The Annals of Probability
1978, Vol. 6, No. 6, 899–929

SPECIAL INVITED PAPER

CENTRAL LIMIT THEOREMS FOR EMPIRICAL MEASURES[1]

BY R. M. DUDLEY
Massachusetts Institute of Technology

Let $(X, \mathcal{Q}, P)$ be a probability space. Let $X_1, X_2, \cdots$, be independent X-valued random variables with distribution P. Let $P_n := n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n})$ be the empirical measure and let $\nu_n := n^{\frac{1}{2}}(P_n - P)$. Given a class $\mathcal{C} \subset \mathcal{Q}$, we study the convergence in law of ν_n, as a stochastic process indexed by $\mathcal{C}$, to a certain Gaussian process indexed by $\mathcal{C}$. If convergence holds with respect to the supremum norm $\sup_{C \in \mathcal{C}}|f(C)|$, in a suitable (usually nonseparable) function space, we call $\mathcal{C}$ a Donsker class. For measurability, X may be a complete separable metric space, $\mathcal{Q}$ = Borel sets, and $\mathcal{C}$ a suitable collection of closed sets or open sets. Then for the Donsker property it suffices that for some m, and every set $F \subset X$ with m elements, $\mathcal{C}$ does not cut all subsets of F (Vapnik-Červonenkis classes). Another sufficient condition is based on metric entropy with inclusion. If $\mathcal{C}$ is a sequence $\{C_m\}$ independent for P, then $\mathcal{C}$ is a Donsker class if and only if for some r, $\Sigma_m(P(C_m)(1 - P(C_m)))^r < \infty$.

1. Introduction. The statistics used in Kolmogorov-Smirnov tests are suprema of normalized empirical measures $n^{\frac{1}{2}}(P_n - P)$ or $(mn)^{\frac{1}{2}}(m + n)^{-\frac{1}{2}}(P_m - Q_n)$ over a class $\mathcal{C}$ of sets, namely the intervals $]-\infty, a]$, $a \in \mathbb{R}$. Donsker (1952) showed here that $n^{\frac{1}{2}}(P_n - P)$ converges in law, in the space $l^\infty(\mathcal{C})$ of all bounded functions on $\mathcal{C}$, to a Gaussian process. Later, Donsker's result was extended to the class of products of intervals parallel to the axes in $\mathbb{R}^k$ (Dudley (1966), (1967a)). Since $l^\infty(\mathcal{C})$ in the supremum norm is nonseparable, some measurability problems (overlooked by Donsker) had to be treated. Recently Révész (1976) proved an iterated logarithm law for a much more general class of sets

$$\bigcap_{1 \leqslant i \leqslant k}\{x : f_i(\{x_j : j \neq i\}) < x_i < g_i(\{x_j : j \neq i\})\}$$

where f_i and g_i have a fixed bound on their partial derivatives of orders $\leqslant k$, and P is the uniform measure on the unit cube. This paper will consider extensions of Donsker's theorem to suitable classes of sets in general probability spaces.

Section 2 will treat countable sequences of sets, with results in particular for independent or disjoint sequences. Sections 3 and 4 treat measurability questions, Section 4 on collections of closed or open sets. Section 5 introduces metric entropy with inclusions, and finds a sufficient condition applicable to bounded collections of convex sets in $\mathbb{R}^2$, or sets with more than $k - 1$ times differentiable boundaries in $\mathbb{R}^k$, if P has a bounded density with respect to Lebesgue measure. Section 6 shows how convergence in law of the one-sample measure $n^{\frac{1}{2}}(P_n - P)$ on a class $\mathcal{C}$

Received June 15, 1977; revised April 25, 1978.
[1]This research was partially supported by National Science Foundation Grant MCS76-07211 A01.
AMS 1970 *subject classifications.* Primary 60F05; Secondary 60B10, 60G17, 28A05, 28A40.
Key words and phrases. Central limit theorems, empirical measures, Donsker classes, Effros Borel structure, metric entropy with inclusion, two-sample case, Vapnik-Červonenkis classes.

extends to the two-sample measure $(mn)^{\frac{1}{2}}(m+n)^{-\frac{1}{2}}(P_m - Q_n)$, where P_m and Q_n are independent empirical measures for P. Section 7 shows that for Vapnik-Červonenkis classes, satisfying suitable measurability conditions, Donsker's theorem holds for all P. Iterated logarithm laws uniformly on classes $\mathcal{C}$ will be treated in a separate paper by J. Kuelbs and the author.

Here are some definitions. Let (S, d) be a metric space (we have in mind a nonseparable space of bounded functions on $\mathcal{C}$ with supremum norm). Let $\mathfrak{B}_b := \mathfrak{B}_b(S, d)$ be the σ-algebra of subsets of S generated by all balls

$$B(x, \varepsilon) := \{y \in S : d(x, y) < \varepsilon\}, \qquad x \in S, \varepsilon > 0.$$

Then $\mathfrak{B}_b$ is a sub-σ-algebra of the σ-algebra $\mathfrak{B}$ of Borel sets, with $\mathfrak{B}_b = \mathfrak{B}$ for S separable. We have $\mathfrak{B}_b \subset \mathfrak{B}$ strictly if the smallest cardinality, γ, of a dense set in S is c, or 2^c, or $2^{2^c}, \cdots$, and hence for all nonseparable S we will treat (Dudley (1967a), proposition and following discussion). If $\gamma = \aleph_\omega$, possibly $\mathfrak{B}_b = \mathfrak{B}$ (Talagrand, 1978).

I will call a probability measure a *law*, defined on $\mathfrak{B}_b$ unless otherwise specified. A sequence μ_n of laws will be said to *converge* to a law μ, $\mu_n \to_{\mathcal{L}} \mu$, if and only if $\int f \, d\mu_n \to \int f \, d\mu$ for every continuous, bounded, $\mathfrak{B}_b$-measurable real-valued f on S (Dudley (1966), (1967a)).

A net X_α of S-valued random variables on a probability space $(\Omega, \mathbb{S}, \Pr)$ is said to converge *almost uniformly* to Y if and only if for every $\varepsilon > 0$ there is a set $A \in \mathbb{S}$ with $\Pr(A) < \varepsilon$ and a β such that for all $\alpha \geqslant \beta$ and $\omega \notin A$, $d(X_\alpha, Y)(\omega) < \varepsilon$. This does not require that $\omega \to d(X_\alpha(\omega), Y(\omega))$ be measurable. Suppose however that for some separable subset $T \subset S$, $\Pr(Y \in T) = 1$. It is easily seen that the metric d is jointly measurable on $(S, \mathfrak{B}_b) \times (T, \mathfrak{B}_b)$, using only balls with centers in a countable dense subset of T. Then $d(X_\alpha, Y)$ is measurable. In this case, a sequence $X_n \to Y$ almost uniformly if and only if $X_n \to Y$ a.s., i.e., $d(X_n, Y) \to 0$ a.s. Wichura (1970) proved that given laws μ_α defined on σ-algebras $\mathcal{Q}_\alpha$ with $\mathfrak{B}_b \subset \mathcal{Q}_\alpha \subset \mathfrak{B}$, such that μ_0 has a separable support, then $\mu_\alpha \to_{\mathcal{L}} \mu_0$ if and only if there exists a probability space $(\Omega, \mathbb{S}, \Pr)$ with random variables X_α such that $X_\alpha^{-1}(B) \in \mathbb{S}$ for all $B \in \mathcal{Q}_\alpha$, $\Pr(X_\alpha^{-1}(B)) = \mu_\alpha(B)$, and $d(X_\alpha, X_0) \to 0$ Pr-almost surely. The usefulness of Wichura's theorem will be seen, e.g., in Section 6.

Now let $(X, \mathcal{Q}, P)$ be a probability space. Let $X_1, X_2, \cdots,$ be independent X-valued random variables with distribution P, defined on a countable product $(X^\infty, \mathcal{Q}^\infty, P^\infty)$ of copies of $(X, \mathcal{Q}, P)$. Let P_n be the random empirical measure

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

where $\delta_x(A) := 1_A(x)$. Let ν_n be the normalized empirical measure $\nu_n := n^{\frac{1}{2}}(P_n - P)$.

Let W_P be the *P-noise* Gaussian process with parameter set $\mathcal{Q}$, $EW_P(A) = 0$ and $EW_P(A)W_P(B) = P(A \cap B)$ for all $A, B \in \mathcal{Q}$. Then W_P has independent values on disjoint sets. For each A let $G_P(A) := W_P(A) - P(A)W_P(X)$. Then G_P is a Gaussian process with parameter space $\mathcal{Q}$ such that $EG_P(A) = 0$ and $EG_P(A)G_P(B) = P(A \cap B) - P(A)P(B)$ for all $A, B \in \mathcal{Q}$.

The central limit theorem in finite-dimensional vector spaces tells us that, at least when restricted to a finite subclass of $\mathcal{A}$, ν_n converges in law to G_P.

Given a subclass $\mathcal{C} \subset \mathcal{A}$, let $l^\infty(\mathcal{C})$ denote the space of all bounded real-valued functions on $\mathcal{C}$, with supremum norm. Then $\nu_n(\cdot)(\omega) \in l^\infty(\mathcal{C})$ for all n and ω.

DEFINITIONS. Given a probability space $(X, \mathcal{A}, P)$ and $A, B \in \mathcal{A}$, let

$$d_P(A, B) := P(A \Delta B), \qquad \text{where} \qquad A\Delta B := (A \setminus B) \cup (B \setminus A).$$

A class $\mathcal{C} \subset \mathcal{A}$ will be called a *$G_P UC$ class* if and only if G_P on $\mathcal{C}$ has a version such that for almost all ω, $C \to G_P(C)(\omega)$ is uniformly continuous on $\mathcal{C}$ for the pseudometric d_P.

Also, $\mathcal{C}$ will be called a $G_P B$ class if and only if G_P has a version which is a.s. bounded on $\mathcal{C}$. If $\mathcal{C}$ is both a $G_P B$ class and a $G_P UC$ class it will be called a $G_P BUC$ class.

Note that $\|1_A - 1_B\|_2 = d_P(A, B)^{\frac{1}{2}}$ where $\|\cdot\|_2$ is the norm in $L^2(X, \mathcal{A}, P)$. Thus these metrics define the same topology and uniform structure on $\mathcal{C}$.

If Y is a Gaussian variable independent of G_P, with $EY = 0$ and $EY^2 = 1$, and $W_P(A) := G_P(A) + P(A)Y$, $A \in \mathcal{A}$, then W_P is a (version of) P-noise.

We recall that if L is a linear map of a Hilbert space H to a space of Gaussian random variables with $EL(x) = 0$ and $EL(x)L(y) = (x, y)$ for all $x, y \in H$, then a set $C \subset H$ is called a *GC-set* (resp. *GB-set*) if and only if L restricted to $\mathcal{C}$ has a version with continuous (resp. bounded) sample functions (Dudley (1967b), (1973)). Let $I_\mathcal{C} := \{1_C : C \in \mathcal{C}\}$.

(1.0). PROPOSITION. *For any $\mathcal{C} \subset \mathcal{A}$, $\mathcal{C}$ is a $G_P BUC$ class if and only if the closure of $I_\mathcal{C}$ in $L^2(X, \mathcal{A}, P)$ is both a GB-set and a GC-set.*

PROOF. In view of the relations between G_P and W_P given above, where we can let $W_P(C) = L(1_C)$, G_P has a version uniformly continuous on a class $\mathcal{C} \subset \mathcal{A}$ if and only if W_P does, and likewise for sample continuity or boundedness. A *GB*-set must be totally bounded (Dudley (1967b), Proposition 3.4, page 295; (1973), Theorem 1.1(c), page 71). Functions on a totally bounded set $I_\mathcal{C}$ extend continuously to its compact closure if and only if they are uniformly continuous. The extension is still a version of the same process since it is continuous in probability. □

(1.1). PROPOSITION. *If $\mathcal{C}$ is countable, then in $l^\infty(\mathcal{C})$, $\mathfrak{B}_b$ equals the smallest σ-algebra $\mathfrak{B}_c$ for which all coordinate evaluations $f \to f(A)$, $A \in \mathcal{C}$, are measurable.*

PROOF. For any $f \in l^\infty(\mathcal{C})$ and $r > 0$, the closed ball

$$B^-(f, r) := \{g : \|g - f\|_\infty \leq r\}$$
$$= \bigcap_{C \in \mathcal{C}} \{g : |g(C) - f(C)| \leq r\}.$$

Thus $\mathfrak{B}_b \subset \mathfrak{B}_c$. Conversely, if $C \in \mathcal{C}$ and $x \in \mathbb{R}$, then

$$\{f : f(C) < x\} = \bigcup_n B(f_n, n)$$

where $f_n(C) := x - n$, $f_n(D) := 0$ for $D \neq C$, $D \in \mathcal{C}$, so $\mathfrak{B}_c \subset \mathfrak{B}_b$. □

For any $\mathcal{C}$, let $C_b(\mathcal{C}, d_P)$ be the space of all bounded real functions on $\mathcal{C}$ continuous for d_P. Let $D_0(\mathcal{C}, P)$ be the linear space of all functions $\phi + \psi$, where $\phi \in C_b(\mathcal{C}, d_P)$ and ψ is a finite linear combination of point masses, $\psi = \Sigma a_i \delta_{x(i)}$. Let $D(\mathcal{C}, P)$ be the closure of $D_0(\mathcal{C}, d_P)$ in $l^\infty(\mathcal{C})$ for the supremum norm.

The space $D(\mathcal{C}, P)$ can be considered as an extension of the usual space $D[0, 1]$ of functions on $[0, 1]$ continuous from the right with limits from the left, where $X = [0, 1]$, $\mathcal{C}$ is the class of all intervals $[0, c]$, $0 \leqslant c \leqslant 1$, and P is Lebesgue measure or any law on $[0, 1]$ with a strictly increasing distribution function.

However, as in this case, functions in D need not have a decomposition into a pure jump part and a continuous part: let $f = 0$ on $[1/(2n + 1), 1/(2n)]$, $f(1/(2n)) = 1/n$, and let f be linear on $[1/(2n), 1/(2n - 1)]$.

Since all our ν_n, and G_P for $G_P BUC$ classes, take values in $D_0(\mathcal{C}, P)$, it will be convenient for us to work in this incomplete space.

For a metric space (S, e) and $T \subset S$, the Borel sets in T are exactly the intersections with T of Borel sets in S. We have

$$\mathfrak{B}_b(T, e) \subset \{A \cap T : A \in \mathfrak{B}_b(S, e)\},$$

where the inclusion may be strict if T is nonseparable, e.g., if $e(x, y) = 1$ for $x \neq y$ in T.

DEFINITION. We say $\mathcal{C}$ is P-EM (*empirically measurable* for P) if and only if for all n, P_n is measurable from the measure-theoretic completion of $(X^\infty, \mathcal{Q}^\infty, P^\infty)$ to $(D_0(\mathcal{C}, P), \mathfrak{B}_b)$.

A countable class $\mathcal{C} \subset \mathcal{Q}$ is always P-EM (by the easy direction of Proposition 1.1). More generally, if $\mathcal{C}$ has a countable subclass $\mathfrak{D}$ such that for all $C \in \mathcal{C}$ there are $D(n) \in \mathfrak{D}$ with $1_{D(n)}(x) \to 1_C(x)$ for all $x \in X$, then $\mathcal{C}$ is P-EM.

EXAMPLE. If P is Lebesgue measure on $[0, 1]$ and $\mathcal{C} = \{\{x\} : x \in E\}$ where E is a nonmeasurable set, then $\mathcal{C}$ is not P-EM, since $\sup_{C \in \mathcal{C}} |P_1(C)|$ is nonmeasurable. Here $\mathcal{C}$ is included in the P-EM class of all singletons.

Note that $\mathfrak{B}_b$-measurability of P_n and ν_n are equivalent. For any P-EM class $\mathcal{C}$, let $\mathcal{L}(\nu_n)$ be the law (probability distribution) of ν_n on $(D_0(\mathcal{C}, P), \mathfrak{B}_b)$.

DEFINITION. A P-EM class $\mathcal{C} \subset \mathcal{Q}$ will be called a *Donsker class* for P, or a *P-Donsker class*, if and only if it is a $G_P BUC$ class and we have convergence of laws $\mathcal{L}(\nu_n) \to \mathcal{L}(G_P)$ in $(D_0(\mathcal{C}, P), \mathfrak{B}_b)$ for the supremum norm, where G_P is taken to have sample functions bounded and d_P-uniformly continuous on $\mathcal{C}$.

There are at least two definitions of convergence of laws, in spaces like $D_0(\mathcal{C}, P)$, different from ours. One the space $D[0, 1]$ of right-continuous functions on $[0, 1]$ with left limits, continuous at 1, Skorohod (1955) and Kolmogorov (1956) defined a complete separable metric topology, now called a "Skorohod topology," for which convergence to a continuous function is equivalent to uniform convergence. See also Billingsley (1968, Chapter 3). Replacing $[0, 1]$ by a cube $[0, 1]^q$, Neuhaus (1971) and Straf (1971) defined a Skorohod topology on a suitable function space $D[0, 1]^q$.

More generally, given a group G of $1-1$ transformations of a probability space X onto itself, with identity element e and a right-invariant metric d, Straf (1971) defines a metric for bounded real functions f, h on X by

$$\rho(f, h) := \inf_{g \in G}(d(e, g) + \sup_x |f(x) - h(g(x))|).$$

$D(X)$ is the ρ-closure of a suitable space of simple functions. Under some conditions, $(D(X), \rho)$ will be separable and topologically complete. But I do not know how to choose a suitable G in the generality of this paper.

Pyke and Shorack (1968) defined weak convergence for processes on $\mathbb{R}$ with bounded sample functions and laws μ_n by $\int f\, d\mu_n \to \int f\, d\mu_0$ for all bounded f which are continuous for the supremum norm and measurable for all μ_n. But the above example $\mathcal{C} = \{\{x\} : x \in E\}$, E nonmeasurable, indicates that there may not be enough such f here.

So, the definition requiring $\mathfrak{B}_b$-measurability will be used. Let $\Omega := X^\infty$ and $\Pr := P^\infty$. Then

$$\Pr^*(Y) := \inf\{\Pr(E) : E \supset Y\} \quad \text{for any set } Y \subset \Omega.$$

Given $\varepsilon > 0$ and $\delta > 0$, let

$$B_{\delta,\varepsilon} := \{f \in D_0(\mathcal{C}, P) : \text{for some } A, B \in \mathcal{C}, d_P(A, B) < \delta \text{ and } |f(A) - f(B)| > \varepsilon\}.$$

Here is a characterization of Donsker classes.

(1.2). THEOREM. *Given a probability space $(X, \mathcal{A}, P)$ and a P-EM class $\mathcal{C} \subset \mathcal{A}$, $\mathcal{C}$ is a Donsker class if and only if both*

(a) *$\mathcal{C}$ is totally bounded for d_P, and*
(b) *for any $\varepsilon > 0$ there is a $\delta > 0$ and an n_0 such that for $n \geqslant n_0$,*

$$\Pr^*\{\nu_n \in B_{\delta,\varepsilon}\} < \varepsilon.$$

PROOF. First assume (a) and (b). Then $\mathcal{C}$ has a countable d_P-dense set $\mathfrak{D}$. Applying the central limit theorem to finite subsets $\mathfrak{D}_m$ of $\mathfrak{D}$, using (b), and letting $\mathfrak{D}_m \uparrow \mathfrak{D}$, shows that $\mathfrak{D}$ is a G_PUC class. Almost all sample functions of G_P then extend uniquely to $\mathcal{C}$ by uniform continuity. It follows that $\mathcal{C}$ is a G_PUC class and a G_PB class.

Let $s(f, g) := \sup_{C \in \mathcal{C}} |f(C) - g(C)|$. For $\gamma > 0$ and a set $K \subset D_0(\mathcal{C}, P)$ let $K^\gamma := \{g : s(f, g) < \gamma \text{ for some } f \in K\}$.

(1.3). LEMMA. 1.2(a) *and* (b) *and* $\mathcal{C}$ P-EM *imply that for any $\varepsilon > 0$ there is a compact set $K \subset C_b(\mathcal{C}, d_P)$ with metric s such that for any $\gamma > 0$, $\Pr\{\nu_n \in K^\gamma\} > 1 - \varepsilon$ for n large enough.*

PROOF. This is a variant of Dudley ((1966), Proposition 2). We may assume $0 < \varepsilon < 1$. By (b), take $\delta > 0$ such that $\Pr^*\{\nu_n \in B_{\delta,\varepsilon/2}\} < \varepsilon/4$ for $n \geqslant n_0 = n_0(\delta, \varepsilon)$. Let $\mathfrak{F}$ be a finite subset of $\mathcal{C}$ such that for all $C \in \mathcal{C}$, $d_P(A, C) < \delta$ for some $A \in \mathfrak{F}$, by (a). Let $\mathfrak{F}$ have k elements. Take M large enough so that

$(M-1)^{-2} < \varepsilon/k$. For each $A \in \mathcal{F}$,

$$\Pr\{|\nu_n(A)| > M-1\} < \varepsilon/(4k)$$

by Chebyshev's inequality. Thus

$$\Pr\{\sup_{A \in \mathcal{F}} |\nu_n(A)| > M-1\} < \varepsilon/4,$$

and

$$\Pr\{\sup_{C \in \mathcal{C}} |\nu_n(C)| > M\} < \varepsilon/2,$$

where the latter event is measurable by the P-EM assumption.

For $m = 1, 2, \cdots$, and $\beta > 0$, let $\varepsilon(m) := \varepsilon/2^m$ and $A_{\beta, m} := B_{\beta, \varepsilon(m)}$. We choose a sequence $\{\beta(m)\}$ of positive numbers satisfying the following two conditions:

(I) $\beta(m+1) < \beta(m)/2$;
(II) For some sequence $\{n_0(m)\}$,

$$\Pr^*\{\nu_n \in A_{\beta(m), m}\} < \varepsilon(m) \quad \text{for all } n \geqslant n_0(m).$$

Let $\delta_m := \beta(m)\varepsilon/(2^{m+1}M)$. Then by (I), $\delta_{m+1} < \delta_m/4$. Let $A_m := A_{\beta(m), m}$. Now if $s(0, f) \leqslant M$, $f \notin A_m$, $C, D \in \mathcal{C}$, and $d_P(C, D) \geqslant \beta(j)$, then $|f(C) - f(D)| \leqslant 2M \leqslant \varepsilon d_P(C, D)/(2^j\delta_j)$, while if $d_P(C, D) < \beta(j)$, then $|f(C) - f(D)| \leqslant \varepsilon(j) = \varepsilon/2^j$. Thus for any $C, D \in \mathcal{C}$,

$$(*) \qquad |f(C) - f(D)| \leqslant \varepsilon 2^{-j} \max(1, d_P(C, D)/\delta_j).$$

Let F_m be the set of all $f \in D_0(\mathcal{C}, P)$ such that $s(0, f) \leqslant M$ and (*) holds for $j = 2, \cdots, m$ and all $C, D \in \mathcal{C}$. Then for $n \geqslant N = N(m)$ large enough, there is a measurable set $E_m \subset X^\infty$ such that $\Pr(E_m) > 1 - \varepsilon$ and for all $\omega \in E_m$, $\nu_n(\cdot)(\omega) \in F_m$.

Now let K be the set of all real functions g on $\mathcal{C}$ such that $s(0, g) \leqslant M$ and for all $j = 1, 2, \cdots, d_P(C, D) < \delta_j/2$ implies $|g(C) - g(D)| \leqslant 3\varepsilon/2^j$. Then $K \subset C_b(\mathcal{C}, d_P)$. Now $(\mathcal{C}, d_P)$ is totally bounded and K is a uniformly bounded, uniformly equicontinuous family of functions, complete for s. Thus by the Arzelà-Ascoli theorem (applied to the completion of $\mathcal{C}$ for d_P), (K, s) is compact.

Given a $\gamma > 0$, choose an integer $m > 1$ such that $\varepsilon/2^m < \gamma/2$. Let us show that $F_m \subset K^\gamma$. Take a maximal set $\mathcal{C}_m \subset \mathcal{C}$ such that $d_P(C, D) \geqslant \delta_m$ for all $C \neq D$ in $\mathcal{C}_m$. Then $\mathcal{C}_m$ is finite by (a). For all $C \in \mathcal{C}$, $d_P(C, D) < \delta_m$ for some $D \in \mathcal{C}_m$.

If $f \in F_m$ and $C, D \in \mathcal{C}_m$, then $|f(C) - f(D)| \leqslant \varepsilon d_P(C, D)/(2^m\delta_m)$, by (*). Then f on $\mathcal{C}_m$ can be extended to a function g on $\mathcal{C}$ with $|g(C) - g(D)| \leqslant \varepsilon 2^{-m} d_P(C, D)/\delta_m$ for all $C, D \in \mathcal{C}$ (McShane (1934)). Taking $\max(-M, \min(g, M))$, we can assume $s(0, g) \leqslant M$. Let us show that $g \in K$. For $i \geqslant m$, since

$$\varepsilon/(2^m\delta_m) = 2M/\beta(m) \leqslant 2M/\beta(i) = \varepsilon/(2^i\delta_i),$$

we have

$$|g(C) - g(D)| \leqslant \varepsilon/2^i \quad \text{for} \quad d_P(C, D) < \delta_i.$$

For $j < m$, given $C, D \in \mathcal{C}$ with $d_P(C, D) < \delta_j/2$, choose $C_m, D_m \in \mathcal{C}_m$ with $d_P(C, C_m) < \delta_m$ and $d_P(D, D_m) < \delta_m$. Then $d_P(C_m, D_m) < 2\delta_m + \delta_j/2 < \delta_j$, and

$$\begin{aligned}|g(C) - g(D)| &\leqslant |g(C) - g(C_m)| \\ &\quad + |f(C_m) - f(D_m)| + |g(D_m) - g(D)| \\ &\leqslant \varepsilon/2^m + \varepsilon/2^j + \varepsilon/2^m < 3\varepsilon/2^j,\end{aligned}$$

using (*) for the middle term. Thus $g \in K$. Now $s(f, g) < \gamma$ since for any $C \in \mathcal{C}$, there is $C_m \in \mathcal{C}_m$ with $d_P(C, C_m) < \delta_m$, and

$$\begin{aligned}|f(C) - g(C)| &\leqslant |g(C) - f(C_m)| + |g(C_m) - g(C)| \\ &\leqslant 2\varepsilon/2^m < \gamma.\end{aligned}$$

So $F_m \subset K^\gamma$ as desired. For $n \geqslant N(m)$, $\Pr^*(\nu_n \notin F_m) < \varepsilon$, so $\Pr(\nu_n \in K^\gamma) > 1 - \varepsilon$, noting that K^γ is a countable union of balls, hence $\mathfrak{B}_b$ measurable. Thus Lemma 1.3 is proved.

Now, using Lemma 1.3, the sequence $\{\mathcal{L}(\nu_n)\}$ of laws on $(D_0(\mathcal{C}, P), \mathfrak{B}_b)$ has a convergent subsequence, as in Dudley ((1966), Theorem 1; (1967a), Theorem). Any limit of such a subsequence is concentrated in a countable union of compact subsets K_n of $C_b(\mathcal{C}, d_P)$, with $\varepsilon = 1/n$ in Lemma 1.3. Now $\bigcup_n K_n$ is separable for s. By the finite-dimensional central limit theorem, these limits of subsequences must all equal $\mathcal{L}(G_P)$. Every subsequence of $\{\mathcal{L}(\nu_n)\}$ has a convergent subsubsequence. (In 1966 I carelessly called this property "precompact"; that should mean "having compact completion" for subsets of uniform spaces.) It follows that $\mathcal{L}(\nu_n) \to \mathcal{L}(G_P)$, i.e., $\mathcal{C}$ is a Donsker class.

Conversely if $\mathcal{C}$ is a Donsker class, hence a G_PB class, then (a) holds by Dudley ((1967b), Proposition 3.4). A version of G_P which has uniformly continuous, hence bounded sample functions on the totally bounded set $\mathcal{C}$ for d_P, gives a law $\mathcal{L}(G_P)$ with a separable support in $D_0(\mathcal{C}, P)$, concentrated in a countable union of compact subsets of $C_b(\mathcal{C}, d_P)$. Thus by Theorem 2 of Wichura (1970), we can assume here that $\nu_n \to G_P$ uniformly on $\mathcal{C}$, so that for any $\delta > 0$,

$$T_{n\delta} := \sup\{|\nu_n(A) - \nu_n(B)| : A, B \in \mathcal{C}, d_P(A, B) \leqslant \delta\}$$

converges almost uniformly (as defined above) for $n \to \infty$ to

$$T_\delta := \sup\{|G_P(A) - G_P(B)| : A, B \in \mathcal{C}, d_P(A, B) \leqslant \delta\}.$$

Here it is not claimed that $T_{n\delta}$ is a measurable random variable, but each T_δ is. For any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\Pr\{T_\delta > \varepsilon/2\} < \varepsilon/2, \qquad \text{and for some} \quad n_0,$$
$$\Pr^*(|T_{n\delta} - T_\delta| > \varepsilon/2) < \varepsilon/2$$

for $n \geqslant n_0$. Thus $\Pr^*(T_{n\delta} > \varepsilon) \leqslant \varepsilon$, proving (b). □

2. Sequences of sets. Let $(X, \mathcal{Q}, P)$ be a probability space.

(2.1). THEOREM. *Let $\{A_m\}_{m=1}^{\infty}$ be any sequence of measurable sets with $P(A_m) = p_m$ such that for some $r < \infty$, $\Sigma_m p_m^r < \infty$. Then $\{A_m\}_{m \geqslant 1}$ is a Donsker class.*

PROOF. We know that any countable collection $\{A_m\}$ is P-EM. Further $\mathcal{C} = \{A_m\}$ is totally bounded for d_P. Hence it remains only to verify condition (b) of Theorem 1.2.

By Theorem 3.1 in Dudley (1967b), $\mathcal{C}$ is a $G_P UC$-class. So for any $\varepsilon > 0$ there is a $\delta = \delta(\varepsilon) > 0$ such that

$$\Pr\{\sup\{|G_P(A_i) - G_P(A_j)| : d_P(A_i, A_j) < 3\delta\} > \varepsilon\} < \varepsilon. \tag{2.2}$$

It will be shown that

(2.3) There exist numbers μ and n_0 such that for all $n \geqslant n_0$,

$$\Pr\{\sup_{m \geqslant \mu} |\nu_n(A_m)| > \varepsilon\} < 2\varepsilon.$$

We define the binomial probabilities

$$b(k, n, p) := n! p^k q^{n-k} / k!\,(n - k)!, \quad \text{where} \quad q := 1 - p,$$

$$E(k, n, p) := \Sigma_{k \leqslant j \leqslant n} b(j, n, p),$$

$$B(k, n, p) := \Sigma_{0 \leqslant j \leqslant k} b(j, n, p)$$

where in B and E, k is not necessarily an integer.

To prove (2.3) it is enough to show that for any $\varepsilon > 0$ there are a μ and an n_0 such that for all $n \geqslant n_0$,

$$\Sigma_{m \geqslant \mu} E\big(np_m + \varepsilon n^{\frac{1}{2}}, n, p_m\big) < \varepsilon \quad \text{and} \tag{2.4}$$

$$\Sigma_{m \geqslant \mu} B\big(np_m - \varepsilon n^{\frac{1}{2}}, n, p_m\big) < \varepsilon. \tag{2.5}$$

We may assume $\frac{1}{2} \geqslant p_m \downarrow 0$ as $m \to \infty$.

To prove (2.5) we use the Chernoff-Okamoto inequality (Okamoto (1958), Lemmas 1, 2b′, or Hoeffding (1963), Theorem 1),

$$B(k, n, p) \leqslant \exp(-(np - k)^2 / 2npq),$$

if $p \leqslant \frac{1}{2}$ and $k \leqslant np$, which implies

$$B\big(np_m - \varepsilon n^{\frac{1}{2}}, n, p_m\big) \leqslant \exp\big(-\varepsilon^2 / 2p_m q_m\big).$$

Let $\delta = 1/r$. Then for some constant $K < \infty$, $p_m \leqslant K m^{-\delta}$ for all m. Thus since $\Sigma_m \exp(-m^{\delta}\varepsilon^2/2K) < \infty$, there is a μ large enough so that (2.5) holds for all n.

For series (2.4), Bernstein's inequality (Bennett (1962) or Hoeffding (1963)) gives

$$E\big(np + \varepsilon n^{\frac{1}{2}}, n, p\big) \leqslant \exp\big(-\varepsilon^2 / \big(2pq + \varepsilon n^{-\frac{1}{2}}\big)\big)$$

$$\leqslant \exp(-\varepsilon^2 / 6pq)$$

if $4pqn^{\frac{1}{2}} \geqslant \varepsilon > 0$ and hence, for $p \leqslant \frac{1}{2}$, if $2pn^{\frac{1}{2}} \geqslant \varepsilon$. Thus

$$\begin{aligned} S_1 &:= \Sigma_m\left\{E\left(np_m + \varepsilon n^{\frac{1}{2}}, n, p_m\right) : 2p_m n^{\frac{1}{2}} \geqslant \varepsilon, m \geqslant \mu\right\} \\ &\leqslant \Sigma_m\left\{\exp\left(-\varepsilon^2/6p_m\right) : 2p_m n^{\frac{1}{2}} \geqslant \varepsilon, m \geqslant \mu\right\} \\ &\leqslant \Sigma_m\left\{\exp\left(-\varepsilon^2 m^\delta/6K\right) : m \geqslant \mu\right\} < \varepsilon/2 \end{aligned}$$

uniformly in n for μ large enough. Let

$$S_2 := \Sigma_m\left\{E\left(np_m + \varepsilon n^{\frac{1}{2}}, n, p_m\right) : 2p_m n^{\frac{1}{2}} < \varepsilon\right\}.$$

The Chernoff-Okamoto inequality (Okamoto (1958), Lemma 1) gives

$$E(k, n, p) \leqslant (np/k)^k (nq/(n-k))^{n-k} \qquad \text{for } k \geqslant np. \tag{2.6}$$

(2.7). LEMMA. *Whenever* $k \geqslant np$, $E(k, n, p) \leqslant (np/k)^k e^{k-np}$.

PROOF. By (2.2), we may prove $(nq/(n-k))^{n-k} \leqslant e^{k-np}$. Let $x := nq/(n-k) \geqslant 1$. Then $x \leqslant e^{x-1}$, giving the result. □

Let $s(n) := n^{\frac{1}{2}}$. Then for $\varepsilon := (k - np)/s(n)$, $E(k, n, p) \leqslant (np/(np + \varepsilon n^{\frac{1}{2}}))^{np+\varepsilon s(n)} e^{\varepsilon s(n)}$, and $\varepsilon n^{\frac{1}{2}} - (np + \varepsilon n^{\frac{1}{2}})\ln(1 + \varepsilon/pn^{\frac{1}{2}}) = \varepsilon n^{\frac{1}{2}}(1 - (x+1)\ln(1 + x^{-1}))$ where $0 < x := n^{\frac{1}{2}}p/\varepsilon < \frac{1}{2}$. The function $f(x) := (1+x)\ln(1+x^{-1})$ is decreasing for all $x > 0$. Thus for $0 < x \leqslant \frac{1}{2}, f(x) \geqslant f(\frac{1}{2}) > 1.5$. Thus

$$\begin{aligned} \varepsilon n^{\frac{1}{2}}(1 - f(x)) &= \varepsilon n^{\frac{1}{2}}(1 - 2f(x)/3) - \varepsilon n^{\frac{1}{2}} f(x)/3 \\ &\leqslant -\varepsilon n^{\frac{1}{2}} f(x)/3. \end{aligned}$$

Hence

$$E(k, n, p) \leqslant \exp\left(-\left(\varepsilon n^{\frac{1}{2}} + np\right)\left(\ln\left(1 + \varepsilon/pn^{\frac{1}{2}}\right)\right)/3\right).$$

Thus

$$S_2 \leqslant \Sigma_m\left\{\left(p_m n^{\frac{1}{2}}/\varepsilon\right)^{\varepsilon s(n)/3} : 2p_m n^{\frac{1}{2}} < \varepsilon\right\}.$$

Since $p_m \leqslant Km^{-\delta}$, we have $S_2 \leqslant S_3 + S_4$ where

$$\begin{aligned} S_3 &:= \Sigma_m\left\{\left(p_m n^{\frac{1}{2}}/\varepsilon\right)^{\varepsilon s(n)/3} : 2Km^{-\delta} n^{\frac{1}{2}} < \varepsilon\right\} \\ &\leqslant \left(Kn^{\frac{1}{2}}/\varepsilon\right)^{\varepsilon s(n)/3} \Sigma_m\left\{m^{-\delta\varepsilon s(n)/3} : m > G\right\} \end{aligned}$$

where $G := G_{nK\varepsilon\delta} := (2Kn^{\frac{1}{2}}/\varepsilon)^{1/\delta}$. Choosing n_1 large enough so that $\delta\varepsilon n_1^{\frac{1}{2}}/3 > 2$, we have for $n \geqslant n_1$

$$\begin{aligned} S_3 &\leqslant \left(Kn^{\frac{1}{2}}/\varepsilon\right)^{\varepsilon s(n)/3} \int_{G-1}^{\infty} x^{-\delta\varepsilon s(n)/3}\, dx \\ &\leqslant \left(Kn^{\frac{1}{2}}/\varepsilon\right)^{\varepsilon s(n)/3}\left(\delta\varepsilon n^{\frac{1}{2}}/3 - 1\right)^{-1}(G-1)^{1-\delta\varepsilon s(n)/3}. \end{aligned}$$

For fixed K, ε and δ, we find that the logarithm of the last expression is asymptotic to $-(\varepsilon n^{\frac{1}{2}}/3)\ln 2 \to -\infty$ as $n \to \infty$, so that $S_3 \to 0$. Thus $S_3 \leqslant \varepsilon/4$ for $n \geqslant n_2$ for some n_2.

Lastly,

$$S_4 := \Sigma_m\Big\{\big(p_m n^{\frac{1}{2}}/\varepsilon\big)^{\varepsilon s(n)/3} : 2p_m n^{\frac{1}{2}} < \varepsilon \leqslant 2Km^{-\delta}n^{\frac{1}{2}}\Big\}$$
$$\leqslant \big(2Kn^{\frac{1}{2}}/\varepsilon\big)^{1/\delta}(1/2)^{\varepsilon s(n)/3} \to 0$$

as $n \to \infty$, so $S_4 < \varepsilon/4$ for $n \geqslant n_3$ for some n_3. Hence for $n \geqslant \max(n_1, n_2, n_3)$, $S_2 < \varepsilon/2$ so (2.4) and hence (2.3) hold. We take μ large enough so that $p_m < \delta$ for all $m \geqslant \mu$. From the finite-dimensional central limit theorem and (2.2) we can find n_4 such that for $n \geqslant n_4$

$$\Pr\big\{\sup\big\{|\nu_n(A_i) - \nu_n(A_j)| : i, j \leqslant \mu \text{ and } d_P(A_i, A_j) < 3\delta\big\} > \varepsilon\big\} < \varepsilon.$$

This and (2.3) then imply condition (1.2b), completing the proof of Theorem 2.1.

(2.8). THEOREM. *If A_m are independent for P and $P(A_m) = p_m$, then $\{A_m\}$ is a Donsker class if and only if for some $r < \infty$, $\Sigma_m(p_m(1 - p_m))^r < \infty$.*

PROOF. Note that a subsequence of the A_m can converge for d_P if and only if their probabilities converge to 0 or 1. Hence, $\{A_m\}$ is Donsker if and only if the collection $\{A_m, X \setminus A_m\}$ of A_m and their complements is Donsker. Thus we may assume $p_m \leqslant \frac{1}{2}$.

Now, "if" follows from Theorem 2.1. Conversely, suppose $\Sigma_m p_m^n = +\infty$ for all n. Then for each n, $\Pr\{P_n(A_m) = 1$ for infinitely many $m\} = 1$ by the Borel-Cantelli lemma. Since $P_n(A_m) = 1$ implies $\nu_n(A_m) = n^{\frac{1}{2}}(1 - p_m) \to \infty$ as $n \to \infty$, $\{A_m\}$ is not a Donsker class. □

Also, any sequence of disjoint measurable sets is a Donsker class by Theorem 2.1 with $r = 1$.

Independent A_m with $P(A_m)\downarrow 0$ as $m \to \infty$ form a G_PBUC class if $P(A_m)\log m \to 0$ as $m \to \infty$ (Dudley (1967b), Proposition 6.7, with $(\int 1_A^2\, dP) = P(A)$). This is weaker than the condition for $\{A_m\}$ to be Donsker, and shows that the binomial upper tail $E(k, n, p)$ for small p becomes substantially larger than the corresponding Gaussian tail.

3. Measurability. A *measurable space* is a pair $(X, \mathfrak{B})$ where X is a set and $\mathfrak{B}$ is a σ-algebra of subsets of X. Then $(X, \mathfrak{B})$ or $\mathfrak{B}$ is called *countably generated* iff there is some countable $\mathcal{C} \subset \mathfrak{B}$ such that $\mathfrak{B}$ is the smallest σ-algebra including $\mathcal{C}$.

Given a measurable space $(X, \mathfrak{B})$, a set $A \subset X$ will be called simply *measurable* if $A \in \mathfrak{B}$. If P is a law defined on $\mathfrak{B}$, A will be called *completion measurable* for P iff $d_P(A, B) = 0$ for some $B \in \mathfrak{B}$, i.e., $A\Delta B \subset C$ for some $C \in \mathfrak{B}$ with $P(C) = 0$. If A is completion measurable for all laws on $\mathfrak{B}$ (where B and C depend on P), it is called *universally measurable*.

If $(Y, \mathcal{E})$ is another measurable space, a function f from X into Y will be called (completion, resp. universally) measurable iff for all $E \in \mathcal{E}, f^{-1}(E)$ is (completion, resp. universally) measurable in X.

A *Polish* space is a topological space metrizable by a complete separable metric. A set A in a metric space is called a *Suslin* set iff there is a continuous function from a Polish space onto A. Note also that any Borel measurable function from a separable metric space into a metric space has separable range (Stone (1962), page 32, Theorem 16). A set in a metric space is Suslin iff it is the range of some Borel function on a Polish space. Thus any Borel set in a Polish space is Suslin.

All Suslin sets are universally measurable (e.g., Federer (1969), Theorem 2.2.12, page 69). Clearly a product of two Suslin sets with product topology is Suslin, a union of countably many Suslin sets is Suslin, and a continuous or Borel image of a Suslin set is Suslin. A countable intersection of Suslin sets is Suslin (e.g., Kuratowski (1966), pages 454, 478). As Suslin proved, a Suslin set whose complement is Suslin is a Borel set (e.g., Kuratowski (1966), pages 485–486), while there exist Suslin non-Borel sets (Kuratowski (1966), page 460).

A measurable space $(X, \mathcal{B})$ will be called *Suslin* iff there is a metric d on X for which (X, d) is Suslin and $\mathcal{B}$ is the σ-algebra of Borel sets.

DEFINITION. If $(X, \mathcal{B})$ and $(\mathcal{C}, \mathcal{S})$ are measurable spaces and $\mathcal{C} \subset \mathcal{B}$, we call $(X, \mathcal{B}; \mathcal{C}, \mathcal{S})$ a *chair*. The chair, or $(\mathcal{C}, \mathcal{S})$, will be called *admissible* iff the $\in$ relation, $\{\langle x, C\rangle : x \in C\}$, is a measurable subset of $X \times \mathcal{C}$ for the product σ-algebra of $\mathcal{B}$ and $\mathcal{S}$. We call $\mathcal{C}$ *admissible* iff there is some σ-algebra $\mathcal{S}$ for which $(\mathcal{C}, \mathcal{S})$ is admissible.

Note that if $(\mathcal{C}, \mathcal{S})$ is admissible, so is $(\mathcal{C}, \mathcal{U})$ for some countably generated $\mathcal{U} \subset \mathcal{S}$.

Suppose $\mathcal{B}$ has a countable set $\mathcal{Q}$ of generators. Let $\mathcal{Q}_0 := \mathcal{Q}$. For each successor ordinal $\alpha + 1$, let $\mathcal{Q}_{\alpha+1}$ be the collection of all complements and countable unions of sets in $\mathcal{Q}_\alpha$. For each limit ordinal $\beta > 0$, let $\mathcal{Q}_\beta := \bigcup_{\alpha<\beta}\mathcal{Q}_\alpha$. The $\mathcal{Q}_\alpha$ are called *Borel classes* (or *Banach classes*, cf. Aumann (1961)).

A collection $\mathcal{C} \subset \mathcal{B}$ will be said to be of *bounded Borel class* iff for some countable set $\mathcal{Q}$ of generators of $\mathcal{B}$, and some countable ordinal α, $\mathcal{C} \subset \mathcal{Q}_\alpha$. This notion does not depend on the choice of $\mathcal{Q}$, although the specific ordinal α does.

We quote a theorem of Aumann ((1961), Theorem D); B. V. Rao ((1971), Theorem 3) gave a shorter proof.

(3.1). THEOREM (Aumann). *For any countably generated measurable space* $(X, \mathcal{B})$ *and* $\mathcal{C} \subset \mathcal{B}$, $\mathcal{C}$ *is admissible iff it is of bounded Borel class.*

Thus in any separable metric space, the collections of all open sets, closed sets, G_δ sets (countable intersections of open sets), F_σ sets (countable unions of closed sets), and countable sets are each admissible, etc. In Section 4, specific admissible σ-algebras will be put on collections of open or closed sets.

In [0, 1], for example, the collection of all Borel sets is not admissible.

We recall that $(\mathcal{C}, \mathcal{S})$ is called a *standard Borel space* iff there is a measurable isomorphism of it with a Polish space carrying its Borel σ-algebra. A metric space (X, d) is called *absolutely Borel* iff X is a Borel set in its completion for d. If X is separable, this property depends on d only through its topology (e.g., Parthasarathy (1967), page 22, Corollary 3.3). A separable metric space with Borel σ-algebra is a standard Borel space iff it is absolutely Borel (Parthasarathy (1967), pages 133–134).

For any probability space $(X, \mathcal{B}, P)$ and $\mathcal{C} \subset \mathcal{B}$, there is a smallest σ-algebra $\mathcal{S}_P$ of subsets of $\mathcal{C}$ for which P is a measurable function, generated by countably many sets $\{A : P(A) < r\}$, r rational. If $(\mathcal{C}, \mathcal{S})$ is admissible, let $\mathcal{U}$ be the smallest σ-algebra including $\mathcal{S}$ and $\mathcal{S}_P$. Then $\mathcal{U}$ is also admissible, and countably generated if $\mathcal{S}$ is; $\mathcal{S}$ always has a countably generated admissible sub-σ-algebra.

DEFINITION. A chair $(X, \mathcal{B}; \mathcal{C}, \mathcal{S})$ is *ϵ-Suslin* iff it is admissible and both $(X, \mathcal{B})$ and $(\mathcal{C}, \mathcal{S})$ are Suslin measurable spaces. Given a law P on $\mathcal{B}$, the chair $(X, \mathcal{B}; \mathcal{C}, \mathcal{S})$ is *$P\epsilon$-Suslin* iff it is ϵ-Suslin and all d_P-open subsets of $\mathcal{C}$ belong to $\mathcal{S}$.

If $(X, \mathcal{B})$ is Suslin, it is countably generated, so $(\mathcal{B}, d_P)$ and $(\mathcal{C}, d_P)$ are always separable.

DEFINITION. Given a probability space $(X, \mathcal{B}, P)$ and $\mathcal{C} \subset \mathcal{B}$, we call $\mathcal{C}$ *strongly* P-EM iff for any n, any real b_i, and independent random variables $X(i)$ in X with law $\mathcal{L}(X(i)) = P$, the map

$$\omega \to \Sigma_{1 \leqslant i \leqslant n} b_i \delta_{X(i)(\omega)}$$

is completion measurable into $(D_0(\mathcal{C}, P), \mathcal{B}_b)$.

(3.2). PROPOSITION. *If $(X, \mathcal{B}; \mathcal{C}, \mathcal{S})$ is $P\epsilon$-Suslin, then $\mathcal{C}$ is strongly P-EM.*

PROOF. For any $x \in X$, the function $C \to \delta_x(C)$ is $\mathcal{S}$-measurable on $\mathcal{C}$ since $(\mathcal{C}, \mathcal{S})$ is admissible. Thus for any $y(1), \ldots, y(k) \in X$ and real $b_1, \cdots, b_k$, the function $C \to \Sigma b_i \delta_{y(i)}(C)$ is $\mathcal{S}$-measurable on $\mathcal{C}$.

Let f be any d_P-continuous function on $\mathcal{C}$. Then for any real t, $\{C \in \mathcal{C} : f(C) > t\}$ is an open set for d_P, thus belongs to $\mathcal{S}$. Hence f is $\mathcal{S}$-measurable. In particular, P is $\mathcal{S}$-measurable. So $P + f + \Sigma b_i \delta_{y(i)}$ is $\mathcal{S}$-measurable.

Now it is enough to show that for independent $X(1), \cdots, X(n)$ with law P, real b_i, and any $\mathcal{S}$-measurable function g on $\mathcal{C}$,

$$\sup_{C \in \mathcal{C}} |\Sigma_{1 \leqslant i \leqslant n} b_i \delta_{X(i)(\cdot)}(C) - g(C)|$$

is measurable. We denote points of X^n by $x = \langle x(1), \cdots, x(n) \rangle$. For any $t \geqslant 0$ we define the set $E_t \subset X^n \times \mathcal{C}$ by

$$E_t := \{\langle x, C \rangle : |\Sigma_{1 \leqslant i \leqslant n} b_i \delta_{x(i)}(C) - g(C)| > t\}.$$

Then

$$E_t := \bigcup_F \{\langle x, C\rangle : x(i) \in C \quad \text{iff} \quad i \in F\}$$
$$\cap \{\langle x, C\rangle : |\Sigma_{i\in F} b_i - g(C)| > t\}$$

where the union runs over all 2^n subsets F of $\{1, \cdots, n\}$.

For each F, $\{\langle x, C\rangle : x(i) \in C$ iff $i \in F\}$ is product measurable since $(X, \mathfrak{B}; \mathcal{C}, \mathcal{S})$ is assumed admissible. Also, $\{C : |\Sigma_{i\in F} b_i - g(C)| > t\}$ is $\mathcal{S}$-measurable since g is an $\mathcal{S}$-measurable function. Thus E_t is jointly measurable.

Since $(X^n, \mathfrak{B}^n)$ and $(\mathcal{S}, \mathcal{C})$ are Suslin spaces, E_t is Suslin and its projection on X^n is Suslin, hence universally measurable. This projection equals

$$\{x : \sup_{C\in\mathcal{C}} |\Sigma_{1\leqslant i\leqslant n} b_i \delta_{x(i)}(C) - g(C)| > t\}. \qquad \square$$

The combination of admissibility and the Suslin property seems not so easy to satisfy. For example, let *Co* be the collection of all countable subsets of $I := [0, 1]$. Then *Co* is of bounded Borel class, hence admissible by 3.1. Also, *Co* has some Suslin measurable structures since its cardinal is c, but I do not know any admissible Suslin structure on *Co*.

For one thing, the structure generated by $\{C \in Co : x \in C\}$, for all $x \in I$, is not countably generated (Szpilrajn-Marczewski (1938)).

Or, take the space I^∞ of all sequences in I, with its standard Borel structure $\mathfrak{B}$. Take the equivalence relation $\equiv$: $\{x_n\} \equiv \{y_n\}$ iff $\{x_n\}$ and $\{y_n\}$ have the same range. Let $\mathfrak{B}/\equiv$ be the factor σ-algebra,

$$\mathfrak{B}/\equiv \; := \{A \in \mathfrak{B} : \quad \text{if} \quad x \equiv y \quad \text{then} \quad x \in A \quad \text{iff} \quad y \in A\}.$$

Then $\mathfrak{B}/\equiv$ is not countably generated (Freedman (1966), Lemma (5)).

For the present, our positive results are for families of open or closed sets (Section 4).

DEFINITION. Given sets X, Y and $E \subset X \times Y$, with projection $\pi_X E \subset X$ where $\pi_X(x, y) := x$, a *selector* for E is a function f from $\pi_X E$ into Y such that $\langle x, f(x)\rangle \in E$ for all $x \in \pi_X E$.

We state for later reference the following extension of a theorem of Lusin and Sierpiński, which is a consequence of Corollary 4.5 of Sion (1960):

(3.3). THEOREM. *Let X and Y be separable metric spaces and E a Borel set in $X \times Y$. Then there is a universally measurable selector f for E.*

4. Spaces of closed or open sets. Let (X, d) be any separable metric space. Then there is a metric e for the d topology of X such that (X, e) is totally bounded (e.g., Kelley (1955), page 125). If X is complete for some metric metrizing the same topology (not for e unless X is compact), X is called a *Polish* space. For any topological space X let $\mathfrak{B}(X)$ denote the Borel σ-algebra generated by the open sets.

For any metric space (X, d), nonempty $A \subset X$, and $\varepsilon > 0$, let

$$A^{\varepsilon} := \{y \in X : d(x,y) < \varepsilon \quad \text{for some} \quad x \in A\}, \qquad \text{and}$$
$$_{\varepsilon}A := \{x \in X : y \in A \quad \text{whenever} \quad d(x,y) < \varepsilon\}.$$

The well-known *Hausdorff* pseudometric h is defined for A and B nonempty by

$$h(A, B) := h_d(A, B) := \inf\{\varepsilon > 0 : A \subset B^{\varepsilon} \quad \text{and} \quad B \subset A^{\varepsilon}\}.$$

Then h_d is finite valued iff (X, d) is bounded. It is a metric on closed sets.

Let $\mathfrak{F}_0$ be the class of all nonempty closed subsets of a separable metric space X, with a totally bounded metric e. Take on $\mathfrak{F}_0$ the Hausdorff metric h_e defined by e. Then $(\mathfrak{F}_0, h_e)$ is a separable metric space; if X is Polish, then $(\mathfrak{F}_0, h_e)$ is Polish (Effros (1965)), although it is not complete unless X is compact. The σ-algebra $\mathfrak{B}(\mathfrak{F}_0)$ of Borel subsets of $\mathfrak{F}_0$ for h_e is called the *Effros* Borel structure on $\mathfrak{F}_0$. It does not depend on the totally bounded metric e except through its topology (Effros (1965)).

Let $\mathfrak{F}$ be the class $\mathfrak{F}_0 \cup \{\phi\}$ of all closed sets in X. If we make ϕ an isolated point of $\mathfrak{F}$, then $\mathfrak{F}$ is also Polish whenever X is. If $(\mathfrak{F}_0, \mathfrak{B}(\mathfrak{F}_0))$ is a Suslin or standard measurable space, so is $(\mathfrak{F}, \mathfrak{B}(\mathfrak{F}))$. We call $\mathfrak{B}(\mathfrak{F})$ the *Effros* Borel structure on $\mathfrak{F}$. It is also generated by all sets $\{F \in \mathfrak{F} : F \subset H\}$ for $H \in \mathfrak{F}$ (Christensen (1971), (1974)) since such collections are h-closed and a countable family of them separates elements of $\mathfrak{F}$. (Note however that for U open, $\{F \in \mathfrak{F} : F \subset U\}$ need not be Effros measurable or even Suslin, e.g., if U is the open unit ball in an infinite-dimensional Hilbert space: Christensen (1971), Theorem 8.) Here a *Suslin* subset of $\mathfrak{F}$ or $\mathfrak{F}_0$ will be the image of a Polish space by a map measurable for the Effros Borel structure of $\mathfrak{F}$ or $\mathfrak{F}_0$.

For any $\mathfrak{G} \subset \mathfrak{F}$ we have the naturally induced Borel structure $\mathfrak{B}(\mathfrak{G})$.

(4.1). PROPOSITION. *For any separable metric space (X, d) and any collection $\mathfrak{G} \subset \mathfrak{F}$ of closed sets, $(\mathfrak{G}, \mathfrak{B}(\mathfrak{G}))$ is admissible and for any law P on $\mathfrak{B}(X)$, P is $\mathfrak{B}(\mathfrak{G})$-measurable.*

PROOF. The set $\epsilon_{\mathfrak{F}} := \{\langle x, F\rangle : x \in F \in \mathfrak{F}\}$ is closed in $X \times \mathfrak{F}$, using the Hausdorff metric h_e on $\mathfrak{F}$. Since $(\mathfrak{F}, h_e)$ is separable, $\epsilon_{\mathfrak{F}}$ belongs to the product σ-algebra generated by rectangles $A \times B$ where $A \in \mathfrak{B}(X)$ and $B \in \mathfrak{B}(\mathfrak{F})$. Thus $(\mathfrak{G}, \mathfrak{B}(\mathfrak{G}))$ is admissible.

For any law P on X and $c > 0$, $\{F \in \mathfrak{F} : P(F) \geqslant c\}$ is closed for h_e. Thus P is upper semicontinuous on $\mathfrak{F}$ and hence Effros measurable on $\mathfrak{F}$ or any subset $\mathfrak{G}$. □

(4.2). PROPOSITION. *For any separable metric (X, e) and law P on X, $(\mathfrak{F}, d_P)$ is separable and all open sets for d_P are Effros measurable.*

PROOF. Since $L^1(X, \mathfrak{B}(X), P)$ is separable, $(\mathfrak{F}, d_P)$ is separable, as is any subset $\mathfrak{G} \subset \mathfrak{F}$. Thus, any open set for d_P is a countable union of open balls. For any fixed closed set F, the function $C \to P(C \setminus F)$ is upper semicontinuous and hence Effros measurable. Also, $C \to P(F \setminus C)$ is lower semicontinuous and Effros measurable. Thus, $C \to P(C \Delta F) = d_P(C, F)$ is Effros measurable. Thus d_P-open balls, and d_P-open sets, are all Effros measurable. □

(4.3). PROPOSITION. *If X is Polish and $\mathcal{G}$ is any Suslin subset of $\mathcal{F}$ (for h_e), with Effros Borel structure $\mathfrak{B}(\mathcal{G})$, then $(X, \mathfrak{B}(X); \mathcal{G}, \mathfrak{B}(\mathcal{G}))$ is $P\epsilon$-Suslin for any law P on $\mathfrak{B}(X)$.*

PROOF. This follows from Propositions 4.1 and 4.2.

Let $\mathcal{U}$ be the class of all open sets in X. On $\mathcal{U} \setminus \{X\}$ we have the metric $h_{(e)}$ defined as the Hausdorff metric of complements:

$$h_{(e)}(U, V) := h_e(X \setminus U, X \setminus V).$$

In the notation defined at the beginning of this section, we have ${}_\varepsilon V \subset U$ iff $(X \setminus U) \subset (X \setminus V)^\varepsilon$. Thus

$$h_{(e)}(U, V) = \inf\{\varepsilon > 0 : {}_\varepsilon U \subset V \quad\text{and}\quad {}_\varepsilon V \subset U\}.$$

Now $h_{(e)}$ has various properties which follow from those of h_e on $\mathcal{F}_0$. For a totally bounded metric e on X, we will call the σ-algebra $\mathfrak{B}(\mathcal{U} \setminus \{X\})$ of Borel sets for $h_{(e)}$ in $\mathcal{U} \setminus \{X\}$ the *Effros* Borel structure, and likewise for the Borel structure induced on $\mathcal{U}$ by making X an isolated point. Then we have an induced measurable structure on any subset $\mathcal{V} \subset \mathcal{U}$. We conclude from Proposition 4.3:

(4.4). PROPOSITION. *If X is a Polish space, and $\mathcal{V}$ is any Suslin subset of $\mathcal{U}$ for $h_{(e)}$, with Effros Borel structure $\mathfrak{B}(\mathcal{V})$, then $(X, \mathfrak{B}(X); \mathcal{V}, \mathfrak{B}(\mathcal{V}))$ is $P\epsilon$-Suslin for any law P on $\mathfrak{B}(X)$.*

Let X be a set and G a collection of real-valued functions on X. Let $\text{pos}(G)$ be the collection of all sets

$$\text{pos}(g) := \{x \in X : g(x) > 0\}, g \in G.$$

Recall that by Lindelöf's theorem any locally compact separable metric space is a countable union of compact sets.

(4.5). PROPOSITION. *Let (X, e) be a locally compact, separable metric space and G a collection of continuous real functions on X. Suppose we are given a σ-algebra $\mathfrak{S}$ of subsets of G such that for each $x \in X$, $g \to g(x)$ is $\mathfrak{S}$-measurable. Then the map $g \to \text{pos}(g)$ is measurable from $(G, \mathfrak{S})$ to $\mathcal{U}$ with Effros Borel structure.*

PROOF. It suffices to show that for any open $V \subset X$, $\{g \in G : V \subset \text{pos}(g)\} \in \mathfrak{S}$. Let $V = \bigcup_{n \geqslant 1} K_n$ with K_n compact. For each n let A_n be a countable dense set in K_n. For any $\varepsilon > 0$, let

$$G(n, \varepsilon) := \{g \in G : g(x) \geqslant \varepsilon \quad\text{for all}\quad x \in A_n\} \in \mathfrak{S}.$$

Now $V \subset \text{pos}(g)$ iff $g \in \bigcap_{n \geqslant 1} \bigcup_{m \geqslant 1} G(n, 1/m)$. □

Under the conditions of Proposition 4.5, if $(G, \mathfrak{S})$ is a Suslin measurable space, then $\text{pos}(G)$ is a Suslin set, and by Proposition 4.4, $(X, \mathfrak{B}(X); \text{pos}(G), \mathfrak{B}(\text{pos}(G)))$ is $P\epsilon$-Suslin.

5. Metric entropy and inclusion.

DEFINITION. Let $(X, \mathcal{Q}, P)$ be a probability space and $\mathcal{C} \subset \mathcal{Q}$. For each $\varepsilon > 0$ let $N_I(\varepsilon) := N_I(\varepsilon, \mathcal{C}, P)$ be the smallest n such that for some $A_1, \cdots, A_n \in \mathcal{Q}$ (not necessarily in $\mathcal{C}$), for every $A \in \mathcal{C}$ there exist i, j with $A_i \subset A \subset A_j$ and $P(A_j \setminus A_i) < \varepsilon$.

Let $N(\varepsilon) := N(\varepsilon, \mathcal{C}, P)$ be the smallest n such that $\mathcal{C} = \bigcup_{1 \leqslant j \leqslant n} \mathcal{C}_j$ for some sets $\mathcal{C}_j$ with $\sup\{d_P(A, B) : A, B \in \mathcal{C}_j\} \leqslant 2\varepsilon$ for each j. Log $N(\varepsilon)$ is called a *metric entropy* and log $N_I(\varepsilon)$ will be called a *metric entropy with inclusion*.

Dehardt (1971) proved in effect that if $N_I(\varepsilon, \mathcal{C}, P) < \infty$ for all $\varepsilon > 0$, and $\mathcal{C}$ is P-EM, then

$$\Pr\{\lim_{n\to\infty} \sup_{A\in\mathcal{C}} |(P_n - P)(A)| = 0\} = 1.$$

We recall that $\mathcal{C}$ is a G_PBUC class if $\int_0^1 (\log N(x^2))^{\frac{1}{2}}\, dx < \infty$ (Dudley (1967b); (1973), page 71). (Note that the L^2 norm of 1_A is $P(A)^{\frac{1}{2}}$, hence the x^2.) If $\mathcal{C}$ is the collection of all finite subsets of [0, 1] and P is Lebesgue measure, $\mathcal{C}$ is not a Donsker class although $d_P(A, B) = 0$ for all $A, B \in \mathcal{C}$. So hypotheses on $N(\varepsilon)$ will not imply the Donsker property. The following sufficient condition on $N_I(\varepsilon)$ is the same as the above condition on $N(\varepsilon)$ for the G_PBUC property. Note however that if $P(A_m) = m^{-\frac{1}{2}}$ with A_m independent for P, then $\mathcal{C} := \{A_m\}_{m\geqslant 1}$ is Donsker by Theorem 2.1, while $N_I(\frac{1}{2}, \mathcal{C}, P) = +\infty$.

(5.1). THEOREM. *If $\mathcal{C}$ is P-EM and $\int_0^1 (\log N_I(x^2))^{\frac{1}{2}}\, dx < \infty$ then $\mathcal{C}$ is a Donsker class.*

PROOF. Since $\mathcal{C}$ is totally bounded for d_P, it is enough to verify Theorem 1.2b.

Let $0 < \varepsilon < 1$. Since $N_I(x)\uparrow$ as $x\downarrow 0$, the hypothesis implies $x \log N_I(x) \to 0$. So there is a $\gamma > 0$ such that

$$N_I(x) \leqslant \exp(\varepsilon^2/(600x)), \qquad 0 < x \leqslant \gamma. \tag{5.2}$$

Take $\alpha > 0$ small enough so that

$$\exp(-\varepsilon^2/(1800\alpha)) < \varepsilon/4. \tag{5.3}$$

The hypothesis on N_I is equivalent to

$$\int_0^1 (\log N_I(y))^{\frac{1}{2}} y^{-\frac{1}{2}}\, dy < \infty$$

and to

$$\Sigma_{i\geqslant 1}(2^{-i} \log N_I(2^{-i}))^{\frac{1}{2}} < \infty.$$

Take u large enough so that

$$\Sigma_{i\geqslant u}(2^{-i} \log N_I(2^{-i}))^{\frac{1}{2}} < \varepsilon/64, \tag{5.4}$$

and such that

$$\Sigma_{i\geqslant 0} \exp(-2^{i+u}\varepsilon^2/(9000(i+1)^4)) < \varepsilon/32. \tag{5.5}$$

Let $\delta_0 := 2^{-r}$ for $r \geqslant u$ and r large enough so that $\delta_0 \leqslant \min(\alpha, \gamma)$. For $k = 0, 1, 2, \cdots$, let $\delta_k := \delta(k) := \delta_0/2^k = 1/2^{k+r}$. Let $m(k) := N_I(\delta_k, \mathcal{C}, P)$ and $b_k := (2^{-k} \log m(k))^{\frac{1}{2}}$.

Take sets $A_{k1}, \cdots, A_{km(k)}$ as in the definition of $N_I(\delta_k, \mathcal{C}, P)$, so that for each $C \in \mathcal{C}$ and $k = 0, 1, 2, \cdots$, there are $r(k) := r(k, C)$ and $s(k) := s(k, C)$ with $A_{kr(k)} \subset C \subset A_{ks(k)}$ and $P(A_{ks(k)} \setminus A_{kr(k)}) < \delta_k$. Let $B_k := B_k(C) := A_{ks(k)} \setminus A_{k+1, s(k+1)}$ and $D_k := D_k(C) := A_{k+1, s(k+1)} \setminus A_{ks(k)}$. Then $P(B_k) < \delta_k$ and $P(D_k) < \delta_{k+1} < \delta_k$.

Let $n_0 := n_0(\varepsilon) := \varepsilon^2/(256\delta_0^2)$. (Note that $\delta_0 \leqslant \alpha < \varepsilon^2/1800$, so that $n_0 > 12{,}000/\varepsilon^2 \to \infty$ as $\varepsilon \downarrow 0$.) For each $n > n_0$ there is a unique $k = k(n)$ such that

$$\tfrac{1}{2} < 8\delta_k n^{\frac{1}{2}}/\varepsilon \leqslant 1. \tag{5.6}$$

Then for each n, $k = k(n)$, $\delta = \delta_k$, each $C \in \mathcal{C}$, $r = r(k, C)$, and $s = s(k, C)$, we have

$$\nu_n(A_{kr}) - \varepsilon/8 \leqslant \nu_n(A_{kr}) - \delta n^{\frac{1}{2}} \leqslant \nu_n(C) \leqslant \nu_n(A_{ks}) + \varepsilon/8. \tag{5.7}$$

We have

$$\begin{aligned} &|\nu_n(A_{ks(k)}) - \nu_n(A_{0s(0)})| \\ &\leqslant \Sigma_{0 \leqslant i < k} |\nu_n(A_{is(i)}) - \nu_n(A_{i+1, s(i+1)})| \\ &\leqslant \Sigma_{0 \leqslant i < k} |\nu_n(B_i)| + |\nu_n(D_i)|. \end{aligned} \tag{5.8}$$

Let β_i be the collection of all sets $B = A_{is} \setminus A_{i+1, t}$ or $A_{i+1, t} \setminus A_{is}$ with $P(B) < \delta_i$. Then for each $C \in \mathcal{C}$, $B_i(C)$ and $D_i(C) \in \beta_i$. The number of sets in β_i is bounded by

$$\operatorname{card}(\beta_i) \leqslant 2m(i)m(i+1). \tag{5.9}$$

We have $\log m(i) = 2^i b_i^2$. Let

$$d_i := \max\big((i+1)^{-2}\varepsilon/32,\ 4b_{i+1}2^{-r/2}\big).$$

Then by (5.4),

$$\Sigma_{i \geqslant 0} d_i < \varepsilon/8. \tag{5.10}$$

For each $i < k = k(n)$, by (5.6), $n^{\frac{1}{2}}\delta_i > n^{\frac{1}{2}}\delta_k > \varepsilon/16$. Thus by (5.10), $d_i \leqslant 2n^{\frac{1}{2}}\delta_i$. Bernstein's inequality (Bennett (1962)) gives for each $B \in \beta_i$

$$\begin{aligned} P_{inB} &:= \Pr\{|\nu_n(B)| > d_i\} \\ &\leqslant 2\exp\Big(-d_i^2\big/\big(2pq + d_i n^{-\frac{1}{2}}\big)\Big), \end{aligned}$$

where $1 - q := p := P(B) < \delta_i$.

Thus $P_{inB} \leqslant 2\exp(-d_i^2/(4\delta_i))$. Let $M_i := 4m(i)m(i+1) \leqslant 4m(i+1)^2 = 4\exp(2^{i+1}b_{i+1}^2)$. Then using (5.9) we have

$$\begin{aligned} P_{in} &:= \Pr\{|\nu_n(B)| > d_i \quad \text{for some} \quad B \in \beta_i\} \\ &\leqslant M_i \exp\big(-d_i^2/(4\delta_i)\big) = M_i \exp\big(-2^i d_i^2/(4\delta_0)\big) \\ &\leqslant 4\exp\big(2^i\big(2b_{i+1}^2 - d_i^2/(4\delta_0)\big)\big). \end{aligned}$$

Now by definition of d_i, $2b_{i+1}^2 \leqslant d_i^2/(8\delta_0)$, and

$$P_{in} \leqslant 4\exp\big(-2^i d_i^2/(8\delta_0)\big)$$
$$\leqslant 4\exp\big(-2^{i+r}\varepsilon^2/\big(8(32)^2(i+1)^4\big)\big).$$

Thus by (5.5),

$$\Sigma_{0\leqslant i\leqslant k} P_{in} < \varepsilon/8. \tag{5.11}$$

Now with $k := k(n)$, $\delta := \delta_k$, let

$$V_n := \sup\{|\nu_n(A_{kr}) - \nu_n(A_{ks})| : A_{kr} \subset A_{ks}, P(A_{ks}\setminus A_{kr}) < \delta, r, s = 1, \cdots, m(k)\}.$$

Let $Q_n := \Pr(V_n > \epsilon/8)$. Then by Bernstein's inequality, and (5.6),

$$Q_n \leqslant m(k)^2 \exp\Big(-\varepsilon^2 64^{-1}/\big(2\delta_k + \varepsilon 8^{-1} n^{-\frac{1}{2}}\big)\Big)$$
$$\leqslant m(k)^2 \exp\big(-\varepsilon^2/(128\delta_k + 128\delta_k)\big)$$
$$\leqslant \exp\big(2^k\big(2b_k^2 - \varepsilon^2/(256\delta_0)\big)\big).$$

Now for $j := k + r$,

$$2b_k^2 = 2^{1-k}\log N_I(2^{-j})$$
$$= 2^{r+1}\big(2^{-j}\log N_I(2^{-j})\big)$$
$$\leqslant 2^r\varepsilon^2/300 \quad \text{by (5.2)}.$$

Thus

$$Q_n \leqslant \exp(-2^{k+r}\varepsilon^2/1800)$$
$$\leqslant \exp(-2^r\varepsilon^2/1800) < \varepsilon/4$$

by (5.3) and choice of r.

If $V_n < \varepsilon/8$ then by (5.7), $|\nu_n(C) - \nu_n(A_{k, s(k, C)})| \leqslant \varepsilon/4$ for all $C \in \mathcal{C}$. Then by (5.8), (5.10) and (5.11),

$$\Pr{}^*\{\sup_{C\in\mathcal{C}}|\nu_n(C) - \nu_n(A_{0s(0, C)})| > \varepsilon/2\} < \varepsilon/2.$$

We also have

$$P_0 := \Pr\big\{\sup\big\{|\nu_n(A_{0i}) - \nu_n(A_{0j})| : P(A_{0i}\Delta A_{0j}) < 3\delta_0\big\} > \varepsilon/4\big\}$$
$$\leqslant m(0)^2 \exp\Big(-\varepsilon^2 16^{-1}/\big(6\delta_0 + \varepsilon 4^{-1} n^{-\frac{1}{2}}\big)\Big).$$

For $n \geqslant n_0$, as in (5.6), $n^{-\frac{1}{2}}\varepsilon/4 \leqslant 4\delta_0$, so

$$P_0 \leqslant m(0)^2 \exp\big(-\varepsilon^2/(160\delta_0)\big).$$

Now by (5.2) and choice of δ_0, we have $m(0)^2 \leqslant \exp(2\varepsilon^2/(600\delta_0))$, so

$$P_0 \leqslant \exp\big(-\varepsilon^2/(250\delta_0)\big) < \varepsilon/4$$

by (5.3) since $\delta_0 \leqslant \alpha$. Thus, for $n \geqslant n_0$,

$$\Pr{}^*\{\sup\{|\nu_n(C) - \nu_n(D)| : C, D \in \mathcal{C}, P(C\Delta D) < \delta_0\} > \varepsilon\} < \varepsilon,$$

proving (1.2b) in this case. □

NOTE. If $\mathcal{C}$ is, e.g., the collection of all half-planes in $\mathbb{R}^2$, the sets A_{kj} cannot be chosen in $\mathcal{C}$. In this case, for suitable P, the A_{kj} can be chosen as intersections of two half-planes.

We recall some definitions from Dudley (1974). Given a continuous function f from the sphere $S^{k-1} := \{x \in \mathbb{R}^k : |x| = 1\}$ into $\mathbb{R}^k$, let $I(f)$ denote the open set of all $y \in \mathbb{R}^k \setminus$ range (f) such that in $\mathbb{R}^k \setminus \{y\}$, f is not homotopic to a constant map. For $\alpha > 0$ and $M > 0$ let $G(k, \alpha, M)$ be the set of all f which with all their partial derivatives of orders $\leqslant \alpha$ are bounded in norm by M (as in Dudley (1974), page 229).

Let $J(f) := I(f) \cup$ range (f). Let $I(k, \alpha, M) := \{I(f) : f \in G(k, \alpha, M)\}$, $J(k, \alpha, M) := \{J(f) : f \in G(k, \alpha, M)\}$. (To correct equation 3.2 in Dudley (1974), put $J(k, \alpha, M)$ in place of $I(k, \alpha, M)$.)

If $|f(\theta) - g(\theta)| < \varepsilon$ for all $\theta \in S^{k-1}$, then ${}_\varepsilon I(f) \subset I(g) \subset J(g) \subset J(f)^\varepsilon$ by Lemma 2.1 of Dudley (1974). Also, $J(f)^\varepsilon \setminus {}_\varepsilon I(f) \subset (\text{range } f)^\varepsilon$.

I understand from R. Pyke that Sun (1976) has found a theorem along the following lines. At this writing I have not seen his precise statements or proofs.

(5.12). THEOREM (Sun). *For any law P on $\mathbb{R}^k$, $k \geqslant 2$, which has a bounded density with respect to Lebesgue measure λ, any $M < +\infty$ and $\alpha > k - 1$, $I(k, \alpha, M)$ and $J(k, \alpha, M)$ are Donsker classes for P.*

PROOF. Since $\alpha > k - 1 \geqslant 1$, the proof of Theorem 3.1 of Dudley (1974) shows that for some $N = N(k, \alpha, M) < +\infty$, $\lambda(J(f)^\varepsilon \setminus {}_\varepsilon I(f)) \leqslant N\varepsilon$ for all $f \in G(k, \alpha, M)$. Let $T := N\, \text{ess.sup}(dP/d\lambda)$.

If $f_1, \cdots, f_r \in G(k, \alpha, M)$ are such that for each $f \in G(k, \alpha, M)$, $\sup\{|f - f_j|(\theta) : \theta \in S^{k-1}\} \leqslant \varepsilon/T$ for some j, let $A_j := {}_{\varepsilon/T}I(f_j)$, $A_{j+r} := J(f_j)^{\varepsilon/T}$, $j = 1, \cdots, r$, to obtain $N_I(\varepsilon, \mathcal{C}, P) \leqslant 2r$, for $\mathcal{C} = I(k, \alpha, M)$ or $J(k, \alpha, M)$.

Thus by Theorem 3 of Clements (1963) as in the proof of Theorem 3.1 of Dudley (1974) the hypothesis on N_I in Theorem 5.1 above holds.

Let d be the usual Euclidean distance on $\mathbb{R}^k$. For the distance $s(f, g) := \sup_\theta d(f(\theta), g(\theta))$, $G(k, \alpha, M)$ is a compact set of functions, for any $\alpha > 0$, by the Arzelà-Ascoli theorem. The map $g \to I(g)$ is continuous from $G(k, \alpha, M)$ onto $I(k, \alpha, M)$ with the metric $h_{(d)}$ as defined after 4.3 above, since if $s(g_m, g) \to 0$ and $\varepsilon > 0$, then for m large, $I(g_m) \supset {}_\varepsilon I(g)$ and ${}_\varepsilon I(g_m) \subset I(g)$ as in Dudley ((1974), Lemma 2.1, proof of Theorem 3.1). (Note however that $g \to I(g)$ is not continuous for the Hausdorff metric h_d.)

So, $I(k, \alpha, M)$ is compact for $h_{(d)}$ and hence absolutely Borel. Note that all sets in $I(k, \alpha, M)$ are included in a fixed compact set, so that there is no need to use a totally bounded metric on all of $\mathbb{R}^k$ to obtain the Effros Borel structure. Then by

Propositions 4.4 and 3.2 above, $I(k, \alpha, M)$ is strongly P-EM. Then by 5.1 it is a Donsker class.

Likewise, $g \to J(g)$ is continuous from $G(k, \alpha, M)$ onto $J(k, \alpha, M)$ with the Hausdorff metric h_d, so $J(k, \alpha, M)$ is compact and absolutely Borel. By Propositions 4.3 and 3.2, $J(k, \alpha, M)$ is strongly P-EM and by 5.1 a Donsker class. □

In the unit cube of $\mathbb{R}^k$ let $R(k, \alpha, M)$ be the class of sets defined by Révész (1976), of the form

$$\{x : f_i(\{x_j\}_{j \neq i}) < x_i < g_i(\{x_j\}_{j \neq i}), i = 1, \cdots, k\},$$

where f_i and g_i, each defined on the unit cube of $\mathbb{R}^{k-1}$, have partial derivatives of all orders $\leqslant \alpha$ bounded by M in absolute value. Aside from differences in the values of M, the classes $\bigcup_M R(k, \alpha, M) := R(k, \alpha)$ and $\bigcup_M I(k, \alpha, M) := I(k, \alpha)$ are different: for $k = 2$, $I(2, 1)$ contains a set bounded by a "figure 8" which is not in $R(2, 1)$. I do not know whether $R(k, \alpha) \subset I(k, \alpha)$. For P with bounded density, one can show as in (5.2) that $R(k, \alpha, M)$ is a Donsker class if $\alpha > k - 1$.

For any set $U \subset \mathbb{R}^k$ let $cnv(U)$ denote the class of all open convex subsets of U. The following result, in the case of the uniform Lebesgue probability on the square I^2, is due to Bolthausen (1976).

(5.13). THEOREM (Bolthausen). *For any law P on $\mathbb{R}^2$ having a bounded density with respect to Lebesgue measure, and any bounded convex open U, $cnv(U)$ is a Donsker class for P.*

PROOF. We apply Theorem 4.1 of Dudley (1974) and its proof, and Theorem 5.1 above and its proof, as in Theorem 5.12. To see that $cnv(U)$ is P-EM we apply 4.4 as in the proof of Theorem 5.12, noting that $cnv(U)$ is compact for $h_{(e)}$. Thus 5.13 is proved.

In $\mathbb{R}^3$, a collection $cnv(U)$ is not $G_P BUC$ (Dudley (1973), page 87, Remark), and a fortiori not Donsker.

6. The two-sample case. We will call $\mathcal{C}$ *2-sample* P-EM iff for any independent empirical measures P_m and Q_n for P, $P_m - Q_n$ is completion measurable into $(D_0(\mathcal{C}, P), \mathfrak{B}_b)$. Note that if $\mathcal{C}$ is strongly P-EM, it is 2-sample P-EM (let $b_1 = \cdots = b_m = 1/m$, $b_{m+1} = \cdots = b_{m+n} = -1/n$). Thus Propositions 3.2, 4.3 and 4.4 give conditions which imply that Suslin classes of closed or open sets are 2-sample P-EM.

(6.1). THEOREM. *Let $(X, \mathcal{A}, P)$ be a probability space and $\mathcal{C}$ a Donsker class, $\mathcal{C} \subset \mathcal{A}$, with $\mathcal{C}$ 2-sample P-EM. Let $\{P_m\}$ and $\{Q_n\}$ be independent empirical measures for P. Then as m and $n \to \infty$,*

$$\mathcal{L}\big((mn)^{\frac{1}{2}}(m+n)^{-\frac{1}{2}}(P_m - Q_n)\big) \to \mathcal{L}(G_P) \quad \text{in} \quad D_0(\mathcal{C}, P).$$

PROOF. Since $\mathcal{C}$ is a Donsker class, it is $G_P BUC$, and $\mathcal{L}(G_P)(T) = 1$ for some separable set $T \subset D_0(\mathcal{C}, P)$. According to Wichura ((1970), Theorem 2), there is a probability space $(\Omega, \mathcal{S}, \mu)$ with random variables A_m and B_n such that the sequence $\{A_m\}$ is independent of $\{B_n\}$, A_m and B_n are measurable into $(D_0(\mathcal{C}, P), \mathcal{B}_b)$, $\mathcal{L}(A_n) = \mathcal{L}(B_n) = \mathcal{L}(n^{\frac{1}{2}}(P_n - P))$ for all n, $A_m \to H$ a.s. as $m \to \infty$ and $B_n \to J$ a.s. as $n \to \infty$ in $(D_0(\mathcal{C}, P), \|\cdot\|_\infty)$ where $\mathcal{L}(H) = \mathcal{L}(J) = \mathcal{L}(G_P)$, and H and J are independent.

As noted in the discussion of almost uniform convergence in the introduction above, since $\mu(H \in T) = 1$, $\sup_{C \in \mathcal{C}}|A_m - H|(C)$ is actually a measurable random variable, converging to 0 a.s. as $m \to \infty$, and likewise for $|B_n - J|$. Let

$$D_{mn} := (m+n)^{-\frac{1}{2}}\left(n^{\frac{1}{2}}A_m - m^{\frac{1}{2}}B_n\right).$$

Then

$$\mathcal{L}\left((mn)^{\frac{1}{2}}(m+n)^{-\frac{1}{2}}(P_m - Q_n)\right) = \mathcal{L}\left((mn)^{\frac{1}{2}}(m+n)^{-\frac{1}{2}}\left(m^{-\frac{1}{2}}A_m + P - n^{-\frac{1}{2}}B_n - P\right)\right) = \mathcal{L}(D_{mn}).$$

Let

$$E_{mn} := (m+n)^{-\frac{1}{2}}\left(n^{\frac{1}{2}}H - m^{\frac{1}{2}}J\right), \quad \text{and}$$

$$F_{mn} := (m+n)^{-\frac{1}{2}}\left(n^{\frac{1}{2}}(A_m - H) - m^{\frac{1}{2}}(B_n - J)\right).$$

Then $D_{mn} = E_{mn} + F_{mn}$, $\mathcal{L}(E_{mn}) = \mathcal{L}(G_P)$ for all m and n, and $F_{mn} \to 0$, uniformly on $\mathcal{C}$ almost surely, and almost uniformly in $l^\infty(\mathcal{C})$, as m and $n \to \infty$. (I do not claim that F_{mn} is measurable.) Now Theorem 6.1 will be a consequence of the following, letting $\alpha = \langle m, n \rangle$, $Y_\alpha = D_{mn}$, and $Z_\alpha = E_{mn}$.

(6.2). LEMMA. *Let (S, d) be any metric space and Y_α, Z_α nets of random variables, measurable into $(S, \mathcal{B}_b)$, with $d(Y_\alpha, Z_\alpha) \to 0$ almost uniformly (we do not assume $d(Y_\alpha, Z_\alpha)$ is measurable) and $\mathcal{L}(Z_\alpha) \to \mu$ as $\alpha \to \infty$, where $\mu(T) = 1$ for some separable $T \subset S$. Then $\mathcal{L}(Y_\alpha) \to \mu$.*

PROOF. Let f be continuous and $\mathcal{B}_b$-measurable on S with $\sup|f| \leqslant 1$. Given $\varepsilon > 0$, for each $x \in T$ take $\delta_x := \delta(x) > 0$ such that whenever $d(x, u) < 3\delta_x$, we have $|f(x) - f(u)| < \varepsilon$. By Lindelöf's theorem, the open cover $\{B(x, \delta_x)\}_{x \in T}$ of T has a countable subcover. So there is a finite set $K \subset T$ such that $\mu(U(K)) > 1 - \varepsilon$ where $U(K) := \bigcup_{x \in K} B(x, \delta_x)$. Let $g(u) := \max(0, \min(1, 2 - \min_{x \in K} d(x, u)/\delta_x))$. Then $1_{U(K)} \leqslant g \leqslant 1_{V(K)}$ where $V(K) := \bigcup_{x \in K} B(x, 2\delta_x)$, and g is continuous and $\mathcal{B}_b$ measurable. Thus $\lim_\alpha Eg(Z_\alpha) = \int g\, d\mu > 1 - \varepsilon$, so for some β,

$$\Pr(Z_\alpha \in V(K)) \geqslant Eg(Z_\alpha) > 1 - \varepsilon$$

for $\alpha \geqslant \beta$. Let $\delta := \min_{x \in K}\delta_x$. For some $\gamma \geqslant \beta$, we have $|Ef(Z_\alpha) - \int f\, d\mu| < \varepsilon$ and $\Pr^*(d(Y_\alpha, Z_\alpha) \geqslant \delta) < \varepsilon$ for $\alpha \geqslant \gamma$. Then, except on some event W with $\Pr(W) < 2\varepsilon$, we have $Z_\alpha \in V(K)$ and $d(Y_\alpha, Z_\alpha) < \delta$, so that for some $x \in K$, $d(Z_\alpha, x) <$

$2\delta_x$, $d(Y_\alpha, x) < 3\delta_x$, and $|f(Y_\alpha) - f(Z_\alpha)| < 2\varepsilon$. Thus $|Ef(Y_\alpha) - Ef(Z_\alpha)| < 6\varepsilon$, so $|Ef(Y_\alpha) - \int f\,d\mu| < 7\varepsilon$. □

If $X = \mathbb{R}^1$ and $\mathcal{C}$ is the class of intervals, $\sup_{A \in \mathcal{C}} G_P(A)$ has the same law for all continuous P. Likewise, $\mathcal{L}(\sup_{A \in \mathcal{C}} |G_P(A)|)$ is the same for all P without atoms. Then, results like Theorem 6.1 apply to testing whether two unknown continuous distribution functions, from which finite samples have been taken, are equal. If $X = \mathbb{R}^k$ for $k > 1$, such laws depend on P, but perhaps they are the same for more restricted classes of laws P.

Limit theorems in the literature for the two-sample case have often been stated under restrictive conditions such as m/n converging to a positive constant. Theorem 6.1 shows that no such restriction is necessary. The proof will give rates of convergence in the two-sample case if one has them (in a suitable form) in the one-sample case.

7. Universal Donsker classes and Vapnik-Červonenkis classes. Given a set X, a collection $\mathcal{C}$ of subsets of X will be called a *universal Donsker class* (UDC) iff it is a Donsker class for every probability measure on the σ-algebra generated by $\mathcal{C}$.

EXAMPLES. If $X = \mathbb{R}^k$, the class of all rectangles $\Pi_{1 \leqslant j \leqslant k}]a_j, b_j]$ is a UDC (Donsker (1952), for $k = 1$; Dudley (1966), for $k > 1$).

DEFINITIONS. Given a class $\mathcal{C}$ of subsets of a set X and a finite set $F \subset X$, let $\Delta^{\mathcal{C}}(F)$ be the number of different sets $C \cap F$ for $C \in \mathcal{C}$. For $n = 1, 2, \cdots$, let $m^{\mathcal{C}}(n) := \max\{\Delta^{\mathcal{C}}(F) : F \text{ has } n \text{ elements}\}$. Let

$$\begin{aligned} V(\mathcal{C}) &:= \inf\{n : m^{\mathcal{C}}(n) < 2^n\} \\ &= +\infty \quad \text{if} \quad m^{\mathcal{C}}(n) = 2^n \ \forall n. \end{aligned}$$

Vapnik and Červonenkis (1971) introduced $\Delta^{\mathcal{C}}$, $m^{\mathcal{C}}$ and $V(\mathcal{C})$. If $m^{\mathcal{C}}(n) < 2^n$ for some n, i.e., if $V(\mathcal{C}) < +\infty$, we will call $\mathcal{C}$ a *Vapnik-Červonenkis class* (VCC).

Here is the main result of this section:

(7.1). THEOREM. *If $\mathcal{C}$ is a* VCC *and for some σ-algebras $\mathcal{Q} \supset \mathcal{C}$ in X and $\mathcal{S}$ in $\mathcal{C}$, $(X, \mathcal{Q}; \mathcal{C}, \mathcal{S})$ is $P\epsilon$-Suslin, then $\mathcal{C}$ is a Donsker class for P.*

Before proving Theorem 7.1 we will go through some other facts. First, recalling Proposition 4.5, here is one way to generate VCC's. It is related to results in Cover (1965).

(7.2). THEOREM. *Let G be any m-dimensional real vector space of real functions on a set X. Then $V(\text{pos}(G)) = m + 1$ (or, if $\text{card}(X) = m$, $\text{pos}(G)$ contains all subsets of X).*

PROOF. First, suppose $A \subset X$ and $\text{card}(A) = m + 1$. Take the map $r : G \to \mathbb{R}^A$ which restricts functions in G to the set A. Then r, as a linear map of an m-dimensional real vector space into one of higher dimension, cannot be onto. For the usual inner product $(\cdot\,,\,\cdot)$ on $\mathbb{R}^A$, let $v \neq 0$ be a vector orthogonal to $r(G)$. We may assume $A_+ := \{x \in A : v(x) > 0\}$ is nonempty, since otherwise we can take $-v$ (thanks to M. Artin for this remark). If $A_+ \in \text{pos}(G)$, take $f \in G$ with $\{f > 0\} = A_+$. Then $(r(f), v) > 0$, a contradiction. So $A_+ \notin \text{pos}(G)$ and $V(\text{pos}(G)) \leqslant m + 1$.

On the other hand, $\dim(G) = m$ implies that for some subset A of X with $\text{card}(A) = m$, $r(G) = \mathbb{R}^A$, so all subsets of A are of the form $B \cap A$, $B \in \text{pos}(G)$, and $m < V(\text{pos}(G)) = m + 1$. □

One example of a finite-dimensional G on $X = \mathbb{R}^k$ is the collection of all polynomials of degree $\leqslant d$ for any fixed $d < \infty$.

Vapnik and Červonenkis (1971) proved inequalities relating $m^{\mathcal{C}}(n)$ to its values for the special case of half-spaces of $\mathbb{R}^k$, as follows. Let

(7.3) $${}_NC_{\leqslant k} := \Sigma_{j=0}^{k}\binom{N}{j}, \quad \text{where} \quad \binom{N}{j} := 0 \quad \text{for } j > N.$$

Recalling the notation ${}_NC_k := \binom{N}{k}$ for the number of k-element subsets of an N-element set, ${}_NC_{\leqslant k}$ is the number of subsets with *at most* k elements.

By *r-flat* I will mean a linear variety of $\mathbb{R}^k$ of dimension r, i.e., a set of the form $\{x \in \mathbb{R}^k : A(x - v) = 0\}$ where $v \in \mathbb{R}$ and A is a $k \times k$ matrix of rank $k - r$. A $(k-1)$-flat in $\mathbb{R}^k$ will be called a *hyperplane*.

Let $\mathcal{H}$ be the collection of all half-spaces $\{x : (v, x) > c\}$ for $x, v \in \mathbb{R}^k$, $v \neq 0$, and $c \in \mathbb{R}$. Let $\mathcal{H}(0)$ be the subcollection of half-spaces bounded by hyperplanes through 0 (i.e., with $c = 0$). Let $\eta_k(N)$ be the maximum number of open regions into which $\mathbb{R}^k$ is decomposed by N hyperplanes $H_1, \cdots, H_N$ of H_N. Then the maximum is attained for $H_1, \cdots, H_N$ in "general position," i.e., if

$$H_j = \{x : (x, v_j) = c_j\}, \qquad j = 1, \cdots, N,$$

then any k or fewer of the v_j are linearly independent. Schläfli (1901) shows that

(7.4) $$\eta_k(N) = {}_NC_{\leqslant k}.$$

Steiner (1826) had proved this for $k \leqslant 3$. If F is a set of N points in $\mathbb{R}^k$, then

(7.5) $$\Delta^{\mathcal{H}}(F) \leqslant 2\Sigma_{r=0}^{k}\binom{N-1}{r} = 2{}_{N-1}C_{\leqslant k},$$

and equality is attained if the points of F are in general position, i.e., no $k + 1$ of them are in any hyperplane (Cover (1965), page 330; Harding (1967); Watson (1969)). One also has

(7.6) $$\Delta^{\mathcal{H}(0)}(F) \leqslant 2{}_{N-1}C_{\leqslant k-1}$$

with equality iff every nonempty subset of F with at most k elements is linearly independent (Schläfli (1901), page 211; Cover (1965), noting varying definitions of "general position" on page 230; Harding (1967)). More generally, if for a fixed j-flat A, $\mathcal{H}(j)$ is the set of all half-spaces bounded by hyperplanes including A, then

(7.7) $$\Delta^{\mathcal{H}(j)}(F) \leqslant 2{}_{N-1}C_{\leqslant k-j-1}$$

(Harding (1967)).

Without using (7.4)–(7.7), but directly from the definition (7.3) and the recurrence relation

(7.8) $${}_{N}C_{\leqslant k} = {}_{N-1}C_{\leqslant k} + {}_{N-1}C_{\leqslant k-1},$$

Vapnik and Červonenkis ((1971), Lemma 1) prove:

(7.9). THEOREM (Vapnik-Červonenkis). *If X is any set, $\mathcal{C}$ any collection of subsets of X, and $V(\mathcal{C}) \leqslant v$, then $m^{\mathcal{C}}(n) < {}_{N}C_{\leqslant v}$ for all $n \geqslant v$.*

They note that ${}_{n}C_{\leqslant k} \leqslant n^{k} + 1$. (Their 1974 book, pages 214–219, shows that $m^{\mathcal{C}}(n) \leqslant {}_{n}C_{\leqslant V(\mathcal{C})-1}$. Note: in the 1971 paper and the 1974 book, pages 97 and 214, are three disagreeing definitions of "$\Phi(k, n)$.") They prove that for $n > k \geqslant 1$, ${}_{n}C_{\leqslant k} \leqslant 1.5n^{k}/k!$. Hence

(7.10) $$\text{for} \quad n > v := V(\mathcal{C}) \geqslant 1,$$
$$m^{\mathcal{C}}(n) \leqslant 1.5n^{v-1}/(v-1)! < n^{v}.$$

For $n < v$, $m^{\mathcal{C}}(n) = 2^{n} \leqslant 2^{v} \leqslant n^{v}$. If $v = 0$, $\mathcal{C}$ is empty. Thus (without using (7.10)) we have:

(7.11) For any collection $\mathcal{C}$ of sets, $m^{\mathcal{C}}(n) \leqslant n^{V(\mathcal{C})}$ for all $n \geqslant 2$, and $m^{\mathcal{C}}(n) \leqslant n^{V(\mathcal{C})} + 1$ for all $n \geqslant 0$.

Now for any sets $A_1, \cdots, A_m$, let $\mathcal{Q}(A_1, \cdots, A_m)$ denote the algebra of subsets of X generated by $A_1, \cdots, A_m$.

(7.12). PROPOSITION. *For any* VCC $\mathcal{C}$ *and any* $k < +\infty$,

$$\mathcal{Q}_k(\mathcal{C}) := \bigcup \{\mathcal{Q}(A_1, \cdots, A_k) : A_1, \cdots, A_k \in \mathcal{C}\} \quad \textit{is a } \text{VCC}.$$

PROOF. By induction, we may assume $k = 2$. Let $\mathfrak{D} := \{A \cap B : A, B \in \mathcal{C}\}$. Then $m^{\mathfrak{D}}(n) \leqslant m^{\mathcal{C}}(n)^2 \leqslant (n^{V(\mathcal{C})} + 1)^2 < 2^n$ for n large, so $\mathfrak{D}$ is a VCC.

We may assume $\phi \in \mathcal{C}$ and $X \in \mathcal{C}$. If $\mathcal{S} := \{A \setminus B : A, B \in \mathcal{C}\}$ then $\mathcal{S}$ is a VCC as above. A finite union of VCC's is likewise a VCC. Now every set in $\mathcal{Q}(A, B)$ is a union of some of the four atoms $A \cap B$, $A \setminus B$, $B \setminus A$, and $(X \setminus A) \setminus B$. Unions of at most four sets can be treated also as above, completing the proof.

(7.13). LEMMA. *If $(X, \mathcal{Q}, P)$ is a probability space, $\mathcal{C} \subset \mathcal{Q}$, $\mathcal{C}$ is a* VCC *and $v := V(\mathcal{C})$, there is a constant $K = K(v)$ (not depending on P) such that for $0 < \varepsilon \leqslant \frac{1}{2}$,*

$$N(\varepsilon, \mathcal{C}, P) \leqslant K\varepsilon^{-v}|\ln \varepsilon|^{v}.$$

PROOF. Suppose $A_1, \cdots, A_m \in \mathcal{C}$, and $P(A_i \Delta A_j) \geqslant \varepsilon$ for $i \neq j$. We may assume $m \geqslant 2$. If $n \geqslant 2$ is so large that $m(m-1)(1-\varepsilon)^n < 2$, then $\Pr\{P_n(A_i \Delta A_j) > 0$ for all $i \neq j\} > 0$. In that case, $m \leqslant m^{\mathcal{C}}(n) \leqslant n^{v}$ by (7.11). If we take the smallest n for which $m^2(1-\varepsilon)^n < 2$, then $m^2(1-\varepsilon)^{n-1} \geqslant 2$ so $n - 1 \leqslant (2 \ln m - \ln 2)/|\ln(1-\varepsilon)|$, $n \leqslant (2 \ln m)/\varepsilon$, and $m \leqslant (2 \ln m)^{v}\varepsilon^{-v}$.

For some $m_0 = m_0(v) < +\infty$, $(2 \ln m)^v \leqslant m^{1/(v+1)}$ for $m \geqslant m_0$, and then $m \leqslant \varepsilon^{-v-1}$, so $\ln m \leqslant (v+1)|\ln \varepsilon|$. Hence

$$m \leqslant K(v)\varepsilon^{-v}|\ln \varepsilon|^v \qquad \text{for} \quad 0 < \varepsilon \leqslant \tfrac{1}{2}$$

if $K(v) = \max(m_0, 2^{v+1}(v+1)^v)$. Thus, choosing at most m points $\geqslant \varepsilon$ apart, the balls of radius ε with these centers cover $\mathcal{C}$, proving Lemma 7.13.

Now to prove Theorem 7.1, given a probability space $(X, \mathcal{Q}, P)$ and VCC $\mathcal{C} \subset \mathcal{Q}$, with $(X, \mathcal{Q}; \mathcal{C}, \mathcal{S})$ $P\epsilon$-Suslin for some $\mathcal{S}$, $\mathcal{C}$ is totally bounded for d_P by Lemma 7.13, and P-EM by Proposition 3.2, so it will suffice to verify 1.2b.

For any $\delta > 0$, let $\mathcal{C}(\delta) := \{A \setminus B : A, B \in \mathcal{C}, P(A \setminus B) \leqslant \delta\}$, $\mathfrak{D}(\delta) := \{\langle A, B\rangle \in \mathcal{C} \times \mathcal{C} : A \setminus B \in \mathcal{C}(\delta)\}$. Note $\mathcal{C}(\delta) \subset \mathcal{C}(1)$.

Let $d_P^{(2)}$ be the product pseudometric on $\mathcal{C} \times \mathcal{C}$, $d_P^{(2)}(\langle A, B\rangle, \langle C, D\rangle) := d_P(A, C) + d_P(B, D)$. Then $\mathfrak{D}(\delta)$ is closed in $\mathcal{C} \times \mathcal{C}$ for $d_P^{(2)}$. By (7.13), $(\mathcal{C}, d_P)$ is separable, so $\mathfrak{D}(\delta)$ is measurable for the product σ-algebra $\mathcal{S} \times \mathcal{S}$.

We can define $\nu_n(\cdot)(\omega)$ for ω in a product X^∞ of copies of $(X, \mathcal{Q}, P)$,

$$\omega = \langle x_1, x_2, \cdots, \rangle, \qquad x(j)(\omega) := x_j.$$

For any $r > 0$, let

$$E_{\delta r} := \{\langle \omega, A, B\rangle : |\nu_n(A \setminus B)(\omega)| > r, \langle A, B\rangle \in \mathfrak{D}(\delta)\}.$$

Then as in the proof of Proposition 3.2, $E_{\delta r}$ is jointly measurable in the Suslin measurable space $(X^\infty \times \mathcal{C} \times \mathcal{C}, \mathcal{Q}^\infty \times \mathcal{S} \times \mathcal{S})$. The projection of $E_{\delta r}$ into X^∞ is measurable for the completion of P^∞. Thus $\sup_{C \in \mathcal{C}(\delta)}|\nu_n(C)|$ is a measurable random variable. By 1.2b, it will suffice to prove that for any $\varepsilon > 0$ there is a $\delta > 0$ and an n_0 such that

$$\Pr\{\sup_{C \in \mathcal{C}(\delta)}|\nu_n(C)| > \varepsilon\} < \varepsilon \qquad \text{for} \quad n \geqslant n_0.$$

By Proposition 7.12, $\mathcal{C}(1)$ is a VCC. By Lemma 7.13, take a w with $V(\mathcal{C}(1)) := v < w < +\infty$ and $N < \infty$ such that $N(\gamma/2, \mathcal{C}(1), P) \leqslant N\gamma^{-w}$ for $0 < \gamma < 1$. Given a δ, $0 < \delta \leqslant 1$, to be chosen later, for each $j = 0, 1, 2, \cdots$, take sets $A_{j1}, \cdots, A_{jm(j)} \in \mathcal{C}(\delta)$ such that for each $A \in \mathcal{C}(\delta)$, $P(A \Delta A_{ji}) \leqslant \delta/2^j$ for some i. Then we can take $m(j) \leqslant N2^{jw}/\delta^w$, $m(0) = 1$, and $A_{01} = \phi$.

For each integer $j \geqslant 1$ and $1 \leqslant i \leqslant m(j)$, take a $k = k(j, i)$ such that

$$P(A_{ji}\, \Delta A_{j-1,k}) \leqslant 2\delta/2^j.$$

Let $\mathcal{C}_j$ be the collection of all sets $A \setminus B$ or $B \setminus A$, $A = A_{ji}$, $B = A_{j-1, k(j,i)}$. There are at most $2m(j)$ such sets, all with probability $\leqslant 2\delta/2^j$. Take $0 < \varepsilon < 1$. Then $P_j := \Pr\{|\nu_n(A)| \geqslant \varepsilon/j^2$ for some $A \in \mathcal{C}_j\} \leqslant 2m(j)\sup\{\Pr\{|\nu_n(A)| \geqslant \varepsilon/j^2 : P(A) \leqslant 2\delta/2^j\}\}$. Given $p \leqslant 2\delta/2^j$, we have by Bernstein's inequality (Bennett (1962), or Hoeffding (1963))

$$\begin{aligned} E_{jnp} &:= E\left(np + \varepsilon n^{\frac{1}{2}}/j^2, n, p\right) \leqslant \exp\left(-\varepsilon^2/\left(2j^4pq + j^2\varepsilon n^{-\frac{1}{2}}\right)\right) \\ &\leqslant \exp\left(-\varepsilon^2 n^{\frac{1}{2}}/\left(4j^4 n^{\frac{1}{2}}\delta 2^{-j} + j^2\varepsilon\right)\right). \end{aligned}$$

We will treat this by cases according to which term in the denominator is larger.

In case $4j^2n^{\frac{1}{2}}\delta > 2^j\varepsilon$, we say $j \in J(n, \varepsilon, \delta)$. Then $E_{jnp} \leqslant \exp(-\varepsilon^2 2^j/8j^4\delta)$. If $j \notin J(n, \varepsilon, \delta)$, then $E_{jnp} \leqslant \exp(-\varepsilon n^{\frac{1}{2}}/2j^2)$. Setting $B_{jnp} := B(np - \varepsilon n^{\frac{1}{2}}/j^2, n, p)$ we have the same upper bounds for B_{jnp} just shown for E_{jnp}. Let $j(n) := [n^{\frac{1}{8}}]$. Then

$$\Sigma_{1 \leqslant j \leqslant j(n)} P_j \leqslant S_1 + S_2 \qquad \text{where}$$

$$S_1 := 2\Sigma_{j \leqslant j(n), j \in J(n, \varepsilon, \delta)} N 2^{jw} \delta^{-w} \exp(-\varepsilon^2 2^j/8j^4\delta),$$

$$S_2 := 2\Sigma_{j \leqslant j(n), j \notin J(n, \varepsilon, \delta)} N 2^{jw} \delta^{-w} \exp(-\varepsilon n^{\frac{1}{2}}/2j^2).$$

Now S_1 is a partial sum of a convergent infinite series, whose value approaches 0 as $\delta \to 0$ for fixed ε and w. Thus for some $\delta_1(\varepsilon) > 0$, $S_1 < \varepsilon/3$ for $0 < \delta \leqslant \delta_1(\varepsilon)$. *We now choose* $\delta = \delta_1(\varepsilon)$. Next, $S_2 \leqslant 2n^{\frac{1}{8}} N 2^{j(n)w} \delta^{-w} \exp(-\varepsilon n^{\frac{1}{4}}/2) \to 0$ as $n \to \infty$, so for some $n_0(\delta)$, $S_2 \leqslant \varepsilon/3$ for $n \geqslant n_0(\delta)$.

Let $D_{ni} := A_{j(n), i}$. Since $A_{01} = \phi$, we have

$$|\nu_n(D_{ni})| \leqslant \Sigma_{r=1}^{j(n)} |\nu_n(C_r)| + |\nu_n(D_r)|$$

for some C_r and $D_r \in \mathcal{C}_r$. Then since $\Sigma_{j \geqslant 1} j^{-2} = \pi^2/6 < 2$, we have

(7.14) $$\Pr\{\max_i |\nu_n(D_{ni})| \geqslant 4\varepsilon\} < 2\varepsilon/3, \qquad n \geqslant n_0(\delta).$$

Let

$$\mathfrak{D}_{n\delta i} := \{\langle A, B\rangle \in \mathfrak{D}(\delta) : P((A \setminus B)\, \Delta D_{ni}) \leqslant \delta/2^{j(n)}\}.$$

The functions $\langle A, B\rangle \to P((A \setminus B) \setminus D_{ni})$, $\langle A, B\rangle \to P(D_{ni} \setminus (A \setminus B))$, and hence $\langle A, B\rangle \to P((A \setminus B)\, \Delta D_{ni})$, are all continuous on the separable pseudometric space $\mathcal{C} \times \mathcal{C}$ for $d_P^{(2)}$, thus $\mathbb{S} \times \mathbb{S}$ measurable, so $\mathfrak{D}_{n\delta i} \in \mathbb{S} \times \mathbb{S}$. For each j and i, the map

$$\langle \omega, A, B\rangle \to \delta_{x(j)(\omega)}((A \setminus B) \setminus D_{ni})$$

is measurable ($\mathcal{Q}^\infty \times \mathbb{S} \times \mathbb{S}$) by the ϵ-Suslin assumption. Thus the function

$$\langle \omega, A, B\rangle \to \nu_n(\omega)((A \setminus B) \setminus D_{ni})$$

is jointly measurable. Likewise, so is

$$\langle \omega, A, B\rangle \to \nu_n(\omega)(D_{ni} \setminus (A \setminus B)).$$

As in the proof of Proposition 3.2, since $\mathfrak{D}_{n\delta i} \in \mathbb{S} \times \mathbb{S}$,

$$\omega \to \sup\{\nu_n(\omega)(D_{ni} \setminus (A \setminus B)) : \langle A, B\rangle \in \mathfrak{D}_{n\delta i}\}$$

is completion measurable for P^∞, and likewise if sup is replaced by inf or $\sup\{|\cdots| : \cdots\}$, or $D_{ni} \setminus (A \setminus B)$ by $(A \setminus B) \setminus D_{ni}$. For each $E \in \mathcal{C}(\delta)$ there is some i such that $d_P(E, D_{ni}) \leqslant \delta/2^{j(n)}$. Let $\mathfrak{B}_{n\delta} := \mathfrak{B}(n, \delta)$ be the collection of all sets $F = E \setminus D_{ni}$ or $F = D_{ni} \setminus E$ where $E \in \mathcal{C}(\delta)$ and $P(F) \leqslant \delta/2^{j(n)}$. Now $\omega \to \sup\{|\nu_n(F)| : F \in \mathfrak{B}_{n\delta}\}$ is P^∞-completion measurable. Let

$$Q_{n\varepsilon} := \Pr\{\sup\{|\nu_n(B)| : B \in \mathfrak{B}_{n\delta}\} \geqslant \varepsilon\}.$$

Then by (7.14),

(7.15) $$\Pr\{\sup_{A\in\mathcal{C}(\delta)}|\nu_n(A)| > 6\varepsilon\} \leqslant 2\varepsilon/3 + Q_{n\varepsilon}.$$

Let

$$C_{n\varepsilon\delta i} := \{\langle\omega, A, B\rangle \in X^\infty \times \mathfrak{D}_{n\delta i} : \nu_n((A \setminus B)\setminus D_{ni}) \geqslant \varepsilon\}.$$

Then $C_{n\varepsilon\delta i}$ is measurable ($\mathcal{Q}^\infty \times \mathcal{S} \times \mathcal{S}$). By Theorem 3.3, let $\omega \to \langle C_{i1}, C_{i2}\rangle(\omega)$ be a universally measurable selector for $C_{n\varepsilon\delta i}$, mapping a universally measurable subset of X^∞ into $\mathfrak{D}_{n\delta i}$. Let $C_i(\omega) := (C_{i1}(\omega) \setminus C_{i2}(\omega)) \setminus D_{ni}$. Then $\omega \to P(C_i(\omega))$ is universally measurable, and so is $\omega \to \nu_n(C_i(\omega))$.

Uhlmann ((1966), Satz 6) and Jogdeo and Samuels (1968) showed that the median of a binomial distribution is within 1 of its mean. Thus for any fixed set $A \in \mathcal{Q}$, $\Pr(\nu_n(A) \leqslant n^{-\frac{1}{2}}) \geqslant \frac{1}{2}$. Let ν_n and ν_n' be two independent copies of the normalized empirical measure $n^{\frac{1}{2}}(P_n - P)$, with $\nu_n(\cdot)(\omega)$ and $\nu_n'(\cdot)(\omega')$ defined for $\langle\omega, \omega'\rangle$ in a product space $X^\infty \times X^\infty$ with product probability. We write "$\exists C_i(\omega)$" iff $C_i(\omega)$ is defined, i.e., iff $\exists\langle A, B\rangle : \langle\omega, A, B\rangle \in C_{n\varepsilon\delta i}$, so that $\nu_n(C_i(\omega)) \geqslant \varepsilon$. Then for each i, $\{\omega : \exists C_i(\omega)\}$ is a universally measurable event. If $\exists i \exists C_i(\omega)$, let $i(\omega)$ be the least such i, and write "$\exists i(\omega)$." Then since ν_n' is independent of ω, we have

(7.16) $$\Pr\{\exists i(\omega) \text{ and } \nu_n'(C_{i(\omega)}(\omega)) \leqslant n^{-\frac{1}{2}}\} \geqslant \Pr\{\exists i(\omega)\}/2.$$

Likewise, if we replace $(A \setminus B)\setminus D_{ni}$ by $D_{nu} \setminus (A \setminus B)$ in the definition of $C_{n\varepsilon\delta i}$, we get another measurable set $D_{n\varepsilon\delta u}$ with a selector $\langle D_{u1}, D_{u2}\rangle(\omega)$. Let $D_u(\omega) := D_{nu} \setminus (D_{u1} \setminus D_{u2})(\omega)$. Let $d(\omega)$ be the least u, if one exists, such that $\exists D_u(\omega)$. Then $\Pr\{\exists d(\omega) \text{ and } \nu_n'(D_{d(\omega)}(\omega)) \leqslant n^{-\frac{1}{2}}\} \geqslant \Pr\{\exists d(\omega)\}/2$.

Replacing "$\geqslant \varepsilon$" by "$\leqslant -\varepsilon$" in the definition of $C_{n\varepsilon\delta i}$, we get a measurable set $E_{n\varepsilon\delta i}$ with a selector $\langle E_{i1}, E_{i2}\rangle$, $E_i(\omega) := (E_{i1} \setminus E_{i2}) \setminus D_{ni}$, and let $e(\omega) := \min\{i : \exists E_i(\omega)\}$. Doing the same for $D_{n\varepsilon\delta u}$ we get $\langle F_{u1}, F_{u2}\rangle$, $F_u(\omega)$, and $f(\omega) := \min\{u : \exists F_u(\omega)\}$. Then

(7.17) $$\Pr\{\exists e(\omega) \text{ and } \nu_n'(E_{e(\omega)}(\omega)) \geqslant -n^{-\frac{1}{2}}\} \geqslant \Pr\{\exists e(\omega)\}/2,$$

and likewise for $f(\omega)$.

Let $X(1), \cdots, X(2n)$ be independent and identically distributed with law P, $\nu_n := n^{\frac{1}{2}}(P_n - P)$, $\nu_n' = n^{\frac{1}{2}}(P_n' - P)$, where $P_n' := n^{-1}\Sigma_{n<j\leqslant 2n}\delta_{X(j)}$. As in Vapnik and Červonenkis (1971), conditional on $P_{2n} = P_n + P_n'$, the distribution of $\langle X(1), \cdots, X(2n)\rangle$ is obtained by averaging over all permutations of the integers $1, 2, \cdots, 2n$. For fixed P_{2n} and a set A with $r_n(A) := 2nP_{2n}(A)$,

$$\Pr\{nP_n(A) \geqslant s, nP_n'(A) \leqslant t | P_{2n}\}$$

is the hypergeometric probability of choosing at least s white balls and at most t black balls in a random sample of $r := r_n(A)$ balls without replacement from an urn containing n white and n black balls. For fixed s and t, this event is equivalent to drawing at least s white balls if $r \leqslant s + t$, or at most t black balls if $r \geqslant s + t$. Thus the probability is maximized when $r = s + t$, and then it is $H := H(t, r, n, 2n) := \Sigma_{j \leqslant t} \binom{n}{r-j}\binom{n}{j} / \binom{2n}{r}$.

On the event in (7.16), $s \geqslant np + \varepsilon n^{\frac{1}{2}}$ and $t \leqslant np + 1$ where $p := P(A)$, so $s - t \geqslant \varepsilon n^{\frac{1}{2}} - 1 > 2$ for n large enough. Hence, by an inequality of Uhlmann (1966), hypergeometric tails are smaller than corresponding binomial tails; specifically $H \leqslant E(s, n, (s + t)/2n)$. Then by Lemma 2.7 above, $H \leqslant ((s + t)/2s)^s \exp((s - t)/2)$. Now

$$s \cdot \ln((s + t)/2s) = s \cdot \ln(1 - (s - t)/2s) \leqslant -(s - t)/2 - (s - t)^2/8s,$$

so $H \leqslant \exp(-(s - t)^2/8s)$. The same inequality holds for $s \geqslant np - 1$ and $t \leqslant np - \varepsilon n^{\frac{1}{2}}$, as in (7.17), if n is large.

By (7.11), $\Delta^{\mathcal{C}}\{X(1), \cdots, X(2n)\} \leqslant (2n)^v$, $\Delta^{\mathfrak{B}(n,\delta)}\{X(1), \cdots, X(2n)\} \leqslant 2(2n)^{2v}$, and using (7.16) and (7.17),

$$\begin{aligned} Q_{n\varepsilon} &= E(\Pr\{\sup\{|\nu_n(B)| : B \in \mathfrak{B}_{n\delta}\} \geqslant \varepsilon | P_{2n}\}) \\ &\leqslant \sup(\Pr\{\sup\{|\nu_n(B)| : B \in \mathfrak{B}_{n\delta}\} \geqslant \varepsilon | P_{2n}\}) \\ &\leqslant 8(2n)^{2v} \sup(H) \leqslant 8(2n)^{2v} \sup\big(\exp\big(-(s - t)^2/(8s)\big)\big) \end{aligned}$$

where the supremum is over values of $p \leqslant \delta/2^{j(n)}$, $j(n) = [n^{\frac{1}{8}}]$, and either $s \geqslant np + \varepsilon n^{\frac{1}{2}}$ and $t \leqslant np + 1$, or $s \geqslant np - 1$ and $t \leqslant np - \varepsilon n^{\frac{1}{2}}$. The function $s \to (s - t)^2/s$ is increasing for $t < s$, so for $n > 2\varepsilon^{-2}$ we have

$$\begin{aligned} Q_{n\varepsilon} &\leqslant 8(2n)^{2v} \exp\Big(-\big(\varepsilon n^{\frac{1}{2}} - 1\big)^2/8\big(np + \varepsilon n^{\frac{1}{2}}\big)\Big) \\ &\leqslant 8(2n)^{2v} \exp\Big(-\varepsilon^2 n/32\big(n\delta/2^{j(n)} + \varepsilon n^{\frac{1}{2}}\big)\Big) \to 0 \end{aligned}$$

as $n \to \infty$ for fixed $\varepsilon > 0$, $\delta > 0$. So for some $n_1 = n_1(\delta, \varepsilon)$, $Q_{n\varepsilon} \leqslant \varepsilon/3$ for $n \geqslant n_1$. Hence by (7.15), for $\delta = \delta_1(\varepsilon)$ and $n \geqslant \max(n_0, n_1)(\delta, \varepsilon)$, $\Pr\{|\nu_n(A)| \geqslant 6\varepsilon$ for some $A \in \mathcal{C}(\delta)\} < \varepsilon$, proving Theorem 7.1.

(7.18). COROLLARY. *If X is a Polish space, $\mathcal{C}$ is a* VCC *in X, and $\mathcal{C}$ is a Suslin measurable collection of closed sets, or of open sets, for the Effros Borel structure, then $\mathcal{C}$ is a universal Donsker class.*

PROOF. We apply Theorem 7.1, Proposition 3.2, and 4.3 or 4.4. □

Note also that the VCC's of open sets given by Theorem 7.2 (and/or their closed complements) are Suslin by Proposition 4.5 if X is locally compact and separable, and G consists of continuous functions.

If P is Lebesgue measure on [0, 1] and $\mathcal{C} = \{\{x\} : x \in E\}$ where E is nonmeasurable, then $\mathcal{C}$ is a VCC but not P-EM. In this case, $\mathcal{C}$ can be enlarged to the set of all singletons, a VCC with good measurability properties. We may ask whether any VCC can be enlarged in such a way. The following shows that closure for the Hausdorff metric may lose the VCC property.

(7.19). PROPOSITION. *In* [0, 1] *there is a* VCC $\mathcal{C}$, *consisting of finite sets, whose closure* $\bar{\mathcal{C}}$ *for the Hausdorff metric contains all (closed) sets in* [0, 1].

PROOF. We enumerate the primes: $p_1 = 2, p_2 = 3, p_3 = 5, \ldots$. Let $\mathcal{C}$ consist of all finite sets $A = \{x_1, \cdots, x_k\}$ with $x_j \in [0, 1]$ such that for each $j = 1, \cdots, k-1$, $x_{j+1} = p_{2j-1}^{m(j)} x_j / p_{2j}^{n(j)}$ for some integers $m(j)$ and $n(j)$. Then for any set E with 3 elements, if $E \subset A \in \mathcal{C}$, then $E = \{x_t, x_u, x_v\}$ for some x_j as above. By unique factorization, whenever $x_t \in B \in \mathcal{C}$ and $x_v \in B$, then $x_u \in B$ so $\{x_t, x_v\} \neq B \cap E$. Hence $\mathcal{C}$ is a VCC.

For each j, the set of rationals of the form p_{2j-1}^m / p_{2j}^n is dense in $]0, \infty[$ (since their logarithms are dense in $\mathbb{R}$: for any irrational ξ, the multiples $m\xi$ are dense mod 1). Thus any finite set $\{y_1, \cdots, y_k\} \subset [0, 1]$ can be approximated as well as desired by some $\{x_1, \cdots, x_k\} \in \mathcal{C}$. □

For classes $\mathcal{C}$ with $V(\mathcal{C}) = +\infty$, it may still happen that $\Delta^{\mathcal{C}}(X_1, \cdots, X_n) < 2^n$ with high probability. Then, even if $\mathcal{C}$ is not a P-Donsker class, one may use $\Delta^{\mathcal{C}}$ in proving that $\sup_{C \in \mathcal{C}} |P_n - P|(C) \to 0$, as in the original work of Vapnik and Červonenkis (1971). Steele (1977) has substantial further results along these lines.

Acknowledgment. Many thanks to a referee for finding an error in an earlier version of Theorem 5.1.

REFERENCES

AUMANN, R. J. (1961). Borel structures for function spaces. *Illinois J. Math.* **5** 614–630.

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

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BOLTHAUSEN, E. (1976). On weak convergence of an empirical process indexed by the closed convex subsets of $\mathbb{R}^2$. (Preprint).

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

CHRISTENSEN, J. P. R. (1971). On some properties of Effros Borel structure on spaces of closed subsets. *Math. Ann.* **195** 17–23.

CHRISTENSEN, J. P. R. (1974). *Topology and Borel Structure*. North-Holland, Amsterdam; American Elsevier, New York.

CLEMENTS, G. F. (1963). Entropies of several sets of real valued functions. *Pacific J. Math.* **13** 1085–1095.

COVER, T. M. (1965). Geometric and statistical properties of systems of linear inequalities with applications to pattern recognition. *IEEE Trans. Elec. Comp.* **EC-14** 326–334.

DEHARDT, J. (1971). Generalizations of the Glivenko-Cantelli theorem. *Ann. Math. Statist.* **42** 2050–2055.

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

DUDLEY, R. M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

DUDLEY, R. M. (1967a). Measures on non-separable metric spaces. *Illinois J. Math.* **11** 449–453.

DUDLEY, R. M. (1967b). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis* **1** 290–330.

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

DUDLEY, R. M. (1974). Metric entropy of some classes of sets with differentiable boundaries. *J. Approximation Theory* **10** 227–236.

EFFROS, E. G. (1965). Convergence of closed subsets in a topological space. *Proc. Amer. Math. Soc.* **16** 929–931.

FEDERER, H. (1969). *Geometric Measure Theory*. Springer, Berlin.

FREEDMAN, D. (1966). On two equivalence relations between measures. *Ann. Math. Statist.* **37** 686–689.

HARDING, E. F. (1967). The number of partitions of a set of N points in k dimensions induced by hyperplanes. *Proc. Edinburgh Math. Soc. (Ser. II)* **15** 285–289.

HAUSDORFF, F. (1937). *Set Theory*, 3rd ed. Transl. by J. Aumann et al. (New York, Chelsea, 1962).

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

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

KELLEY, J. L. (1955). *General Topology*. Van Nostrand, Princeton.

KOLMOGOROV, A. N. (1956). On Skorohod convergence. *Theor. Probability Appl.* **1** 215–222. *(Teor. Verojatnost. i Primenen.* **1** 239–247, in Russian.)

KOLMOGOROV, A. N. and TIHOMIROV, V. M. (1959). ε-entropy and ε-capacity of sets in functional spaces. *Uspehi Mat. Nauk* **14**, #2 (86), 3–86 (in Russian); (1961) *Amer. Math. Soc. Transl.* (Ser. 2), **17** 277–364.

KURATOWSKI, K. (1966). *Topology*, **1**. Academic Press, New York.

MCSHANE, E. J. (1934). Extension of range of functions. *Bull. Amer. Math. Soc.* **40** 837–842.

NEUHAUS, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.

OKAMOTO, MASASHI (1958). Some inequalities relating to the partial sum of binomial probabilities. *Ann. Inst. Statist. Math.* **10** 29–35.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

PHILIPP, W. (1973). Empirical distribution functions and uniform distribution mod 1. In *Diophantine Approximation and its Applications* (C. F. Osgood, ed.) 211–234. Academic Press, New York.

PYKE, R. and SHORACK, G. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755–771.

RAO, B. V. (1971). Borel structures for function spaces. *Colloq. Math.* **23** 33–38.

RÉVÉSZ, P. (1976). Three theorems of multivariate empirical process. *Lecture Notes in Math.* **566** 106–126.

SCHLÄFLI, LUDWIG (1901, posth.). *Theorie der vielfachen Kontinuität*, in *Gesammelte Math. Abhandlungen I* (Basel, Birkhäuser, 1950).

SION, M. (1960). On uniformization of sets in topological spaces. *Trans. Amer. Math. Soc.* **96** 237–245.

SKOROHOD, A. V. (1955). On passage to the limit from sequences of sums of independent random variables to a homogeneous random process with independent increments. *Dokl. Akad. Nauk. SSSR* **104** 364–367 (in Russian).

STEELE, J. M. (1978). Empirical discrepancies and subadditive processes. *Ann. Probability* **6** 118–127.

STEINER, J. (1826). Einige Gesetze über die Theilung der Ebene und des Raumes. *J. Reine Angew. Math.* **1** 349–364.

STONE, A. H. (1962). Non-separable Borel sets. *Rozprawy Mat.* **28** (41 pp.).

STRAF, M. L. (1972). Weak convergence of stochastic processes with several parameters. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 187–221. Univ. of California Press.

SUN, TZE-GONG (1976). Ph.D. dissertation, Dept. of Mathematics, Univ. of Washington, Seattle.

SZPILRAJN, E. (1938). Ensembles indépendants et mesures non séparables. *C. R. Acad. Sci. Paris* **207** 768–770.

TALAGRAND, M. (1978). Les boules peuvent elles engendrer la tribu borélienne d'un espace métrisable non séparable? (Preprint).

UHLMANN, W. (1966). Vergleich der hypergeometrischen mit der Binomial-Verteilung. *Metrika* **10** 145–158.

VAPNIK, V. N. and ČERVONENKIS, A. YA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Probability Appl.* **16** 264–280. (*Teor. Verojatnost. i Primenen.* **16** 264–279, in Russian.)

VAPNIK, V. N. and ČERVONENKIS, A. YA. (1974). *Theory of Pattern Recognition* (in Russian). Nauka, Moscow.

WATSON, D. (1969). On partitions of *n* points. *Proc. Edinburgh Math. Soc.* **16** 263–264.

WICHURA, MICHAEL J. (1970). On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *Ann. Math. Statist.* **41** 284–291.

DEPARTMENT OF MATHEMATICS
ROOM 2-245, M.I.T.
CAMBRIDGE, MASSACHUSETTS 02139