Probab. Th. Rel. Fields 75, 379–423 (1987)

Probability Theory and Related Fields

Rates of Growth and Sample Moduli for Weighted Empirical Processes Indexed by Sets*

Kenneth S. Alexander**

University of Washington, Department of Statistics, Seattle, WA 98195, USA

Summary. Probability inequalities are obtained for the supremum of a weighted empirical process indexed by a Vapnik-Červonenkis class $\mathscr{C}$ of sets. These inequalities are particularly useful under the assumption $P(\cup\{C\in\mathscr{C}: P(C)<t\})\to 0$ as $t\to 0$. They are used to obtain almost sure bounds on the rate of growth of the process as the sample size approaches infinity, to find an asymptotic sample modulus for the unweighted empirical process, and to study the ratio P_n/P of the empirical measure to the actual measure.

I. Introduction

Let $X_1, X_2, \ldots$ be i.i.d. random variables with law P taking values in a space $(X, \mathscr{A})$, and let $\mathscr{C}\subset\mathscr{A}$ be a class of sets. Define the n-th empirical measure and process:

$$P_n = n^{-1}\sum_{i=1}^{n}\delta_{X_i}, \qquad \nu_n = n^{\frac{1}{2}}(P_n - P).$$

ν_n may be viewed as a stochastic process indexed by $\mathscr{C}$. When $\mathscr{C}$ is $\mathscr{D}_d = \{(-\infty, x]: x\in\mathbb{R}^d\}$, ν_n becomes a normalized empirical distribution function; we call this the *d.f. case*. Properties of ν_n in this case, and in the related case where $\mathscr{C}$ is the class $\mathscr{I}_d$ of all subrectangles of $\mathbb{R}^d$, have been extensively studied. Recently, attention has been given to more general classes of sets or functions, both in the theory (Dudley, 1978; Giné and Zinn, 1984; Le Cam, 1983; Pollard, 1982, 1984; and Vapnik and Červonenkis, 1971, 1981) and in the statistical applications, primarily to nonparametric regression (Breiman et al., 1984), density estimation (Alexander 1985; Pollard 1984; Yukich 1985); and projection pursuit (Diaconis and Freedman 1982; Huber 1985). It is in the more general setting that we work here.

* Research supported under an NSF Postdoctoral Fellowship grant No. MCS83-11686, and in part by NSF grant No. DMS-8301807

** *Current address*: Department of Mathematics, University of Southern California, Los Angeles, CA 90089, USA

To obtain detailed information about the behavior of ν_n on small sets, it is often helpful to weight ν_n at each set C by a function of $\sigma^2(C)=P(C)(1-P(C))$, the variance of $\nu_n(C)$. Since $\sigma^2(C)\sim P(C)$ as $P(C)\to 0$, this is often equivalent to weighting by the same function of $P(C)$; when convenient we do the latter. In particular, given a nonnegative nondecreasing function $q\in C[0,1]$ and sequences $\gamma_n\to 0$ and (α_n), we may ask for a finite R and a sequence (b_n) such that

$$\limsup_n \sup\{|\nu_n(C)|/b_n q(\sigma^2(C)): C\in\mathscr{C},\ \gamma_n\leqq\sigma^2(C)\leqq\alpha_n\}=R \quad \text{a.s.} \tag{1.1}$$

or such that

$$\mathbb{P}[|\nu_n(C)|>b_n q(\sigma^2(C)) \quad \text{for some } C\in\mathscr{C} \text{ with } \sigma^2(C)\geqq\gamma_n]\to 0, \tag{1.2}$$

perhaps at a particular rate.

From another angle, we may ask for a function q such that (1.1) or (1.2) hold with $b_n\equiv b$ for some $0<b<\infty$, i.e. such that $\{|\nu_n(C)|/q(\sigma^2(C)): C\in\mathscr{C}, \gamma_n\leqq\sigma^2(C)\leqq\alpha_n\}$ remains bounded, a.s. or in probability, as $n\to\infty$. Such a q (or more precisely, the function $q(t^2)$) acts as a sort of asymptotic sample modulus for ν_n on $\mathscr{C}$. Thirdly, we might ask for the best range $[\gamma_n,\alpha_n]$ of "sizes" of sets $C\in\mathscr{C}$ for which (1.1) or (1.2) is valid.

Many special cases of these questions have been considered. For example, Shorack and Wellner (1982) proved that for P uniform on $[0,1]$ and $\varepsilon>0$,

$$\limsup_n \sup\{|\nu_n(C)|/(P(C)Ln)^{\frac{1}{2}}: C\in\mathscr{I}_1,\ P(C)\geqq\varepsilon n^{-1}Ln\}<\infty \quad \text{a.s.} \tag{1.3}$$

and for $-\infty<\beta<1$,

$$\limsup_n \sup\left\{\frac{|\nu_n(C)|}{P(C)^{\frac{1}{2}}}\frac{LLn}{(Ln)^{1-\beta/2}}: C\in\mathscr{I}_1,\ P(C)\geqq n^{-1}(Ln)^{\beta}\right\}<\infty \quad \text{a.s.} \tag{1.4}$$

where Lx denotes $\log(\max(x,e))$. Stute (1982a) showed that if $\gamma_n\downarrow 0$ with $n\gamma_n\uparrow\infty, n^{-1}Ln=o(\gamma_n)$, and $LLn=o(L\gamma_n^{-1}), P$ is uniform on $[0,1]$, and $0<\beta\leqq\theta<\infty$, then

$$\limsup_n \{|\nu_n(C)|/(2P(C)L\gamma_n^{-1})^{\frac{1}{2}}: C\in\mathscr{I}_1,\ \beta\gamma_n\leqq P(C)\leqq\theta\gamma_n\}=1 \quad \text{a.s.} \tag{1.5}$$

This is generalized to d dimensions in Stute (1984). (1.5) was used to obtain exact rates of convergence for kernel density estimators (Stute 1982b). Note that (1.5) is equivalent to the statement that

$$\limsup_n \{|\nu_n(C)|/(2P(C)\ LP(C)^{-1})^{\frac{1}{2}}: C\in\mathscr{I}_1,\ \beta\gamma_n\leqq P(C)\leqq\theta\gamma_n\}=1 \quad \text{a.s.} \tag{1.6}$$

which essentially says that the function $t(Lt^{-2})^{\frac{1}{2}}$ is a local asymptotic modulus of continuity for ν_n on $\mathscr{I}_1$. A precise definition of "local asymptotic modulus" will be given in Sect. 4.

Van Zuijlen (1982) showed that in the d-dimensional d.f. case, for each $\delta>0$ there exists K and n_0 such that

$$\mathbb{P}[\sup\{|\nu_n(C)|/\sigma(C)Ln: C\in\mathscr{D}_d,\ 3n^{-1}\leqq P(C)\leqq 1-3n^{-1}\}>K]<n^{-(1+\delta)} \tag{1.7}$$

and

$$\mathbb{P}[P_n(C)\leqq(KLn)^{-1}P(C) \quad \text{for some } C\in\mathscr{D}_d \text{ with } P_n(C)\neq 0]<n^{-(1+\delta)} \tag{1.8}$$

for all $n \geqq n_0$. These results were applied to the asymptotic theory of rank statistics. Breiman et al. (1984) showed that if $\mathscr{C}$ is a Vapnik-Červonenkis, or "VC", class of sets (defined below), then for $\delta, \varepsilon > 0$ there exist K and n_0 such that

$$\mathbb{P}[\sup\{|P_n(C)/P(C)-1|: P(C) \geqq Kn^{-1}Ln,\ C \in \mathscr{C}\} > \varepsilon] = O(n^{-(1+\delta)}). \tag{1.9}$$

(To express this in the form of (1.2), take $b_n = n^{\frac{1}{2}}$.)

In Alexander (1985) the upper bounds in (1.3)–(1.9) were extended to more general classes of sets and functions, including VC classes. The growth constants and cutoff levels (the (b_n), (γ_n), and (α_n) in (1.1) or (1.2)) remain the same for VC classes as they are in the special cases (1.3)–(1.9). These extensions, however, do not give the full story, for the growth constants implicit in (1.3)–(1.9) are only upper bounds, except for (1.5). They are not sharp for all $\mathscr{C}$ and P, as the following shows.

For $\tau > 0$ define β_τ to be the solution $\beta > 1$ of $\beta(\log\beta - 1) = (1-\tau)/\tau$, and set $\beta_\infty = 1$. Then

$$\begin{aligned} &\beta_\tau \sim (\tau L\tau^{-1})^{-1} && \text{as} && \tau \to 0, \\ &\beta_\tau - 1 \sim (2\tau^{-1})^{\frac{1}{2}} && \text{as} && \tau \to \infty, \\ &\beta_\tau - 1 \geqq (2\tau^{-1})^{\frac{1}{2}} && \text{for all} && \tau > 0, \end{aligned} \tag{1.10}$$

and

$$(\beta_\tau - 1)\, h_1(\beta_\tau - 1) = \tau^{-1} \tag{1.11}$$

where h_1 is given by

$$h_1(\lambda) = (1+\lambda^{-1})\log(1+\lambda) - 1, \qquad \lambda \geqq 0 \tag{1.12}$$

(see Shorack and Wellner 1982).

Csáki (1977) established that in the one-dimensional d.f. case with P uniform on $[0,1]$,

$$\limsup_n \sup\{|\nu_n(C)|/b_n\sigma(C): C \in \mathscr{D}_1,\ \sigma^2(C) \geqq \gamma_n\} = R \quad \text{a.s.} \tag{1.13}$$

where

$$\begin{aligned} &R = (2(a+1))^{\frac{1}{2}} \text{ and } b_n = (LLn)^{\frac{1}{2}} \text{ if } n^{-1}LLn = o(\gamma_n) \text{ and } LL\gamma_n^{-1}/LLn \to a \\ &R = \max(2, \tau^{\frac{1}{2}}(\beta_\tau - 1)) \text{ and } b_n = (LLn)^{\frac{1}{2}} \text{ if } \gamma_n = rn^{-1}LLn \text{ for all } n \\ &R = 1 \text{ and } b_n = LLn/((n\gamma_n)^{\frac{1}{2}} L(LLn/n\gamma_n)) \quad \text{if} \\ &\qquad \gamma_n = o(n^{-1}LLn) \text{ and } LLn/L(LLn/n\gamma_n) \uparrow \infty. \end{aligned} \tag{1.14}$$

Wellner (1978) showed that, for P uniform on $[0,1]$, if $n^{-1} = o(\gamma_n)$ then

$$\sup\{|P_n(C)/P(C)-1|: C \in \mathscr{D}_1,\ P(C) \geqq \gamma_n\} \to 0 \text{ in probability}, \tag{1.15}$$

while if $n^{-1}LLn = o(\gamma_n)$, then

$$\sup\{|P_n(C)/P(C)-1|: C \in \mathscr{D}_1,\ P(C) \geqq \gamma_n\} \to 0 \quad \text{a.s.} \tag{1.16}$$

Note that the growth rates (b_n) in (1.13) differ from those in (1.3)–(1.6) and the cutoff levels (γ_n) in (1.15) and (1.16) differ from those in (1.9); we would like to understand such differences from a general point of view. In this paper we present an approach which unifies all the results (1.3)–(1.9), (1.13), and (1.15)–(1.16). We will extend them to all VC classes of sets, including extension to higher dimensions of the d.f. and interval cases, and show how to choose optimal q, (b_n), and (γ_n) in general.

The underlying idea is that the right local asymptotic modulus for v_n, call it ψ_1, should be the oscillation modulus for the Gaussian process G_P which is the weak limit of v_n on $\mathscr{C}$. That is, $q_1(t)=\psi_1(t^{\frac{1}{2}})$ should satisfy

$$0<\limsup_{t\to 0}\sup\{|G_P(C)|/q_1(\sigma^2(C)): C\in\mathscr{C},\ \sigma^2(C)\leqq t\}<\infty \quad \text{a.s.} \quad (1.17)$$

The problem of finding such a q_1 was considered in Alexander (1986). The main result can be summarized as follows.

Given a class $\mathscr{C}$ and a law P on $(X, \mathscr{A})$, define for $t\geqq 0$:

$$\mathscr{C}_t=\{C\in\mathscr{C}:\sigma^2(C)\leqq t, P(C)\leqq\tfrac{1}{2}\}\cup\{C^c: C\in\mathscr{C},\ \sigma^2(C)\leqq t,\ P(C)>\tfrac{1}{2}\},$$

$$\mathscr{C}_{t,s}=\{C\backslash D: C, D\in\mathscr{C}_t,\ \sigma^2(C\backslash D)\leqq s\},$$

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

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

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

We may assume E_t is measurable; if not, then replace it throughout by a measurable $F_t\supset E_t$ with $P(F_t)=P^*(E_t)$. We call g the *capacity function* of $\mathscr{C}$ (for P). This is because $g(t)$ can be thought of roughly as the number of disjoint sets of size t which will "fit" in $\mathscr{C}: a(t)$ is the space available, and t is the approximate space needed for each set C with $P(C)\approx\sigma^2(C)=t$. Thus when we approximate all sets in $\mathscr{C}_t$ using a finite subcollection, $g(t)$ should give a lower bound on the number needed. This is quantified in Sect. 3 using the concept of a *full* class $\mathscr{C}$. Note that since $v_n(C)=-v_n(C^c)$, one can often simplify matters, especially the definition of $\mathscr{C}_t$, by imposing the condition that $P(C)\leqq\frac{1}{2}$ for all $C\in\mathscr{C}$, then considering separately those $C\in\mathscr{C}$ with $P(C)\leqq\frac{1}{2}$ and the complement of those with $P(C)>\frac{1}{2}$. This means $\sigma^2(C)$ is of the order of $P(C)$ for those sets of principal interest, i.e. those with small $\sigma^2(C)$. For a reasonably regular VC class $\mathscr{C}$,

$$q_1(t)=(2t(Lg(t)+LLt^{-1}))^{\frac{1}{2}} \quad (1.18)$$

satisfies (1.17).

Suppose $q\leqq q_1$ is given, and we wish to find growth constants (b_n) such that (1.1) or (1.2) holds. Since v_n is less well-behaved on smaller sets, we might expect that roughly

$$\sup\{|v_n(C)|q(\sigma^2(C)): C\in\mathscr{C},\ \gamma_n\leqq\sigma^2(C)\leqq\alpha_n\}$$

$$\approx\sup\left\{\frac{|v_n(C)|}{q_1(\sigma^2(C))}\,\frac{q_1(\sigma^2(C))}{q(\sigma^2 C))}: C\in\mathscr{C},\ \sigma^2(C)=\gamma_n\right\}\approx R q_1(\gamma_n)/q(\gamma_n)$$

for some $0<R<\infty$. This, along with the standard LIL, leads us to expect (b_n) in (1.1) to be on the order of $q_1(\gamma_n)/q(\gamma_n)$ or $(LLn)^{\frac{1}{2}}$, whichever is greater. For example, when $q(t)=t^{\frac{1}{2}}$ we expect b_n to be on the order of

$$(\max(Lg(\gamma_n),\ LL\gamma_n^{-1},\ LLn))^{\frac{1}{2}}, \quad (1.19)$$

which is true provided γ_n is not too small, as Theorem 3.1 shows.

Upper bounds for (b_n) can be obtained by replacing $g(t)$ with its upper bound

t^{-1}; this is equivalent to disregarding the possibility that E_t may be only a small part of the whole space X. This is reasonable in the non-i.i.d. case, where $g(t)$ might vary with n in a complicated way, and is the reason the present results improve on those in Alexander (1985). In the one-dimensional d.f. case, for example, $g(t) \equiv t$, so if $L\gamma_n^{-1} \sim Ln$ then the rate (1.19) is $(LLn)^{\frac{1}{2}}$ (cf. (1.13)) while the upper bound rate is $(Ln)^{\frac{1}{2}}$. The present approach remedies this deficiency and gives exact results in a large number of cases.

II. An Inequality

The Vapnik-Červonenkis property of a class $\mathscr{C}$ of sets has proven to be a useful tool in the study of empirical processes indexed by $\mathscr{C}$ (Alexander 1984, 1985; Dudley 1978; Giné and Zinn 1984; Pollard 1982; Vapnik and Červonenkis 1971). The definition is as follows: a class $\mathscr{C}$ of subsets of a set X is called a *Vapnik-Červonenkis*, or *VC* class if

$$\sup\{\mathrm{card}\{F \cap C: C \in \mathscr{C}\}: \mathrm{card}(F) = n, F \subset X\} < 2^n$$

for some $n \geqq 1$. The least such n is called the *index* of $\mathscr{C}$ and denoted $V(\mathscr{C})$. Examples in $\mathbb{R}^d$ include the classes of all rectangles, all ellipsoids, all lower orthants $(-\infty, x]$, or all polyhedra with at most k sides (k fixed). Any subset of a VC class is a VC class, and $\{C \setminus D: C, D \in \mathscr{C}\}$ or $\{C \triangle D: C, D \in \mathscr{C}\}$ is a VC class if $\mathscr{C}$ is one. See Dudley (1978, 1984) for more about VC classes.

The supremum of v_n over an uncountable class $\mathscr{C}$ need not be measurable in general. This sometimes necessitates use of the outer probability measure $\mathbb{P}^*$. To avoid further measurability difficulties, we assume throughout this paper that $\mathscr{C}_t$ and $\mathscr{C}_{t,s}$ are deviation measurable for P (as defined in Alexander 1984) for all $s, t > 0$. For this it suffices that $\mathscr{C}_t$ and $\mathscr{C}_{t,s}$ be separable for all t, s for the topology of pointwise convergence (i.e. the topology in which $C_i \to C$ if and only if $1_{C_i} \to 1_C$ pointwise). Further, we assume that the r.v.'s X_i are *canonically formed*, i.e. that they are defined on the probability space $(X^\infty, \mathscr{A}^\infty, \mathbb{P})$ with X_i the ith coordinate function, where $\mathbb{P} = P^\infty$. This terminology comes from Gaenssler (1983).

Given the law P, define a *P-stratified VC class* to be an ordered pair of functions $(\mathscr{C}(\cdot), \zeta(\cdot))$ on an interval $[\gamma, \alpha]$, with ζ nonnegative and nondecreasing, and $\mathscr{C}(t)$ a deviation-mesurable (for P) VC class for all t, satisfying

$$\mathscr{C}(s) \subset \mathscr{C}(t) \qquad \text{for } s \leqq t,$$

and

$$P(C) \leqq \tfrac{1}{2},\ \sigma^2(C) \leqq \zeta(t) \qquad \text{for all } C \in \mathscr{C}(t).$$

Given a P-stratified VC class, we define

$$\tilde{E}_t = \bigcup_{C \in \mathscr{C}(t)} C,$$

$$\tilde{a}(t) = P(\tilde{E}_t) \vee \zeta(t),$$

$$\tilde{g}(t) = \tilde{a}(t)/\zeta(t).$$

(As with E_t we may assume $\tilde{E}_t$ measurable.) This notation is used to suggest the canonical example of a P-stratified VC class: $\mathscr{C}(t)=\mathscr{C}_t$ and $\zeta(t)=t$. The other example of interest here is, given a function ζ, to take $\mathscr{C}(t)=\mathscr{C}_{t,\zeta(t)}$.

The key to our results is a bound on probabilities like those in (1.2), analogous to Theorems 1.1 and 1.5 in Alexander (1985) and Inequality 1.2 in Shorack and Wellner (1982). We need a regularity condition for the weight function q: define

$$Q=\{q\in C[0,1]: q\geqq 0, q\uparrow,\ q(t)/t\downarrow\}.$$

As a convention, for monotone functions f on $[0,1/4]$ we define

$$f^{-1}(t)=\begin{cases}\sup\{s\in[0,1/4]: f(s)\leqq t\} & \text{if } f\ \uparrow,\\ \sup\{s\in[0,1/4]: f(s)\geqq t\} & \text{if } f\ \downarrow,\end{cases}$$

taking $\sup\phi$ to be 0.

Theorem 2.1. *Let $(\mathscr{C}(t),\ \zeta(t))$, $t\in[\gamma,\alpha]$, be a P-stratified VC class, and let $q\in Q$, $n\geqq 1$, and $p,b,u\geqq 0$. Set $z(t)=q(t)/\zeta(t)$ and suppose that $z(t), \tilde{g}(t)$, and $\zeta(t)/t$ are nonincreasing. Define*

$$\begin{aligned} r&=[(\tilde{a}^{-1}(p/n)\wedge z^{-1}(2n^{\frac{1}{2}}/b))\vee q^{-1}(u/b)\vee\gamma]\wedge\alpha\\ s&=\inf\{t\geqq 0: z(t)\leqq 2n^{\frac{1}{2}}/b\}.\end{aligned}\tag{2.1}$$

There exists a constant $K=K(V(\mathscr{C}(\alpha)))$ such that if

$$q^2(t)/\zeta(t)L\tilde{g}(t)\geqq Kb^{-2}\quad \textit{for all } r\vee s\leqq t\leqq\alpha,\tag{2.2}$$

$$q(t)/L(n\tilde{a}(t))\geqq Kn^{-\frac{1}{2}}b^{-1}\quad \textit{for all } r\vee s\leqq t\leqq\alpha,\tag{2.3}$$

$$q(r)\geqq Kn^{-\frac{1}{2}}b^{-1}\quad \textit{if } r<s \textit{ and } r\leqq\alpha,\tag{2.4}$$

$$q(t)L(bq(t)/n^{\frac{1}{2}}\zeta(t))/L(n\tilde{a}(t))\geqq Kn^{-\frac{1}{2}}b^{-1}\quad \textit{for all } t\in[r,s)\cap[r,\alpha],\tag{2.5}$$

then

$$\begin{aligned}&\mathbb{P}[|\nu_n(C)|>bq(t)+u\quad \textit{for some } \gamma\leqq t\leqq\alpha \textit{ and } C\in\mathscr{C}(t)]\\ &\leqq p+36\int_{r/2}^{\alpha} t^{-1}\exp(-b^2q^2(t)/512\zeta(t))\,dt\\ &\quad +68\exp(-bq(r)n^{\frac{1}{2}}/256).\end{aligned}\tag{2.6}$$

If also

$$q(t)^{\beta}L\tilde{g}(t)\quad\textit{and}\quad q(t)^{\beta}L(bq(t)/n^{\frac{1}{2}}\zeta(t))\uparrow\quad \textit{on}\quad [\gamma,\alpha]\tag{2.7}$$

for some $0\leqq\beta<1$, and if

$$bq(r)n^{\frac{1}{2}}L(\tilde{g}(r)\wedge(bq(r)/n^{\frac{1}{2}}\zeta(r)))\geqq(1-\beta)^{-1},\tag{2.8}$$

then (2.6) may be improved to

$$\mathbb{P}[|v_n(C)|>bq(t)+u \quad \textit{for some} \quad \gamma \leqq t \leqq \alpha \quad \textit{and} \quad C \in \mathscr{C}(t)]$$
$$\leqq p+36 \int_{r/2}^{\alpha} t^{-1} \exp(-b^2 q^2(t)/512\zeta(t))\,dt$$
$$+68 \exp(-2^{-8} bq(r) n^{\frac{1}{2}} L\tilde{g}(r))$$
$$+36 \exp(-2^{-8} bq(r) n^{\frac{1}{2}} L(bq(r)/n^{\frac{1}{2}}\zeta(r))). \quad \square \tag{2.9}$$

The heuristics of (2.6) and (2.9) are as follows, for "regular" cases: the probability that there is a $C \in \mathscr{C}$ with $P(C) < r$ and $P_n(C) > 0$ is at most p, giving the first term. The integral term arises from sets C with $r \vee s \leqq \sigma^2(C) \leqq \alpha$. For these C, $\mathbb{P}[|v_n(C)| > bq(\sigma^2(C)) + u]$ can be approximated by a Gaussian probability, and (2.2) and (2.3) are used. The last term(s) arise from sets C with $r \leqq \sigma^2(C) < s$, where a Poisson rather than a Gaussian approximation is valid, and (2.4) and (2.5) are used.

In this paper we are interested only in the case $p = u = 0$, so we tactitly assume these values whenever Theorem 2.1 is cited henceforth.

Remark 2.2. In the results that follows, the only intrinsic properties of the capacity function g actually used are that $a(t)/t \leqq g(t) \leqq 1/t$ and that $tg(t)$ increases. Hence for a given $\mathscr{C}$ and P, all results remain valid if g is replaced throughout by a function $g_o \geqq g$ with $g_o(t) \leqq 1/t$ and $tg_o(t)$ increasing. The same applies to $\tilde{g}(t)$, $\tilde{a}(t)$, and $\zeta(t)$ (in place of $g(t)$, $a(t)$, and t) for Theorem 2.1. Thus the following loses us little generality: we tacitly assume henceforth that g (or $\tilde{g}$) is nonincreasing. The possibility of modifying $\tilde{g}$ makes (2.7) a very mild condition. $\square$

III. The Square Root Weight Function

The results in this section may be compared to Csáki's result (1.13).

The key to obtaining upper bounds for the lim sup in (1.1), with $q(t) = t^{\frac{1}{2}}$, lies in the behavior of the metric entropies of the classes $\mathscr{C}_t$. Define for $u > 0$:

$$N_2(u, \mathscr{C}, P) = \min\{k \geqq 1: \quad \text{there exist} \quad C_1, \ldots, C_k \in \mathscr{C} \quad \text{such that}$$
$$\min_{i \leqq k} P(C \triangle C_i) < u^2 \quad \text{for all} \quad C \in \mathscr{C}\}.$$

The function $\log N_2(\cdot, \mathscr{C}, P)$ is called the *metric entropy* of $\mathscr{C}$ in $L^2(P)$. (Note $P(C \triangle D) = \|1_C - 1_D\|^2_{L^2(P)}$.) It measures the size of $\mathscr{C}$, telling us "how totally bounded" $\mathscr{C}$ is.

Define probability measures P_t $(0 < t \leqq \frac{1}{4})$ by $P_t(A) = P(A \cap E_t)/P(E_t)$. If $C, D \in \mathscr{C}_t$ then $P_t(C \triangle D) = P(C \triangle D)/P(E_t)$. Hence using Lemma 2.7 of Alexander (1984), which is based on Lemma 7.13 of Dudley (1978), we have for any $\mathscr{D}_t \subset \mathscr{C}_t$ and $u \in (0,1)$:

$$N_2(ut^{\frac{1}{2}}, \mathscr{D}_t, P) = N_2(ut^{\frac{1}{2}} P(E_t)^{-\frac{1}{2}}, \mathscr{D}_t, P_t)$$
$$\leqq 2(16g(t)u^{-2} L(8g(t)u^{-2}))^{V(\mathscr{C})-1}$$
$$\leqq K(g(t)/u^2)^{V(\mathscr{C})-1+\delta} \tag{3.1}$$

for some constant $K = K(\delta, V(\mathscr{C}))$ for each $\delta > 0$.

For lower bounds on the lim sup in (1.1), the following concept will be useful: we say the class $\mathscr{C}$ is *full* (for P) of for every sufficiently small $\lambda > 0$ there is a $0 < \varepsilon_\lambda < 1$ such that for each sufficiently small $t > 0$ there are $k \geqq \varepsilon_\lambda g(t)^{1-\lambda}$ sets $C_1, \ldots, C_k \in \mathscr{C}$ with

$$\sigma^2(C_i) = t \quad \text{and} \quad P\left(C_i \cap \left(\bigcup_{j \neq i} C_j\right)\right) \leqq \lambda P(C_i).$$

Roughly, $\mathscr{C}$ is full if it contains a large number of nearly-disjoint sets of any given small size. We say $\mathscr{C}$ is *spatially full* if the C_i's can always be chosen disjoint. In some of our proofs we will tacitly assume for convenience that "sufficiently small" above means "at most $\frac{1}{4}$", but it should always be clear that this is not a necessary restriction.

Theorem 3.1. *Let $\mathscr{C}$ be a VC class, let $\gamma_n \downarrow 0$, and define*

$$\begin{aligned} w_n &= Lg(\gamma_n) \vee LLn, \\ y_n &= w_n/((n\gamma_n)^{\frac{1}{2}} L(w_n/n\gamma_n)), \\ c_1 &= \limsup_n w_n^{-1} Lg(\gamma_n), \; c_2 = \limsup_n w_n^{-1} LL\gamma_n^{-1}, \\ c_3 &= \limsup_n w_n^{-1} LLn. \end{aligned} \tag{3.2}$$

Suppose that for some $\varrho, \eta < \infty$,

$$N_2(ut^{\frac{1}{2}}, \mathscr{C}_t \setminus \mathscr{C}_{(1-u^2/4)t}, P) \leqq A u^{-\eta} g(t)^{\varrho+\delta} \tag{3.3}$$

for some $A = A(\delta) < \infty$ for all $t, \delta > 0$ and $u \in (0,1)$.

(A) *Then*

$$\limsup_n \sup\{|\nu_n(C)|/b_n \sigma(C) : C \in \mathscr{C}, \sigma^2(C) \geqq \gamma_n\} = R \; a.s. \tag{3.4}$$

where

(i) *if $n^{-1} w_n = o(\gamma_n)$ and $n^{-1} w_n^{\frac{1}{2}}$ decreases, then $b_n = w_n^{\frac{1}{2}}$ and $R \leqq (2(\varrho c_1 + c_2 + c_3))^{\frac{1}{2}}$;*

(ii) *if $\gamma_n \sim \tau n^{-1} w_n$ for some $\tau > 0$ and $n^{-1} w_n^{\frac{1}{2}}$ decreases, then $b_n = w_n^{\frac{1}{2}}$ and $R \leqq \max((2(\varrho c_1 + c_2 + c_3))^{\frac{1}{2}}, \tau^{\frac{1}{2}}(\beta_{\theta\tau} - 1))$, where $\theta = (\varrho c_1 + c_3)^{-1}$;*

(iii) *if $\gamma_n = o(n^{-1} w_n)$, $n^{-1} y_n$ decreases, and*

$$L(w_n/n\gamma_n) = o(w_n), \tag{3.5}$$

then $b_n = y_n$ and $R \leqq \varrho c_1 + c_3$.

(B) *If $\varrho \leqq 1$, $\mathscr{C}$ is full, the* lim sups *in* (3.2) *are actually limits, and (for* (i) *and* (ii) *only) $Lg(t)/Lt^{-1}$ is nondecreasing, then the upper bounds for R in* (i)–(iii) *above are also lower bounds, so (i)–(iii) give the true values of R in* (3.4).

(C) *If the assumptions in* (B) *hold, $\mathscr{C}$ is spatially full, and $c_3 = 0$, then* "lim sup" *may be replaced by* "lim" *in* (3.4) *for each of* (i)–(iii). □

By (3.1), (3.3) is always valid with $\eta/2 = \varrho = V(\mathscr{C}) - 1$. Later examples, however, will show that this need not be optimal. In fact, we will have $\varrho = 1$ with $V(\mathscr{C})$ arbitrarily large.

Remark 3.2. The condition (3.3) is related to the "relative metric entropy" condition (1.25) in Alexander (1985). In fact, it is easy to verify that

$$N_2(ut^{\frac{1}{2}}, \mathscr{C}_t \setminus \mathscr{C}_{(1-u^2/4)t}, P) \leqq N_2^R(u/2, \mathscr{C}, P)$$

for all $0 < u < \frac{1}{2}$ and $t > 0$, where N_2^R is defined in Alexander (1985).

Qualitatively, (3.3) with $\varrho = 1$ can be interpreted as follows, in sufficiently regular cases. $g(t) = P(E_t)/t$ is the maximum number of disjoint sets of probability near t which can fit in $\mathscr{C}$. Therefore $g(t)$ should be a lower bound on the number of sets required to approximate all sets in $\mathscr{C}$ of probability near t to within a given fraction of t. If $\varrho = 1$, then (3.3) says $g(t)$ is not far from also being an upper bound on this number of sets. □

The dependence of Theorem 3.1 and later results on the behavior of γ_n relative to w_n/n is rooted in the quality of the Gaussian approximation to $\mathbb{P}[|\nu_n(C)|/b_n \sigma(C) > R]$. The approximation is good for all C with $\sigma^2(C) \geqq \gamma_n$ if $w_n/n = o(\gamma_n)$, good up to a constant in the exponent if $w_n/n = O(\gamma_n)$, and not good in general if $\gamma_n = o(w_n/n)$.

The condition (3.5) will ensure that, by a $0-1$ law, the lim sup in (3.4) is some fixed constant a.s.

The heuristics of the value of R in (3.4) are as follows. We need only consider a finite number of sets in $\mathscr{C}$. c_1 comes from the number of sets C, with a given fixed value of $\sigma^2(C)$, which must be considered; this is related to the metric entropy. c_2 comes from the number of fixed values of $\sigma^2(C)$ which must be considered. c_3 comes from the requirement that sums of certain probabilities, for geometrically increasing subsequences of values of n, must be finite, as in some proofs of the LIL.

If $\mathscr{C}$ is full, it is clear that ϱ must be at least 1 if g is unbounded, hence in particular if $c_1 > 0$. Therefore there is no ambiguity in Theorem 3.1 (B) arising from the fact that (3.3) may hold for multiple values of ϱ.

In Theorem 3.1 and in fact throughout this paper, any requirement that a sequence, say (λ_n) (or a function, say $\varphi(t)$), be monotone may be weakened to a requirement that $\lambda_n \sim \theta_n$ as $n \to \infty$ (or $\varphi(t) \sim \xi(t)$ as $t \to 0$) for some monotone sequence (θ_n) (or monotone function $\xi(t)$).

When

$$LL\gamma_n^{-1} = o(LLn) \quad \text{and} \quad Lg(\gamma_n) = o(LLn), \tag{3.6}$$

we see from Theorem 3.1 (i) that the value of R in (3.4) is $2^{\frac{1}{2}}$. This is true, for example, in the one-dimensional uniform d.f. case if $\gamma_n = (Ln)^{-\alpha}$ for some $a > 0$. By the ordinary law of the iterated logarithm, this is the same value achieved at each individual $C \in \mathscr{C}$. Thus no "small" subclass of $\mathscr{C}$, corresponding to the tails in the d.f. case, is the sole determiner of the rate of growth of the weighted empirical process when (3.6) holds.

Example 3.3. When $\mathscr{C} = \mathscr{D}_1$ and P is uniform on $[0,1]$ (the one-dimensional uniform d.f. case), we have $P(E_t) = t$, so $g(t) \equiv 1$, and (3.3) is valid for $\varrho = 0$. Hence in Theorem 3.1 we have $w_n = LLn$, $y_n = LLn/((n\gamma_n)^{\frac{1}{2}} L(LLn/n\gamma_n))$, $c_1 = 0$, and $c_3 = 1$. Under the assumptions in (1.14) we have (3.5) holding. Clearly $\mathscr{D}_1$ is full. Thus (1.13) is a special case of Theorem 3.1. □

Example 3.4. When $\mathscr{C}=\mathscr{D}_d$ and P is uniform on $[0,1]^d$ with $d>1$ (the d-dimensional uniform d.f. case), it is easy to check that $P(E_t)\sim t(Lt^{-1})^{d-1}/(d-1)!$, so $g(t)\sim(Lt^{-1})^{d-1}/(d-1)!$ We will show in the proof of Corollary 3.5 below that (3.3) is valid for $\varrho=1$. Hence if $LL\gamma_n^{-1}/LLn\to a$ for some $a\geqq 0$, in Theorem 3.1 (i) we have

$$c_3>0,\, w_n\sim c_3^{-1}\,LLn,\ c_1=(d-1)\,ac_3,\, c_2=ac_3,$$

and

$$\varrho c_1+c_2+c_3=c_3(1+ad).$$

In (ii), if $\gamma_n\sim\lambda n^{-1}\,LLn$, we get

$$w_n\sim(d-1)\,LLn,\, c_1=1,\ c_2=c_3=(d-1)^{-1},$$

$$\tau=\lambda(d-1)^{-1},\ \varrho c_1+c_2+c_3=(d+1)/(d-1),\quad \text{and}\quad \theta=(d-1)/d.$$

In (iii), if (3.5) holds, i.e. if $n^{-1}(Ln)^{-\varepsilon}=0(\gamma_n)$ for all $\varepsilon>0$, the values are

$$w_n\sim(d-1)\,LLn,\, c_1=1,\, c_3=(d-1)^{-1},\quad \text{and}\quad (d-1)(\varrho c_1+c_3)=d.$$

To see that $\mathscr{D}_d$ is full, fix $0<\lambda<1$ and $0<t<1$, and let $\mu=\lambda d^{-1}$ and

$$\mathscr{F}=\Big\{[0,x]:\ x=(f(t)\,\mu^{-j_1},\,\mu^{j_2},\dots,\mu^{j_d})\in[0,1]^d \text{ for some integers } j_i\geqq 0 \text{ with } j_1=\sum_{i=2}^{d} j_i\Big\} \tag{3.7}$$

where $f(t)\leqq\frac{1}{2}$ is given by $f(t)(1-f(t))=t$. Then

$$\begin{aligned}|\mathscr{F}|&=|\{(j_2,\dots,j_d)\in\mathbb{Z}_+^{d-1}:\ \sum_{i=2}^{d} j_1\leqq(\log f(t)^{-1}/(\log\mu^{-1})\}|\\ &\geqq\varepsilon_d((\log t^{-1})/(\log\mu^{-1}))^{d-1}\vee 1\\ &\geqq\varepsilon_{d\lambda}\,g(t)\end{aligned} \tag{3.8}$$

for some constants ε_d and $\varepsilon_{d\lambda}$, where $\mathbb{Z}_+$ denotes the nonnegative integers. Since $P\Big(C\cap\Big(\bigcup_{D\in\mathscr{F},D\neq C} D\Big)\Big)\leqq d\mu P(C)$ for $C\in\mathscr{F}$, this shows that $\mathscr{D}_d$ is full.

This establishes (up to the proof of (3.3)) the next corollary for $d>1$. □

Corollary 3.5. *Let P be the uniform law on $[0,1]^d\,(d\geqq 1)$ and let $\gamma_n\downarrow 0$ so that $LL\gamma_n^{-1}/LLn\to a$ for some $a\geqq 0$. Then*

$$\limsup_n\,\sup\{|v_n(C)|/b_n\sigma(C):C\in\mathscr{D}_d,\ \sigma^2(C)\geqq\gamma_n\}=R\quad a.s.$$

where

$$R=(2(1+ad))^{\frac{1}{2}}\ \ \textit{and}\ \ b_n=(LLn)^{\frac{1}{2}}\quad \textit{if}\ n^{-1}\,LLn=o(\gamma_n)$$

$$R=\max((2(1+d))^{\frac{1}{2}},\lambda^{\frac{1}{2}}(\beta_{\lambda/d}-1)\ \ \textit{and}\ \ b_n=(LLn)^{\frac{1}{2}}\quad \textit{if}\ \gamma_n\sim\lambda n^{-1}\,LLn$$

$$R=d\ \ \textit{and}\ \ b_n=LLn/((n\gamma_n)^{\frac{1}{2}}L(LLn/n\gamma_n))\quad \textit{if}\ \gamma_n=o(n^{-1}\,LLn)$$

$$\textit{and}\ \ n^{-1}(Ln)^{-\varepsilon}=o(\gamma_n)\quad \textit{for all}\ \varepsilon>0.\ □$$

Example 3.6. Let P be a nondegenerate normal law on $\mathbb{R}^d$, let $C_{bv} = \{x \in \mathbb{R}^d: x \cdot v \geqq b\}$ for v in the sphere S^{d-1} and $b \in \mathbb{R}$, and let $\mathscr{C} = \{C_{bv}: b \in \mathbb{R}, v \in S^{d-1}\}$ be the class of all closed half spaces in $\mathbb{R}^d$. Then there is an affine map $A: \mathbb{R}^d \to \mathbb{R}^d$ with $\mathscr{L}(A)$ the normal law $N(0, I)$. Since A preserves all the relevant structure of $\mathscr{C}$, we may assume $P = N(0, I)$.

Let Φ be the d.f. on $\mathbb{R}$ of $N(0,1)$, and let χ_d^2 be a chi-squared r.v. with d degrees of freedom. Now for $t \leqq \frac{1}{2}$, $\sigma^2(C_{bv}) \leqq t(1-t)$ if and only if $|b| \geqq \Phi^{-1}(1-t)$, so $E_{t(1-t)} = \{x \in \mathbb{R}^d: |x| \geqq \Phi^{-1}(1-t)\}$. Let $r_t = \Phi^{-1}(1-t)$; since $r_t \sim (2Lt^{-1})^{\frac{1}{2}}$ and $r_t^{-1} \exp(-r_t^2/2) \sim (2\pi)^{\frac{1}{2}} t$ as $t \to 0$, it follows that as $t \to 0$,

$$P(E_{t(1-t)}) = \mathbb{P}[|X_1|^2 \geqq r_t^2] = \mathbb{P}[\chi_d^2 \geqq r_t^2] \sim K_1 r_t^{d-2} \exp(-r_t^2/2) \sim K_2 t (Lt^{-1})^{(d-1)/2}$$

where K_1 and K_2 are constants depending on d, so $g(t) \sim K_2 (Lt^{-1})^{(d-1)/2}$.

We will show in the proof of the next corollary that $\mathscr{C}$ is full and (3.3) is valid with $\varrho = 1$, though $V(\mathscr{C}) \geqq d+2$. Suppose $LL\gamma_n^{-1}/LLn \to a$ for some $a \geqq 0$. Then in Theorem 3.1 (i) we have

$$c_3 > 0,\ w_n \sim c_3^{-1} LLn,\ c_1 = (d-1) a c_3/2,\ c_2 = a c_3,$$

and

$$\varrho c_1 + c_2 + c_3 = c_3 (1 + a(d+1)/2).$$

In (ii), if $\gamma_n \sim \lambda n^{-1} LLn$ we get

$$c_2 = c_3 > 0,\ w_n \sim c_3^{-1} LLn,\ c_1 = (d-1) c_3/2,\ \tau = \lambda c_3,\ \theta = 2/c_3(d+1),$$

and

$$\varrho c_1 + c_2 + c_3 = (d+3) c_3/2.$$

In (iii), we find

$$c_3 > 0,\ w_n \sim c_3^{-1} LLn,\ c_1 = (d-1) c_3/2, \quad \text{and} \quad \varrho c_1 + c_3 = (d+1) c_3/2.$$

The next corollary summarizes this. □

Corollary 3.7. *Let P be a nondegenerate normal law on $\mathbb{R}^d$ and let $\mathscr{C}$ be the class of all closed half spaces in $\mathbb{R}^d$ $(d \geqq 1)$. Let $\gamma_n \downarrow 0$ so that $LL\gamma_n^{-1}/LLn \to a$ for some $a > 0$. Then*

$$\limsup_n \limsup_n \sup\{|\nu_n(C)|/b_n \sigma(C): C \in \mathscr{C},\ \sigma(C) \geqq \gamma_n\} = R \quad a.s.$$

where

$$R = (2(1 + a(d+1)/2))^{\frac{1}{2}} \quad \text{and} \quad b_n = (LLn)^{\frac{1}{2}} \quad \text{if } n^{-1} LLn = o(\gamma_n)$$

$$R = \max((d+3)^{\frac{1}{2}}, \lambda^{\frac{1}{2}}(\beta_{2\lambda/(d+1)} - 1)) \quad \text{and} \quad b_n = (LLn)^{\frac{1}{2}} \quad \text{if } \gamma_n \sim \lambda n^{-1} LLn$$

$$R = (d+1)/2 \quad \text{and} \quad b_n = LLn/((n\lambda_n)^{\frac{1}{2}} L(LLn/n\gamma_n))$$

$$\text{if } \gamma_n = o(n^{-1} LLn) \quad \text{and} \quad n^{-1}(Ln)^{-\varepsilon} = o(\gamma_n) \quad \text{for all } \varepsilon > 0. \quad \square$$

Example 3.8. Let P be the uniform law on $[0,1]^d$ $(d \geqq 1)$, recall that $\mathscr{I}_d$ is the class of all subrectangles of $[0,1]^d$, and let $\mathscr{C} = \{C \in \mathscr{I}_d: P(C) \leqq \frac{1}{2}\}$. The bound on $P(C)$ is assumed for convenience, to avoid the technicalities of dealing with complements of large rectangles. Then $P(E_t) \equiv 1$, so $g(t) = t^{-1}$. It is clear that $\mathscr{C}$ is spatially full. (3.3) with $\varrho = 1$ will be checked in the proof of the next corollary, using the techniques of Theorem 2.1 of Orey and Pruitt (1973). Suppose $LLn/L\gamma_n^{-1} \to a$ for some $0 \leqq a \leqq \infty$. If $a \leqq 1$, the constants in Theorem 3.1 are

$$c_1 = 1,\ c_2 = 0,\ c_3 = a,\ w_n \sim L\gamma_n^{-1}$$

while if $a > 1$ they are

$$c_1 = a^{-1},\ c_2 = 0,\ c_3 = 1,\ w_n \sim LLn.$$

By Theorem 3.1 (i), if $1 < a < \infty$ we have

$$\limsup_n \sup\{|v_n(C)|/(\sigma^2(C)\, LLn)^{\frac{1}{2}} : C \in \mathscr{C},\ \sigma(C) \geqq \gamma_n\} = (2(1 + a^{-1}))^{\frac{1}{2}} \quad \text{a.s.}$$

which is equivalent to

$$\limsup_n \sup\{|v_n(C)|/(\sigma^2(C)\, L\gamma_n^{-1})^{\frac{1}{2}} : C \in \mathscr{C},\ \sigma^2(C) \geqq \gamma_n\} = (2(1 + a))^{\frac{1}{2}} \quad \text{a.s.}$$

In Theorem 3.1 (ii), if $\gamma_n \sim \lambda n^{-1} Ln$ the constants are

$$a = 0,\ \tau = \lambda,\ \theta = 1,\ w_n \sim Ln,\ c_1 = 1,\ c_2 = c_3 = 0$$

and we can make use of (1.10). In (iii), if (3.5) holds, i.e. if $n^{-(1+\varepsilon)} = o(\gamma_n)$ for all $\varepsilon > 0$, we get

$$a = 0,\ c_1 = 1,\ c_3 = 0.$$

This is summarized in the following corollary. □

Corollary 3.9. *Let P be the uniform law on $[0,1]^d$ $(d \geqq 1)$ and let $\gamma_n \downarrow 0$ so that $n^{-1}(L\gamma_n^{-1})^{\frac{1}{2}}$ decreases and $LLn/L\gamma_n^{-1} \to a$ for some $0 \leqq a \leqq \infty$. Then*

$$\limsup_n \sup\{|v_n(C)|/b_n\sigma(C) : C \in \mathscr{I}_d,\ P(C) \leqq \tfrac{1}{2},\ \sigma^2(C) \geqq \gamma_n\} = R \quad a.s. \tag{3.9}$$

where

$R = 2^{\frac{1}{2}}$ *and* $b_n = (LLn)^{\frac{1}{2}}$ *if* $a = \infty$ *(i.e. if* $(Ln)^{-\varepsilon} = o(\gamma_n)$ *for all* $\varepsilon > 0$*)*

$R = (2(1+a))^{\frac{1}{2}}$ *and* $b_n = (L\gamma_n^{-1})^{\frac{1}{2}}$ *if* ${}^{-1}Ln = o(\gamma_n)$ *and* $0 < a < \infty$ (3.10)

$R = 2^{\frac{1}{2}}$ *and* $b_n = (L\gamma_n^{-1})^{\frac{1}{2}}$ *if* $n^{-1}Ln = o(\gamma_n)$ *and* $a = 0$ (3.11)

$R = \lambda^{\frac{1}{2}}(\beta_\lambda - 1)$ *and* $b_n = (Ln)^{\frac{1}{2}}$ *if* $\gamma_n \sim \lambda n^{-1} Ln$ (3.12)

$R = 1$ *and* $b_n = Ln/((n\gamma_n)^{\frac{1}{2}} L(Ln/n\gamma_n)$ *if* $c_n = O(n^{-1}Ln)$ *and* $n^{-(1+\varepsilon)} = o(\gamma_n)$ *for all* $\varepsilon > 0$, (3.13)

provided in each case that $n^{-1}b_n$ decreases. For (3.11)–(3.13), "lim sup" *may be replaced by* "lim" *in* (3.9). □

Taking $\gamma_n = n^{-1}(Ln)^\beta$ $(-\infty < \beta < 1)$ in Corollary 3.9, we see that the lim sup in (1.4) is $(1-\beta)$ a.s.

For $d = 1$, (3.11) is Stute's result (1.5), and (3.10) and (3.12) are due to Mason et al. (1983). For $d > 1$ (3.11) is related to other results of Stute (1984).

IV. The Asymptotic Modulus of the Empirical Process

We will call a nondecreasing function ψ on $[0, 1/4]$ an *asymptotic modulus of continuity* for (v_n) on $\mathscr{C}$ if

$$\psi(0)=0, \tag{4.1}$$

$$\psi(s+t) \leqq \psi(s)+\psi(t) \quad \text{for all } s, t \in [0, \tfrac{1}{4}] \tag{4.2}$$

there exist sequences $\gamma_n, \alpha_n \downarrow 0$ satisfying $n\alpha_n \downarrow$, $\gamma_n \leqq \alpha_n$, and (4.3)

$$n^{-1} Ln = o(\alpha_n) \tag{4.4}$$

such that

$$\limsup_n \sup \left\{ \frac{|\nu_n(C)-\nu_n(D)|}{\psi(\sigma(C \triangle D))} : C, D \in \mathscr{C}, P(C \triangle D) \leqq \frac{1}{2}, \gamma_n \leqq \sigma^2(C \triangle D) \leqq \alpha_n \right\} < \infty \text{ a.s.} \tag{4.5}$$

ψ is a *local asymptotic modulus at* ϕ (the empty set) for (ν_n) on $\mathscr{C}$ if (4.1)–(4.3) hold with (4.5) replaced by

$$\limsup_n \sup \left\{ \frac{|\nu_n(C)|}{\psi(\sigma(C))} : C \in \mathscr{C}, \gamma_n \leqq \sigma^2(C) \leqq \alpha_n \right\} < \infty \quad \text{a.s.} \tag{4.6}$$

The requirement (4.4) ensures that, in (4.5) and (4.6), we are not considering only sets so small that P_n need not be much like P.

Define $g_1(t)$ by

$$Lg_1(t) = Lg(t) + LLt^{-1},$$

and set

$$\psi_1(t) = t(Lg_1(t^2))^{\frac{1}{2}} = q_1(t^2), \quad \psi_0(t) = t(Lt^{-2})^{\frac{1}{2}}.$$

Theorem 4.1. *If $\mathscr{C}$ is a VC class then $\psi_0(t)$ is an asymptotic modulus of continuity, and $\psi_1(t)$ a local asymptotic modulus at ϕ, for (ν_n).* □

It follows that in all dimensions, $t(LLt^{-1})^{\frac{1}{2}}$ is a local asymptotic modulus at ψ both for the uniform d.f. case (Example 3.4) and for the half space case with normal law (Example 3.6), and $t(Lt^{-1})^{\frac{1}{2}}$ is an asymptotic modulus of both for the uniform interval case (Example 3.8). The latter fact is a variant of Stute's (1982a) result (1.6).

More detail can be obtained in some cases, as the next theorem shows.

Theorem 4.2. *Let $\mathscr{C}$ be a VC class and let $\gamma_n, \alpha_n \downarrow 0$ with $n\alpha_n \uparrow$, $\gamma_n \leqq \alpha_n$,*

$$n^{-1} Lg_1(\gamma_n) = o(\gamma_n), \quad \textit{and} \quad \liminf_n Lg_1(\alpha_n)/LLn > 0. \tag{4.7}$$

Define

$$\begin{aligned} c_1 &= \limsup_{t\to 0} Lg(t)/Lg_1(t), \\ c_2' &= \limsup_{t\to 0} LLt^{-1}/Lg_1(t), \\ c_2'' &= \limsup_n LL(\gamma_n^{-1}\alpha_n)/Lg_1(\alpha_n), \\ c_2 &= c_2' \wedge c_2'', \\ c_3 &= \limsup_n LLn/Lg_1(\alpha_n). \end{aligned} \tag{4.8}$$

(i) *Suppose the entropy condition* (3.3) *holds for some $\varrho \geqq 0$. Then*

$$\limsup_n \sup \{ |\nu_n(C)|/\psi_1(\sigma(C)) : C \in \mathscr{C}, \gamma_n \leqq \sigma^2(C) \leqq \alpha_n \} \leqq (2(\varrho c_1 + c_2 + c_3))^{\frac{1}{2}} \quad \textit{a.s.} \tag{4.9}$$

(ii) *If* $\mathscr{C}$ *is full,* $\varrho \leqq 1$, $Lg(t)/Lt^{-1}$ *is nondecreasing, and the* lim sups *in* (4.8) *are actually limits, then equality holds in* (4.9).

(iii) *If* $\mathscr{C}$ *is spatially full,* $\varrho \leqq 1$, *and* $c_3 = 0$, *then*

$$\limsup_n \{|v_n(C)|/\psi_1(\sigma(C)) : C \in \mathscr{C},\ \gamma_n \leqq \sigma^2(C) \leqq \alpha_n\} = 2^{\frac{1}{2}} \quad a.s. \tag{4.10}$$

In the d-dimensional uniform interval case (Example 3.8), Theorem 4.2 (iii) tells us that (4.10) holds provided $\gamma_n, \alpha_n \downarrow$, $n\alpha_n \uparrow$, $n^{-1} Ln = o(\gamma_n)$, $\gamma_n \leqq \alpha_n$, and $LLn = o(L\alpha_n^{-1})$. Since here $\psi_1(t) \sim t(Lt^{-2})^{\frac{1}{2}}$, this generalizes Stute's (1982a) result (1.9).

Dudley (1978) proved that v_n, indexed by a VC class $\mathscr{C}$, converges weakly (under some measurability conditions) to a Gaussian process G_P on $\mathscr{C}$ with the same covariance as v_n. Lemma 7.13 of Dudley (1978) and Theorem 2.1 of Dudley (1973) show that ψ_0 is a sample modulus for G_P on $\mathscr{C}$. For $\mathscr{C} = \mathscr{D}_d$ or $\mathscr{I}_d$ (the d.f. or interval cases) and P uniform, $G_P(C)$ is the increment of a tied-down Brownian sheet over the rectangle C. Results of Orey and Pruitt (1973) tell us that

$$\limsup_{\varepsilon \to 0} \{|G_P(C) - G_P(D)|/\psi_0(\sigma(C \vartriangle D)) : C, D \in \mathscr{D}_d,\ \sigma^2(C \vartriangle D) < \varepsilon\} = (2d)^{\frac{1}{2}} \quad \text{a.s.} \tag{4.11}$$

$$\limsup_{\varepsilon \to 0} \{|G_P(C)|/\psi_1(\sigma(C)) : C \in \mathscr{I}_d,\ P(C) \leqq \tfrac{1}{2},\ \sigma^2(C) < \varepsilon\} = 2^{\frac{1}{2}} \quad \text{a.s.} \tag{4.12}$$

$$\limsup_{\varepsilon \to 0} \{|G_P(C)|/\psi_1(\sigma(C)) : C \in \mathscr{D}_d,\ P(C) \leqq \tfrac{1}{2},\ \sigma^2(C) < \varepsilon\} = 2^{\frac{1}{2}} \quad \text{a.s.} \tag{4.13}$$

(4.12) and (4.13) are also obtained as special cases of results for general set-indexed Gaussian processes in Alexander (1986). (4.11) and Theorem 4.1 tell us that the best possible sample modulus for G_P is also an asymptotic modulus of continuity for (v_n). (4.12) may be compared with (4.10), and (4.13) may be compared to the following corollary of Theorem 4.2. (Recall $\mathscr{D}_d$ was proved to be full in Example 3.4.)

Corollary 4.3. *Let* $\mathscr{C}$ *be* $\mathscr{D}_d$ $(d \geqq 1)$, *let P be the uniform law on* $[0,1]^d$, *and let* $\gamma_n \downarrow 0$ *and* $\alpha_n \downarrow 0$ *so that*

$$n^{-1} LLn = o(\gamma_n),\ \gamma_n \leqq \alpha_n,\ n\alpha_n \uparrow,\ LL\alpha_n^{-1}/LLn \to a,$$

and

$$LL(\gamma_n^{-1}\alpha_n)/LL\alpha_n^{-1} \to b \quad \textit{for some } a, b > 0.$$

Then $\psi_1(t) \sim t(d\,LLt^{-1})^{\frac{1}{2}}$ *and*

$$\limsup_n \sup\{|v_n(C)|/\psi_1(\sigma(C)) : C \in \mathscr{D}_d,\ P(C) \leqq \tfrac{1}{2},\ \gamma_n \leqq \sigma^2(C) \leqq \alpha_n\}$$
$$= (2(d - 2 + (1 \wedge b) + a^{-1})/d)^{\frac{1}{2}}. \quad \square$$

Our final theorem in this section is an in-probability version of Theorem 4.1.

Theorem 4.4. *Let* $\mathscr{C}$ *be a VC class and let* $\gamma_n \to 0$ *with* $n^{-1} Lg(\gamma_n) = o(\gamma_n)$. *Then*

$$\sup\{|v_n(C)|/\psi_1(\sigma(C)) : C \in \mathscr{C},\ \sigma^2(C) \geqq \gamma_n\}$$

is bounded in probability. $\square$

V. The Ratio P_n/P

By taking $q(t)=t$ and $b_n=n^{\frac{1}{2}}$ in (1.1) and (1.2), we can study the behavior of

$$\sup\left\{\left|\frac{P_n(C)}{P(C)}-1\right| : C\in\mathscr{C},\ P(C)\geqq\gamma_n\right\}$$

as $n\to\infty$. As with $q(t)=t^{\frac{1}{2}}$, analogs of Wellner's (1978) results (1.15) and (1.16) will depend on the behavior of $P(E_t)$ as $t\to 0$.

Theorem 5.1. *Let $\mathscr{C}$ be a VC class and let $\gamma_n\to 0$.* If

$$n^{-1}Lg(\gamma_n)=o(\gamma_n) \tag{5.1}$$

then as $n\to\infty$,

$$\sup\left\{\left|\frac{P_n(C)}{P(C)}-1\right| : C\in\mathscr{C},\ P(C)\geqq\gamma_n\right\}\to 0 \quad \textit{in probability.} \tag{5.2}$$

If also

$$n^{-1}LLn=o(\gamma_n) \tag{5.3}$$

then the convergence in (5.2) is a.s.

Conversely if $\mathscr{C}$ is full and either

$$\gamma_n=O(n^{-1}Lg(\gamma_n)) \quad \textit{or} \quad \gamma_n=O(n^{-1}LLn) \tag{5.4}$$

then

$$\limsup_n \sup\left\{\left|\frac{P_n(C)}{P(C)}-1\right| : C\in\mathscr{C},\ P(C)\geqq\gamma_n\right\}>0 \quad a.s. \quad \square \tag{5.5}$$

The next theorem includes (1.7) as a special case.

Theorem 5.2. *Let $\mathscr{C}$ be a VC class and let γ_n^* be the solution γ of $\gamma/Lg(\gamma)=n^{-1}$. Then for each $\varepsilon, A>0$ there exist $R<\infty$ and n_0 such that for all $n\geqq n_0$,*

$$\mathbb{P}^*\left[P_n(C)\leqq(RLg(\gamma_n^*))^{-1}P(C) \quad \textit{for some} \quad C\in\mathscr{C} \quad \textit{with} \quad P_n(C)\neq 0\right]<g(\gamma_n^*)^{-A}\wedge e^{-A} \tag{5.6}$$

and

$$\mathbb{P}^*\left[\sup\left\{\frac{P_n(C)}{P(C)} : C\in\mathscr{C},\ P(C)\geqq\varepsilon\gamma_n^*\right\}>R\right]<g(\gamma_n^*)^{-A}\wedge e^{-A}. \tag{5.7}$$

If g is bounded, then $\sup_{\mathscr{C}} P_n(C)/P(C)$ is bounded in probability (i.e. *we can take $\varepsilon=0$ above*). $\square$

γ_n^* is well-defined in Theorem 5.2 since $g(t)$ is assumed nonincreasing. In the d-dimensional uniform d.f. case (Example 3.4), $g(\gamma_n^*)\sim(Ln)^{d-1}/(d-1)!$ if $d>1$, and g is bounded if $d=1$. In the d-dimensional interval case (Example 3.8), $g(\gamma_n^*)\sim n/Ln$ for all $d\geqq 1$.

By Remark 2.2, Theorem 5.2 remains valid if g is increased. This weakens the lower bound on P_n/P in the event in (5.6), but improves the bound on the probability of that event. For example, taking $g(t)=t^{-1}$ yields (1.7).

VI. Proof of the Inequality

The key to the proof of Theorem 2.1 will be an inequality from Alexander (1984). Recall from (1.12) that

$$h_1(\lambda) = (1 + \lambda^{-1}) \log(1 + \lambda) - 1, \; \lambda > 0.$$

It is readily shown that

$$h_1(\lambda)\uparrow, \quad \lambda^{-1} h_1(\lambda)\downarrow, \tag{6.1}$$

$$h_1(\lambda) \sim \begin{cases} \lambda/2 & \text{as } \lambda \to 0 \\ L\lambda & \text{as } \lambda \to \infty \end{cases}, \tag{6.2}$$

$$h_1(\lambda) \geqq \frac{\lambda}{2}(1 - \lambda) \quad \text{for all } \lambda > 0, \tag{6.3}$$

and

$$h_1(\lambda) \geqq \begin{cases} \lambda/4 & \text{if } \lambda \leqq 4 \\ (L\lambda)/2 & \text{if } \lambda \geqq 4 \end{cases}. \tag{6.4}$$

Bennett's inequality (Bennett, 1962) tells us that

$$\mathbb{P}[|v_n(C)| > M] \leqq 2 \exp(-M n^{\frac{1}{2}} h_1(M/n^{\frac{1}{2}} \sigma^2(C))) \tag{6.5}$$

for all $M \geqq 0$ and all C. Hence (6.4) and Theorem 2.8 of Alexander (1984) give us the following.

Proposition 6.1. *Let $\mathscr{C}$ be a VC class of sets, $n \geqq 1$, $M > 0$, and $\alpha \geqq \sup_{\mathscr{C}} \sigma^2(C)$ There exists a constant $K_o = K_o(V(\mathscr{C}))$ such that if either* (i)

$$M^2 \geqq K_o \alpha L(n/\alpha) \quad \textit{and} \quad M \geqq K_o L(n/\alpha)/n^{\frac{1}{2}} L(M/n^{\frac{1}{2}}\alpha) \tag{6.6}$$

or (ii)

$$M^2 \geqq K_o \alpha L \alpha^{-1} \quad \textit{and} \quad M \geqq K_o n^{-\frac{1}{2}} Ln \tag{6.7}$$

then

$$\mathbb{P}\left[\sup_{\mathscr{C}} |v_n(C)| > M\right] \leqq 16 \exp(-M^2/8\alpha) + 16 \exp(-\tfrac{1}{4} M n^{\frac{1}{2}} L(M/n^{\frac{1}{2}}\alpha)). \quad \square \tag{6.8}$$

The first term in (6.8) corresponds to a Gaussian approximation, the second to a Poisson. If $M/n^{\frac{1}{2}}\alpha$ is small, the Gaussian approximation is dominant; if it is large, the Poisson approximation dominates.

Proof of Theorem 2.1. Suppose first that (2.7) and (2.8) hold. Set

$$t_0 = \alpha, \; t_j = q^{-1}(2^{-j} q(\alpha)) \vee r \quad \text{for all } j \geqq 1,$$
$$N = \min\{j \geqq 0 : t_{j+1} \leqq r\}, \quad \text{and} \quad r' = (\tilde{a}^{-1}(p/n) \wedge z^{-1}(n^{\frac{1}{2}}/b)) \vee \gamma.$$

Then

$$\begin{aligned}
&\mathbb{P}[|v_n(C)|>bq(t)+u \quad \text{for some } \gamma \leqq t \leqq \alpha \quad \text{and} \quad C\in\mathscr{C}(t)] \\
&\leqq \mathbb{P}[|v_n(C)|>bq(t) \quad \text{for some } \gamma \leqq t < r' \quad \text{and} \quad C\in\mathscr{C}(t)] \\
&\quad + \sum_{j=0}^{N} \mathbb{P}[|v_n(C)|>bq(t_{j+1}) \quad \text{for some } C\in\mathscr{C}(t_j)] \\
&\equiv \mathbb{P}^{(0)} + \sum_{j=0}^{N} \mathbb{P}_j^{(1)}.
\end{aligned} \tag{6.9}$$

We begin with $\mathbb{P}^{(0)}$. Suppose $\gamma \leqq t < r'$, $C\in\mathscr{C}(t)$, and $P_n(C)=0$. Then

$$|v_n(C)| = n^{\frac{1}{2}} P(C) \leqq 2n^{\frac{1}{2}} \sigma^2(C) \leqq 2n^{\frac{1}{2}}\, q(t)/z(t) \leqq bq(t).$$

It follows that

$$\begin{aligned}
\mathbb{P}^{(0)} &\leqq \mathbb{P}[P_n(C)>0 \quad \text{for some } C\in\mathscr{C}(t) \quad \text{and} \quad \gamma \leqq t < r'] \\
&\leqq nP(E_{r'}) \leqq p.
\end{aligned} \tag{6.10}$$

Turning next to the $\mathbb{P}_j^{(1)}$, we have

$$\begin{aligned}
\mathbb{P}_j^{(1)} &\leqq \mathbb{P}\left[\sup_{\mathscr{C}(t_j)} |v_n(C)| > \tfrac{1}{2} bq(t_j) \,\middle|\, |v_n(\tilde{E}_{t_j})| \leqq \tfrac{1}{8} bq(t_j)\, \tilde{g}(t_j)^{\frac{1}{2}}\right] \\
&\quad + \mathbb{P}[|v_n(\tilde{E}_{t_j})| > \tfrac{1}{8} bq(t_j)\, \tilde{g}(t_j)^{\frac{1}{2}}] \\
&\equiv \mathbb{P}_j^{(2)} + \mathbb{P}_j^{(3)}.
\end{aligned} \tag{6.11}$$

Fix j and let k_1, k_2 be nonnegative integers such that

$$|v_n(\tilde{E}_{t_j})| \leqq \tfrac{1}{8} bq(t_j)\, \tilde{g}(t_j)^{\frac{1}{2}} \quad \text{if and only if } k_1 \leqq nP_n(\tilde{E}_{t_j}) \leqq k_2. \tag{6.12}$$

Then

$$\mathbb{P}_j^{(2)} \leqq \max_{k_1 \leqq k \leqq k_2} \mathbb{P}\left[\sup_{\mathscr{C}(t_j)} |v_n(C)| > \tfrac{1}{2} bq(t_j) \,|\, nP_n(\tilde{E}_{t_j}) = k\right]. \tag{6.13}$$

Define a new probability measure P^{t_j} by

$$P^{t_j}(\cdot) = P(\cdot \,|\, \tilde{E}_{t_j})$$

and let $\mathbb{P}_{t_j} = (P^{t_j})^\infty$ be the corresponding product measure on $(X^\infty, \mathscr{A}^\infty)$. ($\mathscr{C}(t_j)$ is easily shown to be deviation-measurable for P^{t_j}, since it is deviation-measurable for P by assumption.) Then

$$\begin{aligned}
&P^{t_j}(C) = P(C)/P(\tilde{E}_{t_j}) \quad \text{and} \\
&\sigma_{t_j}^2(C) \equiv P^{t_j}(C)\,(1-P^{t_j}(C)) \leqq \sigma^2(C)/P(\tilde{E}_{t_j}) \leqq \tilde{g}(t_j)^{-1} \quad \text{for } C\in\mathscr{C}(t_j).
\end{aligned} \tag{6.14}$$

Fix $k, k_1 \leqq k \leqq k_2$, and define

$$\mu_k = k^{\frac{1}{2}} (P_k - P^{t_j}).$$

By (6.12), for $C\in\mathscr{C}(t_j)$, if $nP_n(E_{t_j})=k$ then

$$\begin{aligned}|n^{\frac{1}{2}}P(C)-kn^{-\frac{1}{2}}P^{t_j}(C)|&=\frac{n^{\frac{1}{2}}P(C)}{P(\tilde{E}_{t_j})}\left|P(\tilde{E}_{t_j})-\frac{k}{n}\right|\\&\leqq 2\tilde{g}(t_j)^{-1}|v_n(\tilde{E}_{t_j})|\\&\leqq \tfrac{1}{4}bq(t_j).\end{aligned}$$

Now $\mathbb{P}[nP_n|_{\mathscr{C}(t_j)}\in\cdot\,|nP_n(\tilde{E}_{t_j})=k]=\mathbb{P}_{t_j}[kP_k\in\cdot]$, where $P_n|_{\mathscr{C}(t_j)}$ is the restriction of P_n to $\mathscr{C}(t_j)$. It follows that

$$\begin{aligned}&\mathbb{P}\left[\sup_{\mathscr{C}(t_j)}|v_n(C)|>\tfrac{1}{2}bq(t_j)\,|\,nP_n(\tilde{E}_{t_j})=k\right]\\&=\mathbb{P}_{t_j}\left[\sup_{\mathscr{C}(t_j)}n^{\frac{1}{2}}\left(\frac{kP_k(C)}{n}-P(C)\right)\right|>\tfrac{1}{2}bq(t_j)\Bigg]\\&=\mathbb{P}_{t_j}\left[\sup_{\mathscr{C}(t_j)}|(k/n)^{\frac{1}{2}}\mu_k(C)+kn^{-\frac{1}{2}}P^{t_j}(C)-n^{\frac{1}{2}}P(C)|>\tfrac{1}{2}bq(t_j)\right]\\&\leqq\mathbb{P}_{t_j}\left[\sup_{\mathscr{C}(t_j)}|\mu_k(C)|>\tfrac{1}{4}(n/k)^{\frac{1}{2}}bq(t_j)\right].\\&\equiv\mathbb{P}_{jk}^{(4)}.\end{aligned}\tag{6.15}$$

We now wish to apply Proposition 6.1 with $M=\frac{1}{4}(n/k)^{\frac{1}{2}}bq(t_j)$, and $\tilde{g}(t_j)^{-1}$ in the role of the α there (see (6.14)).

Case 1.

$z(t_j)\leqq 2n^{\frac{1}{2}}/b$, i.e.

$$bq(t_j)\leqq 2n^{\frac{1}{2}}\zeta(t_j).\tag{6.16}$$

Combining this with (6.12), we get

$$\frac{k}{n}\leqq\tilde{a}(t_j)+\tfrac{1}{8}n^{-\frac{1}{2}}bq(t_j)\,\tilde{g}(t_j)^{\frac{1}{2}}\leqq 2\tilde{a}(t_j)=2\zeta(t_j)\tilde{g}(t_j).\tag{6.17}$$

Hence by (2.2),

$$M^2\geqq b^2q^2(t_j)/32\zeta(t_j)\,\tilde{g}(t_j)\geqq\tfrac{1}{32}K\tilde{g}(t_j)^{-1}L\tilde{g}(t_j)\tag{6.18}$$

while by (2.3),

$$Mk^{\frac{1}{2}}\geqq\tfrac{1}{4}n^{\frac{1}{2}}bq(t_j)\geqq\tfrac{1}{4}KL\,(n\tilde{a}(t_j))\geqq\tfrac{1}{8}KLk.$$

If K is large enough we can now apply Proposition 6.1 (ii), and use (6.17) and (6.18) to get for the $\mathbb{P}_{jk}^{(4)}$ of (6.15):

$$\begin{aligned}\mathbb{P}_{jk}^{(4)}\leqq\;&16\exp(-b^2q^2(t_j)/2^8\zeta(t_j))\\&+16\exp(-2^{-4}n^{\frac{1}{2}}bq(t_j)L(bq(t_j)/8n^{\frac{1}{2}}\zeta(t_j))),\end{aligned}\tag{6.19}$$

so by (6.13), for the $\mathbb{P}_j^{(2)}$ of (6.11),

$$\begin{aligned}\mathbb{P}_j^{(2)} &\leqq 16 \exp(-b^2 q^2(t_j)/2^8 \zeta(t_j)) \\ &\quad + 16 \exp(-2^{-4} n^{\frac{1}{2}} bq(t_j)\, L(bq(t_j)/n^{\frac{1}{2}} \zeta(t_j)).\end{aligned} \tag{6.20}$$

Next, by (6.5) and (6.4) we have for the $\mathbb{P}_j^{(3)}$ of (6.11)

$$\begin{aligned}\mathbb{P}_j^{(3)} &\leqq 2 \exp(-\tfrac{1}{8} bq(t_j)\, \tilde{g}(t_j)^{\frac{1}{2}} n^{\frac{1}{2}} h_1(bq(t_j)/8n^{\frac{1}{2}} \zeta(t_j)\, \tilde{g}(t_j)^{\frac{1}{2}})) \\ &\leqq 2 \exp(-b^2 q^2(t_j)/2^8 \zeta(t_j))\end{aligned} \tag{6.21}$$

since, by (6.16), the argument of h_1 above is at most $\frac{1}{4}$. Combining this with (6.11) and (6.20), we see that for the $\mathbb{P}_j^{(1)}$ of (6.9),

$$\begin{aligned}\mathbb{P}_j^{(1)} &\leqq 18 \exp(-b^2 q^2(t_j)/2^8 \zeta(t_j)) + 18 \exp(-2^{-8} n^{\frac{1}{2}} bq(t_j)\, L(bq(t_j)/n^{\frac{1}{2}} \zeta(t_j))) \\ &\quad + 34 \exp(-2^{-8} n^{\frac{1}{2}} bq(t_j)\, L\tilde{g}(t_j)).\end{aligned} \tag{6.22}$$

(The last term in (6.22) is superfluous now but will be used later.)

Case 2.

$z(t_j) > 2n^{\frac{1}{2}}/b$. Then $r \leqq t_j < s$, so (2.4) holds, and

$$bq(t_j) > 2n^{\frac{1}{2}} \zeta(t_j). \tag{6.23}$$

We wish to use Proposition 6.1 (i). To prove (6.6) it is sufficient to show that, for the constant K_o of that proposition,

$$M^2 \tilde{g}(t_j) \geqq \tfrac{1}{4} M k^{\frac{1}{2}} L(M\tilde{g}(t_j)/k^{\frac{1}{2}}) \geqq K_o L(k\tilde{g}(t_j)). \tag{6.24}$$

We need two subcases, according to which of the two terms added in (6.17) is the larger.

Case 2a.

$$bq(t_j) \leqq 8n^{\frac{1}{2}}\, \zeta(t_j)\tilde{g}(t_j)^{\frac{1}{2}}. \tag{6.25}$$

Then (6.17) is again valid. By (6.17), (6.23), and (2.4),

$$\begin{aligned}M\tilde{g}(t_j)/k^{\frac{1}{2}} &= n^{\frac{1}{2}} bq(t_j)\, \tilde{g}(t_j)/4k \\ &\geqq bq(t_j)/8n^{\frac{1}{2}} \zeta(t_j) \qquad (6.26) \\ &\geqq (\tfrac{1}{4}) \vee (K/8n\zeta(t_j)). \qquad (6.27)\end{aligned}$$

By (6.27), $M\tilde{g}(t_j)/k^{\frac{1}{2}} \geqq \frac{1}{4} L(M\tilde{g}(t_j)/k^{\frac{1}{2}})$, and the first inequality in (6.24) follows. The second follows, if K is large enough, from the following inequalities, which are consequences of (6.17), of (2.4) and (6.27), and of (6.26) and (2.5), respectively.

$$\begin{aligned}L(k\tilde{g}(t_j)) &\leqq L(2n\tilde{a}(t_j)\tilde{g}(t_j) \leqq L(2/n\zeta(t_j)) + 2L(n\tilde{a}(t_j)) \\ Mk^{\frac{1}{2}} L(M\tilde{g}(t_j)/k^{\frac{1}{2}}) &\geqq \tfrac{1}{4} KL(2/n\zeta(t_j)) \\ Mk^{\frac{1}{2}} L(M\tilde{g}(t_j)/k^{\frac{1}{2}}) &\geqq \tfrac{1}{4} n^{\frac{1}{2}} bq(t_j)\, L(bq(t_j)/8n^{\frac{1}{2}} \zeta(t_j)) \\ &\geqq \tfrac{1}{4} KL(n\tilde{a}(t_j)).\end{aligned} \tag{6.28}$$

Thus (6.24) holds as desired. Proposition 6.1 (i), (6.17), and (6.28) now tell us that (6.19), and therefore (6.20), are again valid. From (6.25) we see that the argument of h_1 in (6.21) is at most 1, so by (6.5) and (6.4), (6.21) again holds. Therefore so does (6.22).

Case 2b.

$$bq(t_j) > 8n^{\frac{1}{2}}\zeta(t_j)\,\tilde{g}(t_j)^{\frac{1}{2}}. \tag{6.29}$$

Here, by (6.12) and (6.29),

$$\frac{k}{n} \leqq \tilde{a}(t_j) + \tfrac{1}{8}n^{-\frac{1}{2}}\,\mathrm{bq}(\mathrm{t}_j)\,\tilde{g}(t_j)^{\frac{1}{2}} \leqq \tfrac{1}{4}n^{-\frac{1}{2}}\,bq(t_j)\,\tilde{g}(t_j)^{\frac{1}{2}}. \tag{6.30}$$

It follows that

$$M\tilde{g}(t_j)/k^{\frac{1}{2}} = n^{\frac{1}{2}}\,bq(t_j)\,\tilde{g}(t_j)/4k \geqq \tilde{g}(t_j)^{\frac{1}{2}} \geqq 1, \tag{6.31}$$

so the first inequality in (6.24) holds. By (6.30), (2.4), and the first inequality in (6.31), if K is large enough then

$$\begin{aligned} K_o L(k\tilde{g}(t_j)) &\leqq K_0 L(\tfrac{1}{4}n^{\frac{1}{2}}\,bq(t_j)) + \tfrac{3}{2}K_0 L\tilde{g}(t_j) \\ &\leqq \tfrac{1}{32}n^{\frac{1}{2}}\,bq(t_j)\,L\tilde{g}(t_j) \\ &\leqq \tfrac{1}{4}Mk^{\frac{1}{2}}L(M\tilde{g}(t_j)/k^{\frac{1}{2}}). \end{aligned}$$

Thus we have (6.24). From Proposition 6.1 (i), (6.24), and (6.31) we conclude that

$$\begin{aligned} \mathbb{P}_{jk}^{(4)} &\leqq 16\,\exp(-\tfrac{1}{8}M^2\tilde{g}(t_j)) + 16\,\exp(-\tfrac{1}{4}Mk^{\frac{1}{2}}L(M\tilde{g}(t_j)/k^{\frac{1}{2}})) \\ &\leqq 32\,\exp(-\tfrac{1}{32}Mk^{\frac{1}{2}}L(M\tilde{g}(t_j)/k^{\frac{1}{2}})) \\ &\leqq 32\,\exp(-2^{-8}n^{\frac{1}{2}}\,bq(t_j)\,L\tilde{g}(t_j)) \end{aligned}$$

so by (6.13),

$$\mathbb{P}_j^{(2)} \leqq 32\exp(-2^{-8}n^{\frac{1}{2}}\,bq(t_j)\,L\tilde{g}(t_j)). \tag{6.32}$$

To bound $\mathbb{P}_j^{(3)}$ we use (6.5) to conclude that

$$\begin{aligned} \mathbb{P}_j^{(3)} &\leqq 2\exp(-\tfrac{1}{8}bq(t_j)\,\tilde{g}(t_j)^{\frac{1}{2}}n^{\frac{1}{2}}h_1(bq(t_j)/8n^{\frac{1}{2}}\zeta(t_j)\tilde{g}(t_j)^{\frac{1}{2}})) \\ &\leqq 2\exp(-\tfrac{1}{32}bq(t_j)\,n^{\frac{1}{2}}L\tilde{g}(t_j)). \end{aligned} \tag{6.33}$$

In the second inequality here we have used (6.29) to tell us that the argument of h_1 in (6.33) is at least 1, along with the fact that $h_1(1) > \frac{1}{4}$. From (6.11), (6.32), and (6.33) it now follows that (6.22) holds.

Having established (6.22) for all j, it remains to sum it over j. Define

$$\begin{aligned} \varphi_1(t) &= 2^{-8}n^{\frac{1}{2}}\,bq(t)\,L(bq(t)/n^{\frac{1}{2}}\zeta(t)), \\ \varphi_2(t) &= 2^{-8}n^{\frac{1}{2}}\,bq(t)\,L\tilde{g}(t). \end{aligned}$$

By monotonicity of $g(t)/\zeta(t)$ and $q(t)^{\beta-1}\varphi_i(t)$ (which follows from (2.7)), we have

$$t_{j+1} \leqq t_j/2 \quad \text{and} \quad \varphi_i(\mathrm{t}_{j+1}) \leqq 2^{-(1-\beta)}\varphi_i(t_j) \quad (j<N, i=1,2) \tag{6.34}$$

Hence, using also the monotonicity of $\zeta(t)/t$,

$$\sum_{j=0}^{N} \exp\left(-b^2 q^2(t_j)/2^8 \zeta(t_j)\right)$$

$$= \sum_{j=0}^{N} 2\frac{t_j - t_j/2}{t_j} \exp\left(-b^2 q^2(t_j)/2^8 \zeta(t_j)\right)$$

$$\leqq 2 \int_{r/2}^{\alpha} t^{-1} \exp\left(-b^2 q^2(t)/2^9 \zeta(t)\right). \tag{6.35}$$

Summing the second term in (6.22) over j, (6.34) and (2.8) may be applied to show, for $i=1,2$,

$$\sum_{j=0}^{N} \exp(-\varphi_i(t_j)) \leqq \sum_{j=0}^{N} \exp\left(-2^{(1-\beta)(N-j)} \varphi_i(t_N)\right)$$

$$\leqq \sum_{j=0}^{N} \exp\left(-(1+\theta(N-j))\, \varphi_i(t_N)\right)$$

$$\leqq 2\exp(-\varphi_i(t_N)) \tag{6.36}$$

where $\theta = 2^{1-\beta} - 1 \geqq (1-\beta)\log 2$. The theorem now follows from (6.9), (6.10), (6.22), (6.35), and (6.36).

When we do not have (2.7) or (2.8), we replace $L(\cdot)$ by its lower bound 1 in (6.22), and observe that the second term in (6.22) is then not needed, since the third term in (6.22) becomes an upper bound for the second term in (6.20). Otherwise the proof remains essentially the same. □

VII. Proofs of the General Results

Throughout this section all inequalities in proofs should be taken to have the unstated qualification that the index n or k (which one will be clear from the context) is sufficiently large.

We begin with a lemma demonstrating that a well-known fact about stopping times τ remains true even if τ is not measurable. It is included solely to avoid unwieldy measurability assumptions and is not central to our arguments.

Lemma 7.1. *Let τ be given on $(X^\infty, \mathscr{A}^\infty, \mathbb{P})$ by $\tau = \min\{m: (X_1, \dots, X_m) \in A_m\}$ for some sets $A_m \subset X^m$, and let $n \geqq 1, 0 < \beta < 1$, and $F \subset X^n$. Suppose for each $(x_1, \dots, x_m) \in A_m$, $m < n$, there is a set $B = B(x_1, \dots, x_m) \subset X^{n-m}$ such that*

$$\mathbb{P}^*[(X_{m+1}, \dots, X_n) \in B] \geqq \beta, \quad \textit{and} \tag{7.1}$$

$$(x_{m+1}, \dots, x_n) \in B\,(x_1, \dots, x_m) \quad \textit{implies} \quad (x_1, \dots, x_n) \in F. \tag{7.2}$$

Then $\mathbb{P}^*[\tau \leqq n] \leqq \beta^{-1}\, \mathbb{P}^*[(X_1, \dots, X_n) \in F]$.

Proof. We may write $\mathbb{P}^*(D)$ for $\mathbb{P}^*((X_1, \dots, X_m) \in D)$ for any m and $D \subset X^m$. Let G and $D_1, \dots, D_n$ be measurable sets with $G \supset F$, $D_m \supset [\tau \leqq m]$, $\mathbb{P}(G) = \mathbb{P}^*(F)$, and $\mathbb{P}(D_m) = \mathbb{P}^*[\tau \leqq m]$. Define τ^* on X^∞ by $\tau^*(\omega) = \min\{m: \omega \in D_m\}$. Then $\tau^* \leqq \tau$ so

$$\mathbb{P}^*[\tau \leqq n] \leqq \mathbb{P}[\tau^* \leqq n] = \sum_{m \leqq n} \mathbb{P}[\tau^* = m] = \sum_{m \leqq n} \mathbb{P}(D_m \setminus D_{m-1}) \tag{7.3}$$

where $D_0=\phi$. Fix M and write $(X^\infty, \mathscr{A}^\infty, \mathbb{P})$ as $(\Omega_1, \mathscr{A}^m, P_1)\times(\Omega_2, \mathscr{A}^\infty, P_2)$ where Ω_1 is a copy of X^m on which $X_1,\dots,X_m$ are defined, Ω_2 is a copy of X^∞ on which $X_{m+1}, X_{m+2},\dots$ are defined, P_1 is P^m, and P_2 is P^∞. For $\omega_1\in\Omega_1$ let $s_A(\omega_1)=\{\omega_2:(\omega_1,\omega_2)\in G\}$ be the section of G over ω_1. Then

$$\begin{aligned}\mathbb{P}[G\cap(D_m\setminus D_{m-1})] &= \int_{\Omega_1} P_2(s_A(\omega_1))\,dP_1(\omega_1)\\ &\geqq \beta P_1[\omega_1: P_2(s_A(\omega_1))\geqq\beta]. \end{aligned}\tag{7.4}$$

Now if $\omega_1\in[\tau\leqq m]\cap(D_m\setminus D_{m-1})$ (viewed as a subset of X^m), then by (7.1) and (7.2),

$$P_2(s_A(\omega_1))\geqq P_2^*\,[(X_{m+1},\dots,X_n)\in B(X_1(\omega_1),\dots,X_m(\omega_1))]\geqq\beta.$$

It follows using (7.4) that

$$\mathbb{P}[G\cap(D_m\setminus D_{m-1})]\geqq\beta P_1^*([\tau\leqq m]\cap(D_m\setminus D_{m-1})).\tag{7.5}$$

If E is a measurable set containing $[\tau\leqq m]\cap(D_m\setminus D_{m-1})$ then

$$\begin{aligned}P_1(E)&\geqq P_1(E\cup D_{m-1})-P_1(D_{m-1})\geqq P_1^*[\tau\leqq m]-P_1(D_{m-1})\\ &=P_1(D_m)-P_1(D_{m-1})=P_1(D_m\setminus D_{m-1}),\end{aligned}$$

so $P_1^*([\tau\leqq m]\cap(D_m\setminus D_{m-1}))\geqq P_1(D_m\setminus D_{m-1})$. Combining this with (7.5) and (7.3) we see that

$$\begin{aligned}\mathbb{P}^*[\tau\leqq n]&\leqq\sum_{m\leqq n}\beta^{-1}\,\mathbb{P}[G\cap(D_m\setminus D_{m-1})]\\ &\leqq\beta^{-1}\,\mathbb{P}(G)=\beta^{-1}\,\mathbb{P}^*[(X_1,\dots,X_n)\in F].\quad\square\end{aligned}$$

The asymptotic upper bounds for the weighted empirical process will be obtained from Theorem 2.1 with the help of the following lemma.

Lemma 7.2. *Let $\mathscr{C}$ be a class of sets, let $q\in Q$, and let (b_n), (u_n), (γ_n), (α_n) be nonnegative sequences with*

$$n^{-1}b_n\downarrow,\ u_n\downarrow,\ \gamma_n\downarrow,\ n\alpha_n\uparrow.\tag{7.6}$$

Define events

$$A_n=[|\nu_n(C)|>b_nq(\sigma^2(C))+u_n \quad \textit{for some } C\in\mathscr{C} \quad \textit{with} \quad \gamma_n\leqq\sigma^2(C)\leqq\alpha_n],$$

$$A_n'(\varepsilon)=[|\nu_n(C)|>(1-\varepsilon)(b_nq(\sigma^2(C))+u_n) \quad \textit{for some } C\in\mathscr{C} \quad \textit{with} \quad \gamma_n\leqq\sigma^2(C)\leqq(1+\varepsilon)\alpha_n].$$

Suppose

$$\inf\{b_nt^{-\frac12}q(t): n\geqq1,\ t\in[\gamma_n,\alpha_n]\}>0,\tag{7.7}$$

and suppose that for some $\varepsilon,\theta>0$,

$$\mathbb{P}^*(A_n'(\varepsilon))=O((Ln)^{-(1+\theta)}).\tag{7.8}$$

Then $\mathbb{P}(A_n\ i.o.)=0$.

Proof. Let r be the infimum in (7.7) and choose $\delta > 0$ small enough so $2\delta^2 + 2r^{-1}\delta < \varepsilon$. Fix m, $(1-\delta^2)n < m \leqq n$. If $(x_1, \ldots, x_m) \in A_m$, then there exists $C = C(x_1, \ldots, x_m) \in \mathscr{C}$ with

$$|v_m(C)| > b_m q(\sigma^2(C)) + u_m \geqq (1-\delta^2)\,(b_n q(\sigma^2(C)) + u_n)$$

and

$$\gamma_n \leqq \gamma_m \leqq \sigma^2(C) \leqq \alpha_m \leqq (1+\delta^2)\,\alpha_n; \tag{7.9}$$

here we have used (7.6). If (7.9) occurs for some C and

$$|(nP_n - mP_m - (n-m)P)\,(C)| \leqq 2(n-m)^{\frac{1}{2}}\sigma(C) \leqq 2n^{\frac{1}{2}}\delta\sigma(C), \tag{7.10}$$

then by the definition of r,

$$\begin{aligned} |(v_n(C)| &\geqq (1-\delta^2)|v_m(C)| - 2\delta\sigma(C) \\ &\geqq (1 - 2\delta^2 - 2r^{-1}\delta)\,(b_n q(\sigma^2(C)) + u_n) \\ &\geqq (1-\varepsilon)\,(b_n q(\sigma^2(C)) + u_n). \end{aligned}$$

Let $B = B(x_1, \ldots, x_m)$ be the event that (7.10) holds for $C = C(x_1, \ldots, x_m)$; it follows from the above that (7.2) holds for $F = A_n'(\varepsilon)$. By Čebyšev's inequality, (7.10) occurs with probability more than 1/2 for any fixed C; thus (7.1) holds with $\beta = 1/2$. Hence by Lemma 7.1,

$$\mathbb{P}^*\left(\bigcup_{(1-\delta^2)n < m \leqq n} A_m\right) = \mathbb{P}^*[\tau \leqq n] \leqq 2\,\mathbb{P}^*(A_n'(\varepsilon)). \tag{7.11}$$

For $n(k) = [(1+\delta^2/2)^k]$ (the integer part), we have by (7.8) that $\sum_{k \geqq 1} \mathbb{P}^*(A_{n(k)}'(\varepsilon)) < \infty$, and the lemma then follows from (7.11) and Borel-Cantelli. □

For asymptotic lower bounds on full classes, our method is modeled somewhat after that of Stute (1982a). Let $b(j,n,p) = \binom{n}{j} p^j (1-p)^{n-j}$ denote the binomial probability,

$$B^+(k,n,p) = \sum_{j=k}^{n} b(j,n,p)$$

the binomial upper tail, and

$$h_2(\lambda) = \lambda/(1+\lambda) - \log(1+\lambda) = -(1+\lambda^{-1})^{-1} h_1(\lambda).$$

The following generalizes Lemma 1 of Kiefer (1972).

Lemma 7.3. *Let* (p_n), (k_n), (l_n) *be nonnegative sequences (with* $n \geqq 1$*) satisfying* $p_n \to 0$, $k_n \to \infty$, *and* $k_n \leqq l_n = o(n)$. *Let* $\lambda_0 > 0$. *Then there exists* $\eta_n \to 0$ *such that*

$$|\log B^+(k,n,p) - k h_2(\lambda)| \leqq \eta_n k h_2(\lambda)$$

whenever

$$k = (1+\lambda)np,\ k_n \leqq k \leqq l_n,\ 0 < p \leqq p_n, \quad \text{and} \quad \lambda \geqq \lambda_0. \ \square \tag{7.12}$$

Proof. Fix n, k, p, and λ satisfying (7.12). Comparison to a geometric series shows

$$b(k,n,p) \leqq B^+(k,n,p) \leqq (1+\lambda^{-1})\,b(k,n,p),$$

so

$$|\log B^{+}(k,n,p) - \log b\,(k,n,p)| \leqq (k_n \lambda_0)^{-1} k. \tag{7.13}$$

Define

$$b_0(k,n,p) = \left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k}, \quad b_1(n,k) = \left(\frac{n}{2\pi k(n-k)}\right)^{\frac{1}{2}}.$$

By Stirling's formula, if n is large,

$$\tfrac{1}{2} b(k,n,p) \leqq b_0(k,n,p)\; b_1(n,k) \leqq 2b(k,n,p). \tag{7.14}$$

Now

$$|2 \log b_1(n,k)| = |\log 2\pi + \log k + \log(1-(1+\lambda)p)| \leqq \eta_n' k \tag{7.15}$$

for some $\eta_n' \to 0$, since $k_n \to \infty$ and $(1+\lambda)p \leqq l_n/n \to 0$. Also

$$\begin{aligned} |h_2(\lambda) - k^{-1} \log b_0(k,n,p)| &= \left| \frac{\lambda}{1+\lambda} - \frac{1-(1+\lambda)p}{(1+\lambda)\mathrm{p}} \log\left(1 - \frac{\lambda p}{1-p}\right)\right| \\ &\leqq \left| \frac{\lambda}{1+\lambda}\left(1 - \frac{1-(1+\lambda)p}{1-p}\right) + \theta\left(\frac{1-(1+\lambda)p}{(1+\lambda)p}\right)\left(\frac{\lambda p}{1-p}\right)^2 \right| \\ &\leqq \eta_n'' \to 0 \quad \text{as} \quad n \to \infty, \end{aligned} \tag{7.16}$$

where $|\theta| \leqq 1$ if n is large, since $p_n \to 0$, $\lambda p \leqq (1+\lambda)p \leqq l_n/n \to 0$, and $|\log(1-x) + x^2 + x^2/2| = o(x^2)$ as $x \to 0$. The lemma now follows from (7.13)–(7.16), since $h_2(\lambda)$ is bounded away from 0 for $\lambda \geqq \lambda_0$. □

Lemma 7.3 can be translated into a statement about $\nu_n(C)$ for easier later use. This we state as the next lemma.

Lemma 7.4. *Let $\lambda_0 > 0$ and let (p_n), (m_n), and (M_n) be nonnegative sequences satisfying $p_n \to 0$, $n^{-1/2} = o(m_n)$, and $M_n = o(n^{1/2})$. Then there exists $\eta_n \to 0$ such that*

$$|-\log \mathbb{P}[\nu_n(C) > M] - Mn^{\frac{1}{2}} h_1(M/n^{\frac{1}{2}}\sigma^2(C))| \leqq \eta_n Mn^{\frac{1}{2}} h_1(M/n^{\frac{1}{2}}\sigma^2(C))$$

whenever $P(C) \leqq p_n$, $M/n^{\frac{1}{2}}\sigma^2(C) \geqq \lambda_0$, and $m_n \leqq M \leqq M_n$. □

The next lemma is a version of Theorem 5.2.2 (iii) of Stout (1974). It covers the case, excluded in Lemma 7.4, when $M/n^{\frac{1}{2}}\sigma^2(C)$ is near 0.

Lemma 7.5. *For each $\theta > 0$ there exist $K, \lambda_0 > 0$ such that*

$$\mathbb{P}[\nu_n(C) > M] \geqq \exp(-(1+\theta)M^2/2\sigma^2(C))$$

whenever $n \geqq 1$, $M^2 \geqq K\sigma^2(C)$, and $M/n^{\frac{1}{2}}\sigma^2(C) \leqq \lambda_0$. □

Proof of Theorem 3.1 (A). It is easily verified that $nb_n^2\gamma_n \to \infty$ in each of (i)–(iii). It follows that

$$\sup\{n^{-\frac{1}{2}}/b_n\sigma(C) : C \in \mathscr{C},\ \sigma^2(C) \geqq \gamma_n\} \leqq (nb_n^2\gamma_n)^{-\frac{1}{2}} = o(1).$$

From the Kolmogorov 0–1 law we then conclude that the lim sup in (3.4) is some constant a.s. By Lemma 7.2, to prove (A) it suffices to show that for each $\delta > 0$,

$$\begin{aligned} \mathbb{P}^*[|\nu_n(C)| > (1+8\delta)\, R_\delta b_n \sigma(C) \quad &\text{for some } \mathrm{C} \in \mathscr{C} \text{ with } \sigma^2(C) \geqq \gamma_n] \\ &= O((Ln)^{-(1+\delta)}) \end{aligned} \tag{7.17}$$

where $R_\delta \downarrow R_0$ as $\delta \to 0$ and R_0 is the upper bound for R given in whichever of (i), (ii), or (iii) we are considering; R_δ will be specified later.

Fix $0<\delta<1/8$. Let (α_n) and (u_n) be nonincreasing sequences, to be specified later, with $\alpha_n \geqq \gamma_n$ and $0<u_n<\delta$. Fix n and set

$$\begin{aligned} t_j &= (1-u_n^2/4)^j \alpha_n, \; j \geqq 0, \\ \mathscr{E}(j) &= \mathscr{C}_{t_j} \setminus \mathscr{C}_{t_{j+1}}, \quad \text{and} \\ N_n &= \min\{j \geqq 0 : t_{j+1} \leqq \gamma_n\}. \end{aligned}$$

By (3.3) there exists $\mathscr{F}(j) \subset \mathscr{E}(j)$ for all $j \leqq N_n$, and $A = A(\delta) < \infty$ such that

$$|\mathscr{F}(j)| \leqq A u_n^{-\eta} g(\gamma_n)^{\varrho+\delta} \leqq A u_n^{-\eta} \exp((\varrho+\delta)(c_1+\delta) w_n) \quad \text{for all } j \leqq N_n, \tag{7.18}$$

and such that for each $C \in \mathscr{E}(j)$, there is a $C_0(C) \in \mathscr{F}(j)$ with $P(C \triangle C_0(C)) \leqq u_n^2 t_j$. Set

$$\mathscr{C}(t) = \{C \setminus D : C, D \in \mathscr{C}_t, \; \sigma^2(C \setminus D) \leqq u_n^2 t\}. \tag{7.19}$$

Since

$$|\nu_n(C)| \leqq |\nu_n(C_0(C))| + |\nu_n(C \setminus C_0(C))| + |\nu_n(C_0(C) \setminus C)|,$$

we have

$$\begin{aligned} &\mathbb{P}^*[|\nu_n(C)| > (1+8\delta) R_\delta b_n \sigma(C) \quad \text{for some } C \in \mathscr{C} \text{ with } \gamma_n \leqq \sigma^2(C) \leqq \alpha_n] \\ &\quad \leqq \mathbb{P}^*[|\nu_n(C)| > (1+8\delta) R_\delta b_n t_{j+1}^{1/2} \quad \text{for some } j \leqq N_n \text{ and } C \in \mathscr{E}(j)] \\ &\quad \leqq \sum_{j=0}^{N_n} \mathbb{P}[|\nu_n(C)| > (1+4\delta) R_\delta b_n t_j^{1/2} \quad \text{for some } C \in \mathscr{F}(j)] \\ &\qquad + \mathbb{P}^*[|\nu_n(C)| > \delta R_\delta n_n t^{1/2} \quad \text{for some } \gamma_n \leqq t \leqq \alpha_n \text{ and } C \in \mathscr{C}(t)] \\ &\quad \equiv \sum_{j=0}^{N_n} \mathbb{P}_j + \mathbb{P}_n^*. \end{aligned} \tag{7.20}$$

We now consider separately the cases (i)–(iii) of (3.4).

Proof of (A) (i). Here we take $\alpha_n \equiv 1/4$, $u_n \equiv u$ for some $0<u<\delta$ to be specified later, and $R_\delta = (2((\varrho+\delta)(c_1+\delta)+c_2+c_3+2\delta))^{\frac{1}{2}}$.

Now $\sigma^2(C) \leqq t_j$ for $C \in \mathscr{F}(j)$, and $\max_{j \leqq N_n} w_n^{1/2}/n^{\frac{1}{2}} t_j^{\frac{1}{2}} \to 0$ as $n \to \infty$ since $n^{-1} w_n = o(\gamma_n)$, so by (6.5), (6.2), and (7.18),

$$\begin{aligned} \mathbb{P}_j &\leqq 2|\mathscr{F}(j)| \exp(-(1+4\delta) R_\delta^2 w_n/2) \\ &\leqq \exp(-[(1+4\delta) R_\delta^2/2 - (\varrho+\delta)(c_1+\delta)] w_n) \\ &\leqq \exp(-(1+\delta)(c_2+c_3+2\delta) w_n). \end{aligned}$$

Since $N_n \leqq (\log(1-u^2/4)^{-1})^{-1} L\gamma_n^{-1}$, it follows that

$$\begin{aligned} \sum_{j=0}^{N_n} \mathbb{P}_j &\leqq N_n \exp(-(1+\delta)(c_2+c_3+2\delta) w_n) \\ &\leqq \exp(-(1+\delta)(c_3+\delta) w_n) \\ &\leqq \exp(-(1+\delta) LLn). \end{aligned} \tag{7.21}$$

To bound $\mathbb{P}_n^*$ we use Theorem 2.1, with $\zeta(t)=u^2 t$, $b=\delta R_\delta w_n^{\frac{1}{2}}$, $\gamma=\gamma_n$, $\alpha=\frac{1}{4}$, $q(t)=t^{\frac{1}{2}}$, and $\mathscr{C}(t)$ from (7.19). In the notation of Theorem 2.1, $\tilde{a}(t)$ is at most $a(t)$, we can take $\tilde{g}(t)$ to be $u^{-2}g(t)$ (see Remark 2.2), $z(t)$ is $u^{-2}t^{-\frac{1}{2}}$, r is γ_n, and s is $w_n/4u^2 n=o(\gamma_n)$, so $r>s$ and (2.4) and (2.5) are vacuous. Let K be the constant from Theorem 2.1; if we take $u\leqq\delta R_\delta K^{-\frac{1}{2}}$ then

$$Lg(t)\leqq Lg(\gamma_n)\leqq b^2/u^2 K \quad \text{for all } \gamma_n\leqq t\leqq\tfrac{1}{4}$$

and (2.2) follows. Since $n^{-1}w_n=o(\gamma_n)$,

$$t^{-\frac{1}{2}}Lg(t)\leqq\gamma_n^{-\frac{1}{2}}Lg(\gamma_n)=o((nw_n)^{\frac{1}{2}})$$

and

$$(nt)^{-\frac{1}{2}}L(nt)\leqq(n\gamma_n)^{-\frac{1}{2}}L(n\gamma_n)=o(1)$$

for all $t\geqq\gamma_n$, so

$$Kt^{-\frac{1}{2}}L(n\tilde{a}(t))\leqq Kt^{-\frac{1}{2}}L(nt)+Kt^{-\frac{1}{2}}Lg(t)\leqq\delta R_\delta(nw_n)^{\frac{1}{2}}$$

for all $t\geqq\delta_n$, and (2.3) follows. Theorem 2.1 now tells us that, if we take $u^2<\delta^2/512$,

$$\begin{aligned}\mathbb{P}_n^* &\leqq 36\int_{\gamma_n/2}^{\frac{1}{4}} t^{-1}\exp(-\delta^2 R_\delta^2 w_n/512u^2)\,dt\\ &\quad+68\exp(-\delta R_\delta(n\gamma_n w_n)^{\frac{1}{2}}/256)\\ &\leqq 140(L\gamma_n^{-1})\exp(-(1+\delta)R_\delta^2 w_n/2)\\ &\leqq 140\exp(-(1+\delta)(c_3+\delta)w_n)\\ &\leqq 140\exp(-(1+\delta)LLn).\end{aligned}$$

With (7.20) and (7.21) this proves (7.17), and (A) (i) follows.

Proof of (A) (ii). Here we take $u_n\equiv u$ for some $0<u<\delta$ to be specified later, and $R_\delta=\max(R_{\delta1},R_{\delta2})$ with $R_{\delta1}=(2((\varrho+\delta)(c_1+\delta)+c_2+c_3+2\delta))^{\frac{1}{2}}$ and $R_{\delta2}=\tau^{\frac{1}{2}}(\beta_{\theta(\delta)\tau}-1)$, where $\theta(\delta)=((\varrho+\delta)(c_1+\delta)+c_3+\delta)^{-1}$. We take $\alpha_n\downarrow 0$ with

$$\gamma_n=o(\alpha_n)\quad\text{and}\quad L(\gamma_n^{-1}\alpha_n)=o(w_n). \tag{7.22}$$

Set $\tilde{w}_n=Lg(\alpha_n)\vee LLn$. Since

$$Lg(\alpha_n)\leqq Lg(\gamma_n)\leqq L(\alpha_n^{-1}a(\alpha_n))+L(\gamma_n^{-1}\alpha_n)=Lg(\alpha_n)+o(w_n),$$

we have $w_n\sim\tilde{w}_n$ and $n^{-1}\tilde{w}_n=o(\alpha_n)$. It is easily then verified that the constants c_1, c_2, c_3 are unchanged if γ_n is replaced by α_n in (3.2). Hence by the above proof of part (A) (i) of the theorem,

$$\begin{aligned}\mathbb{P}^*[|v_n(C)|>(1+8\delta)R_{\delta1}w_n^{\frac{1}{2}}\sigma(C)\quad&\text{for some } C\in\mathscr{C}\ \text{with}\ \sigma^2(C)\geqq\alpha_n]\\ &=O((Ln)^{-(1+\delta)}).\end{aligned} \tag{7.23}$$

Hence to obtain (7.17) it suffices to bound the left side (or therefore the right side) of (7.20) by $O((Ln)^{-(1+\delta)})$.

Define $\lambda_t=(1+4\delta)R_{\delta2}w_n^{\frac{1}{2}}/(nt)^{\frac{1}{2}}$. Then by (6.5) and (6.1), since $\sigma^2(C)\leqq t_j$ for $C\in\mathscr{F}(j)$,

$$\begin{aligned}\mathbb{P}_j &\leqq \mathbb{P}[|v_n(C)| > (1+4\delta) R_{\delta 2} w_n^{\frac{1}{2}} t_j^{\frac{1}{2}} \quad \text{for some } C \in \mathscr{F}(j)] \\ &\leqq 2|\mathscr{F}(j)| \exp(-(1+4\delta)^2 R_{\delta 2}^2 w_n \lambda_{t_j}^{-1} h_1(\lambda_{t_j})) \\ &\leqq 2|\mathscr{F}(j)| \exp(-(1+4\delta)^2 R_{\delta 2}^2 w_n \lambda_{\gamma_n}^{-1} h_1(\lambda_{\gamma_n})). \end{aligned} \tag{7.24}$$

Set $\tau_n = n\gamma_n w_n^{-1}$ and $\xi = \beta_{\theta(\delta)\tau} - 1$. Since $\tau_n \to \tau$, we have $\lambda_{\gamma_n}^{-1} = (1+4\delta) R_{\delta 2} \tau_n^{-\frac{1}{2}} \geqq (1+3\delta)\xi$. From (1.11) we know that $\xi h_1(\xi) = (\theta(\delta)\tau)^{-1}$. It follows that

$$\begin{aligned}\lambda_{\gamma_n}^{-1} h_1(\lambda_{\gamma_n}) &\geqq \lambda_{\gamma_n}^{-2}(1+3\delta)\xi h_1(\xi) \\ &= (1+3\delta)\tau_n/(1+4\delta)^2 R_{\delta 2}^2 \theta(\delta)\tau \\ &\geqq (1+2\delta)/(1+4\delta)^2 R_{\delta 2}^2 \theta(\delta)\end{aligned}$$

so by (7.24) and (7.18),

$$\begin{aligned}\sum_{j=0}^{N_n} \mathbb{P}_j &\leqq \sum_{j=0}^{N_n} 2|\mathscr{F}(j)| \exp(-(1+2\delta)\theta(\delta)^{-1} w_n) \\ &\leqq 2\mathrm{Au}^{-\eta} N_n \exp(-(1+2\delta)(c_3+\delta) w_n) \\ &\leqq \exp(-(1+\delta)(c_3+\delta) w_n) \\ &\leqq \exp(-(1+\delta) LLn)\end{aligned} \tag{7.25}$$

since $LN_n = O(LL(\gamma_n^{-1}\alpha_n)) = o(w_n)$ by (7.22).

To bound $\mathbb{P}_n^*$ we use Theorem 2.1, as in the proof of (A) (i), with $\zeta(t) = u^2 t$, $b = \delta R_{\delta 2} w_n^{\frac{1}{2}}$, $\gamma = \gamma_n$, $\alpha = \alpha_n$, $q(t) = t^{\frac{1}{2}}$, and $\mathscr{C}(t)$ from (7.19). Again $\tilde{a}(t)$ is at most $a(t)$, $\tilde{g}(t)$ is $u^{-2} g(t)$, $z(t) = u^{-2} t^{-\frac{1}{2}}$, r is γ_n, and s is $s_n = \delta^2 R_{\delta 2}^2 w_n/4u^4 n \sim \delta^2 R_{\delta 2}^2 \gamma_n/4u^4\tau > 2\gamma_n$ provided we take $u < (\delta^2 R_{\delta 2}^2/8\tau)^{\frac{1}{4}}$. Thus $r \vee s = s_n$. If we take $u < \delta R_{\delta 2} K^{-\frac{1}{2}}$ then

$$L\tilde{g}(t) \leqq L(u^{-2} g(\gamma_n)) \leqq b^2/u^2 K \quad \text{for all } t \geqq s_n$$

and (2.2) follows. Also

$$t^{-\frac{1}{2}} Lg(t) \leqq s_n^{-\frac{1}{2}} Lg(s_n) \leqq (2u^2/\delta R_{\delta 2})(nw_n)^{\frac{1}{2}}$$

and

$$(nt)^{-\frac{1}{2}} L(nt) \leqq (n\gamma_n)^{-\frac{1}{2}} L(n\gamma_n) = o(1) \tag{7.26}$$

for all $t \geqq s_n$, so if $u^2 < \delta^2 R_{\delta 2}^2/4K$,

$$\begin{aligned}Kt^{-\frac{1}{2}} L(n\tilde{a}(t)) &\leqq Kt^{-\frac{1}{2}} L(nt) + Kt^{-\frac{1}{2}} Lg(t) \\ &\leqq (4Ku^2/\delta R_{\delta 2})(nw_n)^{\frac{1}{2}} \leqq \delta R_{\delta 2}(nw_n)^{\frac{1}{2}}\end{aligned} \tag{7.27}$$

and (2.3) follows. (2.4) and (2.8) follow from the fact that $n\gamma_n w_n \sim \tau w_n^2 \to \infty$. To establish (2.5), by the first inequality in (7.27) it suffices to show

$$2KL(nt) + 2KLg(t) \leqq \delta R_{\delta 2}(nw_n t)^{\frac{1}{2}} L(\delta^2 R_{\delta 2}^2 w_n/nu^4 t), \; t \geqq \gamma_n. \tag{7.28}$$

The first term on the left of (7.28) is handled using (7.26); for the second it suffices to consider $t = \gamma_n$ only, since $Lg(t)$ decreases and the right side of (7.28) increases in t. If we take u small enough so $L(\delta^2 R_{\delta 2}^2/2u^4\tau) \geqq 8K/\delta R_{\delta 2}\tau^{\frac{1}{2}}$, then

$$\delta R_{\delta 2}(nw_n\gamma_n)^{\frac{1}{2}} L(\delta^2 R_{\delta 2}^2 w_n/nu^4\gamma_n) \geqq \tfrac{1}{2}\delta R_{\delta 2}\tau^{\frac{1}{2}} w_n L(\delta^2 R_{\delta 2}^2/2u^4\tau) \geqq 4KLg(\gamma_n)$$

and (7.28), and then (2.5), follow. Since $g(t)$ and $t^{-\beta/2}Lt$ decrease and $a(t)$ increases, we have $t^{\beta/2}L\tilde{g}(t)=u^{-\beta}a(t)^{\beta/2}\,(u^{-2}g(t))^{-\beta/2}\,L(u^{-2}g(t))$ increasing, and (2.7) follows. Theorem 2.1 and (7.22) can now be applied, and the result is that

$$\begin{aligned}\mathbb{P}_n^* &\leqq 36\int_{\gamma_n/2}^{\alpha_n} t^{-1}\exp(-\delta^2 R_{\delta 2}^2 w_n/512u^2)\,dt\\ &\quad +68\exp(-2^{-9}\tau^{\frac{1}{2}}\delta R_{\delta 2} w_n L u^{-2})\\ &\quad +36\exp(-2^{-9}\tau^{\frac{1}{2}}\delta R_{\delta 2} w_n L(\delta R_{\delta 2}/2\tau^{\frac{1}{2}}u^2))\\ &\leqq 140\, L(2\gamma_n^{-1}\alpha_n)\exp(-(1+2\delta)(c_3+\delta)w_n)\\ &\leqq 140\exp(-(1+\delta)(c_3+\delta)w_n)\\ &\leqq 140\exp(-(1+\delta)\,\mathrm{LLn})\end{aligned}$$

provided $u\leqq u_0$ for some $u_0(\tau,\delta)>0$. In combination with (7.20) and (7.25) this proves (7.17), and (A) (ii) is proved.

Proof of (A) (iii). This time we take $u_n=(n\gamma_n/w_n)^{\mu}$ for some (large) $\mu>0$ to be specified later, and take $\alpha_n=n^{-1}w_n$ and

$$R_\delta=(\varrho+\delta)(c_1+\delta)+c_3+\delta.$$

Set $\tilde{w}_n=Lg(\alpha_n)\vee LLn$. Since $\alpha_n\geqq\gamma_n$, we have $n^{-1}\tilde{w}_n\leqq n^{-1}w_n=\alpha_n$. Since $\tilde{w}_n^{\frac{1}{2}}=o(y_n)$, it follows from the proofs of parts (A) (i) and (A) (ii) that

$$\begin{aligned}\mathbb{P}^*[|v_n(C)|>(1+8\delta)R_\delta y_n\sigma(C)\quad&\text{for some } C\in\mathscr{C}\ \text{ with }\ \sigma^2(C)\geqq\alpha_n]\\ &=O((Ln)^{-(1+\delta)}).\end{aligned}$$

Hence as in the proof of part (A) (ii) it suffices to bound the right side of (7.20) by $O((Ln)^{-(1+\delta)})$.

Analogously to (7.24), setting $\lambda_n=(1+4\delta)R_\delta y_n/(n\gamma_n)^{\frac{1}{2}}$, we get

$$\begin{aligned}\mathbb{P}_j &\leqq 2|\mathscr{F}(j)|\exp(-(1+4\delta)^2R_\delta^2 y_n^2\lambda_n^{-1}h_1(\lambda_n))\\ &\leqq 2|\mathscr{F}(j)|\exp(-(1+3\delta)R_\delta y_n(n\gamma_n)^{\frac{1}{2}}L(w_n/n\gamma_n))\\ &=2|\mathscr{F}(j)|\exp(-(1+3\delta)R_\delta w_n)\end{aligned}\tag{7.29}$$

since $h_1(\lambda_n)\sim L\lambda_n\sim L(w_n/n\gamma_n)$ by (6.2). Now $Lu_n^{-1}=o(w_n)$ and $N_n=O(Lu_n^{-2}+LL(\gamma_n^{-1}\alpha_n))=o(w_n)$ by (3.5), so as in (7.25), (7.29) and (7.18) give

$$\sum_{j=0}^{N_n}\mathbb{P}_j\leqq\exp(-(1+\delta)LLn).$$

Once again Theorem 2.1 will provide the needed bound on $\mathbb{P}_n^*$. As before we take $\zeta(t)=u_n^2t$, $b=\delta R_\delta y_n$, $\gamma=\gamma_n$, $\alpha=\alpha_n$, $q(t)=t^{\frac{1}{2}}$, and $\mathscr{C}(t)$ as in (7.19), so $\tilde{a}(t)\leqq a(t)$, $\tilde{g}(t)=u_n^{-2}g(t)$, $z(t)=u_n^{-2}t^{-\frac{1}{2}}$, r is γ_n, and s is $s_n=\delta^2R_\delta^2y_n^2/4u_n^4>\alpha_n$, so (2.2) and (2.3) are vacuous. (2.4) and (2.8) follow from (3.5). As in the proof of (A) (ii), to establish (2.5) it suffices to show

$$2KL(nt)+2KLg(t)\leqq\delta R_\delta(nt)^{\frac{1}{2}}y_nL(w_n/nu_n^4t),\ \alpha_n\geqq t\geqq\gamma_n.\tag{7.30}$$

The first term in (7.30) is handled by noting that, since $\gamma_n \leqq t \leqq \alpha_n = n^{-1} w_n$,

$$\begin{aligned}(nt)^{-\frac{1}{2}} L(nt) &\leqq (n\gamma_n)^{-\frac{1}{2}} L(n\gamma_n) = y_n w_n^{-1} L(w_n/n\gamma_n) L(n\gamma_n)\\ &\leqq y_n w_n^{-1} (4\mu)^{-1} L(u_n^{-4}) L w_n \leqq (4\mu)^{-1} y_n L(w_n/n u_n^4 t).\end{aligned}$$

For the second term,

$$\begin{aligned}(nt)^{-\frac{1}{2}} Lg(t) &\leqq (n\gamma_n)^{-\frac{1}{2}} Lg(\gamma_n) \leqq y_n L(w_n/n\gamma_n)\\ &\leqq (4\mu)^{-1} y_n L(w_n/n u_n^4 t).\end{aligned}$$

Thus (7.30) holds if μ is large enough. (2.7) is established as in the proof of (A) (ii). We now apply Theorem 2.1 and use the fact that by (3.5), $LL(2\gamma_n^{-1}\alpha_n) = LL(2\gamma_n^{-1} n^{-1} w_n) = o(w_n)$, to obtain, if μ is large enough,

$$\begin{aligned}\mathbb{P}_n^* &\leqq 36 \int_{\gamma_n/2}^{\alpha_n} t^{-1} \exp(-\delta^2 R_\delta^2 y_n^2/512 u_n^2)\, dt\\ &\quad + 68 \exp(-2^{-8} \delta R_\delta y_n (n\gamma_n)^{\frac{1}{2}} L(u_n^{-2}))\\ &\quad + 36 \exp(-2^{-8} \delta R_\delta y_n (n\gamma_n)^{\frac{1}{2}} L(\delta R_\delta y_n/(n\gamma_n)^{\frac{1}{2}} u_n^2))\\ &\leqq 140\, L(2\gamma_n^{-1}\alpha_n) \exp(-2^{-7} \delta R_\delta \mu y_n (n\gamma_n)^{\frac{1}{2}} L(w_n/n\gamma_n))\\ &\leqq 140 \exp(-(1+\delta)(c_3+\delta) w_n)\\ &\leqq 140 \exp(-(1+\delta) LLn).\end{aligned}$$

The result now follows as in the proof of (A) (ii). □

We introduce now some notation and do preliminary calculations for use in the proofs of the next three propositions. Let $R > 0$ and $\delta, \lambda, \mu \in (0,1)$ be constants to be specified later. Let $\mathscr{C}$ be a full VC class. For each $t \in (0, \frac{1}{4}]$ let

$$\mathscr{D}_t \subset \mathscr{C} \quad \text{with} \quad \varepsilon_\lambda g(t)^{1-\lambda} \leqq |\mathscr{D}_t| \leqq \varepsilon_\lambda g(t)^{1-\lambda} + 1, \quad \text{and}$$

$$\sigma^2(C) = t,\ P(C) \leqq \tfrac{1}{2}, \quad \text{and} \quad P\Big(C \cap \Big(\bigcup_{D \in \mathscr{D}_t,\, D \neq C} D\Big)\Big) \leqq \lambda P(C) \quad \text{for all } C \in \mathscr{D}_t,$$

where $\varepsilon_\lambda \leqq \frac{1}{8}$ is the constant in the definition of "full". Let (γ_n), (α_n), and (b_n) be nonnegative sequences with $\gamma_n \leqq \alpha_n \leqq \frac{1}{4}$ and

$$n^{\frac{1}{2}} b_n q(\gamma_n) \to \infty. \tag{7.31}$$

Let $(n(k), k \geqq 0)$ be a strictly increasing sequence of integers with $n(0) = 0$, and set

$$\begin{aligned}m(k) &= n(k) - n(k-1)\\ Y_k(C) &= \sum_{i=n(k-1)+1}^{n(k)} 1_C(X_i)\\ S_k(C) &= Y_k(C) - m(k) P(C) = n(k)^{\frac{1}{2}} v_{n(k)}(C) - n(k-1)^{\frac{1}{2}} v_{n(k-1)}(C)\\ t_{kj} &= \alpha_{n(k)} \mu^j,\\ N_k &= \min\{j \geqq 0 : t_{k,j+1} < \gamma_{n(k)}\},\\ N_k' &= \min\{j \geqq 0 : t_{kj} \leqq \gamma_{n(k)}^{1-\delta/16}\},\\ I_k &= \{(j,i) : N_k' \leqq j \leqq N_k,\ 1 \leqq i \leqq |\mathscr{D}_{t_{kj}}|\},\\ r(k,j) &= |\mathscr{D}_{t_{kj}}|,\end{aligned}$$

and observe that

$$\mathscr{D}_{t_{kj}} = \{C_{kji} : 1 \leqq i \leqq r(k,j)\}$$

for some sets C_{kji}. Note the k indexes the number of sample points, j indexes the sizes of the sets, and i indexes the collection of sets corresponding to each k and j. The C_{kji} are nearly disjoint for fixed k and j; we wish to replace them with fully disjoint sets D_{kji}. Define

$$\begin{aligned}
G'_{kji} &= C_{kji} \cap \left(\bigcup_{m \neq i} C_{kjm} \right), \\
D'_{kji} &= C_{kji} \setminus G'_{kji}, \\
G''_{kji} &= D'_{kji} \cap \left(\bigcup_{l>j} \bigcup_{m \leqq r(k,l)} C_{klm} \right), \\
G_{kji} &= G'_{kji} \cup G''_{kji}, \\
D_{kji} &= C_{kji} \setminus G_{kji}, \quad \text{and} \\
H_{kj} &= \bigcup_{i \leqq r(k,j)} D_{kji}.
\end{aligned}$$

Thus for fixed k, D_{kji} is obtained from C_{kji} by throwing out any intersection G_{kji} with other sets of equal or smaller size. Since $\mathscr{C}$ is full, $P(D'_{kji}) \geqq (1-\lambda) P(C_{kji})$. $\{D_{kji} : j \geqq 0, i \leqq r(k,j)\}$ and $\{D'_{kji} : i \leqq r(k,j)\}$ are each disjoint collections, so

$$P\left(\bigcup_i D'_{kji} \right) \geqq r(k,j)\,(1-\lambda)\, t_{kj} \geqq \varepsilon_\lambda (1-\lambda)\, a(t_{kj})^{1-\lambda} t_{kj}^\lambda$$

while

$$\begin{aligned}
P\left(\bigcup_i G''_{kji} \right) &\leqq \sum_{l>j} 2 t_{kl}\, r(k,l) \\
&\leqq \sum_{l>j} (2\varepsilon_\lambda a(t_{kl})^{1-\lambda} t_{kl}^\lambda + 2 t_{kl}) \\
&\leqq \sum_{l>j} (2\varepsilon_\lambda a(t_{kj})^{1-\lambda} t_{kj}^\lambda \mu^{\lambda(l-j)} + 2 t_{kj} \mu^{\lambda(l-j)}) \\
&\leqq 2(\varepsilon_\lambda + 1)\, \mu^\lambda (1-\mu^\lambda)^{-1}\, a(t_{kj})^{1-\lambda} t_{kj}^\lambda .
\end{aligned}$$

If, as we henceforth assume, μ is chosen small enouth so $2/(\varepsilon_\lambda + 1)\, \mu^\lambda (1-\mu^\lambda)^{-1} \leqq \lambda \varepsilon_\lambda (1-\lambda)/2$, it follows that $P\left(\bigcup_i G''_{kji} \right) \leqq \lambda P\left(\bigcup_i D''_{kji} \right)/2$. Hence for fixed k,j, for at least half the values of i we have $P(G''_{kji}) \leqq \lambda P(D'_{kji})$. By reducing ε_λ and $|\mathscr{D}_{t_{kj}}|$ by half if necessary, we may assume this is valid for all i; it then follows from $\mathscr{C}$ being full that

$$\begin{aligned}
&P(D_{kji}) \geqq (1-\lambda)\, P(C_{kji}) - \lambda P(D'_{kji}) \geqq (1-2\lambda)\, P(C_{kji}), \quad \text{so} \\
&\sigma^2(D_{kji}) \geqq (1-2\lambda)\, \sigma^2(C_{kji}), \quad \text{and} \quad P(G_{kji}) \leqq 2\lambda P(C_{kji}).
\end{aligned} \tag{7.32}$$

Observe also that

$$\begin{aligned}
&C_{kji} = D_{kji} \cup G_{kji} \quad \text{as a disjoint union, and} \\
&G_{kji} \cap \left(\bigcup_{l \leqq j} \bigcup_{m \leqq r(k,j)} D_{klm} \right) = \phi .
\end{aligned} \tag{7.33}$$

We now define events

$$A_{kji} = [S_k(C_{kji}) \geqq (1-2\delta)\, Rn(k)^{\frac{1}{2}}\, b_{n(k)}\, q(\sigma^2(C_{kji}))]$$
$$A'_{kji} = [S_k(D_{kji}) \geqq (1-\delta)\, Rn(k)^{\frac{1}{2}}\, b_{n(k)}\, q(\sigma^2(C_{kji}))]$$
$$A''_{kji} = [S_k(G_{kji}) \geqq -\delta Rn(k)^{\frac{1}{2}}\, b_{n(k)}\, q(\sigma^2(C_{kji}))]$$
$$E_{kji} = [\nu_{n(k-1)}(C_{kji}) \geqq -\delta R(n(k)/n(k-1))^{\frac{1}{2}}\, b_{n(k)}\, q(\sigma^2(C_{kji}))]$$
$$F_k = \bigcup_{(j,i)\in I_k} E^c_{kji}$$
$$B_k = \bigcup_{(j,i)\in I_k} A_{kji}$$
$$B'_k = \bigcup_{(j,i)\in I_k} A'_{kji}.$$

Note that A'_{kji} and A''_{kji} together imply A_{kji}, and that A_{kji} and E_{kji} together imply that $\nu_{n(k)}(C_{kji}) \geqq (1-3\delta)\, Rb_{n(k)}\, q(\sigma^2(C_{kji}))$. Thus

$$\limsup_n \sup\{\nu_n(C)/b_{n(k)}\, q(\sigma^2(C)) : C\in\mathscr{C},\ \gamma_n \leqq \sigma^2(C) \leqq \alpha_n\} \geqq (1-3\delta)R \quad \text{a.s.} \tag{7.34}$$

provided

$$\mathbb{P}(B_k \text{ i.o.}) = 1, \qquad \mathbb{P}(F_k \text{ i.o.}) = 0. \tag{7.35}$$

Since the S_k are independent, (7.35) will follow if

$$\sum \mathbb{P}(B_k) = \infty, \qquad \sum \mathbb{P}(F_k) < \infty, \tag{7.36}$$

so we wish to bound $\mathbb{P}(B_k)$ from below. Define events

$$U_{kj} = [Y_k(H_{kj}) \leqq z_{kj}]$$

where z_{kj} is given by

$$16\lambda\, t_{kj}(z_{kj} - m(k)\, P(H_{kj})) = \delta Rn(k)^{\frac{1}{2}}\, b_{n(k)}\, q(t_{kj})/2. \tag{7.37}$$

Fix k and define stopping times for $\omega \in B'_k$ by

$$T_1(\omega) = \min\{j : \omega \in A'_{kji} \quad \text{for some } i\},$$
$$T_2(\omega) = \min\{i : \omega \in A'_{kT_1(\omega)i}\},$$

and let $T_1 = T_2 = \infty$ off B'_k. If S_k is large on D_{kji}, it is probably also large on C_{kji}, because D_{kji} is most of C_{kji} by (7.32). To make this precise, we will show that

$$\mathbb{P}(B_k \mid B'_k, U_{kj}, (T_1, T_2) = (j,i)) \geqq \mathbb{P}(A''_{kji} \mid B'_k, U_{kj}, (T_1, T_2) = (j,i)) \geqq \tfrac{1}{2}. \tag{7.38}$$

Once (7.38) is established, we have

$$2\,\mathbb{P}(B_k, T_1 = j) \geqq 2\,\mathbb{P}(B_k, U_{kj}, T_1 - j) \geqq \mathbb{P}(B'_k, U_{kj}, T_1 = j)$$
$$\geqq \mathbb{P}(B'_k, T_1 = j) - \mathbb{P}(U^c_{kj})$$

so that

$$2\,\mathbb{P}(B_k) \geqq \mathbb{P}(B'_k) - \sum_{j=N'_k}^{N_k} \mathbb{P}(U^c_{kj}). \tag{7.39}$$

The first inequality in (7.38) follows directly from the definitions. To prove the second, observe that by (7.33), $S_k(G_{kji})$ is conditionally independent of $(S_k(D_{kjm}),$

$m \leqq r(k,j)$) given $S_k(H_{kj})$ (or equivalently, given $Y_k(H_{kj})$). It follows that

$$\begin{aligned}
&\mathbb{P}(A''_{kji} \mid B'_k, U_{kj}, (T_1, T_2) = (j, i)) \\
&\quad = \sum_{l \leqq z_{kj}} \mathbb{P}(A''_{kji} \mid T_1 \geqq j, Y_k(H_{kj}) = l)\; \mathbb{P}(Y_k(H_{kj}) = l \mid B'_k, U_{kj}, (T_1, T_2) = (j, i)) \\
&\quad \geqq \sum_{l \leqq z_{kj}} \mathbb{P}(A''_{kji} \mid Y_k(H_{kj}) = l)\; \mathbb{P}(Y_k(H_{kj}) = l \mid B'_k, U_{kj}, (T_1, T_2) = (j, i)) \\
&\quad \geqq \min_{l \leqq z_{kj}} \mathbb{P}(A''_{kji} \mid Y_k(H_{kj}) = l).
\end{aligned} \tag{7.40}$$

Fix i, j, k, and $l, l \leqq z_{kj}$. Given $Y_k(H_{kj}) = l$, $Y_k(G_{kji})$ has a binomial distribution with parameters (N, p), where $N = m(k) - l$ and $p = P(G_{kji})/P(H^c_{kj})$. Since the median of a binomial distribution is within one of the mean (Uhlmann 1966; Jogdeo and Samuels 1968), it follows that

$$\mathbb{P}[Y_k(G_{kji}) \geqq Np - 1 \mid Y_k(H_{kj}) = l] \geqq \tfrac{1}{2}.$$

Thus (7.38) will follow from (7.40) once we establish that

$$Np - 1 \geqq m(k)\; P(G_{kji}) - \delta R n(k)^{\frac{1}{2}}\, b_{n(k)}\, q(\sigma^2(C_{kji})) \tag{7.41}$$

whenever the right side of (7.41) is positive. (Note that $(A''_{kji})^c = \phi$ if it is not positive.) Now recalling we took $\varepsilon_\lambda \leqq \frac{1}{8}$, we get $P(H_{kj}) \leqq r(k,j)\;\max_i P(D_{kji})$ $\leqq 2(\varepsilon_\lambda g(t_{kj}) + 1)\, t_{kj} \leqq 2\varepsilon_\lambda + 2\alpha_{n(k)} \leqq \frac{3}{4}$, so by (7.33), 7.37), and (7.31), on the event $[Y_k(H_{kj}) = l]$,

$$\begin{aligned}
m(k) P(G_{kji}) - Np &= P(H^c_{kj})^{-1}\; P(G_{kji}) S_k(H_{kj}) \\
&\leqq 16\lambda\sigma^2(C_{kji})\,(z_{kj} - m(k)\; P(H_{kj})) \\
&\leqq \delta R n(k)^{\frac{1}{2}}\, b_{n(k)}\, q(t_{kj})/2 \\
&\leqq -1 + \delta R n(k)^{\frac{1}{2}}\, b_{n(k)}\, q(\sigma^2(C_{kji})).
\end{aligned}$$

(7.41), then (7.38) and (7.39), now follow.

It is clear that for any m and M and any collection $J_o, \ldots, J_m$ of disjoint sets, $\mathbb{P}[Y_k(J_o) \leqq M \mid Y_k(J_i) = l_i$ for all $1 \leqq i \leqq m]$ is monotone increasing in each l_i. From this "negative dependence" of Y_k on disjoint sets, it follows that

$$\mathbb{P}((B'_k)^c) \leqq \prod_{(j,i) \in I_k} \mathbb{P}((A'_{kji})^c) \leqq \prod_{j=N'_k}^{N_k} (1 - \min_{i \leqq r(k,j)} \mathbb{P}(A'_{kji}))^{r(k,j)}. \tag{7.42}$$

For the remainder of the proof it is necessary to split into cases. We state each of these as a separate proposition. In each proof we will specify $(n\,k))$, R, δ, and λ, then continue the present calculation.

Proposition 7.9. *Let $\mathscr{C}$ be a full VC class and $q \in Q$, and suppose*

$$q(t)/t^{\frac{1}{2}} \downarrow,\; q(t)/(t\,Lt^{-1})^{\frac{1}{2}} \uparrow,\; Lg(t)/Lt^{-1} \uparrow. \tag{7.43}$$

Let (γ_n), (α_n), (b_n) be nonnegative sequences satisfying

$$\gamma_n \leqq \alpha_n$$

and

$$b_n q(\gamma_n)/n^{\frac{1}{2}}\gamma_n \to 0. \tag{7.44}$$

Suppose the following limits exist and are finite:

$$c_1 = \lim_n \gamma_n Lg(\gamma_n)/b_n^2 q^2(\gamma_n)$$

$$c_2 = \lim_n \gamma_n LL(\gamma_n^{-1}\alpha_n)/b_n^2 q^2(\gamma_n)$$

$$c_3 = \lim_n \gamma_n LLn/b_n^2 q^2(\gamma_n). \tag{7.45}$$

Then

$$\limsup_n \sup\{|v_n(C)|/b_n\, q(\sigma^2(C)): C\in\mathscr{C},\ \gamma_n \leqq \sigma^2(C) \leqq \alpha_n\} \geqq (2(c_1+c_2+c_3))^{\frac{1}{2}} \quad a.s. \quad \square \tag{7.46}$$

Proof. Let $0<\delta<1$ and $V=16\delta^{-2}$; we continue our calculation taking

$$R=(2(c_1+c_2+c_3))^{\frac{1}{2}},\ \lambda=\delta/16, \quad \text{and} \quad n(k)=[V^k].$$

We may assume $R>0$. By (7.44), Lemma 7.5, (7.32), and (7.43),

$$\begin{aligned}\mathbb{P}(A'_{kji}) &\geqq \mathbb{P}[v_{m(k)}(D_{kji}) \geqq (1-\delta)\,(1+2V^{-1})^{\frac{1}{2}}\, Rb_{n(k)}\, q(\sigma^2(C_{kji}))]\\ &\geqq \exp(-(1-\delta/2)R^2 b_{n(k)}^2 q^2(\sigma^2(C_{kji}))/2\,(1-2\lambda)\sigma^2(C_{kji}))\\ &\geqq \exp(-(1-\delta/4)\,R^2 b_{n(k)}^2 q^2(\gamma_{n(k)})/2\gamma_{n(k)})\\ &\equiv p_k.\end{aligned} \tag{7.47}$$

Set

$$u(k)=\varepsilon_\lambda g(\gamma_{n(k)})^{1-\delta/8},\ N(k)=N_k-N'_k+1.$$

Since $tg(t)$ and $Lg(t)/Lt^{-1}$ increase, for $j\geqq N'_k$ we have

$$r(k,j)\geqq \varepsilon_\lambda g(t_{kj})^{1-\lambda} \geqq \varepsilon_\lambda g(\gamma_{n(k)}^{1-\delta/16})^{1-\delta/16} \geqq u(k).$$

Hence by (7.42) and (7.47),

$$\mathbb{P}((B'_k)^c) \leqq (1-p_k)^{u(k)N(k)} \leqq (1-N(k)\,u(k)p_k/2) \vee \tfrac{1}{2}$$

so

$$\mathbb{P}(B'_k) \geqq (N(k)\,u(k)p_k/2) \wedge (\tfrac{1}{2}). \tag{7.48}$$

If $N'_k>0$ then

$$\gamma_{n(k)}^{1-\delta/16} < t_{k,N'_k-1} = \alpha_{n(k)}\mu^{N'_k-1}$$

so

$$N'_k \leqq 1+(\log\mu^{-1})^{-1}\,(1-\delta/16)\log(\gamma_{n(k)}^{-1}\alpha_{n(k)})$$

while similarly

$$N_k > -1+(\log\mu^{-1})^{-1}\log(\gamma_{n(k)}^{-1}\alpha_{n(k)}).$$

Hence

$$N(k) \geqq (-1 + (\delta/16 \log \mu^{-1}) \log(\gamma_{n(k)}^{-1} \alpha_{n(k)})) \vee 1 .$$

Since $\gamma_{n(k)}^{-1} \alpha_{n(k)} \to \infty$ if $c_2 > 0$, it follows that

$$LN(k) \geqq c_2 (1 - \delta/4)\, b_{n(k)}^2 q^2(\gamma_{n(k)})/\gamma_{n(k)}. \tag{7.49}$$

Similarly, since $c_2 \leqq R^2/2 < (1 - \delta/4) R^2$,

$$LN(k) \leqq (1 - \delta/4)\, R^2 b_{n(k)}^2 q^2(\gamma_{n(k)})/\gamma_{n(k)} = \log p_k^{-2} . \tag{7.50}$$

For $j \geqq N_k'$ we have $t_{kj} \leqq \gamma_{n(k)}^{1-\delta/16} \leqq \gamma_{n(k)}^{\frac{1}{2}}/2$. Using this and Theorem 1 of Hoeffding (1963) we get for $N_k' \leqq j \leqq N_k$:

$$\begin{aligned} \mathbb{P}(U_{kj}^c) &\leqq \mathbb{P}[v_{m(k)}(H_{kj}) \geqq R b_{n(k)} q(t_{kj})/2\, t_{kj}] \\ &\leqq \exp(-R^2 b_{n(k)}^2 q^2(t_{kj})/2\, t_{kj}^2) \\ &\leqq \exp(-2R^2 b_{n(k)}^2 q^2(\gamma_{n(k)})/\gamma_{n(k)}) \\ &\leqq p_k^4 . \end{aligned} \tag{7.51}$$

Hence by (7.50),

$$\sum_{j=N_k'}^{N_k} \mathbb{P}(U_{kj}^c) \leqq N(k)\, p_k^4 \leqq p_k^2 \leqq (N(k)\, u(k) p_k/4) \wedge (\tfrac{1}{4}) .$$

Combining this with (7.48) and (7.39) we see that

$$8\, \mathbb{P}(B_k) \geqq N(k)\, u(k) p_k \wedge 1 . \tag{7.52}$$

Using (7.49) we obtain

$$\begin{aligned} N(k) u(k) p_k &= \varepsilon_\lambda N(k) g(\gamma_{n(k)})^{1-\delta/8} \exp(-(1-\delta/4)\, R^2 b_{n(k)}^2 q^2(\gamma_{n(k)})/2\gamma_{n(k)}) \\ &\geqq \varepsilon_\lambda \exp(-(1-\delta/4)\, c_3 b_{n(k)}^2 q^2(\gamma_{n(k)})/\gamma_{n(k)}) \\ &\geqq \varepsilon_\lambda \exp(-(1-\delta/8)\, LLn(k)) \\ &\geqq \varepsilon_\lambda k^{-(1-\delta/16)}. \end{aligned} \tag{7.53}$$

With (7.52) this shows $\sum \mathbb{P}(B_k) = \infty$.

To establish (7.36) it remains to bound $\mathbb{P}(F_k)$. By (7.43), for $\gamma_{n(k)} \leqq t \leqq \varepsilon_k \equiv \gamma_{n(k)}^{1-\delta/16}$,

$$\begin{aligned} \frac{q^2(t)}{tLg(t)} &\geqq \frac{q^2(\gamma_{n(k)})}{\gamma_{n(k)} L\gamma_{n(k)}^{-1}} \frac{L\varepsilon_k^{-1}}{Lg(\varepsilon_k)} \geqq \frac{q^2(\gamma_{n(k)})}{\gamma_{n(k)} Lg(\gamma_{n(k)})} \frac{L\varepsilon_k^{-1}}{L\gamma_{n(k)}^{-1}} \\ &\geqq (c_1 + R^2/2)^{-1} (1-\delta/16)\, b_{n(k)}^{-2} \geqq (2c_1 + R^2)^{-1}\, b_{n(k)}^{-2} \end{aligned}$$

and

$$\frac{q^2(t)}{t} \geqq \frac{q^2(\gamma_{n(k)})}{\gamma_{n(k)}} \frac{L\varepsilon_k^{-1}}{L\gamma_{n(k)}^{-1}} \geqq \frac{q^2(\gamma_{n(k)})}{2\gamma_{n(k)}} .$$

Also

$$N(k) \leqq N_k + 1 \leqq (\log \mu^{-1})^{-1} \log(\gamma_{n(k)}^{-1} \alpha_{n(k)}) + 1 .$$

Combining these facts with (6.5), (6.4), (7.44), and (7.43) we obtain

$$\begin{aligned}\mathbb{P}(F_k) &\leqq \sum_{(j,i)\in I_k} 2\exp(-\delta^2 VR^2 b_{n(k)}^2 q^2(\sigma^2(C_{kji}))/4\sigma^2(C_{kji}))\\ &\leqq \sum_{j=N_k'}^{N_k} 4g(t_{kj})\exp(-4R^2 b_{n(k)}^2 q^2(t_{kj})/t_{kj})\\ &\leqq \sum_{j=N_k'}^{N_k} 4\exp(-(R^2+2c_2+2c_3)b_{n(k)}^2 q^2(t_{kj})/t_{kj})\\ &\leqq 4N(k)\exp(-(R^2/2+c_2+c_3)b_{n(k)}^2 q^2(\gamma_{n(k)})/\gamma_{n(k)})\\ &\leqq 4\exp(-(R^2/4+c_3)b_{n(k)}^2 q^2(\gamma_{n(k)})/\gamma_{n(k)})\\ &\leqq 4\exp(-(\tfrac{5}{4})LLn(k))\end{aligned}$$

and (7.36), and then (7.34), follow. Since δ may be arbitrarily small, this proves the proposition. □

Remark 7.10. From the above proof it is apparent that the assumption that the limits in (7.45) exist is stronger than needed. In fact we have shown that if the lim sup of each sequence in (7.45) is finite, then the lim sup in (7.46) is at least $(2c_4)^{\frac{1}{2}}$, where

$$c_4 = \liminf_n \gamma_n(Lg(\gamma_n) + LL(\gamma_n^{-1}\alpha_n) + LLn)/b_n^2 q^2(\gamma_n).$$

Similar considerations apply in the next two propositions; it follows (see the proof of Theorem 3.1 (B) below) that the limits of the sequences in (3.2) need not exist, and ϱ need not be at most one, for us to obtain some lower bound on the R in (3.4). As long as the lim sups c_i exist, the corresponding lim infs, say c_i', provide the lower bound $(2(c_1'+c_2'+c_3'))$ for R. □

Proposition 7.11. *Let $\mathscr{C}$ be a full VC class. Let γ_n, w_n, c_1, and c_3 be as in Theorem 3.1 and $\theta=(c_1+c_3)^{-1}$, and suppose $\gamma_n \sim \tau n^{-1} w_n$ for some $\tau>0$. Suppose the* lim sups *in* (3.2) *are actually limits. Then*

$$\begin{aligned}\limsup_n \sup\{|v_n(C)|/w_n^{\frac{1}{2}}\sigma(C): C\in\mathscr{C},\ \sigma^2(C)=\gamma_n\}\\ \geqq \tau^{\frac{1}{2}}(\beta_{\theta\tau}-1) \quad a.s.\end{aligned} \tag{7.54}$$

Proof. Let $0<\delta<1$ and take $n(k)=[V^k]$, $b_n=w_n^{\frac{1}{2}}$, $\alpha_n=\gamma_n$, $R=\tau^{\frac{1}{2}}(\beta_{\theta\tau}-1)$, $\lambda=\delta/16$, $q(t)=t^{\frac{1}{2}}$ where $V\geqq 8\delta^{-1}$ is large enough so

$$\delta R\tau^{\frac{1}{2}} h_1(\delta VR/2\tau^{\frac{1}{2}}) \geqq 6(c_1+c_3). \tag{7.55}$$

We may assume $R>0$, so $c_1+c_3=\theta^{-1}>0$. Note that $N_k=N_k'=0$, so j is always 0. By Lemma 7.4 and (1.11), similarly to (7.47),

$$\begin{aligned}\mathbb{P}(A_{kji}') &\geqq \mathbb{P}[v_{m(k)}(D_{kji}) \geqq (1-\delta/2)R(w_{n(k)}\gamma_{n(k)})^{\frac{1}{2}}]\\ &\geqq \exp(-(1-\delta/4)w_{n(k)}R\tau^{\frac{1}{2}}h_1(R\tau^{-\frac{1}{2}}))\\ &= \exp(-(1-\delta/4)(c_1+c_3)w_{n(k)})\\ &\equiv p_k.\end{aligned}$$

As in Proposition 7.9 (cf. (7.48)) it follows that

$$\mathbb{P}(B_k') \geqq (u(k)\, p_k/2) \wedge (\tfrac{1}{2})$$

for $u(k) = \varepsilon_\lambda g(\gamma_{n(k)})^{1-\lambda}$. Analogously to (7.51)–(7.53), we obtain $\mathbb{P}(U_{kj}^c) \leqq p_k^4$ and then

$$8\,\mathbb{P}(B_k) \geqq u(k)\, p_k \wedge 1 \geqq \varepsilon_\lambda\, k^{-(1-\delta/8)}$$

so $\sum \mathbb{P}(B_k) = \infty$. By (6.5) and (7.55),

$$\begin{aligned}\mathbb{P}(F_k) &\leqq 2r(k,0)\, \exp(-\delta R(n(k)\, w_{n(k)}\, \gamma_{n(k)})^{\frac{1}{2}}\, h_1(\delta V R/2\tau^{\frac{1}{2}}))\\ &\leqq 4g(\gamma_{n(k)})\, \exp(-5(c_1+c_3)\, w_{n(k)})\\ &\geqq 4\, \exp(-3(c_1+c_3)\, w_{n(k)})\\ &\leqq 4\, \exp(-2\, LLn(k))\end{aligned}$$

and, as in Proposition 7.9, the desired result follows. □

Proposition 7.12. *Let $\mathscr{C}$ be a full VC class. Let γ_n, w_n, y_n, c_1, and c_3 be as in Theorem 3.1, and suppose*

$$\gamma_n = o(n^{-1} w_n) \quad \textit{and} \tag{7.56}$$

$$L(w_n/n\gamma_n) = o(w_n). \tag{7.57}$$

Then

$$\limsup_n \sup\{|\nu_n(C)|/y_n \sigma(C) : C \in \mathscr{C},\ \sigma^2(C) = \gamma_n\} \geqq c_1 + c_3 \quad \textit{a.s.} \quad \square$$

Proof. Let $0 < \delta < 1$ and this time continue the calculation preceding (7.42) taking $b_n = y_n$, $\alpha_n = \gamma_n$, $R = c_1 + c_3$, $\lambda = \delta/16$, and $q(t) = t^{\frac{1}{2}}$. If $c_3 > 0$, take $n(k) = \exp(kLk)$. If $c_3 = 0$, then since $w_{n(k)} \leqq L\gamma_{n(k)}^{-1} LLn(k) \leqq 2Ln(k)$, by (7.56) and (7.57) we can inductively take $n(k)$ large enough so

$$L(n(k)/n(k-1)) \geqq w_{n(k)}/4 \geqq 3\delta^{-2}\, L(w_{n(k)}/n(k)\,\gamma_{n(k)})$$

and

$$w_{n(k)} \geqq 4Lk. \tag{7.58}$$

Again we may assume $R > 0$, and j is always 0. By Lemma 7.4, (6.2), and (7.56),

$$\begin{aligned}\mathbb{P}(A_{kji}') &\geqq \mathbb{P}[\nu_{m(k)}(D_{kji}) \geqq (1-\delta/2)\, R y_{n(k)} (\gamma_{n(k)}\, n(k)/m(k))^{\frac{1}{2}}]\\ &\geqq \exp(-(1-\delta/2)\, R y_{n(k)}\, (n(k)\,\gamma_{n(k)})^{\frac{1}{2}}\, L(R y_{n(k)}/(n(k)\,\gamma_{n(k)})^{\frac{1}{2}}))\\ &\geqq \exp(-(1-\delta/4)\, (c_1+c_3)\, w_{n(k)})\\ &\equiv p_k.\end{aligned}$$

As in Proposition 7.11 it follows from this that

$$8\,\mathbb{P}(B_k) \geqq u(k) p_k \wedge 1$$

where $u(k) = \varepsilon_\lambda g(\gamma_{n(k)})^{1-\lambda}$. If $c_3 = 0$ then $u(k) p_k \geqq 1$ so $\sum \mathbb{P}(B_k) = \infty$. If $c_3 > 0$ then

$$\begin{aligned}u(k) p_k &\geqq \varepsilon_\lambda \exp(-(1-\delta/4)\, c_3\, w_{n(k)}) \geqq \exp(-(1-\delta/8)\, LLn(k))\\ &\geqq \exp(-(1-\delta/16)\, Lk)\end{aligned}$$

and again $\sum \mathbb{P}(B_k) = \infty$.

By (6.5), (6.2), and (7.56),

$$\begin{aligned}\mathbb{P}(F_k) &\leqq 2r(k,0)\exp(-\delta R y_{n(k)}(n(k)\gamma_{n(k)})^{\frac{1}{2}} h_1(\delta R y_{n(k)} n(k)^{\frac{1}{2}}/n(k-1)\gamma_{n(k)}^{\frac{1}{2}})) \\ &\leqq 4g(\gamma_{n(k)})\exp(-\delta^2 R y_{n(k)}(n(k)\gamma_{n(k)})^{\frac{1}{2}} L(n(k)/n(k-1))).\end{aligned} \tag{7.59}$$

If $c_3 > 0$ then $c_3 w_{n(k)} \sim LLn(k)$ so using (7.57),

$$\begin{aligned}L(n(k)/n(k-1)) &\geqq Lk \geqq (LLn(k))/2 \geqq c_3 w_{n(k)}/4 \\ &\geqq 4\delta^{-2} L(w_{n(k)}/n(k)\gamma_{n(k)})\end{aligned}$$

Then by (7.59),

$$\begin{aligned}\mathbb{P}(F_k) &\leqq 4g(\gamma_{n(k)})\exp(-4(c_1+c_3)w_{n(k)}) \\ &\leqq 4\exp(-3c_3 w_{n(k)}) \leqq 4\exp(-2LLn(k)) \leqq 4k^{-2}\end{aligned}$$

so $\sum \mathbb{P}(F_k) < \infty$. If $c_3 = 0$ then by (7.58) and (7.59),

$$\begin{aligned}\mathbb{P}(F_k) &\leqq 4g(\gamma_{n(k)})\exp(-3Rw_{n(k)}) \\ &\leqq 4\exp(-w_{n(k)}) \leqq 4k^{-2}\end{aligned}$$

so once more $\sum \mathbb{P}(F_k) < \infty$. As in the previous two propositions, the desired result now follows. □

Proof of Theorem 3.1 (B). For (i) and (iii) the result is immediate from Propositions 7.9 and 7.12 respectively. For (ii) we have $R \geqq \tau^{\frac{1}{2}}(\beta_{\theta\tau} - 1)$ by Proposition 7.11; to get $R \geqq (2(c_1 + c_2 + c_3))^{\frac{1}{2}}$ we can take a sequence (α_n) as in the proof of Theorem 3.1 (A) (ii). That is, $\gamma_n \leqq \alpha_n$, and the c_i in (3.2) are unchanged but (i) applies, if γ_n is replaced by α_n. □

Proof of Theorem 3.1 (C). We return now to the notation of the calculation preceding Propositions 7.9, 7.11, and 7.12, but with the following changes: now $S_k \equiv k^{\frac{1}{2}} v_k$ and $D_{kji} \equiv C_{kji}$.

We specify

$$\alpha_n \equiv \gamma_n,\ n(k) \equiv k,\ \lambda = \delta/8$$

so that $N_k' = N_k = 0$. Observe that c_1 must be 1 and c_2 must be 0. Since $\mathscr{C}$ is spatially full, we take the C_{k0i} to be disjoint for distinct i and fixed k. To prove the theorem it suffices to show that, whatever δ may be, $\mathbb{P}(B_k^c \text{ i.o.}) = 0$. As in (7.42), since the C_{k0i} are disjoint,

$$\mathbb{P}(B_k^c) \leqq (1 - \min_{i \leqq r(k,0)} \mathbb{P}(A_{k0i}))^{r(k,0)} \tag{7.60}$$

while as in the proofs of Propositions 7.9, 7.11, and 7.12,

$$\begin{aligned}\min_{i \leqq r(k,0)} \mathbb{P}(A_{k0i}) &\geqq \exp(-(1-\delta/4)w_k) \\ &= \exp(-(1-\delta/4)Lg(\gamma_k)) \equiv p_k.\end{aligned}$$

Since $c_3 = 0$,

$$\begin{aligned}r(k,0)p_k &\geqq \varepsilon_\lambda g(\gamma_k)^{1-\lambda}\exp(-(1-\delta/4)Lg(\gamma_k)) \\ &\geqq \varepsilon_\lambda \exp((\delta/8)\mathrm{L}g(\gamma_k)) \geqq 4Lk\end{aligned}$$

so since $p_k \to 0$, (7.60) gives

$$\mathbb{P}(B_k^c) \leqq (1-p_k)^{r(k,0)} \leqq \exp(-p_k r(k,0)/2) \leqq k^{-2}$$

and the theorem follows. □

Theorem 4.1 will be proved after Theorem 4.2, and Corollaries 3.5, 3.7, and 3.9 will be proved in Sect. VIII.

Proof of Theorem 4.2. The proof of (i) is like that of Theorem 3.1 (A) (i), so we will omit details which are similar. Recall that $q_1(t) = \psi_1(t^{\frac{1}{2}})$. If $\lambda < 1$ and $g(t) \geqq e$ then $\lambda t\, Lg(\lambda t) \leqq \lambda t\, L(\lambda^{-1} g(t)) \leqq t\, Lg(t)$; it follows that ψ_1 is increasing (at least for small t, which is clearly all that matters) so we may assume $q_1 \in Q$. By Lemma 7.2, to prove (i) it suffices to show that for each $\delta > 0$,

$$\begin{aligned} \mathbb{P}^*[|v_n(C)| > (1+8\delta)\, R_\delta q_1(\sigma^2(C)) \quad &\text{for some } C \in \mathscr{C} \\ \text{with} \quad \gamma_n \leqq \sigma^2(C) \leqq \alpha_n] &= 0\,((Ln)^{-(1+\delta)}), \end{aligned} \tag{7.61}$$

where $R_\delta = (2((\varrho+\delta)(c_1+\delta)+c_2+c_3+2\delta))^{\frac{1}{2}}$.

Fix n and $0 < \delta < \frac{1}{8}$, then $0 < u < \delta$ small enough so

$$\delta^2 R_\delta^2 / 512\, u^2 \geqq 2(c_3+2). \tag{7.62}$$

Let $t_j = (1-u)^j \alpha_n$ and let $\mathscr{E}(j)$, $\mathscr{F}(j)$, N_n, and $\mathscr{C}(t)$ be as in the proof of Theorem 3.1 (A). As in (7.20),

$$\begin{aligned} &\mathbb{P}^*[|v_n(C)| > (1+8\delta) R_\delta q_1(\sigma^2(C)) \quad \text{for some } C \in \mathscr{C} \quad \text{with} \quad \gamma_n \leqq \sigma^2(C) \in \alpha_n] \\ &\leqq \sum_{j=0}^{N_n} \mathbb{P}[|v_n(C)| > (1+4\delta)\, R_\delta q_1(t_j) \text{for some } C \in \mathscr{F}(j)] \\ &\quad + \mathbb{P}^*[|v_n(C)| > \delta R_\delta q_1(t) \qquad \text{for some } \gamma_n \leqq t \leqq \alpha_n \quad \text{and } C \in \mathscr{C}(t)] \\ &\equiv \sum_{j=0}^{N_n} \mathbb{P}_j + \mathbb{P}_n^*. \end{aligned}$$

By (3.3), we can take

$$|\mathscr{F}(j)| \leqq K g(t_j)^{\varrho+\delta}$$

for some $K = K(\delta, u) < \infty$. Since $q_1(t_j)/n^{\frac{1}{2}} t_j \leqq ((Lg_1(\gamma_n))/n\gamma_n)^{\frac{1}{2}} \to 0$, we obtain using (6.5) and (6.2) that

$$\begin{aligned} \mathbb{P}_j &\leqq 2\,|\mathscr{F}(j)| \exp(-(1+4\delta) R_\delta^2 (Lg_1(t_j))/2 \\ &\leqq 2K \exp(-(1+4\delta)(c_2+c_3+2\delta) Lg_1(t_j)) \\ &\leqq 2K \exp(-(1+4\delta)(c_2+\delta) Lg_1(t_j) - (1+4\delta) LLn). \end{aligned} \tag{7.63}$$

If $c_2 = c_2'$ then

$$\begin{aligned} &\sum_{j=0}^{N_n} \exp(-(1+4\delta)(c_2+\delta) Lg_1(t_j)) \\ &\leqq \sum_{j=0}^{\infty} \exp(-(1+4\delta)\, LLt_j^{-1}) \\ &\leqq \sum_{j=0}^{\infty} (j \log(1-u)^{-1} + L\alpha_n^{-1})^{-(1+4\delta)} \\ &= o(1) \quad \text{as} \quad n \to \infty \end{aligned}$$

so by (7.63),

$$\sum_{j=0}^{N_n} \mathbb{P}_j = O((Ln)^{-(1+4\delta)}). \tag{7.64}$$

If $c_2 = c_2''$ then since $N_n = O(L(\gamma_n^{-1}\alpha_n))$,

$$\sum_{j=0}^{N_n} \exp(-(1+4\delta)(c_2+\delta) Lg_1(t_j))$$
$$\leqq N_n \exp(-(1+4\delta)(c_2+\delta) Lg_1(\alpha_n)) = o(1)$$

so again (7.64) holds.

As in the proof of Theorem 3.1 (A) (i), we obtain from Theorem 2.1, (7.62), and (4.7) that

$$\mathbb{P}_n^* \leqq 36 \int_{\gamma_n/2}^{\alpha_n} t^{-1} \exp(-\delta^2 R_\delta^2 Lg_1(t)/512 u^2)\, dt$$
$$+ 68 \exp(-\delta R_\delta (n\gamma_n Lg_1(\gamma_n))^{\frac{1}{2}}/256)$$
$$\leqq 36 \int_{\gamma_n/2}^{\alpha_n} t^{-1} \exp(-2 LLt^{-1} - 2(c_3+1) Lg_1(\alpha_n))\, dt$$
$$+ 68 \exp(-2(c_3+1) Lg_1(\gamma_n))$$
$$= O(\exp(-2 LLn)),$$

and (7.61), and then (i), follow.

For (ii), fix $\delta > 0$ and set $\gamma_n^* = \gamma_n \vee \alpha_n^{1+\delta}$. We wish to apply Proposition 7.9. Consider the sequences in (7.45): since $Lg(t)/Lt^{-1}$ increases, $Lg_1(\gamma_n^*) \leqq (1+\delta) Lg_1(\alpha_n)$ so

$$\lim_n \gamma_n^* Lg(\gamma_n^*)/q_1^2(\gamma_n^*) = c_1$$
$$\liminf_n \gamma_n^* LL((\gamma_n^*)^{-1}\alpha_n)/q_1^2(\gamma_n^*) \geqq (1+\delta)^{-1} c_2$$
$$\liminf_n \gamma_n^* LLn/q_1^2(\gamma_n^*) \geqq (1+\delta)^{-1} c_3.$$

Since δ is arbitrary, Proposition 7.9 and Remark 7.10 prove (ii).

For (iii), by increasing γ_n we may assume $\gamma_n = \alpha_n$. The proof is then just like that of Theorem 3.1 (C), since $c_2 = 0$ and $c_1 = 1$ whenever $c_3 = 0$. □

Proof of Theorem 4.1. It follows readily from Theorem 4.2 that ψ_1 is a local asymptotic modulus at ϕ for (v_n).

To show ψ_0 is an asymptotic modulus of continuity, let $\gamma_n, \alpha_n \downarrow 0$ with $n\alpha_n \uparrow$, $\gamma_n \leqq \alpha_n$, $n^{-1} Ln = o(\gamma_n)$, and $LLn = O(L\alpha_n^{-1})$. It suffices in (4.5) to consider C, D satisfying $P(C \setminus D) \geqq P(C \triangle D)/2$. Let $\mathscr{D} = \{C \setminus D : C, D \in \mathscr{C}, \gamma_n/2 \leqq \sigma^2(C \setminus D) \leqq \alpha_n\}$ and $\mathscr{E} = \{C \triangle D : C, D \in \mathscr{C}, \gamma_n \leqq \sigma^2(C \triangle D) \leqq \alpha_n\}$. We may take the capacity function of $\mathscr{D}$ or $\mathscr{E}$ to be $\tilde{g}(t) = t^{-1}$ (see Remark 2.2). Since $|v_n(C) - v_n(D)| \leqq |v_n(C \triangle D)| + 2|v_n(C \setminus D)|$, we have

$$\sup\left\{\frac{|v_n(C) - v_n(D)|}{\psi_0(\sigma(C \triangle D))} : C, D \in \mathscr{C},\ P(C \triangle D) \leqq \tfrac{1}{2},\ \gamma_n \leqq \sigma^2(C \triangle D) \leqq \alpha_n\right\}$$
$$\leqq \sup\{|v_n(C)|/\psi_0(\sigma(C)) : C \in \mathscr{E},\ \gamma_n \leqq \sigma^2(C) \leqq \alpha_n\}$$
$$+ 2 \sup\{|v_n(C)|/\psi_0(\sigma(C)) : C \in \mathscr{D},\ \gamma_n/2 \leqq \sigma^2(C) \leqq \alpha_n\}$$

so the result follows from Theorem 4.2. □

Proof of Theorem 4.4. We use Theorem 2.1, with $\mathscr{C}(t)=\mathscr{C}_t$, $\zeta(t)=t$, $q(t)=\psi_1(t^{\frac{1}{2}})$, $\gamma=\gamma_n$, and $\alpha=\frac{1}{4}$. (As always, we use $p=0$.) Then $r=\gamma_n>s$ in (2.1), since $n^{-1}Lg(\gamma_n)=o(\gamma_n)$. If b is large enough then (2.2) is clear, (2.3) follows easily from the observation that

$$L(na(t))\leqq L(nt)+Lg(t), \tag{7.65}$$

and (2.4) and (2.5) are vacuous. Hence (2.6) holds. If b is large then $\exp(-b^2q^2(t)/512t)\leqq b^{-1}(Lt^{-1})^2$, so the second term on the right side of (2.6) can be made small. Since $q(\gamma_n)n^{\frac{1}{2}}\geqq(n\gamma_n)^{\frac{1}{2}}\to\infty$ as $n\to\infty$, the third term is small for large n, and the theorem follows. □

Proof of Theorems 5.1 and 5.2. For simplicity we assume $P(C)\leqq\frac{1}{2}$ for all C; Proposition 6.1 easily handles any larger sets. Observe that for $M>0$,

$$\begin{aligned}&\mathbb{P}\left[\sup\left\{\left|\frac{P_n(C)}{P(C)}-1\right|: C\in\mathscr{C},\ P(C)\geqq\gamma_n\right\}>M\right] \\ &\leqq\mathbb{P}[|v_n(C)|>Mn^{\frac{1}{2}}\sigma^2(C)\quad\text{for some } C\in\mathscr{C}\ \text{ with }\ \sigma^2(C)\geqq\gamma_n/2].\end{aligned}\tag{7.66}$$

Thus to prove the desired results we use Theorem 2.1 with $\mathscr{C}(t)=\mathscr{C}_t$, $\zeta(t)=t$, $\gamma=\gamma_n/2$, $\alpha=\frac{1}{4}$, $q(t)=t$, and $b=Mn^{\frac{1}{2}}$. If (2.2)–(2.5) hold then (2.6) bounds the right side of (7.66) by

$$\begin{aligned}&36\int_{\gamma_n/2}^{\alpha}t^{-1}\exp(-M^2nt/512)\,dt+68\exp(-Mn\gamma_n/256)\\ &=O(\exp(-M^2n\gamma_n/1024))+O(\exp(-Mn\gamma_n/256)).\end{aligned}\tag{7.67}$$

In (2.1) the values are

$$r=\gamma_n,\quad s=\begin{cases}0 & \text{if}\quad M\leqq 2\\ \infty & \text{if}\quad M>2\end{cases}. \tag{7.68}$$

To prove (5.2), we take M fixed but arbitrarily small in (7.66). Then (5.1), (7.65), and the fact that $n\gamma_n\to\infty$ establish (2.2) and (2.3). (2.4) and (2.5) are vacuous by (7.68), so (5.2) follows from (7.66) and (7.67). If (5.3) holds, then this same proof shows the right side of (7.66) is $O((Ln)^{-2})$, and a.s. convergence in (5.2) follows from Lemma 7.2.

To prove (5.6), observe that if $R\geqq 2$, since $P_n(C)\geqq n^{-1}$ whenever $P_n(C)\neq 0$,

$$\begin{aligned}&\mathbb{P}[P_n(C)\leqq(RLg(\gamma_n^*))^{-1}P(C)\quad\text{for some } C\in\mathscr{C}\ \text{ with }\ P_n(C)\neq 0]\\ &\leqq\mathbb{P}\left[\sup\left\{\left|\frac{P_n(C)}{P(C)}-1\right|: C\in\mathscr{C},\ P(C)\geqq R\gamma_n^*\right\}>\frac{1}{2}\right].\end{aligned}\tag{7.69}$$

so we use $M=\frac{1}{2}$ and $\gamma_n=R\gamma_n^*$ this time in (7.65). If R is large enough then (2.2) and (2.3) follow from (7.65), (7.68), and the observations that $n\gamma_n^*\geqq 1$ and $(nt)^{-1}Lg(t)\leqq 2R^{-1}$ for $t\geqq R\gamma_n^*/2$. (2.4) and (2.5) are vacuous. Hence (7.69), (7.66), and (7.67) bound the left side of (5.6) by $O(\exp(-2^{-13}RLg(\gamma_n^*))$, and (5.6) follows. (5.7) is proved similarly, except that now it is (2.2) and (2.3) that are vacuous.

If g is bounded, say $g(t)\leqq\lambda$ for all t, then $\sup\{P_n(C)/P(C): C\in\mathscr{C}, P(C)\geqq\varepsilon\gamma_n^*\}$ is bounded in probability for all $\varepsilon>0$ by (5.7), while for $R>0$,

$$\begin{aligned}&\mathbb{P}[\sup\{P_n(C)/P(C): C\in\mathscr{C},\ P(C)<\varepsilon\gamma_n^*\}>R]\\&\quad\leqq \mathbb{P}[\sup\{P_n(C): C\in\mathscr{C},\ P(C)<\varepsilon\gamma_n^*\}>0]\\&\quad\leqq na(\varepsilon\gamma_n^*)\leqq n\lambda\varepsilon\gamma_n^*\leqq\lambda\varepsilon L\lambda\to 0\quad\text{as}\quad \varepsilon\to 0,\end{aligned}$$

and the last statement in Theorem 5.2 follows.

It remains to prove (5.5) when $\mathscr{C}$ is full and (5.4) holds. By (5.4) we have $\gamma_n\leqq\tau n^{-1}(\mathrm{Lg}(\gamma_n)\vee LLn)$ for all n for some $\tau>0$, so if we define γ_n' to be the solution γ of

$$\gamma=\tau n^{-1}(Lg(\gamma)\vee LLn),$$

then $\gamma_n\leqq\gamma_n'$. Set $w_n'=Lg(\gamma_n')\vee LLn$. By Proposition 7.11 and Remark 7.10, for some $0<\theta<1$ we have infinitely often

$$\begin{aligned}&\sup\left\{\left|\frac{P_n(C)}{P(C)}-1\right|: C\in\mathscr{C},\ P(C)\leqq\tfrac{1}{2},\ \sigma^2(C)\geqq\gamma_n\right\}\\&\quad\geqq\sup\{|\nu_n(C)|/2(n\gamma_n')^{\frac{1}{2}}\,\sigma(C): C\in\mathscr{C},\ \sigma^2(C)=\gamma_n'\}\\&\quad\geqq(\beta_{\theta\tau}-1)/4.\quad\square\end{aligned}$$

VIII. Proofs for Examples

Define $f(t)\leqq\frac{1}{2}$ by $f(t)\,(1-f(t))=t$, so $f(\sigma^2(C))=P(C)$ *if* $P(C)\leqq\frac{1}{2}$. Observe that by (3.1) it suffices to prove (3.3) for small t, say $t\leqq\frac{1}{8}$, to prove it for all $t>0$.

Proof of Corollary 3.5. From Example 3.4 we see that if suffices to prove (3.3) for $\mathscr{C}=\mathscr{D}_d$, $\varrho=1$, and some $\eta<\infty$. Fix $u\in(0,1)$ and $t\in(0,\frac{1}{8}]$, and set

$$\begin{aligned}\mathscr{C}'&=\{C\in\mathscr{D}_d:(1-u^2/4)\,t<\sigma^2(C)\leqq t,\ P(C)\leqq\tfrac{1}{2}\}\\\mathscr{C}''&=\{C^c: C\in\mathscr{D}_d,(1-u^2/4)\ t<\sigma^2(C)\leqq t,\ P(C)>\tfrac{1}{2}\}\end{aligned}$$

so $\mathscr{C}_t\setminus\mathscr{C}_{(1-u^2/4)t}=\mathscr{C}'\cup\mathscr{C}''$; we will consider $\mathscr{C}'$ and $\mathscr{C}''$ separately.

For $\mathscr{C}'$, let $\mu=1-u^2/8d$ and let $\mathscr{F}$ be as in (3.7). We have (cf. (3.8))

$$|\mathscr{F}|\leqq K_1\,(\log\mu^{-1})^{-(d-1)}(Lt^{-1})^{d-1}\leqq K_2\,u^{-2(d-1)}g(t)$$

where K_1 and K_2 depend only on d.

Fix $C\in\mathscr{C}'$. There exists $C_1=[0,y]$ with $C\subset C_1,\sigma^2(C_1)=t$, $C_1^c\in\mathscr{C}''$, and $P(C_1\setminus C)=f(t)-f(\sigma^2(C))\leqq 2(t-\sigma^2(C))\leqq u^2t/2$. Clearly there then exists $C_2=[0,x]\in\mathscr{F}$ with $\mu x_i\leqq y_i\leqq\mu^{-1}x_i$ for all $i\leqq d$, so $P(C_1\,\triangle\,C_2)\leqq 2d(\mu^{-1}-1)$ $P(C_1)\leqq u^2t/2$. Hence $P(C\,\triangle\,C_2)\leqq u^2t$ and $\mathscr{C}'$ is taken care of.

For $\mathscr{C}''$ the proof is somewhat similar. Let N be an integer between $20\,du^{-2}$ and $21\,du^{-2}$ and set

$$\mathscr{G}=\left\{[0,x]: x_i=1-n_i f(t)/N\quad\text{for some } n_i\leqq N\right.\\\left.\text{for each }\ i\leqq d-1,\ \prod_{i=1}^{d}x_i=1-f(t)\right\}.$$

Fix C with $C^c\in\mathscr{C}''$. There exists $C_1=[0,y]$ with $C\supset C_1$, $C_1^c\in\mathscr{C}''$, $\sigma^2(C_1)=t$, and $P(C\setminus C_1)\leqq u^2t/2$. Then we can find $C_2=[0,x]\in\mathscr{G}$ with $x_i-f(t)/N\leqq y_i\leqq x_i$ for each $i\leqq d-1$. Now $\prod_{i=1}^{d-1}x_i\leqq\left(\prod_{i=1}^{d-1}y_i\right)(1+4df(t)/N)$ so $x_d\leqq y_d\leqq x_d\,(1+4df(t)/N)$. Thus

$$P(C^c \triangle C_2^c) = P(C \triangle C_2) \leqq P(C \setminus C_1) + P(C_1 \triangle C_2)$$
$$\leqq u^2 t/2 + \sum_{i=1}^{d} |x_d - y_d|$$
$$\leqq u^2 t/2 + 5\, df(t)/N \leqq u^2 t.$$

Since

$$|\mathscr{G}| \leqq (N+1)^{d-1} \leqq (22d)^{d-1} u^{-2(d-1)} g(t),$$

$\mathscr{C}''$ is now taken care of, and (3.3) follows. □

Proof of Corollary 3.7. We need to show that $\mathscr{C}$ is full and that (3.3) holds with $\varrho = 1$. Example 3.6 shows we may assume $P = N(0, I)$. We use the notation of Example 3.6.

Fix $u \in (0,1)$ and $t \in (0, \frac{1}{8}]$, set

$$r = \Phi^{-1}(1 - f(t)),\ r^* = \Phi^{-1}(1 - f((1 - u^2/4)\, t))$$

and fix $\theta > 0$ to be specified later. Then

$$\mathscr{C}_t \setminus \mathscr{C}_{(1-u^2/4)t} = \{C_{bv} : r \leqq b < r^*,\ v \in S^{d-1}\} \cup \{C_{bv}^c : -r^* < b \leqq -r,\ v \in S^{d-1}\} \tag{8.1}$$

Let V be a maximal subset of S^{d-1} such that the angle between any two vectors in V is at least θ, and $\mathscr{F} = \{C_{rv} : v \in V\}$. Then

$$\delta \theta^{-(d-1)} \leqq |V| = |\mathscr{F}| \leqq M \theta^{-(d-1)} \tag{8.2}$$

for some δ, M depending only on d.

We now prove (3.3). Let us specify $\theta = u^2/16r$. Let $C \in \mathscr{C}_t \setminus \mathscr{C}_{(1-u^2/4)t}$. It is clear that there is a $w \in S^{d-1}$ for which $C \subset C_{rw}$ and

$$P(C \triangle C_{rw}) \leqq P(C_{rw}) - P(C_{r*w}) = f(t) - f((1 - u^2/4)\,\mathrm{t}) \leqq u^2 t/2. \tag{8.3}$$

Since V is maximal, there is a $v \in V$ making an angle $\alpha \leqq \theta$ with w. Suppose we can show that

$$P(C_{rv} \triangle C_{rw}) = 2P(C_{rv} \setminus C_{rw}) \leqq u^2 t/2. \tag{8.4}$$

With (8.3) and (8.2) this shows that

$$N_2(ut^{\frac{1}{2}}, \mathscr{C}_t \setminus \mathscr{C}_{(1-u^2/4)t}, P) \leqq M \theta^{-(d-1)} \leqq 16^{d-1} M u^{-2(d-1)} r^{d-1}.$$

Since $\Phi^{-1}(1 - f(t)) \sim (2Lf(t)^{-1})^{\frac{1}{2}}$ as $t \to 0$, there exists $K = K(d)$ such that

$$K^{-1} g(t) \leqq r^{d-1} \leqq Kg(t) \tag{8.5}$$

and (3.3) follows.

The equality in (8.4) is clear. Since $P(C_{rv} \setminus C_{rw})$ depends only on r and the angle α, we may assume $d = 2$, $v = (0, 1)$, and $w = (-\sin\alpha, \cos\alpha)$ in proving the inequality in (8.4). Let l_1 be the boundary of C_{rw}, $l_2 = \{x : x \cdot w = r\cos\alpha\}$ the line parallel to l_1 through rv, T the strip between l_1 and l_2, $\mathrm{H} = \{x : x \cdot w \leqq r\cos\alpha\}$ the closed half plane bounded by l_2 and disjoint from C_{rw}, and W the wedge $H \cap C_{rv}$ with vertex at rv. Then

$$P(C_{rv} \setminus C_{rw}) \leqq P(T) + P(W). \tag{8.6}$$

Let t_0 satisfy $\Phi^{-1}(1 - f(t_0)) = 1$. For t bounded away from 0, (3.3) follows from Lemma 7.13 of Dudley (1978), so we may assume $t < t_0$. Then $r > 1$ and

$$
\begin{aligned}
P(T) &= \Phi(r) - \Phi(r\cos\alpha) \\
&\leqq r(1-\cos\alpha)\exp(-r^2(\cos^2\alpha)/2) \\
&\leqq r\theta^2 \exp(r^2\theta^2/2)\exp(-r^2/2)/2 \\
&\leqq u^2 r^{-1}\exp(-r^2/2)/16 \leqq u^2(1-\Phi(r))/8 \leqq u^2 t/4 .
\end{aligned}
$$

Using polar coordinates centered at rv we obtain

$$
\begin{aligned}
P(W) &= \int_0^\infty \int_0^\alpha (2\pi)^{-1}\exp(-(r^2+s^2+2rs\sin\beta)/2)\, s\,d\beta\, ds \\
&\leqq \alpha \exp(-r^2/2)\int_0^\infty (2\pi)^{-1}\exp(-s^2/2)\, s\, ds \\
&\leqq u^2 r^{-1}\exp(-r^2/2)/16 \leqq u^2(1-\Phi(r))/8 \leqq u^2 t/4 . \qquad (8.8)
\end{aligned}
$$

Combining (8.6), (8.7), and (8.8) proves (8.4), and (3.3) follows.

To show $\mathscr{C}$ is full we use similar ideas, but change θ to $(16\,dLr)^{\frac{1}{2}}/r$. Fix $\lambda \in (0,1)$ and take $b = b(d)$ large enough so

$$P[\{x: \|x\| > r\}] \leqq b r^{d-2}\exp(-r^2/2) \quad \text{for all } r \geqq 1 . \qquad (8.9)$$

We may assume t is small enough (i.e. r large enough) so

$$r \geqq 1 \quad \text{and} \quad 1 \geqq \theta^2 \geqq 16 r^{-2} L(\lambda^{-1}(2r)^{d-1}) . \qquad (8.10)$$

If $x \in C_{rv} \cap C_{rw}$ for some distinct vectors $v, w \in V$, then since the angle between v and w is at least θ, we have $\|x\|^2 \geqq r^2 + r^2\tan^2(\theta/2)$, so $\|x\| \geqq r(1+\theta^2/16)$. It follows using (8.9) and (8.10) that

$$
\begin{aligned}
&P\Big(C_{rv} \cap \Big(\bigcup_{w\in V,\, w\neq v} C_{rw}\Big)\Big) \\
&\qquad \leqq P(\{x: \|x\| \geqq r(1+\theta^2/16)\}) \\
&\qquad \leqq b(2r)^{d-2}\exp(-r^2\theta^2/16)\exp(-r^2/2) \\
&\qquad \leqq \lambda(2r)^{-1}\exp(-r^2/2) \\
&\qquad \leqq \lambda P(C_{rv}) .
\end{aligned}
$$

Since by (8.2) and (8.5),

$$|\mathscr{F}| = |V| \geqq \delta\theta^{-(d-1)} \geqq \delta r^{d-1}/(16\,dLr)^{(d-1)/2} \geqq \varepsilon g(t)^{1-\lambda}$$

for some constant $\varepsilon = \varepsilon(\lambda, d)$, it follows that $\mathscr{C}$ is full. □

Proof of Corollary 3.9. We must verify (3.3) with $\varrho = 1$. Fix $u \in (0,1)$ and $t \in (0, \frac{1}{8}]$ and set $\tau = \log(2t^{-1})$ and $r = 20\,du^{-2}$. Let $\mathbb{Z}_+$ denote the nonnegative integers. For each $j, k \in \mathbb{Z}_+^d$ with $k_i \leqq r e^{j_i/r}$ for all $i \leqq d$ and $\sum j_i \leqq \tau r$, define $a^{jk}, b^{jk} \in [0,1]^d$ by

$$a_i^{jk} = \frac{k_i}{r} e^{-j_i/r}, \quad b_i^{jk} = \left[\frac{k_i+1}{r} e^{-j_i/r} + e^{-j_i/r}\right] \wedge 1 .$$

The number of rectangles $[a^{jk}, b^{jk}]$ is at most

$$\sum_{j:\sum j_i \leqq \tau r} (r+1)^d \exp\left(\sum j_i/r\right)$$
$$\leqq (\tau r)^d (r+1)^d e^{\tau} \leqq K_1 u^{-4d} t^{-1} (Lt^{-1})^d \leqq K_2 u^{-4d} g(t)^{1+\delta}$$

for some constants $K_i = K_i(d, \delta)$.

Fix $[v, w] \in \mathscr{I}_d$ with $P([v, w]) = \frac{1}{2}$ and $(1 - u^2/4)\, t < \sigma^2([v, w]) \leqq t$, and let

$$j_i = \max\{j : e^{-j/r} \geqq w_i - v_i\}, \; k_i = \max\left\{k : \frac{k}{r} e^{-j_i/r} \leqq v_i\right\}$$

for each $i \leqq d$. Then

$$\sum j_i \leqq r \log P([v, w])^{-1} \leqq r\tau \quad \text{and} \quad k_i \leqq r e^{j_i/r}.$$

Now

$$v_i - 2r^{-1}(v_i - w_i) \leqq a_i^{jk} \leqq v_i \quad \text{and} \quad w_i \leqq b_i^{jk} \leqq w_i + 3r^{-1}(w_i - v_i),$$

so $[v, w] \subset [a^{jk}, b^{jk}]$ and

$$P([a^{jk}, b^{jk}]) \leqq (1 + 5r^{-1})^d P([v, w]) \leqq P([v, w]) + u^2 t.$$

(3.3) now follows. □

References

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

Alexander, K.S. Rates of growth for weighted empirical processes. In: LeCam, L., Olshen, R. (eds.) Proceedings of the Berkeley Conference in honor of Jerzy Neyman and Jack Kiefer, vol. 2. Belmont, CA: Wadsworth 1985

Alexander, K.S.: Sample moduli for set-indexed Gaussian processes. Ann. Probab. **14**, 598–611 (1986)

Bennett, G.: Probability inequalities for the sum of independent random variables. J. Am. Statist. Assoc. **57**, 33–45 (1962)

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression trees. Belmont, CA: Wadsworth 1984

Csáki, E.: The law of the iterated logarithm for normalized empirical distribution function. Z. Wahrscheinlichkeitstheor. Verw. Geb. **38**, 147–167 (1977)

Diaconis, P., Freedman, D.: Asymptotics of graphical projection pursuit. Ann. Statist. **12**, 793–815 (1984)

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

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

Dudley, R.M.: A course on empirical processes. Ecole d'été de probabilités de St.-Flour, 1982. Lect. Notes Math. **1097**, 1–142. Berlin Heidelberg New York: Springer, 1982

Gaenssler, P.: Empirical processes. IMS Lecture Notes – Monograph Series **3** (1983)

Giné, E., Zinn, J.: Some limit theorems for empirical processes. Ann. Probab. **12**, 929–989 (1984)

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

Huber, P.: Projection pursuit. Ann. Statist. **13**, 435–475 (1985)

Jogdeo, K., Samuels, S.M.: Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation. Ann. Math. Statist. **39**, 1191–1195 (1968)

Kiefer, J.: Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. Proc. 6th Berkeley Sympos. Math. Statist. Probab. **1**, 227–244. Berkeley: Univ. of Calif. Press (1972)

LeCam, L.: A remark on empirical measures. In: A Festschrift for Erich L. Lehmann in Honor of His 65th Birthday (Bickel, P., Doksum, K., Hodges, J. (eds). Belmont, CA: Wadsworth (1983)

Mason, D.M., Shorack, G.R., Wellner, J.A.: Strong limit theorems for oscilation moduli of the uniform empirical process. Z. Wahrscheinlichkeitstheor. Verw. Geb. **65**, 83–97 (1983)
Orey, S., Pruitt, W.E.: Sample functions of the N-parameter Wiener process. Ann. Probab. **1**, 138–163 (1973)
Pollard, D.: A central limit theorem for empirical processes. J. Austral. Math. Soc. (Ser. A) **33**, 235–248 (1982)
Pollard, D.: Convergence of stochastic processes, New York Heidelberg Berlin: Springer 1984
Shorack, G.R., Wellner, J.A.: Limit theorems and inequalities for the uniform empirical process indexed by intervals. Ann. Probab. **10**, 639–652 (1982)
Stout, W.: Almost sure convergence. New York: Academic Press 1974
Stute, W.: The oscillation behavior of empirical processes. Ann. Probab. **10**, 86–107 (1982a)
Stute, W.: A law of the logarithm for kernel density estimators. Ann. Probab. **10**, 414–422 (1982b)
Stute, W.: The oscillation behavior of empirical processes: the multivariate case. Ann. Probab. **12**, 361–379 (1984)
Uhlmann, W.: Vergleich der hypergeometrischen mit der Binomial-Verteilung. Metrika **10**, 145–158 (1966)
Vapnik, V.N., Červonenkis, A.Ya.: 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) (1971)
Vapnik, V.N., Červonenkis, A.Ya.: Necessary and sufficient conditions for the uniform convergence of means of their expectations. Theor. Probab. Appl. **26**, 532–553 (1981)
Wellner, J.A: Limit theorems for the ratio of the empirical distribution function to the true distribution function. Z. Wahrscheinlichkeitstheor. Verw. Geb. **45**, 73–88 (1978)
Yukich, J.: Laws of large numbers for classes of functions. J. Multivar. Analysis **17**, 245–260 (1985)
Zuijlen, M.C.A. van: Properties of the empirical distribution function for independent non-identically distributed random vectors. Ann. Probab. **10**, 108–123 (1982)

Received March 16, 1984; in revised form September 12, 1986