(P(C)) \colon P(C) \leq t\} > 1 + 4\varepsilon\right] \\
&\quad \leq \sum_{j=0}^{\infty} \sum_{k=0}^{k(j)} \mathbb{P}\left[\sup_{\mathscr{C}(j,k)} |W_P(C)| > (1+4\varepsilon)\psi_{jk}\right] \\
&\quad \leq \sum_{j=0}^{\infty} \sum_{k=0}^{k(j)} \mathbb{P}\left[\sup_{\mathscr{F}(j,k)} |W_P(C)| > (1+2\varepsilon)\psi_{jk}\right] \\
&\qquad + \sum_{j=0}^{\infty} \sum_{k=0}^{k(j)} \mathbb{P}\left[\sup_{\mathscr{E}(j,k)} |W_P(C)| > \varepsilon\psi_{jk}\right] \\
&\quad := (\mathrm{I}) + (\mathrm{II}).
\end{aligned}$$

[Summands for which $\mathscr{C}(j,k)$ is empty should be interpreted as 0.]

Suppose first $\tilde{\rho}(\mathscr{C}) > 0$. (I) is handled easily because $\mathscr{F}(j,k)$ is finite. Fix $0 < \delta < \tilde{\varepsilon}\rho(\mathscr{C})$. Presuming, as we may, that u is small enough (depending on δ), there exists an $A = A(\delta, u)$ such that the following holds: if $t_{jk} \in I_j \cap J_k$ then

$$\begin{aligned} N\big(ut_{jk}, \mathscr{C}(j,k), P\big) &\le N\big(ut_{jk}/2, \mathscr{C}_{t_{jk}}, P\big) \\ &\le Ag\big(t_{jk}\big)^{\rho_1(\mathscr{C})+\delta} \le A\big(\mu^{k-j-1}t^{-1}\big)^{\rho(\mathscr{C})+\delta}, \end{aligned}$$

while if $t_{jk} \notin I_j \cap J_k$ then

$$\begin{aligned} N\big(ut_{jk}, \mathscr{C}(j,k), P\big) &\le N\big(ut_{jk}/2, \mathscr{C}^*_{t_{jk}}, P\big) \\ &\le Ag^*\big(t_{jk}\big)^{\rho_2(\mathscr{C})+\delta} \le A\big(\mu^{k-j-1}t^{-1}\big)^{\rho(\mathscr{C})+\delta}. \end{aligned}$$

Further,

$$\begin{aligned} \psi_{jk}^2/2t_{jk} &\ge \mu\Big(\tilde{\rho}(\mathscr{C})L\big(\mu^{k+1-j}t^{-1}\big) + LL\big(\mu^j t\big)^{-1}\Big) \\ &\ge \mu^3\Big(\tilde{\rho}(\mathscr{C})L\big(\mu^{k-j-1}t^{-1}\big) + LL\big(\mu^j t\big)^{-1}\Big). \end{aligned} \tag{2.4}$$

Define $\gamma := \mu^{k(j)-j-1}t^{-1}$, so $1 \le \gamma < \mu^{-1}$. It follows from all this that

$$\begin{aligned} \text{(I)} &\le \sum_{j=0}^{\infty}\sum_{k=0}^{k(j)} |\mathscr{F}(j,k)| \exp\big(-(1+4\varepsilon)\psi_{jk}^2/2t_{jk}\big) \\ &\le \sum_{j=0}^{\infty}\sum_{k=0}^{k(j)} A\big(\mu^{k-j-1}t^{-1}\big)^{\rho(\mathscr{C})+\delta} \\ &\qquad \times \exp\Big(-(1+2\varepsilon)\Big(\tilde{\rho}(\mathscr{C})L\big(\mu^{k-j-1}t^{-1}\big) + LL\big(\mu^j t\big)^{-1}\Big)\Big) \\ &\le A\sum_{j=0}^{\infty}\sum_{k=0}^{k(j)} \exp\Big(-\varepsilon\tilde{\rho}(\mathscr{C})L\big(\mu^{k-j-1}t^{-1}\big) - (1+2\varepsilon)LL\big(\mu^j t\big)^{-1}\Big) \\ &\le A\sum_{j=0}^{\infty}\sum_{l=0}^{\infty} \exp\big(-\varepsilon\tilde{\rho}(\mathscr{C})L(\gamma\mu^{-l})\big)\exp\Big(-(1+2\varepsilon)LL\big(\mu^j t\big)^{-1}\Big) \\ &\le A\sum_{j=0}^{\infty}\big(1-\mu^{\varepsilon\tilde{\rho}(\mathscr{C})}\big)^{-1}\big(\log t^{-1} + j\log\mu^{-1}\big)^{-(1+2\varepsilon)} \\ &\to 0 \quad \text{as } t \to 0. \end{aligned} \tag{2.5}$$

For (II), for each j, k we apply a "chain argument" of the type used by Dudley (1973). Fix j and k and set

$$\delta_0 := ut_{jk}, \qquad \delta_i := \delta_0\beta^i \quad \text{for } i \ge 1,$$

$$\eta_i^2 := \eta_0^2(i+1)\beta^i.$$

For each $i \ge 1$ let $\mathscr{H}(i) \subset \mathscr{E}(j,k)$ such that

$$|\mathscr{H}(i)| = N\big(\delta_i, \mathscr{E}(j,k), P\big)$$

and such that for each $C \in \mathscr{E}(j,k)$ there is an $H_i(C) \in \mathscr{H}(i)$ with

$P(C \triangle H_i(C)) < \delta_i$. Set $\mathscr{H}(0) := \{\phi\}$ and $H_0(C) := \phi$ for all C. Now W_P is sample-continuous by (1.3) and (1.5), so $W_P(H_i(C)) \to W_P(C)$ for all $C \in \mathscr{E}(j,k)$ with probability one. Hence

$$|W_P(C)| \le \sum_{i=0}^{\infty} |W_P(H_{i+1}(C)) - W_P(H_i(C))|.$$

It follows using (2.2) that

$$\begin{aligned}
&\mathbb{P}\Big[\sup_{\mathscr{E}(j,k)} |W_P(C)| > \varepsilon\psi_{jk}\Big] \\
&\le \sum_{i=0}^{\infty} \mathbb{P}\Big[\sup_{\mathscr{E}(j,k)} |W_P(H_{i+1}(C)) - W_P(H_i(C))| > \varepsilon\eta_i\psi_{jk}\Big] \\
&\le \sum_{i=0}^{\infty} |\mathscr{H}(i+1)||\mathscr{H}(i)| \\
&\quad \times \max_{\mathscr{E}(j,k)} \mathbb{P}\big[|W_P(H_{i+1}(C)) - W_P(H_i(C))| > \varepsilon\eta_i\psi_{jk}\big].
\end{aligned} \tag{2.6}$$

Suppose first $t_{jk} \in I_j \cap J_k$. Now

$$\begin{aligned}
Lg(t_{jk}) &\le L\Big(\mu^{-2} \inf_{s \in I_j \cap J_k} g(s)\Big) \\
&\le \mu^{-2} b \inf_{s \in I_j \cap J_k} (\psi^2(s)/2s) \\
&\le \mu^{-3} b \psi_{jk}^2 / 2t_{jk}.
\end{aligned} \tag{2.7}$$

As in (1.6), using (1.5) for the second inequality, (2.7) for the fourth, and (2.3) for the last, we have for $i \ge 1$:

$$\begin{aligned}
|\mathscr{H}(i)| &= N\big(\delta_i/P(E_{t_{jk}}), \mathscr{E}(j,k), P_{t_{jk}}\big) \\
&\le N\big(\delta_i/4P(E_{t_{jk}}), \mathscr{C}_{t_{jk}}, P_{t_{jk}}\big)^2 \\
&\le \big(64a(t_{jk})\delta_i^{-1} L(32a(t_{jk})\delta_i^{-1})\big)^{2(V(\mathscr{C})-1)} \\
&\le \big(64g(t_{jk})u^{-1}\beta^{-i}\big)^{4V(\mathscr{C})} \\
&\le \exp\big(4iV(\mathscr{C})\mu^{-3}(L(64/\mu\beta) + b)\psi_{jk}^2/2t_{jk}\big) \\
&\le \exp\big(\eta_0^2\varepsilon^2 i \psi_{jk}^2/16ut_{jk}\big).
\end{aligned} \tag{2.8}$$

Since $P(H_i(C) \triangle H_{i+1}(C)) \le \delta_i + \delta_{i+1} \le 2\delta_i$ for all C,

$$\begin{aligned}
&\mathbb{P}\big[|W_P(H_{i+1}(C)) - W_P(H_i(C))| > \varepsilon\eta_i\psi_{jk}\big] \\
&\quad \le 2\exp\big(-\varepsilon^2\eta_i^2\psi_{jk}^2/4\delta_i\big) \\
&\quad = 2\exp\big(-\varepsilon^2\eta_0^2(i+1)\psi_{jk}^2/4ut_{jk}\big).
\end{aligned}$$

Combining this with (2.6), (2.8), (2.4), and (2.3) we see that

$$\mathbb{P}\left[\sup_{\mathscr{E}(j,k)} |W_P(C)| > \varepsilon\psi_{jk}\right]$$
$$\leq \sum_{i=0}^{\infty} 2\exp\left(-\varepsilon^2\eta_0^2(i+1)\psi_{jk}^2/8ut_{jk}\right)$$
$$\leq 4\exp\left(-\varepsilon^2\eta_0^2\psi_{jk}^2/8ut_{jk}\right)$$
$$\leq 4\exp\left(-\psi_{jk}^2/t_{jk}\right).$$

If $t_{jk} \notin I_j \cap J_k$, the same bound can be obtained by using $g^*(t_{jk})$ instead of $g(t_{jk})$ in the bound on $|\mathscr{H}(i)|$. As in (2.5), this shows (II) $\to 0$ as $t \to 0$, and (2.1) and (1.9) follow.

When $\tilde{\rho}(\mathscr{C}) = 0$, we have $Lg(s) \leq KLLs^{-1}$ for all $0 < s \leq 1$ for some constant K. We can then obtain (2.1) as above, with the following changes: the probability in (2.1) is broken up into a sum over j only (not j and k). We take $0 < \delta < \varepsilon/K$, which enables us [in an analog of (2.5)] to absorb the metric entropy term [corresponding to the $|\mathscr{F}(j,k)|$ of (2.5)] into the term $(1+2\varepsilon)LL(\mu^j t)^{-1}$ in the exponent [instead of into the $(1+2\varepsilon)\tilde{\rho}(\mathscr{C})L(\mu^{k-j-1}t^{-1})$ term as done in (2.5)]. Thus (1.9) is proved in all cases.

To prove (1.10), it suffices to show

$$\text{(2.9)}\qquad \mathbb{P}\left[\sup\{W_P(C)/\varphi_r(P(C)) \colon P(C) \leq t\} < 1 - 3\lambda\right] \to 0 \quad \text{as } t \to 0$$

for all $\lambda > 0$ and all $r \geq 0$ such that $\mathscr{C}$ is r-full, where

$$\varphi_r(t) := \left(2t\left(rLg(t) + LLt^{-1}\right)\right)^{1/2}.$$

Fix such an r, and $\lambda < 1/4$, and $t > 0$ small enough so the definition of r-full applies, and $\mu \in (0, \lambda^2)$. For each $j \geq 0$ there exists a collection $\{C_{ji}\colon 1 \leq i \leq m(j)\} \subset \mathscr{C}$ with $m(j) \geq \varepsilon_\lambda g(\mu^j t)^r$, $P(C_{ji}) = \mu^j t$, $P(C_{ji} \cap C_{jk}) \leq \lambda\mu^j t$ for all $i \neq k$. Let X and $\{Y_{ji}\colon j \geq 0, 1 \leq i \leq m(j)\}$ be independent mean-0 normal random variables with

$$EX^2 = \lambda t, \qquad EY_{ji}^2 = (1-\lambda)\mu^j t,$$

and set $Z_{ji} := \mu^{j/2}X + Y_{ji}$. Then

$$EW_P(C_{ji})W_P(C_{kl}) \leq \begin{Bmatrix} \mu^j t & \text{if } j = k,\ i = l \\ \lambda\mu^j t & \text{if } j = k,\ i \neq l \\ \mu^{(j+k)/2}\lambda t & \text{if } j \neq k \end{Bmatrix} = EZ_{ji}Z_{kl}$$

for all j, i, k, l, with equality when $j = k$, $i = l$. Set $M := M(t) := \inf_{s \leq t} \lambda^{1/2}\varphi_r(s)/s^{1/2}$ and suppose t is small enough (i.e., M large enough) so

$$\text{(2.10)}\qquad 1 - \Phi(u) \geq \exp\left(-(1+\lambda)u^2/2\right) \quad \text{for all } u \geq M.$$

Observe that

$$\text{(2.11)}\qquad Ms^{1/2}\lambda^{1/2} \leq \lambda\varphi_r(s) \quad \text{for all } s \leq t.$$

By Slepian's (1962) inequality and (2.11),

$$
\begin{aligned}
&\mathbb{P}\big[\sup\{W_P(C)/\varphi_r(P(C)) \colon P(C) \le t\} < 1 - 3\lambda\big] \\
&\quad \le \mathbb{P}\big[\sup\{W_P(C_{ji})/\varphi_r(\mu^j t) \colon j \ge 0, i \le m(j)\} < 1 - 3\lambda\big] \\
&\quad \le \mathbb{P}\big[Z_{ji} \le (1 - 3\lambda)\varphi_r(\mu^j t) \text{ for all } j \ge 0, i \le m(j)\big] \\
&\quad \le \mathbb{P}\big[Y_{ji} \le (1 - 3\lambda)\varphi_r(\mu^j t) + M\mu^{j/2}t^{1/2}\lambda^{1/2} \text{ for all } j \ge 0, i \le m(j)\big] \\
&\qquad + \mathbb{P}\big[X < -Mt^{1/2}\lambda^{1/2}\big] \qquad (2.12)\\
&\quad \le \prod_{j \ge 0} \mathbb{P}\big[Y_{j1} \le (1 - 2\lambda)\varphi_r(\mu^j t)\big]^{m(j)} + 1 - \Phi(M(t)) \\
&\quad \le \exp\Big(-\sum_{j \ge 0} m(j)p_j\Big) + 1 - \Phi(M(t)),
\end{aligned}
$$

where $p_j := \mathbb{P}[Y_{j1} > (1 - 2\lambda)\varphi_r(\mu^j t)] = 1 - \Phi(u_j)$ and $u_j := (1 - 2\lambda)\varphi_r(\mu^j t)/((1 - \lambda)\mu^j t)^{1/2}$.

Now $u_j \ge M$ by (2.11), so by (2.10),

$$
\begin{aligned}
m(j)p_j &\ge \varepsilon_\lambda g(\mu^j t)^r \exp\big(-(1 + \lambda)u_j^2/2\big) \\
&\ge \varepsilon_\lambda g(\mu^j t)^r \exp\Big(-(1 - \lambda)\big(rLg(\mu^j t) + LL(\mu^j t)^{-1}\big)\Big) \qquad (2.13)\\
&\ge \varepsilon_\lambda \exp\Big(\lambda rLg(\mu^j t) - (1 - \lambda)LL(\mu^j t)^{-1}\Big), \\
&\ge \varepsilon_\lambda\big(Lt^{-1} + jL\mu^{-1}\big)^{-(1-\lambda)}.
\end{aligned}
$$

Hence $\Sigma m(j)p_j = \infty$, so (2.9) and then (1.10) follow from (2.12).

To prove the last statement in the theorem, repeat the above calculation but use $j = 0$ only. By (1.11) we may assume $r > 0$. As in (2.12) we obtain

$$
\begin{aligned}
&\mathbb{P}\big[\sup\{W_P(C)/\varphi_r(P(C)) \colon P(C) = t\} < 1 - 3\lambda\big] \qquad (2.14)\\
&\qquad \le \exp(-m(0)p_0) + 1 - \Phi(M(t)).
\end{aligned}
$$

By (1.11) and the third inequality in (2.13), if t is small then

$$
m(0)p_0 \ge \varepsilon_\lambda \exp(\lambda rLg(t)/2) \ge \exp(2LLt^{-1}),
$$

while

$$
\begin{aligned}
1 - \Phi(M(t)) &\le \exp\big(-M(t)^2/2\big) \le \exp\Big(-\lambda r \inf_{s \le t} Lg(s)\Big) \\
&\le \exp(-2LLt^{-1}).
\end{aligned}
$$

Therefore the sum of (2.14) over n with $t = t_n$ converges whenever $n = O(Lt_n^{-1})$, and Borel–Cantelli finishes the proof. □

The proof above of the lower bound (1.10) is based on finding enough "sufficiently disjoint" sets in $\mathscr{C}$ to ensure that the supremum over these sets is not small. What we mean by "sufficiently disjoint" is made precise by our concept of "r-full". Other definitions of "sufficiently disjoint" could give rise to alternative lower bounds which at times might improve on (1.10). For example, suppose there exist $\mathscr{C}_1, \dots, \mathscr{C}_k$ such that if $C_i \in \mathscr{C}_i$ for all i then $\bigcup_{i \le k} C_i$ is a disjoint union and is in $\mathscr{C}$. Suppose there is a function φ_0 such that (1.10) holds with $\varphi = \varphi_0$ for each $\mathscr{C}_i$. Roughly, given t we can by (1.10) find $C_i \in \mathscr{C}_i$ with $P(C_i) = t/k$ and $W_P(C_i) \approx \varphi_0(t/k)$, so $C = \bigcup_{i \le k} C_i$ satisfies $W_P(C) \approx k\varphi_0(t/k)$. Thus (1.10) holds for $\mathscr{C}$ with $\varphi(t) := k\varphi_0(t/k)$. This is essentially the method used by Orey and Pruitt (1973) to obtain the lower bound in (1.16), which we could not obtain by our methods. No one definition of "sufficiently disjoint" will work best in all cases.

Acknowledgment. The author would like to thank the referee for some corrections to the main proof.

REFERENCES

ALEXANDER, K. S. (1984a). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067.

ALEXANDER, K. S. (1984b). Rates of growth and sample moduli for weighted empirical processes indexed by sets. Preprint.

ALEXANDER, K. S. and PYKE, R. (1986). A uniform central limit theorem for set-indexed partial-sum processes with finite variance. *Ann. Probab.* **14** 582–597.

BASS, R. and PYKE, R. (1984). Functional law of the iterated logarithm and uniform central limit theorem for partial-sum processes indexed by sets. *Ann. Probab.* **12** 13–34.

DONSKER, M. D. (1951). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* **6** 1–12.

DONSKER, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **23** 277–281.

DUDLEY, R. M. (1973). Sample functions of the Gaussian process. *Ann. Probab.* **1** 66–103.

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

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

LÉVY, P. (1948). *Processus Stochastiques et Mouvement Brownien.* Gauthier-Villars, Paris.

MUELLER, C. (1981). A unification of Strassen's law and Lévy's modulus of continuity. *Z. Wahrsch. verw. Gebiete* **56** 163–179.

OREY, S. and PRUITT, W. E. (1973). Sample functions of the N-parameter Wiener process. *Ann. Probab.* **1** 138–163.

OSSIANDER, M. and PYKE, R. (1984). Lévy's Brownian motion as a set-indexed process. Technical Report 42, Dept. Mathematics, Univ. Washington.

PYKE, R. (1982). A uniform central limit theorem for partial-sum processes indexed by sets. *London Math. Soc. Lect. Note Series* **79** 219–240.

SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.* **41** 463–501.

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.

WENOCUR, R. S. and DUDLEY, R. M. (1981). Some special Vapnik–Chervonenkis classes. *Discrete Math.* **33** 313–318.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195