CONVERGENCE OF EMPIRICAL PROBABILITY MEASURES

by

Joseph Yukich
B.A., Oberlin College
(1978)

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS

IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1982

Signature of Author..
Department of Mathematics, July 6, 1982

Certified by..
Richard M. Dudley
Thesis Supervisor

Accepted by..
Nesmith C. Ankeny
Chairman, Departmental Graduate Committee

CONVERGENCE OF EMPIRICAL PROBABILITY MEASURES

by

Joseph Yukich

Submitted to the Department of Mathematics on July 6, 1982 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Mathematics

ABSTRACT

Let $(X,\mathcal{A})$ be a measurable space and $\mathcal{F}$ a collection of real valued measurable functions on X. For each $x \in X$, let $F := F_{\mathcal{F}}(x) := \sup\{|f(x)| : f \in \mathcal{F}\}$ be the envelope function for $\mathcal{F}$. Let P be a probability measure on $\mathcal{A}$ and P_n the empirical measures for P. For each f let $\nu_n(f) = \sqrt{n} \int f(dP_n - dP)$. Under certain metric entropy conditions on $\mathcal{F}$ and certain restrictions on F, exponential bounds for $\sup_{f \in \mathcal{F}} |\nu_n(f)|$ are proved.

Given $(X,\mathcal{A},P)$, $\mathcal{F}$, and $\varepsilon > 0$ let $N_I^{(q)}(\varepsilon) := N_I^{(q)}(\varepsilon,\mathcal{F},P) := \min\{m: \exists\, f_1,\ldots,f_m \in \mathcal{L}^2(X): \forall f \in \mathcal{F} \,\exists\, i,j: f_i \le f \le f_j$ and $\{\int (f_j - f_i)^q dp\}^{1/q} < \varepsilon\}$. The relationship between $N_I^{(q)}(\varepsilon)$ and Donsker classes of functions is explored and it is shown that for all q, $q \neq q$, $N_I^{(q)}(\varepsilon)$ does not give sharp results.

Define the metric d_g on $\mathcal{P}(\mathbb{R}^k)$ where $d_g(P,Q) := \sup_{x \in \mathbb{R}^k} |\int g(x-y)(dP-dP)(y)|$ and where g is any uniformly continuous density on $\mathbb{R}^k$ such that the Fourier transform $\hat{g}$

has countable zeroes. d_g is generalized to a metric on $\mathcal{P}(G)$ where G is an arbitrary locally compact group. Let P_n be the empirical measures for P. Under suitable restrictions on P and g we obtain a central limit theorem for $d_g(P_n,P)$ and under stronger conditions an invariance principle.

Under certain conditions on sequences $\mathcal{G}_n$ of classes of functions, $\sup_{g \in \mathcal{G}_n} |\int g(dP_n - dP)| \xrightarrow{n} 0$ a.s., where P is a measure on $\mathbb{R}^k$ and P_n are the empirical measures. The results are applied to kernel density estimation. Let X_i, $i \geq 1$ be iid random variables in $\mathbb{R}^k$ with a common density $f(x)$. For each $g \in \mathcal{G}_n$ define the kernels $\hat{g}(x) = \frac{1}{n} \sum_{i=1}^{n} g(x - X_i)$. Then sufficient conditions on g and f are found so that $\sup_x |\hat{g}(x) - f(x)| \xrightarrow{n} 0$ a.s.

Thesis Supervisor: Richard M. Dudley
Title: Professor of Mathematics

Acknowledgments

First and foremost I wish to thank my advisor, Richard Dudley, for his numerous suggestions, generous help, and sustained interest in my research. His thorough and repeated readings of the manuscript were greatly appreciated and indeed, crucial to its completion. I am also indebted to Walter Philipp for his many helpful suggestions and to Phyllis Ruby for her neat typing.

Had I not met my fiancée Arati the thesis would have been completed sooner; fortunately the first event did not happen. Finally I wish to thank my parents who made it all possible.

Table of Contents

Chapter 1

Introduction

Let $(X,\mathcal{A})$ be a measurable space and $\mathcal{F}$ a collection of real valued measurable functions on X. For each $x \in X$ let $F := F_{\mathcal{F}}(x) := \sup\{|f(x)|: f \in \mathcal{F}\}$. F is called the envelope function for $\mathcal{F}$.

Let P be a probability measure on $\mathcal{A}$. Let $(X^{\infty},\mathcal{A}^{\infty},P^{\infty})$ be a countable product of copies of $(X,\mathcal{A},P)$ with coordinates $\xi_j = \xi_{(j)}$, so that the ξ_j are independent, identically distributed random variables with values in X and distribution P. Let $P_n = \frac{1}{n}(\delta_{\xi_{(1)}} + \ldots + \delta_{\xi_{(n)}})$ $n = 1,2,\ldots$ where δ_x is the unit mass at x. P_n are the empirical measures for P. For each $f \in \mathcal{F}$ let $\nu_n(f) = \sqrt{n}\int f(dP_n - dP)$. In Chapter 2 we will be concerned with the suprema of $|\nu_n|$ over the collection $\mathcal{F}$.

As in Pollard [30] we have

<u>Def. 1.1.1</u> Given $\mathcal{F}$, F, and a finite subset $S \subset X$, let $N(\delta,S,\mathcal{F}) := \inf\{m: \exists\ f_1,\ldots,f_m \in \mathcal{F}$ such that $\min_i \sum_{x \in S} (f(x)-f_i(x))^2 < \delta^2 \sum_{x \in S} (f(x))^2$ for every $f \in \mathcal{F}\}$.

Let $N(\delta,\mathcal{F}) := \sup_S N(\delta,S,\mathcal{F})$.

Note that $N(\delta,F) := \inf\{m: \forall n$ and $\forall$ values of P_{2n} there are $f_1,\ldots,f_m \in F$ such that $\forall f \in F \ \exists\, i \le m$ such that $\int (f-f_i)^2 dP_{2n} < \delta^2 ||F||_{2n}^2$ where $||F||_{2n}^2 := \frac{1}{2n} \sum_{i=1}^{2n} F(\xi_i)^2\}$.

Suppose $N(\delta,F) \le \exp(\frac{C}{\delta^{2-\varepsilon}})$ for some constants C,ε where $C \ge 1$, $0 < \varepsilon < 1$, for all δ, $0 < \delta < 1$. If $F \equiv 1$ then in Theorem 2.2.6 we obtain exponential bounds for $\sup_{f\in F} |\nu_n(f)|$ for all $n \ge 1$. If $F \in L^p$, $p > 2$, then in Theorem 2.2.14 we obtain exponential bounds for $\sup_{f\in F} |\nu_n(f)|$. In Theorem 2.2.24 we impose a slightly different metric entropy condition and obtain an exponential bound for $\sup_{f\in F} |\nu_n(f)|$ when $F \equiv 1$.

<u>Def. 1.1.2</u> Given a class C of subsets of a set X and a finite set $F \subset X$, let $\Delta^C(F)$ be the number of different sets $C \cap F$ for C C. For $n = 1,2,\ldots$ let $m^C(n) := \max\{\Delta^C(F): F$ has n elements$\}$. Let

$$v := V(C) = \inf\{n: m^C(n) < 2^n\}$$
$$= +\infty \text{ if } m^C(n) = 2^n \ \forall n.$$

If $v := V(C) < \infty$ we will call C a Vapnik-Červonenkis class (VCC).

Let $f \in L^2(X,A,P)$, C a VCC, and $F := \{f1_C: C \in C\}$.

Then in Theorem 2.2.1 we establish exponential bounds for $\sup_{f \in \mathcal{F}} |\nu_n(f)|$. We shall need the following definitions. We rely heavily upon [13] and [15], c.f. also [16].

Given a class $\mathcal{F} \subset L^2(X,A,P)$ of functions, we say that it is a Donsker class of functions (DCOF) iff there is a G_p process which has uniformly continuous and bounded sample functions on $\mathcal{F}$ and such that $\exists$ $Y_1, Y_2, \ldots$ iid copies of G_p such that

$$\sup_{f \in \mathcal{F}} \max_{k \le n} \frac{1}{\sqrt{n}} \left| \delta_{X_1} + \ldots + \delta_{X_k} - kP - (Y_1 + \ldots + Y_k)(F) \right| \xrightarrow{n} 0$$

in probability where X_j and Y_j are defined on the probability space $(X^\infty, A^\infty, P^\infty) \times ([0,1], \mathcal{B}, \lambda)$ where λ is Lebesgue measure and $\mathcal{B}$ are Lebesgue measurable sets. $\mathcal{F}$ is P-EM if $\mathcal{F}$ is empirically measurable for P; see [13] for details.

In [13] Dudley shows in effect that a P-EM class $\mathcal{F}$ is a DCOF iff

(a) $\mathcal{F}$ is totally bounded for ρ_p where $\rho_p(f,g) := e_p(f-Ef, g-Eg)$ and

(b) $\forall\ \varepsilon > 0\ \exists\ \delta > 0$ such that for $n \ge n_0$ $\Pr^*\{\exists\ f,g \in \mathcal{F}: \rho_p(f,g) < \delta,\ |\nu_n(f-g)| > \varepsilon\} < \varepsilon$.

Def. 1.1.3 Given (X,A,P), F, $\varepsilon > 0$, and $q \geq 1$ let $N_I^{(q)}(\varepsilon) := N_I^{(q)}(\varepsilon,F,P) := \min\{m: \exists f_1,\ldots,f_m \in L^q(X): \forall f \in F \ \exists \ i,j: f_i \leq f \leq f_j$ and $(\int(f_j-f_i)^q dP)^{1/2} < \varepsilon\}$. $N_I^{(q)}(\varepsilon)$, which we shall call q-norm metric entropy with bracketing, is a generalization of metric entropy with bracketing as discussed in [13]. Chapter 2 closes with a discussion of the relationship between $N_I^{(q)}(\varepsilon)$ and DCOF.

In chapter 3 we define the metric d_g on $P(\mathbb{R}^k)$ where

$$d_g(P,Q) := \sup_{x \in \mathbb{R}^k} \left| \int g(x-y)(dP-dQ)(y) \right|$$

and where g is any uniformly continuous density on $\mathbb{R}^k$ such that the Fourier transform $\hat{g}$ has countable zeroes. d_g is generalized to a metric on $P(G)$ where G is an arbitrary locally compact abelian group. The results of chapter 2 give a bounded LIL for $d_g(P_n,P)$ where P_n are empirical measures. Under suitable restrictions on P and g we obtain an invariance principle for $dg(P_n,P)$, a central limit theorem, and the asymptotic distribution.

In chapter 4 general conditions on sequences G_n of classes of functions are found so that $\sup_{g \in G_n} \left| \int g(dP_n - dP) \right| \xrightarrow{n} 0$ a.s., where P_n are the empirical measures for $P \in P(\mathbb{R}^k)$.

The results are applied to kernel density estimation where for each $g \in G_n$ the kernels are $\hat{g}(x) := \frac{1}{n} \sum_{i=1}^{n} g(x-X_i)$, and where X_i are iid random variables with a common density $f(x)$. These kernels generalize the classic kernels $\frac{1}{h_n} g(\frac{x}{h_n})$ considered by Birkel and Rosenblatt [3] and Devroye and Wagner [7] among others. The strong uniform consistency of $\hat{g}(x)$ is investigated, i.e., sufficient conditions on g and f are found so that $\sup_x |\hat{g}(x)-f(x)| \to 0$ as $n \to \infty$ a.s. The result generalizes that in [7].

Chapter 2

Empirical Processes Indexed by Classes of Functions

§1. Introduction

In this chapter we establish exponential bounds and limit theorems for empirical processes indexed by classes of functions. Throughout, let $\mathcal{F}$ be a class of real-valued functions on a probability space $(X,\mathcal{A},P)$. F will denote the envelope for $\mathcal{F}$.

The techniques employed rely heavily upon those developed by Pollard [30] and generalize those used by Alexander [1] for the case when $\mathcal{F}$ is a class of sets.

The chapter closes with a discussion of the relationship between $N_I^{(q)}$ and DCOF.

§2. Exponential Bounds

The techniques used in the following theorem will be generalized in subsequent results. We have

Theorem 2.2.1 Let $(X,\mathcal{A},P)$ be a probability space and $\mathcal{C} \subset \mathcal{A}$ a VCC. Let g be any bounded measurable function on X with $\mathcal{L}^2$ norm $||g||_2 := L$, $||g||_{sup} \leq \ell < \infty$, and let $\mathcal{F} := \{g1_C : C \in \mathcal{C}\}$. If $\mathcal{F}$ is P-EM and 2-sample P-EM, then

$$\Pr\{\sup_{g\in F}|\nu_n(f)| > M + 2L\} \le K \exp\{\frac{-(M+2)^2}{5\ell^2}\}$$

for all $n \ge 1$ for $M \ge M_0$ where K and M_0 are constants which depend only on $v := V(C)$ and ℓ.

Proof: We first formalize the method of randomization in [30]. Let $Y := \{-1,1\}$ and $\mathcal{B} = \mathcal{P}(Y)$ be the set of all subsets of Y. Let $Q(\{-1\}) = Q(\{1\}) = \frac{1}{2}$ and ε_i, $i \ge 1$, the i^{th} coordinate function on Y^∞. Fix $n \ge 1$ and define $\sigma: Y^\infty \times \{1,\ldots,2n\} \to \{1,\ldots,2n\}$ as follows. For $i = 1,\ldots,n$, let

$$\sigma(y,2i-1) = \begin{cases} i & \text{if } \varepsilon_i(y) = 1 \\ n+i & \text{if } \varepsilon_i(y) = -1, \text{ and} \end{cases}$$

$$\sigma(y,2i) = \begin{cases} n+i & \text{if } \varepsilon_i(y) = 1 \\ i & \text{if } \varepsilon_i(y) = -1. \end{cases}$$

Define $G: (X^\infty \times Y^\infty, \mathcal{A}^\infty \times \mathcal{B}^\infty, P^\infty \times Q^\infty) \to (X^\infty, \mathcal{A}^\infty, P^\infty)$ by

$$G(x,y)_i := \begin{cases} X_{\sigma_Y^{-1}(i)} & i \le 2n \\ X_i & i > 2n \end{cases}, \quad \text{where } \sigma_Y(\cdot) := \sigma(y,\cdot).$$

Let ξ_i be the i^{th} coordinate function on X^∞. We now have

$$(1) \quad \sum_{i=1}^{n} \varepsilon_i(y)(\delta_{\xi_{2i-1}(x)} - \delta_{\xi_{2i}(x)}) = \sum_{i=1}^{n} \delta_{\xi_i(G(x,y))} - \sum_{i=n+1}^{2n} \delta_{\xi_i(G(x,y))}.$$

Note that σ serves to change the indices in the LHS of (1) so that δ_{ξ_i} has a positive coefficient iff $i \leq n$.

Now define $P_n': (X^\infty, A^\infty, P^\infty) \to D_0((C,P),B_b)$ by $P_n' := \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i} \circ G$ and P_n'' by $P_n'' := \frac{1}{n}\sum_{i=n+1}^{2n} \delta_{\xi_i} \circ G$. Then P_n' and P_n'' are independent empirical measures for P, and $P_n' \circ G$ and $P_n'' \circ G$ are also independent and have the same laws as P_n' and P_n'' respectively. Define $\nu_n' := \sqrt{n}(P_n' - P)$, $\nu_n'' := \sqrt{n}(P_n'' - P)$, and $\nu_n^0 := \nu_n' - \nu_n'' = \sqrt{n}(P_n' - P_n'')$. Also, let $Pr := P^\infty \times Q^\infty$ and $\mathbb{P} := P^\infty$. We will write ν_n', ν_n'', etc. for $\nu_n' \circ G$, $\nu_n'' \circ G$, etc. using Pr and $\mathbb{P}$ to make clear which is meant. The following is adapted from [30], Lemma 2.3.

<u>Lemma 2.2.2</u> $Pr\{\sup_{f \in F} |\nu_n''(f)| > M + 2L\} \leq \frac{4}{3}\mathbb{P}\{\sup_{f \in F} |\nu_n^0(f)| > M\}$.

<u>Proof</u>: For $x \in X^\infty$, let $\omega_1(x) = (x_1, \ldots, x_n)$ and $\omega_2(x) = (x_{n+1}, \ldots, x_{2n})$. By Chebyshev's inequality and since $E(\nu_n''(f))^2 = \int f^2 dP - (\int f dP)^2 \leq ||f||_2^2 = L^2$, we have $\mathbb{P}\{|\nu_n''(f)| < 2L\} \geq \frac{3}{4}$ for all f in F. So

$$\mathbb{P}\{\sup|\nu_n^0(f)| > M\} = \int_{X^{2n}} P^{2n}\{\sup|\nu_n^0(f)| > M | \sigma(\xi_1, \ldots, \xi_n)\} dP^{2n}$$

$$= \int_{X^n}\int_{X^n} 1_{\sup|\nu_n^0(f)| > M}\, dP^n(\omega_1)\, dP^n(\omega_2).$$

Suppose $\omega_2 \in \{|\nu_n''(f)| > M + 2L\}$ and $f \in F$ is arbitrary. Then

$$\int_{X^n} 1_{\sup|\nu_n^0(f)| > M}\, dP^n(\omega_1) \geq \int_{X^n} 1_{|\nu_n'(f)| < 2L}\, dP^n(\omega_1) \geq \frac{3}{4}.$$

Since f is arbitrary,

$$\mathbb{P}\{\sup|\nu_n^0(f)| > M\} \leq \int_{\sup|\nu_n''(f)|>M+2L} \frac{3}{4} dP^n(\omega_2),$$

and the lemma follows. Q.E.D.

In the remainder of the proof write $||f||_{2n}$ for the $L^2(P_{2n})$ semi-norm of f. Theorem 5.1 of [30] shows that there exist A and $W \geq 1$ depending only on $v := V(C)$ such that $N(\delta, F) \leq A\delta^{-W}$.

Let $m_j := N(2^{-j}, F) \leq A2^{jW}$, $j \geq 1$. Suppose we are given the ordered $2n$-tuple $S := \langle x_1, \ldots, x_{2n} \rangle$. Then for all $j \geq 1$ we find $F_j := \{f_{j1}, \ldots, f_{jm_j}\}$ such that for all $f \in F$,

$$(2) \qquad \min_{i \leq m_j} [\int (f - f_{ji})^2 dP_{2n}]^{1/2} \leq 2^{-j} ||f||_{2n}.$$

Thus, for all $j \geq 1$ and $f \in F$ we may define $f_j(X)$ to be one of $f_1, \ldots, f_m$ such that
$\int (f - f_j(S))^2 dP_{2n} \leq 2^{-2j} ||f||^2_{2n}$.

Def. 2.2.3 Define $\eta_1 > 0$, $j \geq 1$, by $3jW \log 2 = \eta_j^2 2^{2j}/72\ell^2$. Note that $\Sigma\eta_j < \infty$.

Def. 2.2.4 Let $\beta := 2\gamma(\log 2)\ell^2$ and $r := r(M) := \left[\frac{M^2}{2W\beta}\right]$, where $[\cdot]$ denotes integer part.

Fact 2.2.5 If $M \geq M_0(\ell,W) := \sqrt{7}\, W^{1/2}(\log_2 86\ell W^{1/2})^{1/2}$,

then $\sum_{j=r+1}^{\infty} \eta_j < 1$.

Proof: By definition, $\eta_j := \sqrt{216jW \log 2}\ \ell \leq K\sqrt{j}2^{-j}$ where $\ell\sqrt{216(\log 2)W} \leq \frac{49}{4}\ell W^{1/2} := K$.

Note that $\eta_j \leq K2^{-j/2}$ $\forall$ $j \geq 1$ and therefore

$\sum_{r+1}^{\infty} \eta_j < 1$ if $K\frac{2^{-(r+1)/2}}{1-2^{-1/2}} < 1$. This last inequality will be satisfied if $2^{-(r+1)/2} < \frac{1}{7K}$, using $\frac{1-2^{-1/2}}{2} > \frac{1}{7}$.

Since $2^{-(r+1)/2} \leq 2^{-M^2/4W\beta}$ it suffices to choose M large enough so that $2^{-M^2/4W\beta} < \frac{1}{7K}$. Clearly, $M > (4W\beta \log_2(7K))^{1/2}$ will do. Note that

$\log_2(7K) \leq \log_2(86\ell W^{1/2})$ and $4W\beta \leq 10\ell^2 W \log 2 \leq 7\ell^2 W$.

We may choose $M_0(\ell,W) := \sqrt{7}\ell W^{1/2}(\log_2(86\ell W^{1/2}))^{1/2}$.

Q.E.D.

From now on consider any M such that $M \geq M_0(\ell,W)$.

For a given $f \in \mathcal{F}$ denote by f_j the function f_{ji} (for least i) in $\mathcal{F}_j$ for which the LHS of (2) achieves its minimum. Notice that $||f_j - f||_{2n} \to 0$ as $j \to \infty$. Thus $\nu_n^0(f_j(S)) \to \nu_n^0(f)$ as $j \to \infty$. Thus for any fixed s,

$$\nu_n^0(f) - \nu_n^0(f_s(S)) = \sum_{j=r+1}^{\infty} \nu_n^0(f_j(S)) - \nu_n^0(f_{j-1}(S)).$$

Omit the S in $f_k(S)$ and just write f_k.

If $M \geq M_0(\ell, W)$ then $\sum_{j=r+1}^{\infty} \eta_j < 1$, and it follows that either $|\gamma_n^0(f_r)| > M - 1$ or $\exists\, j > r$ such that $|\gamma_n^0(f_j) - \gamma_n^0(f_{j-1})| > \eta_j$. Note that

$$\begin{aligned} \int (f_j - f_{j-1})^2 dP_{2n} &\leq (||f - f_j||_{2n} + ||f - f_{j-1}||_{2n})^2 \\ &\leq ||f||_{2n}^2 (2^{-j} + 2^{-(j-1)})^2 \leq 9\ell^2 2^{-2j}. \end{aligned} \tag{3}$$

It follows that for our fixed values of $\langle x_1, \ldots, x_{2n} \rangle$,

$$\begin{aligned} Q^\infty[\sup|\nu_n^0(f)| > M] \leq\; & Q^\infty[\max_{i \leq m_r} |\nu_n^0(f_{ri})| > M - 1] + \\ & + \sum_{j=r+1}^{\infty} Q^\infty[\max_{\substack{i \leq m_j \\ k \leq m_{j-1}}}{}' |\nu_n^0(f_{ji}) - \nu_n^0(f_{(j-1)k})| > \eta_j], \end{aligned}$$

where max' denotes max for $||f_{ji}-f_{(j-1)k}||_{2n}^2 \leq 9\cdot 2^{-2j}\ell^2$. The LHS of (4) is

$$(5)\quad \leq m_r \max_{i\leq m_r} Q^\infty[|\nu_n^0(f_{ri})| > M-1] +$$

$$+ \sum_{j=r+1}^{\infty} m_j m_{j-1} \max_{i\leq m_j,\ k\leq m_{j-1}}{}' Q^\infty[|\nu_n^0(f_{ji}) - \nu_n^0(f_{(j-1)k})| > \eta_j].$$

Now $\nu_n^0(f_{ji}-f_{(j-1)k})$ can be written as $\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n} h_\ell$ where $h_\ell := (f_{ji}-f_{(j-1)k})(\xi_{2\ell}) - (f_{ji}-f_{(j-1)k})(\xi_{2\ell-1})$. By Theorem 2 of Hoeffding [20],

$$(6)\quad Q^\infty[|\nu_n^0(f_{ji}-f_{(j-1)k})| > \eta_j] \leq 2\exp\{-2n\eta_j^2 \frac{1}{4\Sigma_{\ell=1}^n h_\ell^2}\}.$$

Observe that $\sum_{\ell=1}^{n} h_\ell^2 \leq 2\sum_{\ell=1}^{2n}(f_{ji}-f_{(j-1)k})^2(\xi_\ell)$

$$(7)\quad \leq 36n\ell^2 2^{-2j},$$

by (3). Thus the second term on the RHS of (4) is

$$(8)\quad \leq 2\sum_{j=r+1}^{\infty} A^2 2^{2jW}\exp\{-\eta_j^2 2^{2j}\frac{1}{72\ell^2}\} = 2\sum_{j=r+1}^{\infty} A^2 2^{-jW}.$$

Applying Hoeffding again to the first term on RHS of (4) we get

$$Q^{\infty}[|\nu_n^0(f_{ri})| > M-1] \le 2\ \exp\{-2n(M-1)^2 \frac{1}{4n||f_{ri}||_{2n}^2}\}$$

$$\le 2\ \exp\{-(M-1)^2 \frac{1}{2\ell^2}\},$$

using $||f_{ri}||_{2n} \le \ell$. Note that this holds for all i, $i \le m_r$. Now $m_r \le A_2^{rW} \le A\ \exp\{\frac{M^2 \log 2}{2\beta}\} = A\ \exp\{\frac{M^2}{4\gamma\ell^2}\}$. It is easily checked that for M large and $\gamma > 1$, that $\frac{M^2}{4\gamma} - \frac{(M-1)^2}{2} = -\frac{(M+2)^2}{4\gamma} \le -\frac{(M+2)^2}{5}$. Thus the first term on the RHS of (4) is bounded by $2A\ \exp\{\frac{-(M+2)^2}{5\ell^2}\}$. The second term is bounded by (8) and is (where we may and do assume $W \ge 2$) $= 2 \sum_{j=r+1}^{\infty} A^2 2^{-jW} \le 4A^2 2^{-(r+1)W}$ $\le 4A^2 2^{-M^2/2\beta} = 4A^2 \exp\{\frac{-M^2}{4\gamma\ell^2}\} \le 4A^2\ \exp\{\frac{-(M+2)^2}{5\ell^2}\}$ if M is large enough. Thus

$$Q^{\infty}[\sup_{f \in F}|\nu_n^0(f)| > M] \le (2A+4A^2)\exp\{\frac{-(M+2)^2}{5\ell^2}\},$$

and therefore by Lemma 2.2.2

$$\Pr\{\sup_{f \in F}|\nu_n(f)| > M+2L\} \le \frac{4}{3}(4A^2+2A)\exp\{\frac{-(M+2)^2}{5\ell^2}\}.$$

Let $K := \frac{4}{3}(4A^2+2A)$ to complete the proof. Q.E.D.

The bound $N(\delta, F) \le A\delta^{-W}$ is crucial to the proof of the above theorem. The next theorem shows that under a stronger restriction on the envelope, the bound may be relaxed.

<u>Theorem 2.2.6</u> Let (X,A,P) be a probability space, F a class of real-valued functions with envelope $F \leq 1$. Suppose that $N(\delta,F) \leq \exp(\frac{C}{\delta^{2-\varepsilon}})$, for some constants C,ε where $C \geq 1$, $0 < \varepsilon < 1$, for all δ, $0 < \delta < 1$. Assuming that F is P-EM and 2-sample P-EM, we have for all $n \geq 1$

$$\Pr\{\sup_{f\in F}|\nu_n(f)| > M\} \leq 8\cdot\exp\{\frac{-M^2}{5}\},$$

where $M \geq M(\varepsilon,C) := 2 + \max(37,(5C)^{1/2}(\frac{120\sqrt{6C}}{\varepsilon})^{2/\varepsilon})$.

<u>Proof</u>. Define ν_n', ν_n'', ν_m^0, Pr, and $\mathbb{P}$ as in the proof of Theorem 2.2.1. Using Lemma 2.2.2 and its proof we have

<u>Lemma 2.2.7</u> $\Pr\{\sup_{f\in F}|\nu_n''(f)| > M+2\} \leq \frac{4}{3}\mathbb{P}\{\sup_{f\in F}|\nu_n^0(f)| > M\}$.

We will need some facts and definitions:

<u>Fact 2.2.8</u> If $\gamma = \frac{9}{8}$ and if $M \geq 37$ then

$$\frac{M^2}{4\gamma} - \frac{(M-1)^2}{2} \leq \frac{-(M+2)^2}{4\gamma} \quad\text{and}\quad \frac{M^2}{4\gamma} \geq \frac{(M+2)^2}{5}.$$

<u>Fact 2.2.9</u> Let $\gamma = \frac{9}{8}$ and $r := r(M) := [\frac{1}{2-\varepsilon}\log_2 \frac{M^2}{4\gamma C}]$. If $M \geq (5C)^{1/2}(\frac{120\sqrt{6C}}{\varepsilon})^{2/\varepsilon}$ then $\frac{2^{-\varepsilon(r+1)/2}(216C)^{1/2}}{1-2^{-\varepsilon/2}} < 1$.

Proof: The inequality will be satisfied if

$2^{-\varepsilon(r+1)/2} < \frac{\varepsilon}{2}(\log 2)(\frac{1-\log 2}{2})(\frac{1}{216C})^{1/2}$ which in turn is satisfied if $2^{-\varepsilon(r+1)/2} < \frac{\varepsilon}{20}(\frac{1}{216C})^{1/2} = \frac{\varepsilon}{120\sqrt{6C}}$.

Since $2^{-\varepsilon(r+1)} \leq 2^{\frac{-\varepsilon}{2-\varepsilon}\log_2(\frac{M^2}{4\gamma C})} = (\frac{M^2}{4\gamma C})^{\frac{-\varepsilon}{2-\varepsilon}}$, this implies $2^{-\varepsilon(r+1)/2} \leq (\frac{M^2}{4\gamma C})^{\frac{-\varepsilon}{2(2-\varepsilon)}}$. Thus it suffices to choose M large enough so that $(\frac{M^2}{4\gamma C})^{\frac{-\varepsilon}{2(2-\varepsilon)}} < \frac{\varepsilon}{120\sqrt{6C}}$. Thus $M \geq (5C)^{1/2}(\frac{120\sqrt{6C}}{\varepsilon})^{2/\varepsilon}$ will work. Q.E.D.

Def. 2.2.10 Let $\eta_j > 0$ be such that $\eta_j^2 := 2^{-\varepsilon j}(216C)$.

Def. 2.2.11 $m_j := N(2^{-j}, F) \leq \exp\{C2^{(2-\varepsilon)j}\}$.

Suppose that we are given the realization of $\xi_1, \ldots, \xi_{2n}$; i.e., we are given the values $\langle x_1, \ldots, x_{2n} \rangle$ which we will call S. Then $\forall j \geq 1$ we may find $F_j := \{f_{j1}, \ldots, f_{jm_j}\}$ such that

$$(9) \qquad \min_{i \leq m_j} [\int (f-f_{ji})^2 dP_{2n}]^{1/2} \leq 2^{-j}$$

for all $f \in F$. From now on consider any M such that $M \geq 2 + \max(37, \sqrt{5C}(\frac{120\sqrt{6C}}{\varepsilon})^{2/\varepsilon})$. Fact 2.2.9 and the definition of r and η_{ij} show that $\sum_{j=r+1}^{\infty}\eta_j < 1$.

Suppose that $|\nu_n^0(f)| > M$ for some $f \in F$. Then for any such f denote by $f_j(S)$ a function $f_{ji} \in F_j$ for which the LHS of (9) achieves its minimum. For any integer s

$$\nu_n^0(f) - \nu_n^0(f_s(S)) = \sum_{j=s+1}^{\infty} \nu_n^0(f_j(S)) - \nu_n^0(f_{j-1}(S)).$$

From now on, omit the S in $f_k(S)$ and just write f_k. Using $\sum_{r+1}^{\infty} \eta_j < 1$ it follows that either $|\nu_n^0(f_r)| > M-1$ or there is a $j > r$ such that $|\nu_n^0(f_j) - \nu_n^0(f_{j-1})| > \eta_j$. As in (3)

$$\int (f-f_{j-1})^2 dP_{2n} \leq 9 \cdot 2^{-2j}.$$

As in the proof of Theorem 2.2.1 it follows that for our fixed values of $\langle x_1,\ldots,x_{2n}\rangle$,

$$(10) \quad Q^{\infty}[\sup|\nu_n^0(f)| > M] \leq m_r \max_{i \leq m_r} Q^{\infty}[|\nu_n^0(f_{ri}) > M-1] +$$

$$+ \sum_{j=r+1}^{\infty} m_j m_{j-1} \max_{1 \leq m_j, k \leq m_{j-1}}{}' Q^{\infty}[|\nu_n^0(f_{ji})-\nu_n^0(f_{(j-1)k})| > \eta_j],$$

where $\max'$ denotes max subject to $||f_{ji} - f_{(j-1)k}||_{2n}^2 \leq 9 \cdot 2^{-j}$. As in (6),

(11) $Q^\infty[|\nu_n^0(f_{ji}-f_{(j-1)k})| > \eta_j] \le 2\exp\{-2n\eta_j^2 \frac{1}{4\sum_{\ell=1}^n h_\ell^2}\}$,

where $\sum_{\ell=1}^{n} h_\ell^2 \le 36n\cdot 2^{-2j}$. Thus the second term on the RHS of (10) is

$$(12) \quad \le 2\sum_{j=r+1}^{\infty} \exp\{2C\cdot 2^{(2-\varepsilon)j}\}\exp\{-\eta_j^2 2^{2j}\,\tfrac{1}{72}\}$$

$$= 2\sum_{j=r+1}^{\infty} \exp\{-C2^{(2-\varepsilon)j}\}.$$

Applying Hoeffding [20] to the first term on the RHS of (10) we get $Q^\infty[|\nu_n^0(f_{ri})| > M-1] \le 2\exp\{-\frac{(M-1)^2}{2}\}$, which holds for all i, $i \le m_r$. Now $m_r \le \exp\{C2^{(2-\varepsilon)r}\} \le$ $\le \exp\{\frac{M^2}{4\gamma}\}$ by the way r was chosen. Using Fact 2.2.8, the first term on the RHS of (10) is thus bounded by $2\exp\{\frac{-(M+2)^2}{5}\}$. The second term on the RHS of (10) is $\le 2\sum_{j=r+1}^{\infty}\exp\{-C2^{(2-\varepsilon)j}\} \le 4\exp\{-C2^{(2-\varepsilon)(r+1)}\}$ $\le 4\exp\{\frac{-M^2}{4\gamma}\} \le 4\exp\{\frac{(M+2)^2}{5}\}$. So on S we have $Q^\infty[\sup_{f\in F}|\nu_n^0(f)| > M] \le 6\exp\{\frac{-(M+2)^2}{5}\}$. Integrating over X^{2n} and applying Lemma 2.2.7 gives the desired result. Q.E.D.

Corollary 2.2.12 Let $(X,\mathcal{A},P)$ be a probability space, $\mathcal{F}$ a class of real valued functions with envelope $F \equiv 1$. Let $N(\delta,\mathcal{F},\sup)$ be the least integer such that $\forall f \in \mathcal{F}$ there exist $f_1,\ldots,f_m \in \mathcal{F}$ such that $||f-f_i||_{\sup} < \delta$ for some $i \leq m$. Suppose that $N(\delta,\mathcal{F},\sup) \leq \exp(\frac{C}{\delta^{2-\varepsilon}})$ for some constants C,ε where $C \geq 1$, $0 < \varepsilon < 1$, and all δ, $0 < \delta < 1$. Assuming that $\mathcal{F}$ is P-EM and 2-sample P-EM,

$$\Pr\{\sup_{f\in\mathcal{F}}|\nu_n(f)| > M\} \leq 8\exp\{\frac{-M^2}{5}\}$$

$n \geq 1$ and for $M \geq 2 + \max(37, \sqrt{5C}(\frac{120\sqrt{6C}}{\varepsilon})^{2/\varepsilon})$.

Proof: Given $f \in \mathcal{F}$ find the f_i from the class $f_1,\ldots,f_m \in \mathcal{F}$ such that $||f-f_i||_{\sup} < \delta$. Then for all values of $P_{2n}(\omega)$, $\int(f-f_i)^2 dP_{2n}(\omega) < \delta^2$. Therefore $N(\delta,\mathcal{F}) \leq m := N(\delta,\mathcal{F},\sup) \leq \exp(\frac{C}{2-\varepsilon})$. The hypotheses of Theorem 2.2.6 are satisfied and the result follows.

Q.E.D.

Here is an example showing that the exponent for δ appearing in Theorem 2.2.6 and Corollary 2.2.13 can not be $2+\varepsilon$, $\varepsilon > 0$.

Following [14] define a class of functions $\mathcal{F}_{d,\alpha,K}$ as follows. Let $\alpha > 0$ and $K > 0$ and let β be the greatest integer $< \alpha$. Let

$$D^p := \partial^{[p]}/\partial x_1^{p_1} \cdots \partial x_d^{p_d},\ [p] := p_1 + \ldots + p_d, \text{ for}$$

p_i integers ≥ 0, $p = (p_1, \ldots, p_d)$. For a function f on $\mathbb{R}^d$ such that $D^p f$ is continuous whenever $[p] \leq \beta$, let

$$||f||_\alpha := \max_{[p]<\beta} \sup\{|D^p f(x)| : x \in \mathbb{R}^d\} +$$

$$+ \max_{[p]=\beta} \sup_{x \neq y}\{|D^p f(x) - D^p f(y)|/|x-y|^{\alpha-\beta}\},$$

where $|u| := (u_1^2 + \ldots + u_k^2)^{1/2}$, $u \in \mathbb{R}^k$.
Let I^d be the unit cube $\{x \in \mathbb{R}^d : 0 \leq x_j \leq 1, j = 1, \ldots, d\}$. As in [26] let $F_{d,\alpha,K} := \{f \text{ on } I^d : ||f||_\alpha \leq K,\ \alpha = q+r,$ $0 < r \leq 1,\ q \text{ some integer}\}$. Let $N_I(\delta, F, \sup) := \inf\{m:$ there exist $f_1, \ldots, f_m \in F$: for all $f \in F$ there are i,j: $f_i \leq f \leq f_j$ and $||f_j - f_i||_{\sup} < \delta\}$. In [26] bounds on metric entropy in the sup norm are established and from this it follows immediately that there are constants $m_{\alpha,d}$ and $M_{\alpha,d}$ such that

$$\frac{m_{\alpha,d}}{\delta^{d/\alpha}} \leq \log N_I(\delta, F_{d,\alpha,K}, \sup) \leq \frac{M_{\alpha,d}}{\delta^{d/\alpha}}.$$

Choose $K = d = 1$. Given $\varepsilon > 0$, choose $\alpha = \frac{1}{2+\varepsilon}$. Then

$\frac{d}{\alpha} = 2+\varepsilon$; i.e. the exponent for δ in Theorem 2.2.6 and Corollary 2.2.13 is $2+\varepsilon$.

However, Theorem 1 in [2] implies that there is a $\gamma = \gamma(1,\alpha) > 0$ such that for all possible values of P_n,

$$\sup_{f\in F_{1,\alpha,1}} \{\int f\,d(P_n-P)\} \geq \gamma n^{-\alpha/d} = \gamma n^{-\alpha}, \text{ and}$$

$$\sup_{f\in F_{1,\alpha,1}} \{\int f\,d\nu_n : ||f||_\alpha \leq 1\} \geq \gamma n^{\frac{1}{2}-\alpha} \to \infty,$$ completing the example.

Here is an example of a class of functions satisfying the metric entropy hypotheses of Corollary 2.2.12.

<u>Example 2.2.13</u> Let $F := \{f \in BL[0,1]$ such that the Lipschitz constant for f is 1 and $||f||_{sup} \leq 1\}$. Then $N(\delta,F) \leq (\frac{2}{\delta}+1)3^{\frac{1}{\delta}+1}$, which may be seen as follows. Consider a grill on $[0,1]\times[-1,1]$ with grill width δ. Considering the intersection points of the verticals and horizontals, construct the class F_δ of piecewise linear Lip 1 functions passing through the intersection points and linear between these points. The number of such functions can not exceed $(\frac{2}{\delta}+1)3^{\frac{1}{\delta}+1}$. Clearly, for all $f \in F$ there is a $f_i \in F_\delta$ such that $||f-f_i||_{sup} < \delta$

for some $i \leq (\frac{2}{\delta} + 1)3^{\frac{1}{\delta}+1}$.

Before considering other examples of classes of functions satisfying metric entropy hypotheses, we first generalize Theorem 2.2.6 to the case where the envelope $F \in L^p$, $p > 2$. Unfortunately, the resulting exponential bounds will only hold for large n.

Theorem 2.2.14 Let (X,A,P) be a probability space, $\mathcal{F}$ a class of real-valued functions with envelope $F \in L^p(X,A,P)$, $p > 2$. Let F have L^2 norm L and L^p norm K. Suppose $N(\delta,\mathcal{F}) \leq \exp(\frac{C}{\delta^{2-\varepsilon}})$ for some constants C,ε where $C \geq 1$, $0 < \varepsilon < 1$, and all δ, $0 < \delta < 1$. Assume that $\mathcal{F}$ is P-EM and 2-sample P-EM.

Then if $M \geq M(\varepsilon,C,L) := 2 + \max\{45, L(17C)^{1/2}(\frac{20L\sqrt{864C}}{\varepsilon})^{2/\varepsilon}\}$

$$\Pr\{\sup_{f\in\mathcal{F}}|\nu_n(f)| > M + 2L\} \leq \frac{28}{3}\exp\frac{-M^2}{17L^2}\},$$

provided that $n \geq n_0(M,p) := [A\exp(\frac{M^2}{17L^2})]^{\frac{2}{t-2}}$ where $t := \min(p,4)$ and A is some suitable large finite constant depending upon p and K.

Proof: We can assume $p \leq 4$. Define ν_n', ν_n'', ν_n^0, Pr, and $\mathbb{P}$ as in the proof of Theorem 2.2.1. Using Lemma 2.2.1 and its proof we have

<u>Lemma 2.2.15</u> $\Pr\{\sup_{f\in F}|\nu_n''(f)| > M + 2L\} \le \frac{4}{3}\mathbb{P}\{\sup_{f\in F}|\nu_n^0(f)| > M\}$.

We will need some facts and definitions:

<u>Fact 2.2.16</u> If $\gamma = \frac{33}{32}$ and $M \geq 45$ then

$$\frac{M^2}{16\gamma} - \frac{(M-1)^2}{8} \le \frac{-M^2}{17}.$$

<u>Fact 2.2.17</u> Let $\gamma = \frac{33}{32}$ and $r := r(M) := \frac{1}{2-\varepsilon}[\log_2(\frac{M^2}{16\gamma CL^2})]$.

If $M \geq (17C)^{1/2}L(\frac{20L\sqrt{864C}}{\varepsilon})^{2/\varepsilon}$ then $\frac{2^{-\varepsilon(r+1)/2}\sqrt{864C}L}{1-2^{-\varepsilon/2}} < 1$.

<u>Proof:</u> As in Fact 2.2.9 the inequality will be satisfied if $2^{-\varepsilon(r+1)/2} < \frac{\varepsilon}{2}(\log 2)(\frac{1-\log 2}{2})\frac{1}{\sqrt{864C}L}$. Since

$2^{-\varepsilon(r+1)/2} \le (\frac{M^2}{16\gamma CL^2})^{\frac{-\varepsilon}{2(2-\varepsilon)}}$ we need only choose M large

enough so that $(\frac{M^2}{16\gamma CL^2})^{\frac{-\varepsilon}{2(2-\varepsilon)}} < \frac{\varepsilon}{20}\frac{1}{\sqrt{864C}L}$. Thus

$M \geq (17C)^{1/2}L(\frac{20L\sqrt{864C}}{\varepsilon})^{2/\varepsilon}$ will work. Q.E.D.

<u>Def. 2.2.18</u> Let $T := \{\omega : ||F||_{2n} \le 2||F||_2\}$ where $||F||_{2n}$ is as on page 7.

<u>Def. 2.2.19</u> Let $\eta_j > 0$ be such that $\eta_j^2 := 2^{-\varepsilon j}864CL^2$.

<u>Def. 2.2.20</u> $m_j := N(2^{-j}, F) \le \exp\{C2^{(2-\varepsilon)j}\}$.

For $\omega \in T$, suppose that we are given the realization of $\xi_1,\ldots,\xi_{2n}$; i.e., we are given $\langle x_1,\ldots,x_{2n}\rangle$ which we will call S. Then for all $j \geq 1$ find $F_j := \{f_{j1},\ldots,f_{jm_j}\}$ such that

$$(13) \quad \min_{i\leq m_j} [\int (f-f_j)^2 dP_{2n}]^{1/2} \leq 2^{-j}||F||_{2n}$$

for all f in F. From now on consider any M such that $M \geq M(\varepsilon,C) := 2 + \max(45,(17C)^{1/2}L(\frac{20L\sqrt{864C}}{\varepsilon})^{2/\varepsilon})$. By Fact 2.2.17 and the definition of r and η_j, we have $\sum_{j=r+1}^{\infty} \eta_j < 1$.

Suppose that $|\nu_n^0(f)| > M$ for some $f \in F$. Then for any such f denote by $f_j(S)$ a function f_{ji} in F_j for which the LHS of (13) achieves its minimum. For any fixed integer s we have

$$\nu_n^0(f) - \nu_n^0(f_s(S)) = \textstyle\sum_{j=s+1}^{\infty} \nu_n^0(f_j(S)) - \nu_n^0(f_{j-1}(S)).$$

From now on omit the S in $f_k(S)$ and just write f_k. Using $\sum_{j=r+1}^{\infty} \eta_j < 1$, it follows that either $|\nu_n^0(f_r)| > M-1$ or there is a $j > r$ such that $|\nu_n^0(f_j) - \nu_n^0(f_{j-1})| > \eta_j$. Given $\omega \in T$ and the fixed realization S we have by (3):

$$\int (f_j - f_{j-1})^2 dP_{2n} \leq 9\cdot 2^{-2j}||F||_{2n}^2 \leq 36L^2 2^{-2j}.$$

Using the same method as in the proof of Theorem 2.2.1, it follows that for the fixed S

(14) $$Q^\infty[\sup|\nu_n^0(f)| > M] \leq m_r \max_{i\leq m_r} Q^\infty[|\nu_n^0(f_{ri}) > M-1] +$$

$$+ \sum_{j=r+1}^{\infty} m_j m_{j-1} \max_{i\leq m_j, k\leq m_{j-1}}{}' \; Q^\infty[|\nu_n^0(f_{ji})-\nu_n^0(f_{(j-1)k})| > \eta_j],$$

where max' denotes max subject to $||f_{ji}-f_{(j-1)k}||_{2n}^2 \leq 36L^2 2^{-2j}$. As in 6,

(15) $$Q^\infty[|\nu_n^0(f_{ji}-f_{(j-1)k}) > \eta_j] \leq 2\exp\{-2n\eta_j^2 \frac{1}{4\sum_{\ell=1}^n h_\ell^2}\},$$

where $\sum_{\ell-1}^{n} h_\ell^2 \leq 4n(36L^2 2^{-2j}) = 144nL^2 2^{-2j}$ on T. Thus the second term on the RHS of (14) is

(16) $$\leq 2\sum_{j=r+1}^{\infty} \exp\{2C\cdot 2^{(2-\varepsilon)j}\}\exp\{-\eta_j^2 2^{2j} \frac{1}{288L^2}\}$$

$$= 2\sum_{j=r+1}^{\infty} \exp\{-C2^{(2-\varepsilon)j}\}.$$

Applying Hoeffding [20] to the first term on the RHS of (14) we get $Q^\infty[|\nu_n^0(f_{ri})| > M-1] \leq 2\exp\{\frac{-2n(M-1)^2}{4n||f_{ri}||_{2n}^2}\}$.

Since $||f_{ri}||^2_{2n} \leq 4L^2$ on T we have

$Q^\infty[|\nu_n^0 f_{ri})| > M-1] \leq 2\exp\{-\frac{(M-1)^2}{8L^2}\}$, which holds for all i, $i \leq m_r$. Now $m_r \leq \exp\{C2^{(2-\varepsilon)r}\} \leq \exp\{\frac{M^2}{16\gamma L^2}\}$ by the way r was chosen. By Fact 2.2.16, the first term on the RHS of (14) is bounded by $2\exp\{\frac{-M^2}{17L^2}\}$. The second term on the RHS of (14) is

$\leq 2\sum_{j=r+1}^{\infty} \exp\{-C2^{(2-\varepsilon)j}\} \leq 4\exp\{-C2^{(2-\varepsilon)(r+1)}\}$ which is

$\leq 4\exp\{\frac{-M^2}{16\gamma L^2}\} \leq 4\exp\{\frac{-M^2}{17L^2}\}$. Therefore, on T

(17) $Q^\infty[\sup_f |\nu_n^0(f)| > M] \leq 6\exp\{\frac{-M^2}{17L^2}\}$.

Before completing the proof we will need

<u>Lemma 2.2.21</u> If $n \geq n_0(M,p) := [A\exp(\frac{M^2}{17L^2})]^{\frac{2}{p-2}}$, $2 < p < 4$, and $A < \infty$ is some suitably large constant depending upon K and p, then $P^\infty(T^c) \leq \exp\{\frac{-M^2}{17L^2}\}$.

For general p it suffices to take $n_0(M,p) :=$ $[A\exp(\frac{M^2}{17L^2})]^{2/t-2}$ where $t := \min(p,4)$.

<u>Proof</u>: We rely heavily upon Theorem 1a of [19].

Theorem: Let X_N for $N = 1,2,\ldots$ be a sequence of independent random variables with finite first absolute moments; let $a_{N,k}$ for $N,k = 1,2,\ldots$ be real numbers. Define $S_N := \sum_{k=1}^{\infty} a_{N,k}(X_k - EX_k)$. Let $\tau > 0$ and $\{\rho_N\}$ a sequence of positive numbers such that $\sum_k |a_{N,k}|^\tau < \rho_N$. If $1 \le \tau \le 2$ and $E|X_k - EX_k|^\tau \le M < \infty$ for all k, then for every $\varepsilon > 0$, $P\{|S_N| > \varepsilon\} = O(\rho_N)$.

Now let $X_\ell := F^2(\xi_\ell)$ for all $\ell = 1,2,\ldots$ and $\bar{S}_{2n} := X_1 + \ldots + X_{2n}$. Then $P^\infty\{\int F^2 dP_{2n} > 4\int F^2 dP\} = P^\infty\{\frac{\bar{S}_{2n}}{2n} > 4EX_1\} \le P^\infty\{|\frac{\bar{S}_{2n} - 2nEX_1}{2n}| > 3EX_1\}$.

Defining $a_{2n,k} := \frac{1}{2n}$ for $k = 1,\ldots,2n$ and 0 otherwise, we have $|\frac{\bar{S}_{2n} - 2nEX_1}{2n}| = |S_{2n}|$ by definition of S_{2n}. Let $\tau := \frac{p}{2}$. By assumption, $1 < \tau \le 2$ and $E|X_\ell - EX_\ell|^\tau \le 2^\tau E|X_\ell|^{p/2} \le 2^\tau K^p$ by definition of K. Letting $\rho_{2n} = \sum_{k=1}^{2n} (2n)^{-\tau} = (2n)^{1-\tau}$ gives $P^\infty\{\int F^2 dP_{2n} > 4\int F^2 dP\} \le P^\infty\{|S_{2n}| > 3EX_1\} = O(n^{1-\tau})$ which is $\le An^{1-\tau}$ for some constant $A < \infty$ depending upon K and p. It thus suffices to find $n_0 = n_0(M,p)$ such that $n \ge n_0$ implies $An^{1-\tau} \le \exp\{\frac{-M^2}{17L^2}\}$. Clearly, we may take $n_0 = [A \exp\{\frac{M^2}{17L^2}\}]^{2/p-2}$. Q.E.D.

Returning to the proof of the theorem, combine Lemma 2.2.15 and 2.2.21 with (17), showing that if $M \geq M(\varepsilon,C)$ and $n \geq n_0(M,p)$ then $\Pr\{\sup_{f\in F}|\nu_n(f)| > M + 2L\} \leq$

$$\frac{4}{3} \Pr\{\sup|\nu_n^0(f)| > M\} = \frac{4}{3} \int_{X^\infty} \int_{Y^\infty} 1_{\{\sup|\nu_n^0(f)|>M\}} dQ^\infty dP^\infty \leq$$

$\frac{28}{3} \exp\{\frac{-M^2}{17L^2}\}$. Q.E.D.

We note that if the hypothesis $p > 2$ of the above theorem is strengthened to $p = 2$ then exponential bounds are still obtainable for $n \geq n_0$, although it is not possible (to my knowledge) to obtain a precise value of n_0.

Now we provide examples of classes F satisfying

$$(18) \qquad N(\delta,F) \leq \exp(\frac{C}{\delta^{2-\varepsilon}})$$

for some constants C,ε where $C \geq 1$, $0 < \varepsilon < 1$, and all δ, $0 < \delta < 1$. Theorems 2.2.6 and 2.2.14 will then apply, depending upon whether $F \equiv 1$ or $F \in L^2(X,A,P)$ respectively.

Example 2.2.22 Let G be any class of functions satisfying (18) with envelope $G \in L^2(X,A,P)$. Let

$F := \{gl_C, g \in G, C \in C\}$ where C is a VCC. Then F satisfies (18).

Proof: Given P_{2n} let gl_C be any arbitrary element of F. For this g find the function g_i where $i \leq N(\frac{\delta}{2}, G)$ such that $\int (g-g_i)^2 dP_{2n} < \frac{1}{4}\delta^2 ||G||_{2n}^2$. Likewise, for l_C find the function l_{C_j} where $j \leq N(\frac{\delta}{2}, \{Gl_C; C \in C\})$ such that $\int (Gl_C - Gl_{C_j})^2 dP_{2n} < \frac{1}{4}\delta^2 ||G||_{2n}^2$.

Now $g_i l_{C_j}$ will serve as the approximating function to gl_C. We have $\int (gl_C - g_i l_{C_j})^2 dP_{2n} =$

$$= \int (gl_C - gl_{C_j} + gl_{C_j} - g_i l_{C_j})^2 dP_{2n}$$

$$\leq 2\int (gl_C - gl_{C_j})^2 dP_{2n} + 2\int (gl_{C_i} - g_i l_{C_j})^2 dP_{2n} = \delta^2.$$

Lastly, the cardinality of the set $\{g_i l_{C_j}\}$ is $\leq A\delta^{-W} \exp(\frac{C}{\delta^{2-\varepsilon}}) \leq \exp(\frac{2C}{\delta^{2-\varepsilon}})$ for small enough δ. Q.E.D.

Example 2.2.23 Let F_1 and F_2 be any two classes of functions satisfying (18) with envelopes F_1 and F_2 respectively. Let $F_1 \vee F_2 := \{f \vee g: f \in F_1, g \in F_2\}$. Then $F_1 \vee F_2$ also satisfies (18) for some C.

<u>Proof</u>: First assume that $N(\delta, F_i) \leq \exp(\frac{C_i}{\delta^{2-\varepsilon}})$ for $i = 1,2$ and some $\varepsilon := \varepsilon_1 \wedge \varepsilon_2$.

Consider any arbitrary element $f \vee g$ of $F_1 \vee F_2$ where f and g are arbitrary elements of F_1 and F_2 respectively. For f find the function f_i with $i \leq N(\frac{\delta}{2}, F_1)$ such that $\int (f-f_i)^2 dP_{2n} < \frac{1}{4}\delta^2 ||F_1||^2_{2n}$. Likewise, for g find the function g_j with $j \leq N(\frac{\delta}{2}, F_2)$ such that $\int (g-g_j)^2 dP_{2n} < \frac{1}{4}\delta^2 ||F_2||^2_{2n}$.

Now $f_i \vee g_j$ will serve as our approximating function to $f \vee g$. Using $|f \vee g - f_i \vee g_j| < |f-f_i| + |g-g_j|$, $\int (f \vee g - f_i - g_j)^2 dP_{2n}$ is

$$\leq 2 \int (f-f_i)^2 dP_{2n} + 2 \int (g-g_j)^2 dP_{2n}$$

$$\leq \frac{1}{2}\delta^2 (||F_1||^2_{2n} + ||F_2||^2_{2n})$$

$$\leq \delta^2 ||F_1 \vee F_2||^2_{2n},$$

where $F_1 \vee F_2$ is the envelope for $F_1 \vee F_2$. Finally, the cardinality of the set $\{f_i \vee g_j\}$ is $\leq \exp(\frac{16(C_1+C_2)}{\delta^{2-\varepsilon}})$.

Q.E.D.

The following theorem is a variation of Theorem 2.2.6, making use of a slightly different metric entropy condition.

Theorem 2.2.24 Let (X,A,P) be a probability space and F a class of real-valued functions on (X,A,P). Assume that F has envelope $F \equiv 1$. Given n, let $j(n) := [\frac{1}{2}\log_2 n] + 1$. Suppose that for all values of P_{2n} (except those in a set A_{2n} with $P(A_{2n}) := p_n \downarrow 0$ as $n \to \infty$) and for all j, $1 \le j \le j(n)$, that there exist functions $f_{j1},\ldots,f_{jm} \in F$, $m := m(j)$, such that for all $f \in F$ there is an $i \le m$ such that

$$(19)\qquad \int |f-f_{ji}|\,dP_{2n} < 2^{-j}.$$

Assume for some ε, $0 < \varepsilon < \frac{1}{2}$, and $C \ge 1$ that

$$(20)\qquad m := m(j) \le \exp\{C2^{(1-\varepsilon)}\} \quad j = 1,2,\ldots$$

Assuming that F is P-EM and 2-sample P-EM, and

$M \ge 2 + \max(37, (5C)^{1/2}(\frac{480\sqrt{C}}{\varepsilon})^{1/\varepsilon})$ we have for all $n \ge 1$

$$(21)\qquad \Pr\{\sup_{f\in F}|\nu_n(f)| > M\} \le 8\exp\{\frac{-M^2}{5}\} + \frac{4}{3}p_n.$$

Proof: Define ν_n', ν_n'', ν_n^0, Pr, and $\mathbb{P}$ as in the proof of Theorem 2.2.1. Using Lemma 2.2.3 and its proof we have

<u>Lemma 2.2.25</u> $\Pr\{\sup_{f\in F}|\nu_n''(f)| > M+2\} \le \frac{4}{3}\mathbb{P}\{\sup_{f\in F}|\nu_n^0(f)| > M\}$.

We will need some facts and definitions:

<u>Fact 2.2.26</u> If $\gamma = \frac{9}{8}$ and $M \ge 37$ then

$\frac{M^2}{4\gamma} - \frac{(M-1)^2}{2} \le -\frac{(M+2)^2}{4\gamma}$ and $\frac{M^2}{4\gamma} \ge \frac{(M+2)^2}{5}$.

<u>Fact 2.2.27</u> Let $\gamma = \frac{9}{8}$ and $r := r(M) := [\frac{1}{1-\varepsilon}\log_2 \frac{M^2}{4\gamma C}]$.

If $M \ge (5C)^{1/2}(\frac{480\sqrt{C}}{\varepsilon})^{1/\varepsilon}$, then $\frac{2^{-\varepsilon(r+1)/2}12\sqrt{C}}{1-2^{-\varepsilon/2}} < \frac{1}{2}$.

<u>Proof:</u> See the proof of Fact 2.2.9. Q.E.D.

<u>Def. 2.2.28</u> Let $\eta_j > 0$ be such that $\eta_j^2 := 2^{-\varepsilon j}(144C)$.

<u>Def. 2.2.29</u> $m_j := m(j) \le \exp\{C2^{(1-\varepsilon)j}\}$.

For $\omega \in A_{2n}^c$ suppose that we are given the realization of $\xi_1,\ldots,\xi_{2n}$; i.e., we are given $\langle x_1,\ldots,x_{2n}\rangle$, which we will call S. Then for all j, $1 \le j \le j(n)$, find $F_j := \{f_{j1},\ldots,f_{jm_j}\}$ such that

$$(22) \qquad \min_{i\le m_j} \int|f-f_{ji}|\,dP_{2n} < 2^{-j}$$

for all $f \in F$. From now on consider any M such that $M \ge M(\varepsilon,C) := 2 + \max(37, (5C)^{1/2}(\frac{480\sqrt{C}}{\varepsilon})^{1/\varepsilon})$. Fact 2.2.27

and the definition of r and η_j imply $\sum_{r+1}^{\infty} \eta_j < \frac{1}{2}$

From now on we will still be considering $\omega \in A_{2n}^c$ and so all equations will hold except with probability $P(A_{2n}) := p_n$.

Suppose that $|\nu_n^0(f)| > M$ for some $f \in F$. Then for any such f denote by $f_j(S)$ a function $f_{ji} \in F_j$ for which the LHS of (22) achieves its minimum. Notice that for any fixed integer s, $s < j(n)$,

$$\nu_n^0(f) - \nu_n^0(f_s(S)) = \sum_{j=s+1}^{j(n)} \nu_n^0(f_j(S)) - \nu_n^0(f_{j-1}(S)) + \tau(S)$$

where $|\tau(S)| \leq |\nu_n^0(f-f_{j(n)}(S))| \leq \sqrt{n}\ 2^{-j(n)} \leq \frac{1}{2}$, by definition of $j(n)$. From now on omit the S in $f_k(S)$ and just write f_k.

Using $\sum_{r+1}^{\infty} \eta_j < \frac{1}{2}$ it follows that either $|\nu_n^0(f_r)| > M-1$ or there is a $j > r$ such that $|\nu_n^0(f_j) - \nu_n^0(f_{j-1})| > \eta_j$.

Using the same method as in the proof of Theorem 2.2.1 it may be shown (using Facts 2.2.26 and 2.2.27 and Hoeffding's inequality) for the fixed S and for $\omega \in A_{2n}^c$ that $Q^{\infty}[\sup|\nu_n^0(f)| > M] \leq 6 \exp\{\frac{-(M+2)^2}{5}\}$. Combining this with

Lemma 2.2.25 and replacing $M+2$ by M will give the desired result. Q.E.D.

§3. A Bounded LIL

Remark 2.3.1. If $\mathcal{F}$ is a class of functions satisfying the hypotheses of Theorem 2.2.6 of Theorem 2.2.14 then

$$(1) \qquad \limsup_{n\to\infty} \sup_{f\in\mathcal{F}} \frac{|\nu_n(f)|}{\sqrt{2\log\log n}} \le \sup_{f\in\mathcal{F}} \sigma(f) \quad \text{a.s.}$$

where $\sigma^2(f) = \int f^2 dP - (\int f dP)^2$. To see this, observe that by Theorem 4.2 of [30], $\mathcal{F}$ is a DCOF. Since the envelope $F \in \mathcal{L}^2$, the hypotheses of Theorem 1.2 of [15] are satisfied. Let S be the space of all bounded real-valued functions on $\mathcal{F}$. If $\psi \in S$, let $||\psi|| := \sup\{|\psi(f)|: f \in \mathcal{F}\}$. By Theorem 1.2 [15], there is a sequence of iid Gaussian random variables $\{Y_j, j \ge 1\}$ on S such that

$$||\sum_{j\le n} f(x_j) - \int f dP - Y_j(f)|| = O((n \log\log n)^{1/2}) \quad \text{a.s.}$$

By [28], $\displaystyle \limsup_{n\to\infty} \sup_{f\in\mathcal{F}} \frac{||n^{-1/2} \sum_{j\le n} Y_j(f)||}{\sqrt{2\log\log n}} \le \sup_{f\in\mathcal{F}} \sigma(f)$ a.s., showing (1).

§4. Metric Entropy and Donsker Classes of Functions (DCOF)

In this section the relationship between $N_I^{(q)}(\varepsilon)$

(see Def. 1.1.3) and DCOF is discussed.

Suppose $q = 1$, $\mathcal{F}$ is a P-EM class with envelope $F \in \mathcal{L}^p(X)$ for some $p > 2$, and that for some γ, $0 < \gamma < 1 - \frac{2}{p}$, and some $M < \infty$, $N_I^{(1)}(\varepsilon,\mathcal{F},P) \leq \exp(M\varepsilon^{-\gamma})$ for ε small enough. Then by [13], $\mathcal{F}$ is a DCOF. However, if $1 - \frac{1}{p-1} < \gamma < 1$ and $2 < p < 3$ then $\mathcal{F}$ may or may not be a DCOF, as illustrated by examples 2.4.2 and 2.4.3. If p is fixed, $2 < p < 4$, and $\gamma < \frac{1}{2}$ then $\mathcal{F}$ may be a DCOF. Also, if $p \downarrow 2$ then $\mathcal{F}$ may not be a DCOF, even when $\gamma \downarrow 0$, $\gamma > 1 - \frac{1}{p-1}$.

Suppose $q = 2$ and $\mathcal{G} := \{\sum_m c_m f_m,\ c_m = 0 \text{ or } 1]$ and $\forall\, m \geq 1$, $f_m \in \mathcal{L}^2(X)$ and f_m have disjoint support. If $\forall\, \varepsilon > 0$, $N_I^{(2)}(\varepsilon,\mathcal{G},P) \leq 2^{M\varepsilon^{-\gamma}}$ where $M < \infty$ and $\gamma < 1$, then $\mathcal{G}$ is a DCOF. Conversely, if $\mathcal{G}$ is a DCOF then

$$\int_0^1 \sqrt{\log N_I^{(2)}(\varepsilon,\mathcal{G},P)}\,d\varepsilon < \infty .$$

Suppose $q > 2$ is fixed and $\mathcal{F}$ has envelope in $\mathcal{L}^p$, $p < \frac{q}{2} + 1$. If $N_I^{(2)}(\varepsilon) \leq 2^{\varepsilon^{-\gamma}}$, $2 < q < \gamma < \infty$, then examples 2.4.7 and 2.4.8 show that there are classes $\mathcal{F}$ of a special type which may or may not be DCOF.

Suppose $q < 2$ is fixed and $\mathcal{F}$ has envelope in $\mathcal{L}^p$, $2 \leq p < 4 - q$. If $N_I^{(q)}(\varepsilon) \leq 2^{\varepsilon^{-\gamma}}$ and $\frac{q(p-2)}{p-q} < \gamma < \frac{q}{2}$

then examples 2.4.9 and 2.4.10 show that there are classes $\mathcal{F}$ of a special type which may or may not be DCOF.

Thus for all q, $q \neq 2$, q-norm metric entropy with bracketing will not, in general, give sharp results. The case $q = 2$ is unresolved.

We turn to the results outlined above, discussing them in that order. Let $(X,\mathcal{A},P)$ be as above. For all $j \geq 1$ let $A_j \subset X$ be disjoint sets with $p_j := P(A_j) = C/j^\beta$ for some $\beta > 1$ and some constant $C < \infty$. Let $a_j := j^\alpha$, $\alpha > 0$.

<u>Def. 2.4.1</u> Let $\mathcal{F}_{\alpha,\beta,C} := \{\sum_j a_j s_j 1_{A_j},\ s_j = 0 \text{ or } 1\}$, a_j and A_j as above.
$\mathcal{F}_{\alpha,\beta,C}$ has envelope $F = \sum a_j 1_{A_j}$ and $EF^p = \sum a_j{}^p p_j$.

<u>Example 2.4.2</u> Here we show that for fixed p, $2 < p < 4$, and $N_I^{(1)}(\varepsilon) \geq 2^{\varepsilon^{-\gamma}}$ where γ is any number $< \frac{1}{2}$, that $\mathcal{F}_{\alpha,\beta,C}$ may be a DCOF.

Now $F \in \mathcal{L}^p$ iff $\beta > p\alpha + 1$ and by [13] $\mathcal{F}$ is a DCOF iff $\sum a_j \sqrt{p_j} < \infty$ iff $\beta > 2\alpha + 2$. So given α, choose β such that $\beta > \max(p\alpha+1, 2\alpha+2)$; i.e., $\mathcal{F}$ is a DCOF with envelope in $\mathcal{L}^p$.

Given $\varepsilon > 0$ we find a lower bound for $N_I^{(1)}(\varepsilon,F,P)$. If the individual terms $a_j p_j \geq \varepsilon$ for all $j \leq j_0 := j_0(\varepsilon)$ then

(1) $$N_I^{(1)}(\varepsilon) \geq 2^{j_0}.$$

Now $a_j p_j \geq \varepsilon$ iff $C_j^{\alpha-\beta} \geq \varepsilon$. So we may take $j_0 \simeq \varepsilon^{\frac{-1}{\beta-\alpha}}$, and if $\gamma = \frac{1}{\beta-\alpha}$ then $N_I^{(1)}(\varepsilon) \geq 2^{\varepsilon^{-\gamma}}$. Given α we wish to maximize γ. Let $\tau > 0$ be arbitrarily small and take $\beta := 2\alpha + 2 + \tau$. If α is chosen so that $\alpha(p-2) < 1 + \tau$ then $p\alpha + 1 < \beta$ and so $\beta > \max(p\alpha+1, 2\alpha+2)$.

<u>Fact</u> Given p, α, and τ as above $p \geq (p-2)(\alpha+2+\tau)$ iff $p \leq \frac{2(\alpha+2+\tau)}{\alpha+1+\tau} = 2 + \frac{2}{1+\alpha+\tau}$.

Now $\gamma = \frac{1}{\beta-\gamma} = \frac{1}{\alpha+2+\tau} \geq \frac{p-2}{p} = 1 - \frac{2}{p}$ by the Fact.

Moreover, if α is any positive number $< \frac{1}{p-2}$ and τ any small positive number, then all of the above calculations remain valid. So we may let the sum $\rho := \alpha + \tau \downarrow 0$, concluding that if $p \leq 2 + \frac{2}{1+\rho} = 4 - \frac{2\rho}{1+\rho} < 4$, then $\gamma \geq 1 - \frac{2}{p}$.

The above shows that for p fixed, $2 < p < 4$, and $N_I^{(1)}(\varepsilon) \geq 2^{\varepsilon^{-\gamma}}$ where γ is any number $< \frac{1}{2}$, that

F may be a DCOF. However, F need not be a DCOF, as illustrated by

<u>Example 2.4.3</u> Take $F := F_{\alpha,\beta,C}$. If $\gamma > 1 - \frac{1}{p-1}$, $2 < p < 3$, and $N_I^{(1)}(\varepsilon) \leq 2^{\varepsilon^{-\gamma}}$, then F may not be a DCOF.

To see this, take $p\alpha + 1 < \beta \leq 2\alpha + 2$; i.e., F is not a DCOF and the envelope $F \in L^p$. In this case we must have

(3) $(p-2)\alpha < 1$.

Given $\varepsilon > 0$, find an upper bound for $N_I^{(1)}(\varepsilon)$. If $j_0 := j_0(\varepsilon)$ and $\sum_{j \geq j_0} a_j p_j < \varepsilon$ then $N_I^{(1)}(\varepsilon) \leq 2 \cdot 2^{j_0}$. To find an upper bound for j_0, note that $\sum_{j \geq j_0} a_j p_j =$ $= \sum_{j \geq j_0} C_j^{\alpha-\beta} \leq \varepsilon$ if for some constant K, $Kj_0^{\alpha-\beta+1} \leq \varepsilon$. Therefore $j_0 \simeq \varepsilon^{\frac{-1}{\beta-\alpha-1}}$ and $N_I^{(1)}(\varepsilon) \leq 2^{\varepsilon^{-\gamma}}$ where $\gamma := \frac{1}{\beta-\alpha-1}$.

Given p, let's minimize α. Let $\beta = 2\alpha + 2$. Let δ be any arbitrarily small number, $0 < \delta < 1$, and let $\alpha = \frac{1-\delta}{p-2}$. Clearly, (3) is satisfied.

Therefore $\gamma = \frac{1}{\alpha+1} = \frac{p-2}{1-\delta+p-2} = 1 - \frac{1-\delta}{p-1-\delta}$ and so $\gamma \downarrow 1 - \frac{1}{p-1}$ as $\delta \downarrow 0$.

Finally, if $\alpha = \frac{1-\delta}{p-2}$ and $p < 3-\delta$ then a simple calculation shows $\gamma < \frac{1}{2}$.

Examples 2.4.2 and 2.4.3 show that if $2 < p < 3$ and $1 - \frac{1}{p-1} < \gamma < \frac{1}{2}$ then F may or may not be a DCOF.

In examining the case $q = 2$ it will be convenient to consider the class $G := \{\sum c_m f_m : c_m = 0 \text{ or } 1\}$ where $\forall\, m \geq 1$, $f_m \in L^2(X)$ and f_m have disjoint support.

There are three theorems.

<u>Theorem 2.4.4</u> G is a DCOF iff $\sum_m ||f_m||_2 < \infty$.

<u>Proof</u>: Suppose $\sum_m ||f_m||_2 < \infty$. Then $E(\nu_n(f_m))^2 =$ $= \int f_m^2 dP - (\int f_m dP)^2$ for all n and f_m, where $\nu_n := n^{1/2}(P_n - P)$. Thus $E|\nu_n(f_m)| \leq ||f_m||_2$ and $\sup_n E\sum_{j\geq m} |\nu_n(f_j)| \to 0$ as $m \to \infty$. So for any $\varepsilon > 0$,

$$\sup_n \Pr\{\sum_{j\geq m} |\nu_n(f_j)| > \varepsilon\} \to 0$$

as $m \to \infty$. So in Theorem 1.3 of [13], condition (b) holds. The other condition holds for G and so G is a DCOF.

Conversely, let G be a DCOF and suppose $\sum_m ||f_m||_2 = \infty$. Then $\sup_{f \in F} |G_p(f)| = +\infty$ a.s. To see this, note that $W_p(f_m)$ are independent and $\sum |W_p(f_m)| = +\infty$ a.s. by the 3-series theorem. So either $\sum W_p(f_m)^+ = +\infty$ in which case we take $c_m = 1_{W_p(f_m)>0}$, or $\sum W_p(f_m)^- = +\infty$ in which case we take $c_m = 1_{W_p(f_m)<0}$. So W_p is unbounded on G and therefore G_p is as well, showing that G is not a DCOF. Q.E.D.

<u>Theorem 2.4.5</u> Given G as above suppose that for all $\varepsilon > 0$, $N_I^{(2)}(\varepsilon) \leq 2^{M\varepsilon^{-\gamma}}$ where $M < \infty$ and $\gamma < 1$. Then G is a DCOF.

<u>Proof</u>: We need only show $\sum_m ||f_m||_2 < \infty$. If $||f_m||_2 \not\to 0$ then G is not totally bounded and for some $\varepsilon > 0$, $N_I^{(2)}(\varepsilon) = +\infty$. So $||f_m||_2 \to 0$ and we may assume that $||f_m||_2 \downarrow$. If there are N terms $||f_m||_2$ which are $> \varepsilon$ then $N_I^{(2)}(\varepsilon) \geq 2^N$.

Let m_1 be any positive integer. Find $\varepsilon_1 := \varepsilon_1(m_1)$ such that $m_1 = M\varepsilon_1^{-\gamma}$. Then $\varepsilon_1 = (\frac{M}{m_1})^{1/\gamma}$. There can only be $M\varepsilon^{-\gamma}$ terms $||f_m||_2$ which are $> \varepsilon_1$.

So if $m > m_1$ then $||f_m||_2 < \varepsilon_1$. In particular, $||f_{m_1+1}|| \leq (\frac{M}{m_1})^{1/\gamma}$. Since m_1 is arbitrary, $\sum_m ||f_m||_2 \leq ||f_1||_2 + M^{1/\gamma}\sum(\frac{1}{m})^{1/\gamma} < \infty$. Q.E.D.

<u>Theorem 2.4.6</u> Given G as above, suppose that G is a DCOF. Then

$$\int_0^1 \sqrt{\log N_I^{(2)}(\varepsilon,G,P)}\,d\varepsilon < \infty.$$

<u>Proof</u>: From Theorem 2.4.4, $\sum_m ||f_m||_2 < \infty$. Following Dudley's proof of Borisov's theorem [5] and [16], it will be enough to show

$$\sum_{k=1}^{\infty} \sqrt{\log N_I^{(2)}(2^{-k},G,P)}/2^k < \infty.$$

We may assume $||f_m||_2 \downarrow$. Let r_j be the number of values of m such that $4^{-j-1} < ||f_m||_2 < 4^{-j}$, $j = 0,1,\ldots$ and $c_j := r_j/4^j$. Then $\sum c_j < \infty$. For each $k = 1,2,\ldots$ let j_k be such that

(4) $$\sum_{j\geq k} c_j/4^j \leq 1/4^k < \sum_{j\geq j_k} c_j/4^j.$$

Let $m_k := \sum_{j=0}^{k} r_j$. Then $\sum_{m>m_k} ||f_m||_2^2 \leq \sum_{j>j_k} r_j/16j \leq 1/4^k$.

Let A_i run over all subsets of $\{f_1,\ldots,f_{m_k}\}$ where $i = 1,\ldots,2^{m_k}$. Let $Y := \{f_m\}_{m=1}^{\infty}$. Let $B_i :=$ $A_i \cup \{f_m \quad Y\colon m > m_k\}$. Identify any subset $F \subset Y$ with the function $f := \{\sum f_m\colon f_m \in F\}$.

Then for any $F \subset Y$ let $A_i = F \cap \{f_1,\ldots,f_{m_k}\}$. Then $A_i \subset F \subset B_i$ and $\int (B_i - A_i)^2 dP = \int (\sum_{m>m_k} f_m)^2 dP$ $\leq \sum_{m>m_k} ||f_m||_2^2 \leq 4^{-k}$. Thus, $N_I^{(2)}(2^{-k}) \leq 2^{m_k+1}$, and it will be enough to prove $\sum_k \frac{m_k^{1/2}}{2^k} < \infty$. This is done in [16], and is as follows. Letting $j(k) := j_k$ we have

$$\sum_k \frac{m_k^{1/2}}{2^k} = \sum_k (\sum_{j=1}^{j(k)} 4^j c_j)^{1/2}/2^k$$

$$\leq \sum_k \sum_{j=1}^{j(k)} 2^{j-k} c_j^{1/2} = \sum_{j=1}^{\infty} c_j^{1/2} \sum_{k:j\leq j(k)} 2^{j-k}.$$

To prove that this converges, since $\sum c_j < \infty$, it is enough by Cauchy's inequality to prove

$$\sum_j (\sum_{k:j\leq j(k)} 2^{j-k})^2 < \infty.$$

Let $k(j) := k_j$ be the smallest k such that $j(k) \geq j$.

Then

$$\sum_{k:j\leq j(k)} 2^{j-k} \leq 2^{j+1-k(j)}.$$

We must now prove that $\sum_j 4^{j-k(j)} < \infty$. Setting $j(0) := 0$ we have

$$\sum_j 4^{j-k(j)} \leq \sum_k \sum_{j:j(k-1)<j\leq j(k)} 4^{j-k(j)} \leq \sum_k 4^{1+j(k)-k}.$$

Now for each k, let $K(k)$ be the smallest K such that $j_k = j_K$. Then from (4) for K, and letting $\mathcal{K}$ denote the range of $K(\cdot)$, we have

$$4^{-K(k)} < \sum_{j\geq j(k)} c_j/4^j.$$

Thus $\sum\limits_k 4^{j(k)-k} \leq \sum\limits_k 4^{j(k)-k+K(k)} \sum\limits_{j\geq j(k)} c_j/4^j$

$$= \sum_j c_j 4^{-j} \sum_{K\in\mathcal{K}, j(K)\leq j} 4^{j(K)+K} \sum_{k:K(k)=K} 4^{-k}$$

$$\leq 2\sum_j c_j 4^{-j} \sum_{K\in\mathcal{K}, j(k)\leq j} 4^{j(K)}$$

Since $j(\cdot)$ is one-to-one on $\mathcal{K}$, the above sum is $\leq 4\sum\limits_j c_j < \infty$. Q.E.D.

Now consider the case $q > 2$. If $N_I^{(q)}(\varepsilon) \geq 2^{\varepsilon^{-\gamma}}$ then the following example shows that F may be a DCOF as $\gamma \uparrow \infty$ and when F has envelope in L^p, $p < \frac{q}{2} + 1$.

Example 2.4.7 Take $F := F_{\alpha,\beta,c}$. Let $\delta > 0$ be some arbitrary small number and take $\beta = 2\alpha + 2 + \delta$. Let $\varepsilon > 0$ and find a lower bound for $N_I^{(q)}(\varepsilon)$.

If the individual terms $a_j^q p_j \geq \varepsilon^q$ for all $j \leq j_0 := j_0(\varepsilon)$ then $N_I^{(q)}(\varepsilon) \geq 2^{j_0}$. Now $a_j^q p_j \geq \varepsilon^q$ iff $j^{q\alpha-\beta} \geq \varepsilon^q$ iff $j_0 \cong \varepsilon^{\frac{-q}{\beta-q\alpha}}$. So if $\gamma \simeq \frac{q}{(2-q)\alpha+2+\delta}$, then $N_I^{(q)}(\varepsilon) \geq 2^{\varepsilon^{-\gamma}}$. Given q, let $\alpha = \frac{2}{q-2} > 0$. Then $\gamma \geq \frac{q}{\delta} \uparrow \infty$ as $\delta \downarrow 0$. These computations hold even when F has envelope in L^p, p such that $p\alpha + 1 < \beta := 2\alpha + 2 + \delta$. In this case $p < \frac{1}{\alpha} + 2 = \frac{q}{2} + 1$.

If $N_I^{(q)}(\varepsilon) \leq 2^{\varepsilon^{-\gamma}}$ where $\gamma > q > 2$ then F may not be a DCOF as illustrated by

Example 2.4.8 Let $F := F_{\alpha,\beta,c}$ and take $\beta \leq 2\alpha + 2$; i.e., F is not a DCOF. Given $\varepsilon > 0$ we find an upper bound for $N_I^{(q)}(\varepsilon)$. If for some $j_0 := j_0(\varepsilon)$

(5) $$\int \Big(\sum_{j \geq j_0} a_j 1_{A_j}\Big)^q dP \leq \varepsilon^q,$$

then $N_I^{(q)}(\varepsilon) \leq 2 \cdot 2^{j_0}$. Since A_j are disjoint, (5) follows from

$$\int \Big(\sum_{j \geq j_0} a_j 1_{A_j} \Big)^q dP = \sum_{j \geq j_0} a_j^q p_j = \sum_{j \geq j_0} j^{q\alpha - \beta} \leq \varepsilon^q ,$$

where the last inequality will be true if $j_0^{q\alpha-\beta+1} \simeq \varepsilon^q$; i.e., $j_0 \simeq \varepsilon^{\frac{-q}{\beta-q\alpha-1}}$. So if $\gamma \simeq \frac{q}{\beta-q\alpha-1}$ then $N_I^{(q)}(\varepsilon) \leq 2^{\varepsilon^{-\gamma}}$. Take $\beta = 2\alpha + 2$. Then $\gamma \simeq \frac{q}{(2-q)\alpha+1}$ where we are free to take α as small as we please. As $\alpha \downarrow 0$, $\gamma \downarrow q$.

These calculations hold even if F has envelope in L^p, p such that $p\alpha + 1 < \beta := 2\alpha + 2$. Moreover, as $\alpha \downarrow 0$ we may let $p \uparrow +\infty$. So if $p < \frac{q}{2} + 1$, $N_I^{(q)}(\varepsilon) \leq 2^{\varepsilon^{-\gamma}}$, and $2 < q < \gamma < \infty$, then examples 2.4.7 and 2.4.8 show that $F := F_{\alpha,\beta,c}$ may or may not be a DCOF.

Finally, consider the case $q < 2$. If $q < 2$ is fixed and $p < 4-q$ then $\frac{p-2}{p-q} < \frac{1}{2}$. Examples 2.4.9 and 2.4.10 show that if $F \in L^p$, $2 \leq p < 4-q$, $N_I^{(q)}(\varepsilon) \leq 2^{\varepsilon^{-\gamma}}$, and $\frac{q(p-2)}{p-q} < \gamma < \frac{q}{2}$, then $F := F_{\alpha,\beta,c}$ may or may not be a DCOF.

Example 2.4.9 Take $F := F_{\alpha,\beta,c}$ and $p\alpha+1 < \beta := 2\alpha+2$; i.e., F is not a DCOF and the envelope $F \in L^p$, $p \geq 2$. If $p > 2$ then clearly $\alpha < \frac{1}{p-2}$. The computations of example 2.4.8 show that if $\gamma \approx \frac{q}{\beta-q\alpha-1} = \frac{q}{(2-q)\alpha+1}$, then $N_I^{(q)}(\varepsilon) \leq 2^{\varepsilon^{-\gamma}}$.

Letting $\alpha \uparrow \frac{1}{p-2}$ shows that $\gamma \downarrow \frac{q(p-2)}{p-q}$. If $p = 2$ we may let $\alpha \uparrow \infty$ and $\gamma \downarrow 0$.

Example 2.4.10 As in example 2.4.7 take $F := F_{\alpha,\beta,c}$ where $\beta := 2\alpha + 2 + \delta$. Assume $F \in L^p$, $p \geq 2$. Then

$$(6) \qquad p\alpha + 1 < \beta := 2\alpha + 2 + \delta.$$

Now the calculations of example 2.4.7 hold for $q < 2$, showing that if $\gamma \approx \frac{q}{(2-q)\alpha+2+\delta}$ then $N_I^{(q)}(\varepsilon) \geq 2^{\varepsilon^{-\gamma}}$. Let $\alpha \downarrow 0$ and $\delta \downarrow 0$. We may let $p \uparrow \infty$ and still satisfy (6). So $\gamma \uparrow \frac{q}{2}$ showing that F may be a DCOF if $N_I^{(q)}(\varepsilon) \geq 2^{\varepsilon^{-\gamma}}$, $\gamma < \frac{q}{2}$.

Chapter 3

The Metric d_g

§1. Introduction

The exponential bounds of Chapter 2, together with a theorem of Dudley [13], will be used to study the properties of the convolution metric d_g on $\mathcal{P}(\mathbb{R}^k)$, where

$$d_g(P,Q) := \sup_{x \in \mathbb{R}^k} \left| \int g(x-y)\, d(P-Q)(y) \right|$$

and where g is any uniformly continuous density on $\mathbb{R}^k$ such that the Fourier transform $\hat{g}$ has a countable set of zeroes.

Let (S,d) be a separable metric space and $S_1, X_2, \ldots$ a sequence of independent identically distributed (iid) S-valued Borel measurable random variables. Let $\mathcal{P}(S)$ be the set of all Borel probability measures on S. Let $\mathcal{L}(X_1) = P$ and define the random (empirical) measures P_n as in Chapter 1.

On $\mathcal{P}(S)$ put the weak-star topology TW^*; i.e., the weakest topology such that the map $p \to \int f dp$ is continuous for each bounded and continuous $f: S \to \mathbb{R}$. The Glivenko-Cantelli theorem states that $P_n \to P$ weak-star as $n \to \infty$ almost surely (a.s.). This generalization of the Glivenko-Cantelli theorem is due to Varadarajan [36]. Weak-star convergence is metrizable and we shall discuss several metrics which induce TW*.

Denote by $BL(S)$ the set of all bounded Lipschitz functions $f: S \to \mathbb{R}$. Then $BL(S)$ is a Banach space with the norm

$$||f||_{BL} := \max\{||f||_\infty, ||f||_L\}$$

where

$$||f||_\infty := \sup_{x \in S} |f(x)|, \text{ and } ||f||_L := \sup_{\substack{x,y \in S \\ x \neq y}} |f(x)-f(y)|/d(x,y).$$

Let $BL(S,1) := \{f \in BL(S): ||f||_{BL} \leq 1\}$. For any $P,Q \in \mathcal{P}(S)$ define

$$\beta(P,Q) := \sup_{f \in BL(S,1)} |\int f(dP-dQ)|.$$

Then β is a metric on $\mathcal{P}(S)$ and the topology it induces is precisely TW* [9]. β was apparently first used by Fortet and Mourier [17], who proved that $\beta(P_n,P) \to 0$ a.s.

Now define the Prokhorov metric. For $x \in S$ and $T \subset S$, let $d(x,T) := \inf\{d(x,y): y \in T\}$. For $\delta > 0$ let $T^\delta := \{x \in S: d(x,T) < \delta\}$. Given any P,Q $\mathcal{P}(S)$, define

$\rho(P,Q) := \inf\{\varepsilon > 0 \colon P(F) \leq Q(F^{\varepsilon}) + \varepsilon$ for all closed $F \subset S\}$. Then ρ is a metric on $\mathcal{P}(S)$ and induces TW*, as proved in section 1.4 of [32] for S complete, and general separable S in section 2 of [10].

If S is complete and P,Q $\mathcal{P}(S)$, then the Prokhorov distance $\rho(P,Q)$ is the minimum distance "in probability" between two random variables distributed according to P and Q [34]. Dudley [10] extends this result with "minimum" replaced by "infimum" to an arbitrary separable metric space.

In $\mathbb{R}^1$, weak-star convergence is equivalent to convergence in the Lévy metric ρ_L where

$$\rho_L(P,Q) := \inf\{h\colon F(x-h)-h \leq G(x) \leq F(x+h) + h\},$$

where F and G are the distribution functions for P and Q respectively.

Weak-star convergence has the advantage that it takes into account the error which is inherent in the measurement of a random variable. For example, for any positive number σ, denote by F^{σ} the distribution of X + Y where X has the distribution F, and Y, independent of X, has a normal $N(0,\sigma^2)$ distribution with

mean 0 and variance σ^2. Following [21], define $\Delta^\sigma(P,Q) := \rho_1(F^\sigma, G^\sigma)$ where F and G are the distributions of P and Q respectively, and where for absolutely continuous distributions $\rho_1(F,G) := \int_\infty^\infty |F'(x)-G'(x)|dx$.
Then $P_n \xrightarrow{L} P$ iff for all $\sigma > 0$ $\Delta^\sigma(P_n,P) \to 0$ as $n \to \infty$.
The metric Δ^σ was apparently first defined by Kolmogorov in 1953 (25) but to my knowledge has been used little since.

If $P,Q \in \mathcal{P}(\mathbb{R})$ with distribution functions F and G respectively, and if ϕ is the density for the normal distribution $N(0,\sigma^2)$, then note that

$$(1)\quad \Delta^\sigma(P,Q) = \int_{-\infty}^{\infty} dx \left| \int_{-\infty}^{\infty} \phi^\sigma(x-y)(dF(y)-dG(y)) \right| .$$

Now define and consider the metric d_g which like Δ^σ, is also obtained by convolving the difference F-G of the distributions with a density g having certain desirable properties.

First define the metric d_g on $\mathcal{P}(G)$ where G is an arbitrary locally compact abelian (LCA) group. Before doing so, we recall from [33] a few facts about LCA groups.

§2. The Metric d_g

A complex function γ on a LCA group G is called a character of G if $|\gamma(x)| = 1$ $\forall\, x \in G$ and if

$\gamma(x+y) = \gamma(x)\gamma(y)\ \forall\ x,y \in G$. The set of all continuous characters of G forms a group Γ, the dual (or character) group of G, if addition is defined by $(\gamma_1+\gamma_2)(x) = \gamma_1(x)\gamma_2(x)\ \forall\ x \in G$ and $\gamma_1,\gamma_2 \in \Gamma$. In view of the duality between G and Γ it is customary to write (x,γ) in place of $\gamma(x)$.

Let dx be Haar measure on G. Then $\forall f \in L^1(G)$, the function $\hat{f}$ defined on Γ by

$$\hat{f}(\gamma) = \int_G f(x)(-x,\gamma)dx, \quad \gamma \in \Gamma$$

is called the Fourier transform of f. Let $B(G)$ be the set of all functions f on G which are finite linear combinations of continuous complex positive definite functions on G. From [33, pp. 21-33] we have

<u>Inversion Theorem</u>: (a) If $f \in L^1(G) \cap B(G)$ then $\hat{f} \in L^1(\Gamma)$. (b) If the Haar measure of G is fixed, then the Haar measure of Γ can be normalized so that the inversion formula $f(x) = \int_\Gamma \hat{f}(\gamma)(x,\gamma)d\gamma \quad x \in G$ is valid for every $f \in L^1(G) \cap B(G)$.

We will also need

<u>Fact 3.2.1</u> Given any separable LCA group G with dual group Γ, there is a real-valued function $h(\gamma)$ on Γ such that h has no zeroes on Γ and $\hat{h}(x)$ is

a uniformly continuous density on G.

Proof. Let $f > 0$ be a real valued continuous density on Γ for a Borel probability measure. Define $\tilde{f}(\gamma) := f(-\gamma)$ and $h := f * \tilde{f}$. Then it may be easily shown that h is continuous, positive-definite, and in $L^1(\Gamma)$, see e.g., [33], p. 18.

This h satisfies the conditions of the Inversion Theorem and using the notation formulated above,

$$h(\gamma) := (f*\tilde{f})(\gamma) = \int_{\Gamma} \widehat{(f*\tilde{f})}(x)(\gamma, x)\,dx$$

where $x \in G$. Thus, $\hat{h}(x) = |\hat{f}(x)|^2 \geq 0$ and $\hat{h}(x) \in L^1(G)$ using Inversion Theorem (a). Therefore $\hat{h}(x)$ is a density on G, normalizing if necessary.

Moreover, since $f*\tilde{f}$ is a denisty for a regular probability measure, $\hat{h}(x)$ is a bounded and uniformly continuous density. Q.E.D.

Having disposed of the preliminary details, we are ready to define the metric d_g on $\mathcal{P}(G)$.

Given the separable group G, let $g(x) = \hat{h}(x)$ where $h(\gamma)$ is as in Fact 3.2.1.

Given any two Borel probability measures P and Q in $\mathcal{P}(G)$, define the distance

(2) $d_g(P,Q) := \sup_{x \in G} |\int g(x-y)\, d(P-Q)(y)|.$

It is easily verified that d_g is a pseudo-metric on $\mathcal{P}(G)$. The following theorem shows that d_g also metrizes convergence in law.

Theorem 3.2.2 Suppose P_n, $P \in \mathcal{P}(G)$. Then $P_n \xrightarrow{L} P$ iff $d_g(P_n,P) \to 0$ as $n \to \infty$.

Proof: Let $\mu \in \mathcal{P}(G)$ be the probability measure in $\mathcal{P}(G)$ having density g with respect to Haar measure on G.

Suppose $d_g(P_n,P) \to 0$ as $n \to \infty$. By an easy application of Sheffé's theorem (see [4], pp. 223-224), we obtain $\mu * P_n \xrightarrow{L} \mu * P$ as $n \to \infty$.

Then $\widehat{\mu * P_n}(y) \to \widehat{\mu * P}(y)$ uniformly over compacta. Since $\hat{\mu}(y)$ is non-vanishing and continuous, $\hat{P}_n(y) \to \hat{P}(y)$ uniformly over compacta and thus $P_n \xrightarrow{L} P$.

Conversely, suppose $P_n \xrightarrow{L} P$ as $n \to \infty$. Since g is uniformly continuous and bounded with respect to Haar measure on G, the class $\mathcal{F} = \{g(x-y): y \quad G\}$ is a uniformly bounded equicontinuous class of functions. Therefore $P_n \to P$ uniformly over $\mathcal{F}$; see, for example [9], Theorem 7. This completes the proof. Q.E.D.

Having defined and shown the existence of the metric d_g on $\mathcal{P}(G)$ for arbitrary LCA groups G, let us consider the case when G is Euclidean space $\mathbb{R}^k$.

Suppose that g is a uniformly continuous density on $\mathbb{R}^k$ and allow $\hat{g}$ to have countable zeroes. Then under these relaxed conditions on g, the metric d_g defined by (2) metrizes convergence in law, as shown by

Theorem 3.2.3 Suppose $P_n, P \in \mathcal{P}(\mathbb{R}^k)$. Suppose g is a uniformly continuous density on $\mathbb{R}^k$ and $\hat{g}$ has a countable set of zeroes. Then $P_n \xrightarrow{L} P$ iff $d_g(P_n,P) \to 0$ as $n \to \infty$.

Proof: Let $Q \in \mathcal{P}(\mathbb{R}^k)$ be the probability measure in $\mathcal{P}(\mathbb{R}^k)$ having density g with respect to Lebesgue measure on $\mathbb{R}^k$.

Suppose $d_g(P_n,P) \to 0$ as $n \to \infty$. As in the proof of Theorem 3.2.2 we have $Q_n * P_n \xrightarrow{L} Q*P$. Therefore, for all $t \in \mathbb{R}^k$, $\widehat{Q*P_n}(t) \to \widehat{Q*P}(t)$. Thus $\hat{P}_n \cdot \hat{Q} \to \hat{P}\hat{Q}$ and $\hat{P}_n(t) \to \hat{P}(t)$ as $n \to \infty$ except possibly on the countable set of zeroes of $\hat{Q}(t)$. Let the countable set be $\{a_k\}$.

For $\{a_k\}$, find a subsequence $\hat{P}_{n_\ell}$ of $\hat{P}_n$ such that $\hat{P}_{n_\ell}(a_k)$ converges. Then $\hat{P}_{n_\ell}(t)$ converges for all $t \in \mathbb{R}^k$, the limit function is continuous around 0, and is thus a characteristic function and everywhere continuous. So $\hat{P}_{n_\ell}(t) \to h(t)$ for some characteristic function $h(t)$. Since $h(t)$ and $\hat{P}(t)$ are continuous, $h(t) = \hat{P}(t)$. Therefore $\hat{P}_{n_\ell} \to \hat{P}$ for all $t \in \mathbb{R}^k$, $P_n \xrightarrow{L} P$, and so $P_n \xrightarrow{L} P$.

The converse is established as in the proof of Theorem 3.2.2. Q.E.D.

Example. Let $g(x) := 0 \vee (1-|x|)$. Then $\hat{g}(t) = \frac{2(1-\cos t)}{t^2}$ which clearly has countable zeroes. So d_g metrizes convergence in law on $\mathcal{P}(\mathbb{R})$. In general, if g is any Lipschitz density on $\mathbb{R}^k$ such that g is any Lipschitz density on $\mathbb{R}^k$ such that $\hat{g}$ has countable zeroes, then d_g metrizes convergence in law on $\mathcal{P}(\mathbb{R}^k)$ and $d_g(P,Q) \leq K\beta(P,Q)$ where K is the Lipschitz constant for g.

Here is a necessary condition on g for d_g to be a metric on $\mathcal{P}(\mathbb{R}^k)$.

Proposition 3.2.4 If d_g is a metric on $\mathcal{P}(\mathbb{R}^k)$ then $\{t: \hat{g}(t) = 0\}$ must have empty interior.

Proof: Suppose $k = 1$ and $\{t: \hat{g}(t) = 0\} \supset \breve{U}$ where $\breve{U} = (-b,-a) \cup (a,b)$, $0 < a < b < \infty$, is a union of two intervals. Then there are characteristic functions h and j, $h \neq j$, such that $\{h \neq j\} \subseteq U$ and so d_g is not a metric. Before constructing h and j we recall Polya's theorem [31, pp. 116-118].

<u>Theorem</u> Let $f(t)$ be defined for all real values of t and suppose

1) $f(t)$ is real-valued and continuous

2) $f(0) = 1$

3) $\lim_{t\to\infty} f(t) = 0$

4) $f(t) = f(-t)$

5) $f(t)$ is convex for $t > 0$.

Then $f(t)$ is a characteristic function.

Now construct h and j, as in [31]. Now $h(t) := e^{-|t|}$ is a characteristic function. Consider the graph of $h(t)$. On U, replace the arc of the graph by its chord and call the resulting continuous and convex function $j(t)$; $j(t)$ is a characteristic function by Polya's theorem. Since $\{h \neq j\} \subseteq U$, this completes the case $k = 1$.

For general k suppose that $\{t: \hat{g}(t) = 0\} \supset U$ where $U \subseteq \mathbb{R}^k$ is open and symmetric about the origin. Again, we find characteristic functions f and g, $f \neq g$, such that $\{f \neq g\} \subset U$. Choose the coordinates $x_1, x_2, \ldots, x_k$ so that for some a and δ, $a > \delta > 0$, we have

$U \supset \{|x-a| < \delta, \ |(x_2,\ldots,x_k)| < \delta\}.$

From the case $k = 1$, there are characteristic functions $h(x_1)$ and $j(x_1)$ such that $h = j$ except on the set $|x_1 \pm a| < \delta$. Let $k(x_2,\ldots,x_k)$ be any characteristic function such that $k(y) = 0$ if $|y| \geq \delta$. Define the characteristic functions f and g by $f := h(x_1)k(x_2,\ldots,x_k)$ and $g := j(x_1)k(x_2,\ldots,x_k)$. Then $f \neq g$ and $\{f \neq g\} \subset U$.

Q.E.D.

§3. Comparison of Uniform Structures and a Bounded LIL

In this section we assume that g has Lipschitz constant K, support on the unit cube, and $\{t: \hat{g}(t) = 0\}$ is discrete.

Assuming that g has support on $[0,1]$, the uniform structure of d_g on $\mathcal{P}(\mathbb{R})$ is strictly weaker than that of β and ρ. Take P_n and $Q_n \in \mathcal{P}(\mathbb{R})$ where $P_n(2j) = Q_n(2j+1) = \frac{1}{n}$, $j = 1,\ldots,n$. Then

$d_g(P_n,Q_n) = \sup_x |\int g(x-y)\,d(P_n-Q_n)(y)| \leq \frac{K}{n}$ and

$\beta(P_n,Q_n) \geq \frac{1}{2}$. This example may be extended to $\mathbb{R}^k$.

On $[0,1]$ the topology of the Kolmogorov-Smirnov statistic $||P-Q||_\infty$ is strictly stronger than TW^*. To see this, take $P = \delta_0$ and $Q_n = \delta_{1/n}$. Then $Q_n \xrightarrow{L} P$, i.e., $d_g(P,Q) \to 0$. However, $||Q_n - P||_\infty \not\to 0$.

The Lévy metric ρ_L and d_g do not give the same uniform structure on $\mathcal{P}(\mathbb{R})$. Take P to be uniform on $[0,n]$ and Q uniform on $[n,2n]$. Then $\sup_x |\int g(x-y)\,dP(y)| \le \frac{1}{n}$ and $\sup_x |\int g(x-y)\,dQ(y)| \le \frac{1}{n}$. So $d_g(P,Q) \le \frac{2}{n}$ and $\rho_L(P,Q) \simeq 1$.

We now consider a generalized version of Δ^σ, as in (1). For all $\sigma > 0$, let $g^\sigma(x)$ be a collection of uniformly bounded (in x) and uniformly continuous probability densities with mean 0 and variance σ^2. Let $P,Q \in \mathcal{P}(\mathbb{R}^k)$ and define

$$(3) \quad \Delta_{g^\sigma}(P,Q) := \int_{\mathbb{R}^k} dx \left| \int_{-\infty}^{\infty} g^\sigma(x-y)\,d(P-Q)(y) \right|.$$

Let g_m be a collection of uniformly continuous probability densities of laws $\gamma_m \xrightarrow{\mathcal{L}} \gamma_0$ as $m \to \infty$. Let

$$(4) \quad \Delta_{g_m}(P,Q) := \int_{\mathbb{R}^k} dx \left| \int_{-\infty}^{\infty} g_m(x-y)\,d(P-Q)(y) \right|.$$

<u>Theorem 3.3.2</u> Let P_n and P $\mathcal{P}(\mathbb{R}^k)$. Then $P_n \xrightarrow{\mathcal{L}} P$ as $n \to \infty$ iff $\Delta_{g_m}(P_n,P) \to 0$ as $n \to \infty$ for all m.

<u>Proof:</u> Suppose $\Delta_{g_m}(P_n,P) \to 0$ as $n \to \infty$ for all m.

Then $g_m * P_n \xrightarrow{L} g_m * P$ as $n \to \infty$ for all m and $P_n \xrightarrow{L} P$ as $n \to \infty$. Conversely, suppose $P_n \xrightarrow{L} P$ as $n \to \infty$. Since P_n are uniformly tight, it follows that $P_n * g_m$ are uniformly tight for any density g_m. That is, if $B(0,M)$ is the ball of radius M centered at the origin of $\mathbb{R}^k$ then for any $\varepsilon > 0$ there is an M large enough so that $\int_{B(0,M)^c} P_n * g_m < \varepsilon$ and $\int_{B(0,M)^c} P * g_m < \varepsilon$ for $n \geq n_0$. Also, for any g_m we have $|(P_n - P) * g_m| \leq \frac{\varepsilon}{(2M)^k}$ for $n \geq n_1$ since $g_m(x - \cdot)$ are a class of uniformly bounded equicontinuous functions, see [9]. For $n \geq \max(n_1, n_0)$ we have $\int_{\mathbb{R}^k} |(P_n - P) * g_m| dx = \int_{B(0,M)} |(P_n - P) * g_m| dx +$ $\int_{B(0,M)^c} |(P_n - P) * g_m| dx \leq 2\varepsilon$, completing the proof. Q.E.D.

Corollary 3.3.3 Let P_n and $P \in \mathcal{P}(\mathbb{R}^k)$. Then $P_n \xrightarrow{L} P$ iff $\Delta_{g^\sigma}(P_n, P) \to 0$ for all $\sigma > 0$, where $\Delta_{g^\sigma}(P_n, P)$ is as in (3).

Finally, using the results of Chapter 2 we have the following a.s. upper bound for $d_g(P_n, P)$. Note that upper bounds for the β metric are less sharp.

Corollary 3.3.4 Let $P \in \mathcal{P}(\mathbb{R}^k)$ be a probability measure with support on the unit cube I^k. Let P_n be the empirical measures. If g is continuous, $|g| \leq 1$, and g has support on I^k, then $\forall\, \varepsilon > 0$

$$\lim_{n\to\infty} \frac{\sqrt{n}\; d_g(P_n,P)}{\sqrt{2 \log\log n}} \leq 1 \quad \text{a.s.}$$

Proof: Let $F := \{g(x-y);\ y \in \mathbb{R}^k\}$. Since the elements of F are non-zero iff $y \in [-1,2]^k$ we have $N(\delta,F) \leq \frac{C}{\delta^k}$ for some finite constant C and for all δ, $0 < \delta \leq 1$. Note that $N(\delta,F)$ is as it appears in Theorem 2.2.6. The hypotheses of Theorem 2.2.6 (or its corollary) are clearly satisfied and Remark 2.3.1 completes the proof. Q.E.D.

§4. An Invariance Principle for the Metric d_g

Theorem 3.4.1 Let g be a bounded density on $\mathbb{R}^k$ and X_i, $i = 1,2,\ldots$ iid random variables with $L(X) = P$. Let $F := \{g(x-\cdot):\ x \in \mathbb{R}^k\}$ and suppose that $\log N_I^{(2)}(x,F,P) \leq Cx^{-\tau}$ for some constants $C > 0$ and $0 \leq \tau < 1$. For any $\theta < \frac{1-\tau}{4\tau}$ there are random variables Y_j such that

$$(1)\quad \sup_{f\in F} \Big| \sum_{j\leq n} f(x_j) - \int f dP - Y_j(f) \Big| = o(n^{1/2}(\log n)^{-\theta}) \quad \text{a.s.}$$

Proof: Since $\int_0^1 \sqrt{\log N_I^{(2)}(x^2)\,dx} < \infty$, an examination of the proof of Theorem 5.1 of [12] shows that F is a DCOF. To see this, replace C by F, sets A in C by functions in F, $\nu_n(A)$ by $\int f d\nu_n$, etc., noting that since g is bounded, Bernstein's inequality still holds, as on pp. 915-916 of [12].

By Theorem 7.1 of [15] for any $\theta < \frac{1-\tau}{4\tau}$ we can choose Y_j in Theorem 1.3 of [15] such that (1) holds. This is true since the proof of Theorem 7.1 does not depend upon the use of sets in Theorem 5.1 of [12].

Q.E.D.

We remark that this result gives an invariance principle for $d_g(P_n,P)$ where P_n are empirical measures for $P \in \mathcal{P}(\mathbb{R}^k)$.

Corollary 3.4.2 Let $X = \mathbb{R}^k$ and $||\cdot||$ the usual Euclidean norm on $\mathbb{R}^k$. Let X_i, $i = 1,2,\ldots$ be iid $\mathbb{R}^k$-valued random variables such that for some $k > 2$ and $\beta > 2$

$$(2) \qquad P(||X|| > L) \leq K\left(\frac{1}{\log L}\right)^{\beta}$$

for all $L > 1$. Let g be a Lipschitz probability density on $\mathbb{R}^k$ such that for $||x|| > e$

$$g(x) \leq K'\left(\frac{1}{\log ||x||}\right)^{\gamma},$$

where $0 < K' < \infty$ and $\gamma > 1$. If $F := \{g(x-\cdot),\ x \in \mathbb{R}^1\}$ then (1) holds.

Proof: Given $\gamma > 1$, we may choose β so that $\gamma > \frac{\beta}{2}$. Let $\beta > \beta' > 2$. Given $\varepsilon > 0$, let $M := M(\varepsilon) := \exp(\frac{2KK'^2}{\varepsilon^4})^{1/\beta}$ and $y_0 := y_0(\varepsilon) := \exp(\frac{1}{\varepsilon^4})^{1/\beta'}$.

Given $a > 0$, let $B(a)$ be the ball centered at the origin with radius a. Let $x \in \mathbb{R}^k$ and

$$h(x) = K' 1_{B(M)^c}(x) + \frac{\varepsilon^2}{2} 1_{B(M)}(x).$$

Given M as above consider only those $\varepsilon > 0$ such that $y_0 \geq 2M$. For these ε observe that if $||y|| > y_0$ then

(3) $$0 < g(x-y) \leq h(x).$$

Now (3) is true since if $||z|| > M$ then

$$g(z) \leq K'(\frac{\varepsilon^4}{2KK'^2})^{\gamma/\beta} \leq \frac{1}{2}\varepsilon^2,$$

by definition of M, K, K', and β. Also, $\int h^2 dP \leq P(M(B)^c) + \frac{\varepsilon^4}{4} \leq \varepsilon^4$ and so $(\int h^2 dP)^{1/2} < \varepsilon^2$.

Now on $B(y_0)$ find $x_i \in \mathbb{R}^k$ dense within $\frac{\varepsilon^2}{2}$ and let

(4) $\qquad g_i^{\pm} = g(x-x_i) \pm \frac{\varepsilon^2}{2}$.

Using (3), (4), and the Lipschitz condition on g, we have for some $C < \infty$: $N_I^{(2)}(\varepsilon^2, F) \leq 2 + C(\frac{y_0}{\varepsilon^2})^k$. By definition of y_0, $N_I^{(2)}(\varepsilon^2) < 2 + C(\frac{1}{\varepsilon})^{2k}\exp\{k\varepsilon^{-4/\beta'}\}$, and for some constant $D < \infty$, $\log N_I^{(2)}(\varepsilon^2) \leq Dk\varepsilon^{-4/\beta'}$. Replacing ε^2 by x we have $\log N_I^{(2)}(x) \leq Dk_x^{-\tau}$ where $\tau = \frac{2}{\beta'} < 1$. The hypotheses of Theroem 3.4.1 are satisfied. completing the proof. Q.E.D.

§5. A CLT for d_g on $P(\mathbb{R}^k)$

In this section the metric d_g is not restricted to the case when g is Lipschitz. Let $d_g(P_n,P) := \sup_{x\in\mathbb{R}^k} |\frac{1}{n}\sum_{i=1}^{n} g(x-X_i) - Eg(x-X_i)|$ and $G(s,\omega) := g(s-X) - Eg(s-X)$.

Let $H^k := \{f: \mathbb{R}^k \to \mathbb{R}$, f any positive, continuous, bounded function such that

(i) if $x,y \in \mathbb{R}^k$ and $||x|| = ||y||$, then $g(x) = g(y)$, and

(ii) if $||x|| \leq ||y||$ then $g(x) > g(y)\}$.

The following results give sufficient conditions for $G(s,\omega)$ to satisfy the central limit theorem (CLT).

<u>Theorem 3.5.1</u> Suppose X is a $\mathbb{R}^1$-valued random variable such that $E\{\log(|X|+2)[\log\log(|X|+3)]^{2\alpha}\} < \infty$ for some $\alpha > 1$. Then $G(s,\omega)$ satisfy the CLT for all Lipschitz g with compact support.

<u>Theorem 3.5.2</u> If $g \in H^k$ then $G(s,\omega)$ satisfy the CLT for all $P \in P(\mathbb{R}^k)$.

<u>Proof of Theorem 3.5.1</u> Assume that g has support on $[0,1]$ and has Lipschitz constant $K \geq 4$. We will rely heavily upon Theorem 1 of [22]. Define $\mathbb{R}^*$ to be the 1-point compactification of $\mathbb{R}$ and $C_0(\mathbb{R}^*) := \{f: f$ continuous and $f(x) \to 0$ as $x \to \pm\infty\}$. Let $C^2 := E||g(s-X) - Eg(s-X)||^2$ where $||\cdot||$ denotes the sup norm on $C_0(\mathbb{R}^*)$.

<u>Lemma 3.5.3</u> For all $s,t \in \mathbb{R}^*$, define

$$e(s,t) := \sup_u \frac{|g(s-u) - g(t-u)|}{C(\log(|u|+3))^{1/2}(\log\log(|u|+16))^{\alpha}} .$$

Then e metrizes the topology of $\mathbb{R}^*$ and for all real x, s, and t

(1) $\frac{1}{C}|g(s-x)-g(t-x)| \leq \log(|x|+3)^{1/2}(\log\log(|x|+16)^{\alpha}e(s,t)$.

Proof of Lemma $e(s,t) = 0$ iff $s = t$, $e(s,t) = e(t,s)$, and

$$\frac{|g(s-u) - g(t-u)|}{(\log(|u|+3))^{1/2}(\log\log(|u|+16))^{\alpha}} \le$$

$$\le \frac{|g(s-u) - g(z-u)|}{(\log(|u|+3))^{1/2}(\log\log(|u|+16))^{\alpha}} +$$

$$+ \frac{|g(z-u) + g(t-u)|}{(\log(|u|+3))^{1/2}(\log\log(|u|+16))^{\alpha}}$$

implying the triangle inequality. So $e(s,t)$ is a metric and (1) is clearly satisfied. Q.E.D.

Returning to the proof of the theorem, note that if g is any continuous density with compact support then $g \in C_0(\mathbb{R}^*)$. For all $s \in \mathbb{R}^*$ let $G_i(s,\omega) :=$ $\frac{1}{C}[g(s-X_i(\omega) - EG(s-X_i(\omega)]$ and write $G(s,\omega)$ for $G_1(s,)$. Note that $EF(G) = 0$ for all $f \in C_0(\mathbb{R}^*)^*$, the dual to $C_0(\mathbb{R}^*)$. By the normalization, $E||G(s,\omega)||^2 = 1$ where $||\cdot||$ denotes sup norm on $C_0(\mathbb{R}^*)$. Note that $G(s,\omega)$ is a $C_0(\mathbb{R}^*)$-valued random variable on some probability space (Ω, F, P).

By Lemma 3.5.3, $|G(s,\omega)-G(t,\omega)| \le 2M(\omega)e(s,t)$, where $M(\omega) := \log(|X|+3)^{1/2}(\log\log(|X|+16))^{\alpha}$.

$M(\omega)$ is non-negative and $EM^2 < \infty$ by hypothesis.

Finally, for $\mathbb{R}^*$ equipped with the metric e, let $N_e(\mathbb{R}^*, 2^{-j})$ denote the minimum number of balls of e-radius $\leq 2^{-j}$ covering $\mathbb{R}^*$. Define $H_e(\mathbb{R}^*, 2^{-j}) :=$ $\log N_e(\mathbb{R}^*, 2^{-j})$.

<u>Lemma 3.5.4</u> Let ρ be such that $1 < \rho < \alpha$. If j is sufficiently large than $N_e(\mathbb{R}^*, 2^{-j}) \leq \frac{2K}{C}2^j(\exp\{2^{2j}(\frac{1}{j})^{2\rho}\})$.

<u>Proof of Lemma 3.5.4</u>: First note that whenever $\alpha > \rho$ and ρ is positive, a calculation shows that for j large,

$$\frac{1}{[2^{2j}(\frac{1}{j})^{2\rho}]^{1/2}(\log_2[2^{2j}(\frac{1}{j})^{2\rho}])^{\alpha}} \leq 2^{-j}.$$

Thus $e(s,\infty) \leq 2^{-j}$ if $s \geq \exp\{2^{2j}(\frac{1}{j})^{2\rho}\}$. Noting that $\ell n(|u|+3)\ell n(\ell n(|u|+16)) > 1$ for all u and using the Lipschitz condition on g we can obviously cover each of the $[s]+1$ intervals $[i,i+1]$, $0 \leq i \leq [s]$, with $\leq \frac{K}{C}2^j$ balls of e-radius $\leq 2^{-j}$. Considering also the closed interval $[-\infty,0]$ as well as $[0,\infty]$, we will clearly need at most $\frac{2K}{C}2^j\exp(2^{2j}(\frac{1}{j})^{2\rho})$ balls of e-radius $\leq 2^{-j}$ in order to cover $\mathbb{R}^*$. Q.E.D.

Finally, returning to the proof of Theorem 3.5.1, observe that $H_e(\mathbb{R}^*, 2^{-j}) \leq \log(\frac{2K}{C}2^j) + 2^{2j}(\frac{1}{j})^{2\rho}$ and

$\sum_j 2^{-j} H_e^{1/2}(\mathbb{R}^*, 2^{-j}) < \infty$, since $\rho > 1$. Applying Theorem 1 of [22], Theorem 3.5.1 follows. Q.E.D.

Consider now Theorem 3.5.2.

Proof of Theorem 3.5.2 We rely heavily upon Theorem 1.9 of [13]:

Theorem. Let $G := \{1_A : A \in C\}$ be a DCOF, $M < \infty$, and $\tilde{F} := \{f: X \to [-M,M],\ f^{-1}([a,b[) \in C$ whenever $a < b\}$. Then $\tilde{F}$ is a DCOF.

Consider $k = 1$ and take the density $g \in H^1$. Assume $g \leq M$ and observe that $g(\cdot - t)^{-1}([a,b[)$ is an interval or a union of two intervals. By [12, section 7], the set of all unions of two intervals is a VCC. So $\{g(\cdot - t) : t \in \mathbb{R}\}$ is a DCOF and $G(s,\omega)$ satisfy the CLT for all $P \in P(\mathbb{R}^1)$.

For the general case, take $g \in H^k$ and assume $g \leq M$. Let B_r^{k-1} be the ball with radius r centered at the origin of $\mathbb{R}^k$. Then $g(\ )^{-1}([a,b[)$ is the difference B_{rs}^{k-1} of two balls in $\mathbb{R}^k$ where $B_{rs}^{k-1} := B_r^{k-1} - B_s^{k-1}$ for some $0 < s < r$. Likewise, $g(\cdot - t)^{-1}([a,b[)$ is some translate $B_{rs}^{k-1} + t$ of some B_{rs}^{k-1}. The set of all B_{rs}^{k-1} and their translates forms a VCC. Applying Theorem 1.9 of [13] shows that $\{g(\cdot - t) : t \in \mathbb{R}^k\}$ is a DCOF and so $G(s,\omega)$ satisfy the CLT for all $P \in P(\mathbb{R}^k)$. Q.E.D.

§6. The Asymptotic Distribution of $d_g(P_n,P)$

In this section I determine the asymptotic distribution of $d_g(P_n,P)$ for special choices of g and where P_n are empirical measures for the measures P on $\mathbb{R}$. I rely heavily on the results of V. I. Dmitrovskiĭ [8].

Let g be a continuous density function on $\mathbb{R}$ and $F := \{\{g_t\}_{t\in\mathbb{R}}\}$ where $g_t(y) := g(t-y)$. Equip F with the e_P metric, where $e_P(f,g) := (\int (f-g)^2 dP)^{1/2}$. Following Dmitrovskiĭ [8], let

$$\zeta(g_t) = \lim_{n\to\infty} \sqrt{n} \int_{-\infty}^{\infty} g(t-y)(dP_n - dP)(y), \qquad g_t \in F$$

where P is a probability measure on $\mathbb{R}$ and P_n the empirical measures.

Denote by $N_{e_P}(F,\varepsilon)$ the minimal number of balls of radius $\leq \varepsilon$ which cover F. Write $N(\varepsilon)$ for $N_{e_P}(F,\varepsilon)$.

Suppose that

(i) $\sup\{E\zeta^2(g_t) : g_t \in F\} = 1$, and

(ii) the Dudley-Fernique conditions holds; i.e., $\int_0^1 [\ln N(x)]^{1/2} dx < \infty$.

Then by Theorem 3 of [8],

$$\Pr\{\sup_{g_t \in F} |\zeta(g_t)| > u\} = \exp\{\frac{-u^2}{2} + u\cdot 0(1)\}, \quad u \to \infty;$$

in other words,

$$\Pr\{\lim_{n\to\infty} \sqrt{n}\; d_g(P_n, P) > u\} = \exp\{\frac{-u^2}{2} + u\cdot 0(1)\}, \; u \to \infty.$$

The following example illustrates an application of Theorem 3.

Example 3.6.1 Let $\zeta(g_x) = \lim_{n\to\infty} \sqrt{n} \int_{-\infty}^{\infty} g(x-y)(dP_n - dP)(y)$

where g is the standard normal density and the measure P satisfies

$$P(|X| > L) \leq K(\frac{1}{\log L})^{\beta}$$

for some $K > 0$, $\beta > \beta' > \frac{1}{2}$, and $L > 1$. Then

(i) $\sup\{E\zeta^2(g_x)\colon g_x \in F\} = 1$, normalizing if necessary,

and

(ii) $\int_0^1 [\ell n\; N(x)]^{1/2} dx < \infty$.

Here is a proof of (ii):

Given K and $\beta > \beta' > \frac{1}{2}$ as above, consider only those

$\varepsilon > 0$ such that $\exp[-3e^{2\cdot(\frac{4K}{\varepsilon})^{1/\beta}}] \leq \frac{\varepsilon}{4}$. For these ε, we show that $f_p(y) := \int_{-\infty}^{\infty} (g(x-y))^2 dP(x) \leq \frac{\varepsilon}{2}$, for $|y| \geq y_0$ where $y_0 = O(\exp(\varepsilon^{-1/\beta'}))$. This will be used to prove (ii). Letting $M := M(\varepsilon) := \exp((\frac{4K}{\varepsilon})^{1/\beta})$, write

$$(6.1) \qquad f_p(y) := \int_{|X| \leq M} (g(x-y))^2 dP(x) + \int_{|X| > M} (g(x-y))^2 dP(x).$$

Since g is bounded by $\frac{1}{\sqrt{2\pi}}$ the second term on the RHS of 6.1 is bounded by $\frac{\varepsilon}{4}$. Also, if y is such that $|y| \geq y_0 := 3M$, $|x| \leq M$, then $g(x-y)^2 = \frac{1}{2\pi} e^{-x^2+2xy-y^2}$ $\leq \frac{1}{2\pi} e^{-3M^2} \leq \frac{\varepsilon}{4}$ by choice of ε. So the first term on the RHS of 6.1 is bounded by $\frac{\varepsilon}{4}$, showing that $f_p(y) \leq \frac{\varepsilon}{2}$. Finally,

$$y_0 := y_0(\varepsilon) = O(3 \exp(\frac{4K}{\varepsilon})^{1/\beta}) = O(\exp(\varepsilon)^{-1/\beta'})$$

as desired. Next, $\forall\, y \in \mathbb{R}$, define $g_y(x) = g(x-y)$. By a simple computation, $|g_y - g_z| \leq \frac{1}{\sqrt{2\pi e}} |y-z|$. So $N(\varepsilon) = 2 + O(2y_0 \frac{1}{\varepsilon}) = O(\frac{1}{\varepsilon} \exp(\varepsilon^{-1/\beta'}))$, proving (ii).

Q.E.D.

Chapter 4

Rates of Convergence for Classes of Functions

§1. Introduction

In this chapter we find general conditions on sequences G_n of classes of functions so that $\sup_{g\in G_n} |\int gd(P_n-P)| \overset{n}{\to} 0$ a.s. We apply this result to kernel density approximation in the following way.

Let $X_1, X_2,\ldots$ be a sequence of iid random variables with values in $\mathbb{R}^k$ with a common and unknown density f. Let $\{g_n\}$ be a sequence of probability densities on $\mathbb{R}^k$. For each g_n define the kernel density of f by

(1) $$\hat{g}_n(x) := \frac{1}{n} \sum_{i=1}^{n} g_n(x-X_i).$$

We find conditions on the g_n such that

(2) $$\sup_x |\hat{g}_n(x) - f(x)| \overset{n}{\to} 0 \quad \text{w.p. } 1.$$

We will need

<u>Def. 4.1.1</u> Let $f,g\colon X \to \mathbb{R}$ and $f(x) \le g(x)$, $x \in X$. Let $[f,g] = \{h\colon X \to \mathbb{R}^1 | f(x) \le h(x) \le g(x),\ x \in X\}$. Given a class of functions F with a measure P on X let

$N_I := N_I(\varepsilon, F, P)$ be a collection of minimal cardinality of of sets $[g^-, g^+]$ consisting of measurable functions with $g^- \leq g^+$ and such that

$$F \subseteq \cup_{[g^-,g^+] \in N_I} [g^-, g^+]$$

and $\int (g^+ - g^-) dP < \varepsilon$. Each set $[g^-, g^+]$ is called a bracket. Let card $N_I(\varepsilon, F, P) := N_I(\varepsilon)$.

Note that given F, X, and P, N_I is not uniquely defined. Also, $2N_I(\varepsilon) \geq N_I^{(1)}(\varepsilon) \geq N_I(\varepsilon)$ where $N_I^{(1)}(\varepsilon)$ is as in Def. 1.1.3.

§2. Rates of convergence

Here is the main result.

Theorem 4.2.1 Let P be a probability measure on $\mathbb{R}^k$ with a uniformly bounded density f. Let G_n be a sequence of classes of probability densities on $\mathbb{R}^k$. Suppose $\exists$ constants $1 < C < \infty$ and $\lambda < 1$ such that $\forall g_n \in G_n$ and all n

(a) $||g_n||_{sup} \leq Cn^\lambda$,

and suppose that for some $\gamma > 0$ we may choose $N_I(\varepsilon, G_n, P)$ such that

(b) $\forall f \in N_I(\varepsilon, G_n, P)$, $\int f^2 dP \leq Cn^\lambda$ and

$N_I(\varepsilon) \leq \exp\{D(\varepsilon) n^{1-\lambda-\gamma}\}$ where $D(\varepsilon) < \infty$ is some constant depending upon ε.

If P_n are the empirical measures for P, then

(3) $$\lim_{n\to\infty} \sup_{g_n \in G_n} \left|\int g_n d(P_n - P)(y)\right| = 0 \quad \text{a.s.}$$

Remark. De Hardt [6, Lemma 1] showed that for a uniformly bounded class F of measurable functions on $\mathbb{R}^k$ with $N_I(\delta, F, P) < \infty \ \forall \delta > 0$,

$$\lim_{n\to\infty} \sup_{f \in F} \left|\int f d(P_n - P)(y)\right| = 0 \quad \text{a.s.}$$

Proof: Assume $f \le M$ for some $M \ge 1$. Note that

(4) $$\int g_n^2 dx \le (\sup g_n) \int g_n dx \le Cn^\lambda .$$

Also, we may assume that $\forall\, [g^-, g^+] \in N_I$

(5) $$-Cn^\lambda \le g^- \le g^+ \le Cn^\lambda .$$

Finally, by (b) we have

(6) $$N_I(1) \le \exp\{D(1) n^{1-\lambda-\gamma}\}.$$

Assume $0 < t < 1$. For each $[g^-, g^+] \in N_I(\varepsilon, G, P)$ find a $\tilde{g}_n \in [g^-, g^+] \cap G_n$. Then

$$\Pr\{\sup_{g_n \in G_n} |\int g_n (dP_n - dP)(y)| \geq t\} \leq$$

(7) $$\leq \Pr\{\max_{[g^-,g^+] \in N_I(1,G_n,P)} |\int \tilde{g}_n (dP_n - dP)(y)| \geq \frac{t}{2}\} +$$

$$+ \Pr\{|\max_{[g^-,g^+] \in N_I(1.G_n,P)} |\int \tilde{g}_n (dP_n - dP)(y)| -$$

$$- \sup_{g_n \in G_n} |\int g_n (dP_n - dP)(y)| \geq \frac{t}{2}\},$$

Our first goal is to bound the first term on the RHS of (7). For a fixed $g_n \in G_n$ we have

$$\Pr\{|\int g_n (dP_n - dP)(y)| \geq \frac{t}{2}\} = \Pr\{\frac{1}{n}|S_n| \geq \frac{t}{2}\}$$

where $S_n := \sum_{i=1}^{n} g_n(X_i) - Eg_n(X_i)$, and $\sigma_n = \sqrt{\operatorname{Var} S_n} \leq \sqrt{MC}\ n^{\frac{1+\lambda}{2}}$.

Using Bernstein's inequality and $M \geq 1$, we have

$$\Pr\{\frac{1}{n}|S_n| \geq \frac{t}{2}\} = \Pr\{|S_n| \geq \frac{t}{2\sqrt{MC}}\ n^{\frac{1-\lambda}{2}} \sigma_n\} \leq 2\exp\{\frac{\frac{-t^2}{4MC} n^{1-\lambda}}{2 + \frac{t}{3M}}\} \leq$$

(8) $$\leq 2\exp\{\frac{-t^2}{12MC} n^{1-\lambda}\}.$$

Multiplying (8) by $N_I(1,G_n)$ gives a bound for the first term on the RHS of (7). That bound is (using 6):

(9) $$2\exp\{D(1)n^{1-\lambda-\gamma}\}\exp\{\frac{-t^2}{12MC}n^{1-\lambda}\}.$$

Finally, note for future use that if $\lambda < 1$ and $\gamma > 0$ then

(10) $$\sum_{n=1}^{\infty}\exp\{D(1)n^{1-\lambda-\gamma}\}\exp\{\frac{-t^2}{12MC}n^{1-\lambda}\} < \infty.$$

Now find a bound for the second term on the RHS of (7). Let $z_n(g) := \int g(dP_n - dP)$, and from now on write g instead of g_n and $\tilde{g}$ instead of $\tilde{g}_n$. Observe that

$$\left| \max_{[g^-,g^+]\in N_I(1)} |z_n(\tilde{g})| - \sup_{g\in G_n} |z_n(g)| \right| \leq$$

$$\leq \max_{[g^-,g^+)\in N_I(1)} \ \sup_{g\in[g^-,g^+)\cap G_n} |z_n(g) - z_n(\tilde{g})|.$$

Let $[g^-,g^+] \in N_I$ and let $g \in [g^-,g^+] \cap G_n$. For any g and j we can find a bracket $[g_j^-,g_j^+] \in N_I(2^{-j},G_n)$ such that $g \in [g_j^-,g_j^+]$. Given t, $0 < t < 1$, choose m such that

(11) $$2^{-m} \leq \frac{t}{4} \leq 2^{-m+1},$$

and note that $\frac{t}{4} \leq \frac{1}{4}$ and so $m \geq 2$. We have

(12) $|z_n(g)-z_n(\tilde{g}) \le |z_n(g_m^-)-z_n(\tilde{g})| + |z_n(g)-z_n(g_m^-)|$

and $z_n(g_m^-) + \int g_m^- dP \le z_n(g) + \int g dP \le z_n(g_m^+) + \int g_m^+ dP$.

Also $\int (g_m^+ - g_m^-) dP \le 2^{-m} \le \frac{t}{4}$. Using this we may bound the second term on the RHS of (12): $z_n(g) - z_n(g_m^-) = \int (g-g_m^-) d(P_n-P) \le \int (g-g_m^-) dP_n \le \int (g_m^+ - g_m^-) dP_n \le |z_n(g_m^+ - g_m^-)| + \frac{t}{4}$. Also, $z_n(g_m^-) - z_n(g) = \int (g-g_m^-) d(P-P_n) \le \int (g_m^+ - g_m^-) dP \le \frac{t}{4}$. Thus, $|z_n(g) - z_n(g_m^-)| \le |z_n(g_m^-) - z_n(g_m^-)| + \frac{t}{4}$

Thus we may rewrite (12) as

(15) $|z_n(g) - z_n(\tilde{g})| \le |z_n(g_m^-) - z_n(\tilde{g})| + |z_n(g_m^+) - z_n(g_m^-)| + \frac{t}{4}$.

Therefore $\Pr\{|\max_{[g^-,g^+]\in N(1,G_n)} |z_n(\tilde{g})| - \sup_{g\in G_n} |z_n(g)|| > \frac{t}{2}\}$

is bounded by

(16) $\Pr\{\max'|z_n(g_m^-) - z_n(\tilde{g})| + \max'|z_n(g_m^-) - z_n(g_m^+)| > \frac{t}{4}\}$,

where max' denotes max over $[g_m^-, g_m^+] \in N_I(2^{-m}, G_n)$. Now (16) is $\le$

(17) $\Pr\{\max'|z_n(g_m^-)-z_n(\tilde{g})| > \frac{t}{8}\} +$

$\Pr\{\max'|z_n(g_m^-) - z_n(g_m^+)| > \frac{t}{8}\}$.

By condition (a) and (5) we have

(18) $||g_m^- - \tilde{g}||_{sup} \leq 2Cn^{\lambda}$ and $||g_m^+ - g_m^-||_{sup} \leq 2Cn^{\lambda}$.

Clearly, by hypotheses (a) and (b) we also have

(19) $\mathrm{Var}(g_m^- - \tilde{g}) \leq 4MCn^{\lambda}$ and $\mathrm{Var}(g_m^+ - g_m^-) \leq 4MCn^{\lambda}$.

Note that (17) is bounded by

(20) $2N_I(2^{-m}, G_n)^2 \Pr\{|z_n(f)| > \frac{t}{8}\}$,

where f has the form $g_m^- - \tilde{g}$ or $g_m^- - g_m^+$. Let $S_n(g) := \sum_{i=1}^{n} g(X_i) - Eg(X_i)$. Then $\sigma_n := \mathrm{Var}\, S_n(g) \leq 2\sqrt{MC}\, n^{\frac{1+\lambda}{2}}$. Let $\theta := \frac{t}{8}$. Using Bernstein's inequality, (18), (20), $M \geq 1$, and $C \geq 1$ we have

(21) $$\Pr\{|z_n(f)| > \theta\} \leq \Pr\{|S_n(f)| > \theta \frac{n^{\frac{1-\lambda}{2}}}{2\sqrt{MC}} \sigma_n\}$$

$$\leq 2 \exp\{\frac{-\theta^2}{12MC} n^{1-\lambda}\}.$$

Since (21) holds for all choices of f, (20) is bounded by

(22) $$4 \exp\{2D(2^{-m}) n^{1-\lambda-\gamma}\} \exp\{\frac{-\theta^2}{12MC} n^{1-\lambda}\}.$$

So the LHS of (7) is bounded by (9) plus (22). Note that

$$\sum_{n=1}^{\infty} \exp\{2D(2^{-m})n^{1-\lambda-\gamma}\}\exp\{\frac{-\theta^2}{12MC}\, n^{1-\lambda}\} < \infty ,$$

whenever $\lambda < 1$, $\gamma > 0$. Finally, (7), (10), (22), and an application of the Borel-Cantelli Lemma combine to show that for all $0 < t \leq 1$,

$$\Pr\{\limsup_{n\to\infty} \sup_{g_n \in G_n} |\int g_n d(P_n - P)(y)| \leq t\} = 1.$$

Since $0 < t \leq 1$ is arbitrary, (3) follows. Q.E.D.

§3. <u>Kernel Density Estimates</u>

We may apply Theorem 4.2.1 to the problem of kernel density approximation. Let $\hat{g}_n(x)$ be as in (1) and $||\cdot||$ the usual Euclidean norm on $\mathbb{R}^k$. Let $||\cdot||_L$ be the Lipschitz norm as defined in chapter 3.1.

<u>Corollary 4.3.1</u> Let $\{g_n\}$ be a sequence of Lipschitz probability densities on $\mathbb{R}^k$ with $||g_n||_{\sup} \leq Cn^{\lambda}$, $C \geq 1$, $\lambda < 1$, and such that $||g_n||_L \leq \exp(n^{\rho})$ where $\rho < 1 - \lambda$.

Suppose that for $||x|| > e$,

$$(23) \qquad g_n(x) \leq K(\frac{1}{\log ||x||})^{\gamma},$$

where $0 < K \leq Cn^{\lambda}$ and $\gamma > \frac{\lambda}{1-\lambda}$. Let X_i, $i = 1,2,\ldots$ be iid $\mathbb{R}^k$-valued random variables with a uniformly bounded and uniformly continuous density f such that

$$(24) \qquad P(||X|| > L) \leq K'\left(\frac{1}{\log L}\right)^{\beta},$$

for some $K' > 2$, $\beta > \frac{\lambda}{1-\lambda}$, and all $L > 1$. Suppose that $\forall \varepsilon > 0 \int_{|x|>\varepsilon} g_n dx \to 0$ as $n \to \infty$. Then

$$(25) \qquad \sup_x |\hat{g}_n(x) - f(x)| \overset{n}{\to} 0 \quad \text{a.s.}$$

Proof: Let $G_n := \{g_n(x-\cdot): x \in \mathbb{R}^k\}$. We find a bound on $N_I(\varepsilon)$ and apply Theorem 4.2.1.

Given $\gamma > \frac{\lambda}{1-\lambda}$, choose β and β' such that $\gamma > \beta > \beta' > \frac{\lambda}{1-\lambda}$. Given $n \geq 1$ and $\varepsilon > 0$, let $M := M(\varepsilon,n) := \exp\left(\frac{2K'Cn^{\lambda}}{\varepsilon}\right)^{1/\beta}$ and $y_0 := y_0(\varepsilon,n) := \exp\left(\frac{n^{\lambda}}{\varepsilon}\right)^{1/\beta'}$.

Given $a > 0$, let $B(a)$ be the ball centered at the origin of $\mathbb{R}^k$ with radius a. Let $x \in \mathbb{R}^k$ and

$$h(x) = Cn^{\lambda} 1_{B(M)^c}(x) + \frac{\varepsilon}{2} 1_{B(M)}(x).$$

Given M as above consider only those $\varepsilon > 0$ such that

$y_0 \geq 2M$. For these ε observe that if $||y|| > y_0$ then

(26) $$0 \leq g_n(x-y) \leq h(x).$$

(26) is true since if $||z|| > M$ then

$$g_n(z) \leq K\left(\frac{\varepsilon}{2K'Cn^\lambda}\right)^{\gamma/\beta} \leq \frac{n^\lambda}{2}\left(\frac{\varepsilon}{n^\lambda}\right) \leq \frac{\varepsilon}{2},$$

by definition of M, K, K', γ, and β. Also

(27) $$\int hdP \leq Cn^\lambda P(B(M)^c) + \frac{\varepsilon}{2} \leq Cn^\lambda K' \frac{\varepsilon}{2K'Cn^\lambda} + \frac{\varepsilon}{2} \leq \varepsilon.$$

Now on $B(y_0)$ find $x_i \in \mathbb{R}^k$ dense within $\frac{\varepsilon}{2}$ and let

(28) $$g_i^{\pm} = g_n(x-x_i) \pm \frac{\varepsilon}{2}.$$

Using (26), (27), (28), and the Lipschitz condition on g we have

$$N_I(\varepsilon) \leq 2 + A\left(\frac{y_0}{\varepsilon} \exp(n^\rho)\right)^k$$

for some $A < \infty$. By the definition of y_0 we have $N_I(\varepsilon, G_n) \leq 2 + A\left(\frac{1}{\varepsilon}\right)^k \exp\{kn^{\lambda/\beta'}\varepsilon^{-1/\beta'}\}\exp\{kn^\rho\}$. Now $\beta' > \frac{\lambda}{1-\lambda}$, $\frac{\lambda}{\beta'} < 1 - \lambda$, and for some $\gamma > 0$

(29) $$N_I(\varepsilon, G_n) \leq A\left(\frac{1}{\varepsilon}\right)^k \exp\{Dn^{1-\lambda-\gamma}\}$$

where $D := D(\varepsilon) := 2k\varepsilon^{-1/\beta'}$.

Also,

(30) $\quad \int h^2 dP \le Cn^{\lambda}$ and $\int g_n^2 dP \le \sup g_n \int g_n dP \le NCn^{\lambda}$,

where N is the uniform bound for the density f. Thus it is possible to choose $N_I(\varepsilon, G_n)$ such that $\forall f \in N_I$, $\int f^2 dP < C'n^{\lambda}$ for some $1 < C' < \infty$. Using (29) and (30), Theorem 4.2.1 shows that

(31) $\quad \sup_x |\hat{g}_n(x) - E\hat{g}_n(x)| \overset{n}{\to} 0$ a.s.

It only remains to show

<u>Lemma 4.3.2</u> $\quad \sup_x |E\hat{g}_n(x) - f(x)| \to 0$ as $n \to \infty$.

<u>Proof</u>: For all $\delta > 0$ we have

$$\sup_x |E\hat{g}_n(x) - f(x)| \le \sup_x \int_{|y| \le \delta} |f(x-y) - f(x)| g_n(y)$$

$$+ \sup_x \int_{|y| > \delta} |f(x-y) - f(x)| g_n(y) dy$$

(32) $$\le \sup_x \sup_{|y| \le \delta} |f(x-y) - f(y)| + 2N \int_{|y| > \delta} g_n(y) dy$$

Let $\eta > 0$ be arbitrarily small. By choosing δ sufficiently small we can make the first term on the RHS of (32) less than $\eta/2$. Having so chosen δ, we can then choose n so large that the second term on the RHS of (32) is also less than $\eta/2$. This proves the lemma and the Corollary. Q.E.D.

<u>Remark</u>: Let g be a Lipschitz probability density on $[-a,a]$, $a < 1$, with Lipschitz constant K. Let $h_n = n^{-\lambda}$. Assume $g \leq C$. The kernels g_n defined by $g_n(\cdot) := g_n(\frac{\cdot}{h_n})(\frac{1}{h_n})$ satisfy the hypotheses of Corollary 4.3.1. This special case is treated by Devroye and Wagner [7] and they prove (25) with $\frac{1}{h_n} = o(\frac{n}{\log n})$ and without restriction (24).

Assuming a unimodal condition on the densities g_n we may relax the hypotheses of Corollary 4.3.1 and obtain

<u>Theorem 4.3.3</u>. Let G_n, $n \geq 1$, be classes of probability densities $\{g_n\}$ such that $\forall\, n \geq 1$

(a) g_n is unimodal $\forall\, g_n \in G_n$,

(b) there is an $a_n = o\sqrt{\frac{n}{\log \log n}}$ such that $\forall\, g_n \in G_n$, $||g_n||_{\sup} < a_n$, and

(c) $\forall\, \varepsilon > 0$, $\sup_{g_n \in G_n} \int_{|x|>\varepsilon} g_n dx \to 0$ as $n \to \infty$.

Let P be a probability measure on $\mathbb{R}$ with a uniformly continuous density f. Then

$$\sup_{g_n \in \mathcal{G}_n} \sup_x |\hat{g}_n(x) - f(x)| \overset{n}{\to} 0 \quad \text{a.s.}$$

<u>Proof</u>: Using a slight modification of Lemma 4.3.2 and its proof we have

$$\sup_{g_n \in \mathcal{G}_n} \sup_x |E\hat{g}_n(x) - f(x)| \overset{n}{\to} 0.$$

So it suffices to show

$$(33) \qquad \sup_{g_n \in \mathcal{G}_n} \sup_x |\int g_n(x-y)\,d(P_n - P)(y)| \overset{n}{\to} 0 \quad \text{a.s.}$$

Define $\tilde{\mathcal{G}}_n := \{g_n(x-\cdot),\ x \in \mathbb{R}, \text{ and } g_n \in \mathcal{G}_n\}$. Then it suffices to show

$$\sup_{h_n \in \tilde{\mathcal{G}}_n} |\int h_n\,d(P_n - P)(y)| \overset{n}{\to} 0 \quad \text{a.s.}$$

Let $\mathcal{H}_n := \{\frac{h_n}{a_n} : h_n \in \tilde{\mathcal{G}}_n\}$ and $\mathcal{H} := \{h$ such that h is unimodal and $||h||_{\sup} \leq 1\}$. Then $\mathcal{H}_n \subseteq \mathcal{H}$. By Theorem 1.9 of [13] the CLT is satisfied uniformly over $\mathcal{H}$. An application of Theorems 1.1 and 1.2 of [15] shows that the LIL is satisfied uniformly over $\mathcal{H}$. For the classes $\tilde{\mathcal{G}}_n$ we have

$$\sup_{h_n \in \tilde{\mathcal{G}}_n} |\nu_n(h_n)| \leq O(a_n\sqrt{\log\log n}) \quad \text{a.s.}$$

or, equivalently,

$$\sup_{h_n \in \tilde{\mathcal{G}}_n} |\int h_n d(P_n - P)(y)| \leq O(a_n\sqrt{\frac{\log\log n}{n}}) \quad \text{a.s.}$$

By definition of a_n, (33) is clearly satisfied. Q.E.D.

Bibliography

[1] Alexander, K. (1982), Ph.D. Thesis, M.I.T.

[2] Bakhvalov, N. S. (1959), On approximate calculation of multiple integrals. Vestnik Mosk. Univ. Ser. Mat. Mekh. Astron. Fiz. Khim. no. 4, 3-18.

[3] Bickel, P. J. and Rosenblatt, M. (1971), On some global measures of the deviations of density function estimates, Annals of Statist, 1, 1071-1095.

[4] Billingsley, P. (1968), Convergence of Probability Measures, Wiley, New York.

[5] Borisov, I. S. (1981), Some limit theorems for empirical distributions, Abstracts of Reports, Third Vilnius Conf. Probability Th. Math. Statist. I, 71-72.

[6] De Hardt, J. (1971), Generalizations of the Glivenko-Cantelli theorem, Ann. Math. Statist. 42, 2050-2055.

[7] Devroye, L. P. and Wagner. T. J. (1980), The Strong Uniform Consistency of Kernel Density Estimates, Mult. Analysis V, 59-77.

[8] Dmitrovksi, V. A. (1980), A Boundedness Condition and Estimates of the Distribution of the Maximum of Random Fields on Arbitrary Sets, Soviet Math Dokl. 22, pp. 59-62.

[9] Dudley, R. M. (1966), Convergence of Baire measures, Studia Math., 27, 251-268.

[10] Dudley, R. M. (1968), Distances of probability measures and random variables, Ann. Math. Statist. 39, 1563-1572.

[11] Dudley, R. M. (1969), The Speed of mean Glivenko-Cantelli convergence, ibid. 40, 40-50.

[12] Dudley, R. M. (1978), Central limit theorems for empirical measures, Ann. Probability 6, 899-929; Correction, ibid. 7 (1979) 909-911.

[13] Dudley, R. M. (1981), Donsker classes of functions, Statistics and Related Topics (Proc. Symp. Ottawa, 1980), North-Holland, New York, Amsterdam, 341-352.

[14] Dudley, R. M. (1982), Empirical and Poisson processes on classes of sets or functions too large for central limit theorems, preprint.

[15] Dudley, R. M. and Philipp, W. (1982), Invariance Principles for sums of Banach space valued random elements and empirical processes, preprint.

[16] Dudley, R. M. (1983?), Empirical Processes, École d'été de probabilités de St.-Flour (1982), to appear in Lecture Notes in Math.

[17] Fortet, R. and Mourier, E. (1971), Convergence de la répartition empirique vers la répartition théorique, Ann. Sci. Ecole Norm. Sup. 70, 266-285.

[18] Gnedenko, B. V. and Kolmogorov, A. N. (1954), Limit Distributions for sums of Independent Random Variables, Addison-Wesley, Reading, Mass.

[19] Hanson, D. L. and Wright, F. T. (1969), Some more results on rates of convergence in the law of large numbers for weighted sums of independent random variables, Trans. Amer. Math. Soc. 141, 443-464.

[20] Hoeffding, W. (1963), Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58, 13-30.

[21] Ibragimov, I. A. and Linnik, Yu V. (1971), Independent and Stationary Sequences of Random Variables, Wolters-Noordhoff, Groningen.

[22] Jain, N. C. and Marcus, M. B. (1975), Central limit theorems for C(S)-valued random variables, J. Functional Analysis 19, 216-231.

[23] Kolcinski, V. I. (1981a), On the central limit theorem for empirical measures, Teor. Verioatnost. Mat. Statist. (Kiev) 1981, no. 24, 63-75.

[24] Kolcinski, V. I. (1981b), On the law of the iterated logarithm in Strassen's form for empirical measures, ibid, no. 25, 40-47.

[25] Kolmogoriv, A. N. (1953), Some recent work in the field of Limit Theorems in Probability Theory, Moscow Vestnik, 10, 29-38.

[26] Kolmogorov, A. N. and Tikhomirov, V. M. (1959), ε-entropy and ε-capacity of sets in functional spaces, Amer. Math. Soc. Transl. (Ser. 2) 17 (1961), 277-364 = Uspekhi Math. Nauk. 14, vyp. 2(86), 3-86.

[27] Kuelbs, J. (1977), Kolmogorov's law of the iterated logarithm for Banach space valued random variables, Ann. Probability 4, 744-771.

[28] Kuelbs, J. and Philipp, W. (1980), Almost sure invariance principles for partial sums of mixing B-valued random variables, Ann. Probability 8, 1003-1036.

[29] Parthasarathy, K. R. (1967), Probability measures on metric spaces, Academic Press, New York.

[30] Pollard, D. (1981), A central limit theorem for empirical processes, to appear in J. Australian Math. Soc.

[31] Pólya, G. (1949), Proc. Berk. Symp. on Math. Statist. and Prob., U. C. Press.

[32] Prohorov, Yu. V. (1956), Convergence of random processes and limit theorems in probability theory, Theor. Prob. Appl. 1, 157-214.

[33] Rudin, W., (1962), Fourier Analysis on Groups, Interscience Publishers, New York.

[34] Strassen, V. (1965), The existence of probability measures with given marginals, Ann. Statist. 36, 423-439.

[35] Vapnik, V. N. and Červonenkis, A. Ya. (1971), On the uniform convergence of relative frequencies of events to their probabilities, Theor. Prob. Appl. 16, 264-280.

[36] Varadarajan, V. S., (1958), On the convergence of sample probability distributions, Sankhya 19, 23-26.