Empirical Processes in Action: A Review

Jon A. Wellner

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

Summary

A review is given of recent applications of empirical process theory and methods to statistics with emphasis on empirical processes indexed by sets and functions. After a brief survey of empirical process theory, we review applications of this theory to estimation (censoring, truncation, biased sampling, regression and density function estimation, minimum distance methods), testing (classical goodness of fit and minimum distance tests, permutation and bootstrap tests, local alternatives and power), pattern recognition, clustering, and classification, bootstrapping of empirical measures, and the delta method. One new theorem on the asymptotic behavior of empirical processes under local alternatives is presented.

Key words: Applications; Biased sampling; Bootstrapping; Bootstrap tests; Censoring; Classification; Clustering; Density function; Differentiable functions; Donsker theorem or CLT; Empirical distribution function; Empirical measure; Empirical process; Estimation Glivenko-Cantelli theorem or SLLN; Inequalities; Local alternatives; Minimum distance; Nonstandard asymptotics; Omnibus tests; Pattern recognition; Permutation tests; Regression function; Trunction; U-processes; Vapnik-Chervonenkis classes.

I Introduction

The theory of empirical measures and processes has developed rapidly over the past 13 years since the key paper by Dudley (1978). This vigorous theoretical development has gone hand in hand with advances in probability theory, notably the theory of Gaussian processes (e.g. Adler (1990) for an introduction) and limit theory for probability distributions on Banach spaces (e.g. the recent book by Ledoux & Talagrand (1991)). For a recent excellent expository paper, see Pollard (1989), and the discussion thereof. This development includes a wealth of tools and techniques for asymptotic theory in statistics.

Applications of general empirical process theory in statistics have developed somewhat more slowly in the past decade, however. The purpose of this paper is to review *applications* of empirical process theory in statistics. As recently as 1984, Pyke ((1984), page 251) wrote:

> 'The asymptotic results that have been obtained during the seventies for empirical processes indexed by families of sets have as yet *not been applied significantly to problems of inference*'.

As we will see in the course of this review, some progress *has* been made in this direction in the meantime—often involving further theoretical developments.

Of course the difficult question

What is an *application* of empirical processes?

rears its head immediately. One possible definition would be to include only papers which involved 'real data'. Unfortunately, this narrow definition would virtually reduce this

review to a null set (in spite of 5 pages with the heading 'empirical' in Science Citation Indices for 1985–1989). In this sense, Pyke's statement in 1984 may still be true! In analogy with the well-known 'Erdös number' associated with each mathematician, any paper on empirical processes might be given a 'real data number' indicating at how many papers removed is the 'real data' motivating the theoretical development. Thus I would assign a 'real data number' of 0 to the recent paper by Olshen, Biden, Wyatt, & Sutherland (1989) since it contains *both* an application of empirical process theory *and* 'real data' (from 'gain analysis'); while the paper of Pakes & Pollard (1989) would receive a 'real data number' of 1 since the 'real data' (involving renewal of patents) was given in Pakes (1986).

I will, however, take a rather broad perspective and answer this question here by defining an application of empirical processes to be any development in statistics resulting in an understanding of the properties of some particular statistical procedure or method *which has used empirical process tools or methods*.

Moreover, this review will emphasize 'applications' in the above sense which use modern empirical process theory—as developed since 1978—for data with values in a possibly high dimensional space. In fact, modern empirical process theory deals with empirical measures and processes for data with values in a *completely arbitrary*, perhaps *infinite-dimensional* sample space. This aspect of the theory will undoubtedly become more important in future applications as statisticians develop methods for dealing with 'function' and 'picture'-valued data such as seismographs, noise level tracings, electrocardiograms, and high-dimensional biomedical data (survival times together with hundreds of covariates).

I believe that one important consequence of the rapid developments in modern empirical process tools and techniques is a shortening of the lag time between the introduction of a new method (e.g. a new estimator or test statistic) in statistics and the development of an understanding of the properties and performance of the method. The following table gives some support to this claim. The table contains the (approximate?) times of introduction of a few selected estimators or methods, together with the date of

Estimator/Problem	Date introduced	Date CLT first established	Lag (years)
empirical df	Cramér (1928)	Donsker (1952)	24
empirical measure	Fortet & Mourier (1953)	Dudley (1978)	25
		Pollard (1982b)	
		Ossiander (1987)	
		Giné & Zinn (1984)	
monotone density	Grenander (1956)	Prakasa Rao (1969)	13
		Groeneboom (1985)	29
		Kim & Pollard (1990)	34
right censoring	Kaplan & Meier (1958)	Breslow & Crowley (1974)	16
		Gill (1983)	25
k-means clustering	MacQueen (1967)	Pollard (1982a)	15
left truncation	Lynden-Bell (1971)	Woodroofe (1985)	14
double censoring	Turnbull (1974)	Chang (1990)	16
Efron's bootstrap	Efron (1979)		
	df:	Bickel & Freedman (1981)	3
	measure:	Giné & Zinn (1990)	11
Rubin's Bayesian bootstrap	Rubin (1981)		
	df:	Lo (1987)	6
	measure:	Praestgaard (1991)	10
censored regr.	Powell (1984)	Pollard (1990)	6
biased sampling	Vardi (1985)	Gill, Vardi & Wellner (1988)	3
simplicial depth process	Liu (1990)	Arcones & Giné (1991)	1

publication of the central limit theorem (or analog thereof) and the resulting lag times. Of course the methods chosen for inclusion in this table reflect my own biases and subjective preferences, and thereof suffer severely from selection bias!

For an interesting review of applications up to 1980, see Dudley (1981). For applications of empirical process theory, and probabilistic methods more generally, to the theory of Banach spaces, see Pisier (1986), (1989) and Ledoux & Talagrand (1991), chapters 14 and 15. Pyke (1992) gives a survey and a variety of open problems connected with empirical processes and related product processes.

II A Brief Review of Empirical Process Theory

II.1 Limit Theorems for Empirical Measures and Processes

Suppose that $X_1, \ldots, X_n, \ldots$ are independent and identically distributed from a distribution P on an arbitrary measurable space $(A, \mathbb{A})$. Here A is the 'sample space', and $\mathbb{A}$ is some sigma-field of subsets of A. Usually $(A, \mathbb{A})$ will be $(R^k, \mathbf{B}^k)$, k-dimensional Euclidean space with its Borel sigma-field for some fixed k, but it could be completely arbitrary. Let δ_x denote the probability measure with mass 1 at $x \in A$, and let

$$\mathbb{P}_n := n^{-1} \sum_{i=1}^{n} \sigma_{X_i} \tag{1}$$

denote the *empirical measure* of the first n of the observations X_i. Thus for any set $B \in \mathbb{A}$

$$\mathbb{P}_n(B) = n^{-1} \sum_{i=1}^{n} \delta_{X_i}(B) = \frac{\#\{k \leqslant n : X_k \in B\}}{n}.$$

Of course, the importance for statistics is that: (i) $\mathbb{P}_n$ is the nonparametric maximum likelihood estimator of P (see e.g. Kiefer & Wolfowitz (1956), Scholz (1980)); (ii) $\mathbb{P}_n$ is sufficient for $P \in \mathbf{M} := \{$all probability distributions on $A\}$ (see Dudley (1984), theorem 10.1.3, page 95); and (iii)

$$\mathbb{P}_n \text{ is 'close' to } P \text{ for } n \text{ large.} \tag{2}$$

These facts, together with the classical central limit theorem lead us to define the *empirical process* $\mathbb{X}_n$ by

$$\mathbb{X}_n := \sqrt{n} \, (\mathbb{P}_n - P). \tag{3}$$

Our goal in this section is to briefly review/survey the available limit theory making the assertion (2) precise, with emphasis on laws of large numbers (Glivenko–Cantelli theorems) and central limit theorems (Donsker theorems).

Modern empirical process theory views the empirical measure $\mathbb{P}_n$ as a stochastic process indexed by a large class of functions $\mathbf{F}$ or sets $\mathbf{C}$ as follows: suppose that $\mathbf{F}$ is a collection of real-valued measurable functions defined on A, and write Pf for $\int f \, dP$. Thus, for $f \in \mathbf{F}$,

$$\mathbb{P}_n f = \int f \, d\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

If $\mathbf{C} \subset A$ is a class of (measurable) subsets of A, then

$$\mathbb{P}_n C = \int 1_C \, d\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} 1_C(X_i) = \frac{1}{n} \#\{k \leqslant n : X_k \in C\}.$$

In either case, the resulting stochastic process, as $f \in \mathbf{F}$ or $C \in \mathbf{C}$ varies, is just

$$\{\mathbb{P}_n f : f \in \mathbf{F}\} \quad \text{or} \quad \{\mathbb{P}_n C : C \in \mathbf{C}\}.$$

If the sample space $(A, \mathbf{A})$ is $(R^1, \mathbf{B}^1)$ and the collection of sets $\mathbf{C}$ is the collection of left orthants, $\mathbf{C} = \{(-\infty, x] : x \in R^1\}$, then

$$\{\mathbb{P}_n C : C \in \mathbf{C}\} = \{\mathbb{P}_n((-\infty, x] : x \in R^1\}$$

is just the usual one-dimensional empirical distribution function (df) for the real-valued data $X_1, \ldots, X_n$. Since the indicator functions of sets C are just particular functions, we will formulate results below for $\mathbb{P}_n$ indexed by classes of functions $\mathbf{F}$.

To get limit theorems for $\mathbb{P}_n$ indexed by $\mathbf{F}$ which are uniform over all $f \in \mathbf{F}$ we need to rule out two potential problems: First, as in the case of the moment hypotheses for the usual strong law of large numbers or central limit theorem, the functions in $\mathbf{F}$ should not be 'too big pointwise.' Second, there should not be 'too many of them.' To make these more precise, we define two more notions: the envelope function F of $\mathbf{F}$, and the entropy of $\mathbf{F}$.

For a given class $\mathbf{F}$ of measurable real-valued functions on A, define a function G by

$$G(x) := \sup_{f \in \mathbf{F}} |f(x)| \quad \text{for} \quad x \in A.$$

Thus $G(x) \geq |f(x)|$ for all $x \in A$ and all $f \in \mathbf{F}$. Since G is *not* nesessarily measurable, we let F be the smallest measurable function above G, often denoted by G^*. Thus

$$F(x) = G^*(x) = \left(\sup_{f \in \mathbf{F}} |f(x)| \right)^*$$

is the least measurable envelope of $\mathbf{F}$, and we refer to it from now on as simply *the envelope of* $\mathbf{F}$.

The second notion is that of an *entropy number* for the size of $\mathbf{F}$. Viewing $\mathbf{F}$ as a subset of a metric space (D, d) (such as $(L_r(P), \|\cdot\|_r)$), for any $\varepsilon > 0$ we can consider covering $\mathbf{F}$ with ε balls $B(f_j, \varepsilon)$ with centers f_j in $\mathbf{F}$. Let $N(\varepsilon, \mathbf{F}, d)$ be the number of such balls required to cover $\mathbf{F}$. Thus $\mathbf{F} \subset \bigcup_{j=1}^m B(f_j, \varepsilon)$ for some $f_j \in \mathbf{F}$ with $m = N(\varepsilon, \mathbf{F}, d)$. If $\mathbf{F} \subset L_2(P)$, define $\rho_P(f, g)$ for $f, g \in \mathbf{F}$ by $\rho_P^2(f, g) = \mathrm{Var}_P (f(X) - g(X))$.

Now a heuristic summary of the two types of limit theorems below is as follows:

SLLN or Glivenko theorem: If an 'appropriate' entropy number $N(\varepsilon, \mathbf{F}, d)$ is finite for every $\varepsilon > 0$ and $\mathbf{F}$ has integrable envelope function F ($F \in L_1(P)$), then

$$\|\mathbb{P}_n - P\|_{\mathbf{F}}^* := \sup_{f \in \mathbf{F}} |\mathbb{P}_n f - Pf| \to_{a.s.} 0 \quad \text{as} \quad n \to \infty.$$

CLT or Donsker theorem: If an 'appropriate' entropy number $N(\varepsilon, \mathbf{F}, d)$ is $\sqrt{\log \bullet}$- integrable

$$\left(\text{i.e.} \int_0^1 \{\log N(u, \mathbf{F}, d)\}^{\frac{1}{2}} \, du < \infty \right),$$

and the envelope function F of $\mathbf{F}$ is square integrable, then

$$\sqrt{n} (\mathbb{P}_n - P) \Rightarrow G_P$$

where G_P is a *P-Brownian bridge process* with ρ_P-continuous sample paths: i.e. G_P is a mean zero Gaussian process on $\mathbf{F}$ with covariance function

$$\mathrm{Cov} (G_P(f), G_P(g)) = P(fg) - PfPg, \quad \text{for} \quad f, g \in \mathbf{F}$$

and G_P takes values in $\mathbf{C}_u(\mathbf{F}, \rho_P)$ the subcollection of $l^\infty(\mathbf{F})$ consisting of all (bounded and) ρ_P-uniformly continuous functions on $\mathbf{F}$. Note that when $\mathbf{F}$ is the collection of indicators of a class of sets $\mathbf{C}$, $\mathbf{F} = \{1_C : C \in \mathbf{C}\}$, this covariance becomes

$$\mathrm{Cov}\,(G_P(B), G_P(C)) = P(B \cap C) - P(B)P(C), \quad \text{for} \quad B, C \in \mathbf{C};$$

and when this is specialized still further to the one-dimensional df case with $\mathbf{C} = \{1_{(-\infty, x]} : x \in R^1\}$, this becomes, with $G_P(-\infty, x] := G_P(x)$, $P(-\infty, x] := H(x)$,

$$\mathrm{Cov}\,(G_P(x), G_P(y)) = H(x \wedge y) - H(x)H(y)$$
$$= H(x) \wedge H(y) - H(x)H(y) \quad \text{for} \quad x, y \in R^1,$$

the covariance function of the usual H-Brownian bridge process in one-dimension $\mathbb{U}(H)$ where $\mathbb{U}$ is a standard Brownian bridge process on $[0, 1]$.

Here are my favorite two examples of each of these two types of limit theorems. To make the statements precise we define two types of entropy numbers as follows.

First, for $r > 0$ and $\varepsilon > 0$, define the $L_r(P)$-*metric entropy with bracketing* by

$$N_{[]}^{(r)}(\varepsilon, \mathbf{F}, P) := \min \left\{ k: \begin{array}{l} \text{there exist } u_1, \ldots, u_k, l_1, \ldots, l_k \in L_r(P) \\ \text{such that for each } f \in \mathbf{F} \text{ there are } i, j \le k \\ \text{such that } l_i \le f \le u_j, \, P(u_j - l_i)^r \le \varepsilon^r \end{array} \right\}.$$

The second type of entropy number is defined in terms of the envelope function F and probability measures Q on A of the same form as the empirical measure as follows: for $r > 0$, $\varepsilon > 0$, and any measure Q on A of the form $Q = k^{-1} \sum_{i=1}^k \delta_{x(i)}$ for some finite collection $x(1), \ldots, x(k)$ of points in A, let $N_F^{(r)}(\varepsilon, \mathbf{F}, Q)$ be the number of balls of radius $\varepsilon \|F\|_{Q,r}$ required to cover $\mathbf{F}$ in $(L_r(Q), \|\cdot\|_{Q,r})$; here $\|f\|_{Q,r}^r = Q(|f|^r)$. Then set

$$N_F^{(r)}(\varepsilon, \mathbf{F}) := \sup_Q N_F^{(r)}(\varepsilon, \mathbf{F}, Q).$$

This is the combinatorial entropy of $\mathbf{F}$ defined by Pollard (1982b) and Kolčinskii (1981).

Here are Glivenko–Cantelli theorems and Donsker theorems formulated in terms of these entropies (and ignoring some measurability issues):

THEOREM 1 (SLLN or Glivenko-Cantelli theorem). *Suppose that:*

    (i) $N_F^{(1)}(\varepsilon, \mathbf{F}) < \infty$ *for every* $\varepsilon > 0$, *or*

    (i') $N_{[]}^{(1)}(\varepsilon, \mathbf{F}, P) < \infty$ *for every* $\varepsilon > 0$.

    (ii) $F \in L_1(P)$.

Then $\mathbf{F} \in \mathrm{SLLN}(P) : \|\mathbb{P}_n - P\|_{\mathbf{F}}^* \to_{\text{a.s.}} 0$ *as* $n \to \infty$.

The first part of Theorem 1—using the hypothesis (i) on the Pollard-Kolčinskii entropy $N_F^{(1)}(\varepsilon, \mathbf{F})$—is due to Pollard (1982b) with refinements by Dudley (1984); see Dudley (1984), theorem 11.1.6. The second part of Theorem 1—using the hypothesis (i') on the entropy with bracketing $N_{[]}^{(1)}(\varepsilon, \mathbf{F})$—is due to Blum (1955) and Dehardt (1971), again with refinement by Dudley (1984); see Dudley (1984), theorem 6.1.5.

THEOREM 2 (CLT or Donsker theorem). *Suppose that:*

    (i) $\displaystyle\int_0^1 \sqrt{\log N_F^{(2)}(u, \mathbf{F})}\, du < \infty$, *or,*

    (i') $\displaystyle\int_0^1 \sqrt{\log N_{[]}^{(2)}(u, \mathbf{F}, P)}\, du < \infty$.

    (ii) $F \in L_2(P)$.

Then $\mathbf{F} \in \mathrm{CLT}(P) : \mathbb{X}_n = \sqrt{n}\,(\mathbb{P}_n - P) \Rightarrow G_P$ *as* $n \to \infty$.

The first part of Theorem 2—using the hypothesis (i) on the Pollard–Kolčinskii entropy $N_F^{(2)}(\varepsilon, \mathbf{F})$—is due to Pollard (1982b). The second part of Theorem 2—using the hypothesis (i') on the entropy with bracketing $N_{[]}^{(2)}(\varepsilon, \mathbf{F})$—is due to Ossiander (1987).

A great variety of different theorems of this type have now been established, notably by Alexander (1987b), Andersen, Giné, Ossiander & Zinn (1988), Bolthausen (1978), Dudley (1978), (1984), (1987), Giné & Zinn (1984), (1986), and Pollard (1990). The influence of techniques from probability in Banach spaces on this development have been deep and profound; see especially Giné & Zinn (1984), (1986), (1990), (1991) and their discussion of Pollard (1989).

Perhaps the most important corollary to note for the present is that all the Vapnik–Červonenkis (VC) classes of sets (satisfying some additional measurability condition) satisfy the integrability hypothesis of Pollard's central limit Theorem 2(i), and hence satisfy the CLT for every fixed P. Moreover, as Pollard (1982b) shows, so does the collection of functions $\mathbf{F} := \{F1_C : C \in \mathbf{C}\}$ for any square integrable function F and VC-class $\mathbf{C}$. Since many classes of sets are indeed VC-classes (for example, in $A = R^d$, the classes of: all closed balls, open rectangles, all half spaces, all polyhedra with at most m faces, . . . are VC classes) these two observations alone yield a large variety of function classes $\mathbf{F}$ which satisfy the CLT.

It is important also to note that the above two theorems are formulated for a single fixed P, and the notation used reflects this: for example, we have written $\mathbf{F} \in \mathrm{CLT}(P)$ to emphasize that the result holds for P, but may not hold for another probability distribution $Q \neq P$. Of course in statistics we are often thinking about whole collections of P's at once, and are trying to 'decide' which P in a given collection (or model) is the 'best' P for the given data, and we therefore want to consider limit theorems for many, or perhaps all, probability distributions P on a given sample space A. Empirical process theory has begun to address these issues in the past few years: statistical motivations have led to the introduction of *universal Glivenko–Cantelli classes* and *universal Donsker classes* of functions $\mathbf{F}$: these are classes for which the SLLN or CLT respectively holds for *every* P on the sample space A; see e.g. Dudley (1978), (1987) for universal Donsker classes and Dudley, Giné & Zinn (1991) for universal Glivenko–Cantelli classes. Moreover, a statistician would also like to know that the limit theorems are also *uniform* in P (i.e. that the large $N(\varepsilon)$ necessary to make the implicit error less than ε does not depend on which P is true), and further about the *rate* of (preferably uniform) convergence to the limit. This has led to the introduction and study of $\mathbf{P}$-*uniform Glivenko–Cantelli classes* $\mathbf{P}$-*uniform Donsker classes* of functions $\mathbf{F}$ for a given collection $\mathbf{P}$ of probability distributions P on A; see e.g. Giné & Zinn (1991), Sheehy & Wellner (1992) for $\mathbf{P}$-*uniform Donsker classes* of functions $\mathbf{F}$, and Dudley, Giné & Zinn (1991) for $\mathbf{P}$-*uniform Glivenko–Cantelli classes* of functions $\mathbf{F}$.

We will briefly review some results concerning rates of convergence in Section II.2 below.

Statisticians are also interested in *local alternatives* both for the study of the power of tests and the regularity of estimators. This raises this issue of central limit theorems for local (or 'contiguous' alternatives); we will return to this issue in Section III.2.

II.2 Inequalities and Rates of Convergence

Often the key to understanding the behavior of some particular statistical method is an inequality. Prime examples of this are the inequalities of Dvoretzky, Kiefer & Wolfowitz (1956), and Kiefer (1961) for the 1-dimensional and d-dimensional empirical distribution

functions respectively:

(d = 1) $\Pr\{\sqrt{n}\,\|\mathbb{F}_n - F\|_\infty > \lambda\} \leq Ce^{-2\lambda^2}$ for all $\lambda > 0,\quad n \geq 1$;

and, for every $\varepsilon > 0$ there is a $C = C_\varepsilon$ so that

(d > 1) $\Pr\{\sqrt{n}\,\|\mathbb{F}_n - F\|_\infty > \lambda\} \leq Ce^{-(2-\varepsilon)\lambda^2}$ for all $\lambda > 0,\quad n \geq 1$.

Massart (1990) shows that $C = 2$ in (d = 1) works. Alexander (1984) refines Kiefer's $d > 1$ inequality to obtain

$$\Pr\{\sqrt{n}\,\|\mathbb{F}_n - F\|_\infty > \lambda\} \leq 16\lambda^{2^{12}(d+1)}e^{-2\lambda^2}\quad\text{for all}\quad \lambda \geq 8,\quad n \geq 1,$$

and this has been further refined by Adler & Brown (1986) to

$$\Pr_F\{\sqrt{n}\,\|\mathbb{F}_n - F\|_\infty > \lambda\} \leq C_F\lambda^{2(d-1)}e^{-2\lambda^2}$$

for all $\lambda > 0$, $n \geq$ some N_λ. But Alexander's (1984) inequalities apply much more generally: one of his special cases is: for any (measurable) Vapnik–Červonenkis classes of sets $\mathbf{C}$

$$\Pr_P^*\{\sqrt{n}\,\|\mathbb{P}_n - P\|_\mathbf{C} > \lambda\} \leq 16\lambda^{2^{12}v}e^{-2\lambda^2}\tag{1}$$

for all $\lambda \geq 8$, $n \geq 1$, where $v = V_\mathbf{C}$ is the Vapnik–Červonenkis index of $\mathbf{C}$. This was improved by Massart (1986) to

$$\Pr_P^*\{\sqrt{n}\,\|\mathbb{P}_n - P\|_\mathbf{C} > \lambda\} \leq C(\varepsilon, \mathbf{C})\lambda^{6v+\varepsilon}e^{-2\lambda^2}\tag{1'}$$

Alexander (1984) and Massart (1986) prove other exponential bounds like (1) and (1'), but with $\mathbf{C}$ replaced by some bounded class of functions $\mathbf{F}$ satisfying an 'appropriate' entropy condition.

Recently Giné & Zinn (1991) proved similar inequalities for arbitrary universal bounded Donsker classes $\mathbf{F}$ ($\mathbf{F}$ is *P-bounded Donsker* if

$$M_P \equiv \sup_{n \geq 1} E_P^* \|\sqrt{n}\,(\mathbb{P}_n - P)\|_\mathbf{F} < \infty;\tag{2}$$

$\mathbf{F}$ is *universal* bounded Donsker if (2) holds for all P). Here is one of their inequalities (which they derive via Gaussian process methods and Borell's inequality (see Adler (1990), page 43): if $\mathbf{F}$ is a (measurable) universal bounded Donsker class of functions with $0 \leq f \leq 1$ for all $f \in \mathbf{F}$, then

$$\Pr_P^*\{\sqrt{n}\,\|\mathbb{P}_n - P\|_\mathbf{F} > \lambda\} \leq 2\exp\{-\lambda^2/2\pi(2 + cM)\}$$

for all $\lambda > 0$, $n \geq 1$, and all P.

Other useful inequalities and limit theorems showing how $\mathbb{P}_n$ differs from P in a ratio sense, have been developed by Breiman, Friedman, Olshen & Stone (1984) in the course of work on classification and partitioning regression methods, and refined and extended by Pollard (1987) and Alexander (1987a). Here is one of Alexander's (1987a) ratio type limit theorems: If $\mathbf{C}$ is a Vapnik–Červonenkis class of sets and $a_n \to 0$ satisfies $n^{-1}\log(1/a_n) = o(a_n)$, then

$$\sup\left\{\left|\frac{\mathbb{P}_n(C)}{P(C)} - 1\right| : C \in \mathbf{C}, P(C) \geq a_n\right\} \to_p 0.\tag{3}$$

If furthermore $na_n/\log\log n \to \infty$, then the convergence in (3) is almost sure.

The best results concerning rates of convergence of empirical processes to their Gaussian limit processes are apparently those of Massart (1986), (1989). Massart (1986)

gives rates of convergence for Pollard's (1982b) CLT under additional hypotheses on the envelope function F and the entropy $N(\varepsilon, \mathbf{F})$, while Massart (1989) focuses on the case $A = R^k$ and imposes additional hypotheses on the entropy with bracketing $N_{[]}^{(1)}(\varepsilon, \mathbf{F}, P)$ of the class $\mathbf{F}$. Dudley & Philipp (1983) give rates of convergence for more general situations.

III Consequences for Statistics

III.1 Estimation

Empirical process methods and techniques have been applied to a wide range of estimation problems. The following rough groupings will help to organize the discussion:

    A. Models for censoring, truncation, biased sampling.
    B. Regression and density function estimation.
    C. Minimum distance estimation.

Here is a brief review of progress in these areas involving empirical process techniques or methods.

A Models. The problem of nonparametric estimation of a survival distribution subject to random right censorship provides a striking example of the application of empirical process methods. The Kaplan–Meier estimator was first derived in 1958, and its asymptotic normality (as a process) was established by Breslow & Crowley (1974) using empirical process tools and methods. The use of martingale methods by Gill (1983) simplified and extended the Breslow & Crowley results, but it is now widely recognized that the martingale methods do not extend to many closely related, but more complicated problems (for example, bivariate censored data) whereas the original proof of Breslow & Crowley (1974), which is now seen as an application of empirical process theory combined with the general 'delta method' (see e.g. Gill (1989)), does apply. We will return to this theme and the delta method in Section III.4.

Important progress on double censoring has recently been made by Chang (1990); he proves weak convergence of the nonparametric maximum likelihood estimator of Turnbull (1974) for doubly censored data.

For work on nonparametric estimation of a distribution function subject to left truncation, and left truncation together with right censoring, see Woodroofe (1985), Wang, Jewell & Tsai (1986), Keiding & Gill (1990), and Lai & Ying (1991). (Woodroofe (1985) gets a real data number of 1 in view of Lynden-Bell (1971).) Some of this development hinges more on martingale theory than empirical process theory per se; however, in just slight extensions of these problems the martingale theory breaks down, and empirical process theory again yields the most convenient tool.

Nonparametric maximum likelihood estimates for quite general biased sampling models was considered by Vardi (1985). He gave conditions for the existence of nonparametric maximum likelihood estimates of a general measure P in an s-sample biased sampling model. The limit theory for Vardi's estimators was established by Gill, Vardi & Wellner (1988).

There are many interesting and practically important truncation and censoring problems for higher dimensional data which have just begun to receive attention; see e.g. Dabrowska (1988), (1989) and Gill (1990) for bivariate censoring problems.

B Regression and density function estimation. Empirical process methods have received considerable use in several recent studies of regression methods in high dimensions. In a study of projection pursuit regression, Diaconis & Freedman (1984) use the original

Vapnik–Červonenkis (1971) inequality to obtain an interesting consistency-Glivenko–Cantelli theorem for half-spaces even for increasing dimension d: their result says that the Kolmogorov distance from empirical to true for half-spaces converges a.s. to zero if $d/n \to 0$ as $n \to \infty$. Huber (1985) uses an exponential inequality of Alexander (1984) to establish a consistency result for projection pursuit regression methods.

Breiman, Friedman, Olshen & Stone (1984) establish and use ratio-type inequalities as discussed briefly in Section II.2 to prove consistency results for their tree-structured regression methods. Pollard (1987) obtained refined inequalities for ratios and used them to prove consistency results for kernel type density and regression function estimators uniform in the bandwidth parameter; and Nolan & Marron (1989) used Pollard's (1987) inequalities to establish consistency results for automatic and adaptive bandwidth density estimators. In some closely related work, Yukich (1989) studies smoothed or 'perturbed' empirical processes quite generally: here the focus is on the smooth empirical measure itself, rather than on its density.

Van de Geer (1990) uses empirical process inequalities and techniques to establish rates of convergence of nonparametric least squares and least absolute deviations regression estimators. The entropy functions for various function classes, together with good exponential bounds, play a key role in her work.

C Minimum distance estimators. Minimum distance estimators and tests have a long history in connection with empirical distributions and measures: see e.g. Bolthausen (1977) for a study of estimators based on $L_2[0, 1]$-distance for real-valued data, and Pollard (1980) for a study of related tests. Beran & Millar (1987) introduce and study 'stochastic' minimum distance estimators (and tests). Pakes & Pollard (1989) study 'simulated optimization estimators' and establish large sample theory for these procedures using empirical process methods. In both Beran & Millar (1987) and in Pakes & Pollard (1989) some extra randomness is introduced in order to make calculations possible and simple in an otherwise complicated situation—and the analysis aims to take explicit account of the extra randomness introduced.

III.2 Testing

Considerable recent progress has been made in the construction of tests based on empirical measures and processes (and the related confidence sets obtained by inversion of the tests). Although empirical process methods are frequently useful for studying other test statistics, in this review I will focus on 'omnibus tests;' i.e. tests which are consistent against *all* alternatives for the particular problem under consideration.

Many of these recent developments are largely due to progress in the study and understanding of 'the bootstrap' (i.e. on reliance on limit theory for bootstrap resampling methods), and hence this section overlaps significantly with Section III.5 below on bootstrapping for empirical measures and processes. The key difficulty—to which the bootstrap methods give a solution— is that Kolmogorov-type statistics, which are 'distribution free' in one dimension under continuity hypotheses, fail to be distribution free in $d \geqslant 2$ dimensions; see e.g. Bickel (1969) or Romano (1988), page 700, for discussions. Our review here will be organized by way of the following categories:

A. Classical goodness of fit and minimum distance tests.
B. Permutation tests.
C. Power under local alternatives.

The theorem established in subsection C for local alternatives is the one new result contained in this review.

A   Classical goodness of fit and minimum distance tests. Two of the first papers to apply modern empirical process theory to testing are those of Pollard (1979), (1980). Pollard (1979) showed how Dudley's (1978) results could be used to allow for data dependent cells in the limit theory of chi-square goodness of fit tests (of course chi-square tests are *not* generally consistent against all alternatives so I have already violated my stated selection criteria; but this *is* a nice application of empirical process theory). Pollard (1980) used empirical process theory to study minimum distance tests of fit for parametric families, and to extend Bolthausen's (1977) results for minimum distance estimates. Pyke (1984) and Pyke & Wilbour (1988) have studied one-sample Kolmogorov statistics for testing a simple null hypothesis. Their statistics are based on translates of a fixed set such as a ball or square, and, since such a class of sets is a determining class as shown by Pyke (1984), are consistent against all alternatives. Pyke & Wilbour (1988) give a very interesting Monte-Carlo study of the power of such tests. It would be of some interest to have available sufficient theory in order to theoretically compute (or at least approximate) the power of their tests; Theorem III.2.1 below is a step in that direction.

In a major advance for tests based on empirical measures for directional data (data from probability distributions on a connected Riemannian manifold, e.g. the sphere S^p, hemisphere H^p, or torus T^p) Giné (1975) proposed classes of invariant tests of uniformity based on Sobolev norm distances from the empirical measure to the hypothesized uniform measure. Giné's proposed tests included many known test statistics as well as a variety of new test statistics consistent against all alternatives. These statistics are based on what might now be called "elliptical Donsker classes" $\mathbf{F}$ of functions.

B   Permutation and bootstrap tests. There is a long history of the use of permutation tests in connection with empirical measures for problems with two or more samples and for tests of independence. Bickel (1969) used the then existing empirical process theory of Dudley (1966) to establish consistency of a two-sample test based on the Kolmogorov static

$$D_{m,n}(\mathbf{O}) := \sup_{C \in \mathbf{O}} \|\mathbb{P}_m(C) - \mathbb{Q}_n(C)\|$$

where $\mathbf{O}$ is the class of lower left orthants. (It is of interest to note that Dudley (1969), page 41, also notes the possibility of using the permutation principle together with empirical measures in two-sample problems, but he is apparently suggesting a test based on either the Prohorov metric or the dual-bounded Lipschitz metric.) This was recently generalized to $k \geq 2$ samples and any Vapnik-Čhervonenkis class of sets $\mathbf{C}$ (replacing the orthants $\mathbf{O}$) by Romano (1989) (proposition 3.4, page 153). Romano (1988), (1989) also considers bootstrap implementations of these tests and a variety of other testing problems, e.g. independence.

Beran & Millar (1987), (1989) study 'stochastic' minimum distance tests based on empirical measures indexed by Vapnik–Čhervonenkis classes of sets; their analysis includes consideration of the extra randomness in the proposed procedure due to the 'stochastic norming' and the bootstrap implementation.

Wellner (1979) extended Giné's (1975) goodness of fit tests to two-sample testing problems using the permutation ideas of Bickel (1969). Jupp & Spurr (1983), (1985) have used this circle of methods and ideas to obtain tests of symmetry and independence for directional data. In all of these problems, the family of test statistics considered includes members which are consistent against all alternatives.

Further development of permutational limit theory for general empirical processes and tests based thereon is contained in the U.W. Ph.D. dissertation of Praestgaard (1991).

C   Local alternatives and power. In this section we give a Donsker theorem for the

empirical process under 'local alternatives' to a fixed probability distribution P_0. The theorem has many corollaries concerning the (local asymptotic) power of tests. Consider a sequence $\{P_n\}$ of measures on $(A, \mathbb{A})$. For each $n = 1, 2, \ldots$ we suppose that $X_{n1}, \ldots, X_{nn}$ are row independent, iid P_n, (A-valued) random variables. We assume that the resulting triangular array is defined on a common probability space

$$(\Omega, \Sigma, \mathrm{Pr}) := (A^1, \mathbb{A}^1, P_1^1) \times \cdots \times (A^n, \mathbb{A}^n, P_n^n) \times \cdots \times ([0, 1], \mathbf{B}, \lambda) \tag{1}$$

where λ denotes Lebesgue measure. We define the empirical measure $\mathbb{P}_n$ of the n random variables in the n-th row of the array by

$$\mathbb{P}_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_{ni}}, \tag{2}$$

and the empirical process by

$$\mathbb{X}_n^{P_n} := \mathbb{X}_n := \sqrt{n}\,(\mathbb{P}_n - P_n). \tag{3}$$

We will assume that the sequence $\{P_n\}$ satisfies

$$\int \{\sqrt{n}\,[(dP_n)^{\frac{1}{2}} - (dP_0)^{\frac{1}{2}}] - \tfrac{1}{2}h(dP_0)^{\frac{1}{2}}\}^2 \to 0 \tag{4}$$

for some $h \in L_2^0(P_0) := \{h \in L_2(P_0) : P_0(h) = 0\}$.

The following theorem (first proved by Sheehy & Wellner (1988)) asserts that the property $\mathbf{F} \in \mathrm{CLT}(P_0)$ is preserved for a sequence of local alternatives satisfying (4) under just the slight additional integrability condition (iii).

THEOREM 1. *Suppose that*:
> (i) $\{P_n\}$ *is a sequence* (*in* $\mathbf{P}$) *satisfying* (4) *for some* $h \in L_2^0(P_0)$.
> (ii) $\mathbf{F}$ *is a* P_0-*Donsker class.*

Then, with $\delta_h \in \mathbf{C}_u(\mathbf{F}, \rho_{P_0})$ *defined by* $\delta_h(f) := \int fh\,dP_0 = P_0(hf)$,

C1: $\mathbb{X}_n^0 := \sqrt{n}\,(\mathbb{P}_n - P_0) \Rightarrow \mathbb{X}_0 + \delta_h$ *under* P_n.

If, moreover,

> (iii) $\mathbf{F}$ *has envelope function* F *satisfying* $\limsup_{n \to \infty} P_n(F^2) < \infty$, *then*:

C2: $\Delta_n := \sqrt{n}\,(\mathbb{P}_n - P_0)$ *satisfies* $\|\Delta_n - \delta_h\|_{\mathbf{F}} \to 0$ *as* $n \to \infty$.

C3: $\mathbf{F} \in \mathrm{CLT}(\{P_n\}_{n \geqslant 0}) : \mathbb{X}_n^{P_n} \Rightarrow \mathbb{X}_0 \sim G_{P_0}$.

An easy corollary of our theorem for 'local alternatives' $\{P_n\}$ to P_0 is the behavior of the (local asymptotic) power of the $\mathbf{F}$-Kolmogorov statistic for testing $H_0 : P = P_0$ versus $H_1 : P \neq P_0$. Among other corollaries emphasized by Sheehy & Wellner (1988) is the *local regularity* of $\mathbb{P}_n$ as an estimator of P in $l^\infty(\mathbf{F})$.

COROLLARY. *Suppose that hypotheses* (i) *and* (ii) *of Theorem 1 hold, and for a fixed number* $0 < \alpha < 1$ *that* $c_\alpha(P_0)$ *satisfies*

$$\mathrm{Pr}\,(\|G_{P_0}\|_{\mathbf{F}} \geqslant c_\alpha(P_0)) = \alpha.$$

Then, with $D_n(\mathbf{F}) := \|\mathbb{P}_n - P_0\|_{\mathbf{F}}$, *under the null hypothesis* H_0,

$$\mathrm{Pr}_{P_0}^* (\sqrt{n}\,D_n(\mathbf{F}) \geqslant c_\alpha(P_0)) \to \mathrm{Pr}\,(\|G_{P_0}\|_{\mathbf{F}} \geqslant c_\alpha(P_0)) = \alpha, \tag{5}$$

and, under the sequence of local alternatives $\{P_n\}$

$$\mathrm{Pr}_{P_n}^* (\sqrt{n}\,D_n(\mathbf{F}) \geqslant c_\alpha(P_0)) \to \mathrm{Pr}\,(\|G_{P_0} + \delta_h\|_{\mathbf{F}} \geqslant c_\alpha(P_0)). \tag{6}$$

Remark 1. A first result in the direction of Theorem 1 and the corollary for general empirical processes is due to Pollard (1980), Section 6, who considered the case when **F** is the collection of indicators of some class of sets **C**. Note that our Theorem 1 and the corollary have a slightly different spirit than Pollard's however: our theorem is more in line with the spirit of Le Cam's third lemma (see e.g. Hájek & Šidak (1967), pages 202–210, or Shorack & Wellner (1986), pages 156, 157, and 165) which asserts that convergence under local alternatives for real valued statistics is *always* true if the statistics converge jointly in law under the null hypothesis. We use contiguity theory to deduce C3 from C1 via C2, whereas Pollard (1980) reverts to a condition involving an entropy calculated under P_n to prove C3, and then obtains a result like C1 from C2 and C3.

Remark 2. It is interesting to note that the conclusion C1 *always holds* for a Donsker class **F** under local alternatives $\{P_n\}$ satisfying (4). The additional integrability hypothesis (iii) is only used to replace P_0 by P_n in the centering and thereby to argue that C1 implies C3 by way of C2.

Proof of Theorem 1. First we prove that (i) and (ii) imply C1. The following argument relies on an extension of Le Cam's third lemma to the Hoffmann–Jørgensen weak convergence theory established in Van der Vaart and Wellner (1990).

First note that (4) implies that $\{Pr_{P_n}\}$ is contiguous to $\{Pr_{P_0}\}$ by Le Cam's first lemma: (4) implies that $\Lambda_n := \log \prod_{i=1}^{n} (p_n/p_0)(X_i)$ satisfies

(a)
$$\Lambda_n - \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i) - \tfrac{1}{2} P_0(h^2) \right) = o_{P_0}(1),$$

and hence that $\Lambda_n \to_d N(-\sigma^2/2, \sigma^2)$ under P_0 with $\sigma^2 = P_0(h^2)$.

Now consider $(\mathbb{X}_n^0, \Lambda_n)$ for $n = 1, 2, \ldots$ as elements of the product space $l^\infty(\mathbf{F}) \times \bar{R}$ equipped with the metric d given by $d((x, r), (y, s)) := \|x - y\|_{\mathbf{F}} \vee \arctan |r - s|$. Since $\mathbf{F} \in \mathrm{CLT}(P_0)$, $\mathbb{X}_n^0 \Rightarrow \mathbb{X}_0 \sim G_{P_0}$ in $l^\infty(\mathbf{F})$ under P_0, and $\{\mathbb{X}_n^0\}$ is tight under P_0. Since $\Lambda_n \Rightarrow \Lambda \sim N(-\sigma^2/2, \sigma^2)$, under P_0, it follows from Lemma 1.5 of Van der Vaart & Wellner (1990) that $\{(\mathbb{X}_n^0, \Lambda_n)\}$ is jointly tight under P_0. Now for $f \in \mathbf{F}$ we have $(\mathbb{X}_n^0(f), \Lambda_n) \Rightarrow N_2(\mu, \Sigma)$ under P_0 with $\mu = (0, -\sigma^2/2)$ and $\Sigma = (\sigma_{ij})$ given by $\sigma_{11} = \mathrm{Var}_{P_0}(f)$, $\sigma_{22} = P_0(h^2) = \mathrm{Var}_{P_0}(h)$, and $\sigma_{12} = \sigma_{21} = P_0(fh) = \mathrm{Cov}_{P_0}(f, h)$ (recall that $P_0(h) = 0$), and similarly for all the finite dimensional laws. Thus it follows that under P_0 we have $(\mathbb{X}_n^0, \Lambda_n) \Rightarrow (\mathbb{X}_0, \mathbb{X}_0(h) - P_0(h^2)) := (\mathbb{X}_0, \Lambda)$. By Le Cam's third lemma (for the Hoffmann–Jørgensen weak convergence theory; see Van der Vaart & Wellner (1990), lemma 1.6), it follows that $\mathbb{X}_n^0 \Rightarrow$ (some Borel measurable $\mathbb{Z}$) under P_n and

$$P(\mathbb{Z} \in B) = E 1_B(\mathbb{X}_0) e^\Lambda.$$

By standard calculations with the finite-dimensional laws, this implies $\mathbb{X}_n^0 \Rightarrow \mathbb{X}_0 + \delta_h$ under P_n where δ_h is as defined in C1, and this proves C1.

Now we prove C2. This goes along the lines of Lemma 5.21 of Van der Vaart (1988); we give the details for completeness. Define μ_n, p_n, p_0, s_n, and s_0 by $\mu_n := P_n + P_0$, $p_n := dP_n/d\mu_n$, $s_n := \sqrt{p_n}$, $p_0 := dP_0/d\mu_n$, and $s_0 := \sqrt{p_0}$. Then for $f \in \mathbf{F}$ and $\{P_n\}$ satisfying (4),

$$\Delta_n(f) - \delta_h(f) = \int f\{\sqrt{n}\,(s_n - s_0) - \tfrac{1}{2} h s_0\}(s_n + s_0)\,d\mu_n$$

(b)
$$+ \tfrac{1}{2} \int f h (s_n - s_0) s_0 \, d\mu_n$$

$$:= A_n(f) + B_n(f)$$

where

(c)
$$|A_n(f)| \le \left\{ 2 \int f^2 \, dP_n + 2 \int f^2 \, dP_0 \right\}^{\frac{1}{2}}$$

$$\cdot \left\{ \int [\sqrt{n} \, (s_n - s_0) - \tfrac{1}{2} h s_0]^2 \, d\mu_n \right\}^{\frac{1}{2}}$$

$$\le \{ 2P_n(F^2) + 2P_0(F^2) \}^{\frac{1}{2}}$$

$$\cdot \left\{ \int [\sqrt{n} \, (dP_n^{\frac{1}{2}} - dP_0^{\frac{1}{2}}) - \tfrac{1}{2} h \, dP_0^{\frac{1}{2}}]^2 \right\}^{\frac{1}{2}}$$

(d) $\qquad\qquad \rightarrow 0 \quad$ uniformly in $\ f \in \mathbf{F} \ $ as $\ n \rightarrow \infty.$

by hypotheses (i) and (iii). Furthermore,

(e)
$$|2B_n(f)| \le \left| \int_{|f| \le n^{1/4}} f h s_0 (s_n - s_0) \, d\mu_n \right|$$

$$+ \left| \int_{|f| > n^{1/4}} h s_0 f (s_n - s_0) \, d\mu_n \right|$$

$$\le \left\{ n^{\frac{1}{4}} \int h^2 \, dP_0 \int (s_n - s_0)^2 \, d\mu_n \right\}^{\frac{1}{2}}$$

$$+ \{ P_0(h^2 1_{[F > n^{1/4}]})(P_n(F^2) + P_0(F^2)) \}^{\frac{1}{2}}$$

(f) $\qquad\qquad \rightarrow 0 + 0 = 0 \quad$ uniformly in $\ f \in \mathbf{F} \ $ as $\ n \rightarrow \infty$

since $P_0(h^2) < \infty$ and by hypotheses (i) and (iii). Combining (b)–(f) yields C2.

By writing $\mathbb{X}_n^{P_n} = \mathbb{X}_n^0 - \sqrt{n} \, (P_n - P_0)$, C1 and C2 together yield C3. $\quad \square$

III.3 *Pattern Recognition; Clustering; Classification*

This subsection overlaps substantially with the regression and density function estimation material in Section II.1. Statistical problems in pattern recognition and classification clearly motivated the key work of Vapnik & Červonenkis (1971), and this is reflected in the books by Vapnik & Červonenkis (1974) and Vapnik (1982). We have already briefly mentioned the work on *Classification and Regression Trees* by Breiman, Friedman, Olshen & Stone (1984), in connection with the inequalities and ratio type limit theorems of Alexander (1987a) and Pollard (1987). Most of the applications of empirical process theory to date in this general area concern consistency results. A notable and striking exception to this (which proves the rule!) is the asymptotic distribution theory obtained by Pollard (1982a) of the cluster centers of MacQueen's (1967) k-means clustering method.

Sheehy (1988) established consistency of a clustering method based on Kullback–Leibler distances.

Empirical process theory and Vapnik–Červonenkis classes have also begun to appear in the study of 'neural networks'; see e.g. Baum (1988), Blumer, Ehrenfeucht, Haussler & Warmuth (1989), Barron (1989), and White (1990). The neural nets methods deserve to be compared more closely with alternative methods from statistics such as projection pursuit regression as mentioned in Section III.1 (Diaconis & Freedman (1984), Huber (1985)).

Considerable scope remains for the application of empirical process methods in this interesting area of statistics, and we cannot really do this large field justice here.

III.4　The Delta Method

The 'delta-method' is a time-honored tool in large sample theory in statistics with many uses, with many of the classical applications being connected with variance stabilizing transformations: for example, if $X_1, \ldots, X_n$ are i.i.d. Bernoulli(p), then $\sqrt{n}\,(\hat{p}_n - p) := \sqrt{n}\,(\bar{X}_n - p) \to_d N(0, p(1-p))$ by the DeMoivre–Laplace central limit theorem, and, with $\phi(x) := 2\arcsin(\sqrt{x})$ it follows by the elementary version of the delta method that

$$\sqrt{n}\,(\phi(\hat{p}_n) - \phi(p)) \to_d \dot{\phi}(p)N(0, p(1-p)) = N(0, 1).$$

(It seems to be less well-known that the same is true with ϕ replaced by $\bar{\phi}(x) := \arcsin(2x - 1)$; of course the range of $\bar{\phi}$ is quite different than the range of ϕ.)

The delta-method—in connection with empirical processes—is currently enjoying a revival, due in large part to the recent paper by Gill (1989), which builds on the earlier work of Reeds (1976). The importance of these developments is that they permit fairly direct and straightforward treatments of the asymptotic theory of non-linear functionals ϕ of the empirical measure $\mathbb{P}_n$ in a clean and intuitive way. We now give a simple generalization of Gill's (1989) theorem 1.

Suppose that $\mathbf{P} \subset l^\infty(\mathbf{F})$ is a subset of all the probability measures $\mathbf{M}$ on $(A, \mathbb{A})$, and that $\mathbf{P}$ is large enough to contain all the empirical measures $\mathbb{P}_n$. Consider $\phi : \mathbf{P} \to \mathbf{B}$ where $\mathbf{B}$ is a Banach space. Then ϕ is *Hadamard differentiable or compactly differentiable at $P \in \mathbf{P}$* if there is a continuous linear function $\dot{\phi} : l^\infty(\mathbf{F}) \to \mathbf{B}$ such that, for any sequence of numbers $\varepsilon_n \to 0$ and any sequence $\{\Delta_n\} \subset l^\infty(\mathbf{F})$ with $\{P + \varepsilon_n \Delta_n\} \subset \mathbf{P}$ and $\|\Delta_n - \Delta\|_\mathbf{F} \to 0$ for some $\Delta \in l^\infty(\mathbf{F})$

$$\frac{\phi(P + \varepsilon_n \Delta_n) - \phi(P)}{\varepsilon_n} \to \dot{\phi}(\Delta) \quad \text{as} \quad n \to \infty \quad \text{in } \mathbf{B}. \tag{1}$$

Furthermore, ϕ is *Fréchet differentiable at $P \in \mathbf{P} \subset l^\infty(\mathbf{F})$* if

$$\phi(P_n) - \phi(P) - \dot{\phi}(P_n - P) = o(\|P_n - P\|_\mathbf{F})$$

for any sequence $\{P_n\} \subset \mathbf{P}$. Here is a simple example of a function ϕ defined on pairs of probability distributions (or, in this case, df's) which is compactly differentiable with respect to the familiar supremum or uniform norm, but which is *not* Fréchet differentiable with respect to this norm.

Example 1. For distribution functions F, G on R, let define ϕ by $\phi(F, G) = \int F\,dG = P(X \leq Y)$ where $X \sim F$, $Y \sim G$ are independent. Let $\|\bar{F} - F_\infty\| := \sup_x |\bar{F}(x) - F(x)| := \|\bar{F} - F\|_\mathbf{F}$ where $\mathbf{F} := \{1_{(-\infty, t]} : t \in R\}$. Then $\phi(F, G)$ is Hadamard differentiable at every pair of df's (F, G) with derivative $\dot{\phi}$ given by

$$\dot{\phi}(\alpha, \beta) = \int \alpha\,dG - \int \beta\,dF. \tag{2}$$

See Gill (1989), lemma 3 for a proof of this. But ϕ is *not* Fréchet differentiable with respect to $\|\cdot\|_\infty$. If ϕ were Fréchet-differentiable, it would have to be true that

$$\phi(F_n, G_n) - \phi(F, G) - \dot{\phi}(F_n - F, G_n - G) = o(\|F_n - F\|_\infty \vee \|G_n - G\|_\infty) \tag{3}$$

for every sequence of pairs of df's $\{(F_n, G_n)\}$ with $\|F_n - F\|_\infty \to 0$ and $\|G_n - G\|_\infty \to 0$. We now exhibit a sequence $\{(F_n, G_n)\}$ for which (3) fails.

By straightforward algebra using (2),

$$\phi(F_n, G_n) - \phi(F, G) - \dot{\phi}(F_n - F, G_n - G) = \int (F_n - F)\,d(G_n - G). \tag{4}$$

Consider the df's F_n and G_n corresponding to the measures which put masses n^{-1} at $0, \ldots, (n-1)/n$ and $1/n, \ldots, 1$ respectively:

$$F_n = n^{-1} \sum_{k=0}^{n-1} \delta_{k/n}, \quad \text{and} \quad G_n = n^{-1} \sum_{k=1}^{n} \delta_{k/n}.$$

Both of these sequences of df's converge uniformly to the uniform $(0, 1)$ df $F(x) := x :=$ $G(x)$, and furthermore $\|F_n - F\|_\infty = \|G_n - G\|_\infty = 1/n$. Now

$$(F_n - F)(x) = \sum_{k=1}^{n} \left(\frac{k}{n} - x\right) 1_{[(k-1)/n, k/n)}(x),$$

$(F_n - F)(1) = 0$, and

$$(G_n - G)(x) = (F_n - F)(x) - \frac{1}{n} = \sum_{k=1}^{n} \left(\frac{k-1}{n} - x\right) 1_{[(k-1)/n, k/n)}(x)$$

with $(G_n - G)(1) = 0$. Thus, separating $G_n - G$ into its discrete and continuous parts,

$$\int (F_n - F) d(G_n - G) = \sum_{k=1}^{n} (F_n - F)\frac{1}{n} + n \int_0^{1/n} \left(\frac{1}{n} - t\right)\{-dt\}$$

$$= n\frac{1}{n}\frac{1}{n} - n\left\{\frac{1}{n}\frac{1}{n} - \frac{1}{2}\left(\frac{1}{n}\right)^2\right\}$$

$$= \frac{1}{2n} = O\left(\frac{1}{n}\right)$$

$$\neq o(\|F_n - F\|_\infty \vee \|G_n - G\|_\infty) = o(1/n).$$

Hence (3) fails and ϕ is not Fréchet-differentiable. [This example was suggested to me by R. M. Dudley in Seattle in February, 1990.]

Dudley (1990b) has found 'bigger' norms for which this ϕ is almost Fréchet differentiable. As far as I know however, $\phi(F, G) = \int F \, dG$ has not yet been shown to be Fréchet differentiable with respect to any norm compatible with the empirical measures $\mathbb{P}_m$ and $\mathbb{Q}_n$ (or *df*'s $\mathbb{F}_n$ and $\mathbb{G}_n$). Here we choose to focus on compact differentiability and refinements thereof.

A particular refinement of Hadamard differentiability which is very useful is as follows: since the limiting P-Brownian bridge process G_P of the empirical process $\mathbb{X}_n$ is in $\mathbf{C}_u(\mathbf{F}, \rho_P)$ with probability one for any $\mathbf{F} \in \mathrm{CLT}(P)$, we say that ϕ is *Hadamard differentiable tangentially to* $\mathbf{C}_u(\mathbf{F}, \rho_P)$ at $P \in \mathbf{P}$ if there is a continuous linear function $\dot{\phi} : \mathbf{C}_u(\mathbf{F}, \rho_P) \to \mathbf{B}$ so that (1) holds for any sequence $\{\Delta_n\}$ such that $\|\Delta_n - \Delta_0\|_\mathbf{F} \to 0$ with $\Delta_0 \in \mathbf{C}_u(\mathbf{F}, \rho_P)$. Then a nice version of the delta-method for nonlinear functions ϕ of $\mathbb{P}_n$ is given by the following theorem:

THEOREM. *Suppose that:*
(i) *ϕ is Hadamard differentiable tangentially to $\mathbf{C}_u(\mathbf{F}, \rho_P)$ at $P \in \mathbf{P}$.*
(ii) *$\mathbf{F} \in \mathrm{CLT}(P)$: $\sqrt{n}\,(\mathbb{P}_n - P) \Rightarrow \mathbb{X}_0$ (where $\mathbb{X}_0$ takes values in $\mathbf{C}_u(\mathbf{F}, \rho_P)$ by definition of $\mathbf{F} \in \mathrm{CLT}(P)$).*

Then

$$\sqrt{n}\,(\phi(\mathbb{P}_n) - \phi(P)) \Rightarrow \dot{\phi}(\mathbb{X}_0). \tag{5}$$

Proof. Define $g_n : \mathbf{P} \subset l^\infty(\mathbf{F}) \to \mathbf{B}$ by

$$g_n(x) := \sqrt{n}\,(\phi(P + n^{-\frac{1}{2}}x) - \phi(P)) 1_\mathbf{P}(P + n^{-\frac{1}{2}}x).$$

Then, by (i), for $\{\Delta_n\}$ in $l^\infty(\mathbf{F})$ with $\|\Delta_n - \Delta_0\|_\mathbf{F} \to 0$ and $\Delta_0 \in \mathbf{C}_u(\mathbf{F}, \rho_P)$,

$$g_n(\Delta_n) \to \dot{\phi}(\Delta_0) := g(\Delta_0).$$

Thus by the extended continuous mapping theorem in the Hoffmann–Jørgensen weak convergence theory (see Van der Vaart & Wellner (1990), proposition 1.5.A), $g_n(\mathbb{X}_n) \Rightarrow g(\mathbb{X}_0) = \dot{\phi}(\mathbb{X}_0)$, and hence (5) holds. $\square$

The immediate corollary for the classical Mann–Whitney form of the Wilcoxon statistic given in Example 1 is:

COROLLARY 1. *If $X_1, \ldots, X_m$ are iid F and independent of $Y_1, \ldots, Y_n$, which are iid G, and $\lambda_N := m/N := m/(m+n) \to \lambda \in (0, 1)$, then*

$$\sqrt{\frac{mn}{N}} \left\{ \int \mathbb{F}_m \, d\mathbb{G}_n - \int F \, dG \right\} = \sqrt{\frac{mn}{N}} \left\{ \phi(\mathbb{F}_m, \mathbb{G}_n) - \phi(F, G) \right\}$$

$$\to_d \sqrt{1-\lambda} \int \mathbb{U}(F) \, dG - \sqrt{\lambda} \int \mathbb{V}(G) \, dF$$

$$\sim N(0, \sigma_\lambda^2(F, G))$$

where $\mathbb{U}$ and $\mathbb{V}$ are two independent Brownian bridge processes and

$$\sigma_\lambda^2(F, G) = (1 - \lambda) \operatorname{Var}(G(X)) + \lambda \operatorname{Var}(F(Y)).$$

This is, of course, well-known, and can be proved in a variety of other ways (by treating $\phi(\mathbb{F}_m, \mathbb{G}_n)$ as a two-sample U-statistic, or as a rank statistic, or by a direct analysis), but the proof via the differentiable functional approach seems instructive and useful.

Other interesting applications of the delta method have recently been given by: Grübel (1988) (who studies the asymptotic theory of the length of the shorth); Pons & Turckheim (1989) (who study bivariate hazard estimators and tests of independence based thereon), and Gill & Johansen (1990) (who prove Hadamard differentiability of the 'product integral'). Gill, Van der Laan, & Wellner (1992) give applications to several problems connected with estimation of bivariate distributions. Arcones & Giné (1990) study the delta method in connection with M-estimation and related methods of Pollard (1985). Van der Vaart (1991b) shows that Hadamard differentiable functions preserve asymptotic efficiency properties of estimators.

III.5 *Bootstrapping empirical measures*

As mentioned in Section III.2 on testing, the bootstrap methods introduced by Efron (1979), (1982) have opened up many possibilities for inference that were previously intractable or unapproachable by other methods. Efron's basic idea of sampling from the empirical measure $\mathbb{P}_n^\omega$ is intuitively appealing and relatively easy to explain to non-statisticians. To discuss recent progress in this area we need some further notation and terminology.

To discuss bootstrapping, we need to include the 'omega' in our notation. Thus we write $\mathbb{P}_n^\omega$ for the empirical measure $\mathbb{P}_n$ for the fixed data $X_1(\omega), X_2(\omega), \ldots, X_n(\omega), \ldots$. If $X_1^\#, X_2^\#, \ldots, X_n^\#$ is a 'bootstrap sample' from $\mathbb{P}_n^\omega$, then $\mathbb{P}_n^\# := n^{-1} \sum_1^n \delta_{X^\#}$ is the *bootstrap empirical measure* and $\mathbb{X}_n^\# = \sqrt{n} \, (\mathbb{P}_n^\# - \mathbb{P}_n^\omega)$ is the *bootstrap empirical process*. It is important to note that since $\mathbb{P}_n^\omega$ is discrete, sampling from $\mathbb{P}_n^\omega$ with replacement just yields a multinomial distribution; and hence if $M_n \sim \operatorname{Mult}_n(n; (1/n, \ldots, 1/n))$ is independent of $X_1, \ldots, X_n$ then we can think of the bootstrap empirical process $\mathbb{X}_n^\#$ as

$$\mathbb{X}_n^\# = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n M_{ni} \delta_{X_i(\omega)} - \mathbb{P}_n^\omega \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (M_{ni} - 1) \delta_{X_i(\omega)}.$$

The first limit theorems justifying the bootstrap for empirical distribution functions in the case $A = R$ were established by Bickel & Freedman (1981). Gaenssler (1986) extended their results for Efron's bootstrap to the empirical process indexed by Vapnik–Červonenkis classes of sets. Meanwhile, Rubin (1981) proposed a 'Bayesian bootstrap' in which the weights or multipliers M_{ni}/n are replaced by D_{ni}, $i = 1, \ldots, n$ where $\{D_{ni}\}$ are the spacings from $n - 1$ independent Uniform$(0, 1)$ random variables. Lo (1987) established asymptotic theory for Rubin's Bayesian bootstrap. Both of these bootstrap methods are special cases of a general *exchangeably weighted bootstrap* based on an exchangeable random vector $W_n = (W_{n1}, \ldots, W_{nn})$ with $\sum_1^n W_{ni} = 1$, $W_{ni} \geq 0$. Then we call

$$\mathbb{P}_n^W := \sum_{i=1}^n W_{ni} \delta_{X_i(\omega)}$$

the *general exchangeable bootstrap empirical measure* and

$$\mathbb{X}_n^W := \sqrt{n} \, (\mathbb{P}_n^W - \mathbb{P}_n^\omega) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (nW_{ni} - 1)\delta_{X_i(\omega)}$$

is the *general exchangeable bootstrap empirical process*.

Inspired by a 'multiplier central limit theorem' of Ledoux & Talagrand (1988) (apparently J. Zinn was also involved in the proof thereof), Giné & Zinn (1990) proved the following elegant characterization of the consistency of Efron's bootstrap:

THEOREM. (Giné & Zinn, 1990). *Suppose that* $\mathbf{F} \in M(P)$ *(a measurability condition).*
    I. *Then the following are equivalent:*
        A. $\mathbf{F} \in \text{CLT}(P)$ *and* $P(F^2) < \infty$.
        B. $\mathbb{X}_n^\# \Rightarrow \mathbb{X}_0^\# \sim G_p$ *for* P^∞-*a.e.* ω.
    II. *Moreover, the following are also equivalent:*
        C. $\mathbf{F} \in \text{CLT}(P)$.
        D. $\mathbb{X}_n^\# \Rightarrow \mathbb{X}_0^\# \sim G_p$ *in* P^∞-*probability*.

For a precise formulation of D, see Giné & Zinn (1990). The multiplier CLT of Ledoux and Talagrand asserts something very much akin to the equivalence of A and B in part I, but with the vector of multipliers $M_{ni} - 1$ replaced by a vector of iid random variables $Y_1, Y_2, \ldots$ satisfying the $L_{2,1}$ condition

$$\int_0^\infty \sqrt{P(|Y_1| > t)} \, dt < \infty.$$

Giné and Zinn prove their theorem by symmetrization and Poissonization to get domination of the (i.e. Efron's) bootstrap empirical process by a process with iid symmetrized Poisson multipliers which easily satisfy the $L_{2,1}$ condition. For a variant of their approach which avoids symmetrization, see Klaassen & Wellner (1991).

Praestgaard (1990) treats exchangeable weights of the form $W_{ni} = Y_i/\sum_1^n Y_i$ for nonnegative iid Y_i's (note that standard exponential Y's correspond to the Rubin's Bayesian bootstrap by way of standard equivalences in law for uniform spacings and exponentials divided by their sums) by fairly direct use of the Ledoux & Talagrand (1988) multiplier CLT. In his U.W. Ph.D. dissertation, Praestgaard (1991) treats the general exchangeable weighted bootstrap empirical process $\mathbb{X}_n^W$ and obtains an analogue of the Giné and Zinn theorem for Efron's bootstrap under an appropriate $L_{2,1}$ condition on the W_{ni}'s; see Praestgaard & Wellner (1992). Praestgaard's (1991) theorem contains all the known limit theorems for bootstrapping general empirical processes (of which I am aware) and gives asymptotic justification of many new exchangeable bootstraps as well.

Beran & Millar (1986), (1987), (1989) give several applications of bootstrap methods to testing, estimation, and the construction of confidence sets. Their joint work with Le Cam (Beran, Le Cam & Millar (1987)) takes explicit account of the extra randomness introduced in the implementation of bootstrap methods. As mentioned in Section III.2, Romano (1988), (1989) studies implementations of nonparametric tests based on both bootstrap and permutation ideas. Pons & Turkheim (1989) apply bootstrap methods to problems of estimating a bivariate hazard function in the presence of right censoring.

In an extension of Bickel & Freedman (1981), Gill (1989) has developed bootstrap methods in connection with the delta method. His results have been further extended to general empirical processes by Sheehy & Wellner (1988), Pons & Turkheim (1989), and Arcones & Giné (1990).

Olshen, Biden, Wyatt & Sutherland (1989) give a nontrivial application of bootstrap methods to "gait analysis." (As mentioned in the introduction, this paper is an excellent example of a paper with a real data number of 0!) This application motivated and was made possible as a result of the development of new Vapnik–Chervonenkis classes of sets by Stengle & Yukich (1989).

We should also mention the important work on refined bootstrap limit theory for one-dimensional empirical and quantile processes by Csörgő & Mason (1989) who show how their methods yield useful asymptotic theory for bootstrapping many other 'empirical functions' including mean residual life, total time on test, and the Lorenz curve and its inverse. They also give an interesting re-analysis of the von Bortkiewitz yearly deaths by horsekicks data (and hence their paper also receives a real data number of 0!).

III.6 Miscellaneous

There have been a number of recent developments in which empirical process methods or extensions thereof are used to treat slightly nonstandard problems arising in applications. I will list only a few of these here, with the selection governed primarily by my estimate of potential for further development and applications. They will be grouped under the loose heading of *nonstandard asymptotics*, *U-statistics*, and *dimension asymptotics*.

Perhaps the most recent work on nonstandard asymptotic theory with clear ties to empirical process theory is that of Kim & Pollard (1990). They focus on problems with 'a sharp edge effect' which leads to $n^{1/3}$ normalizations to obtain limit distributions rather than the usual (smooth case) normalizations of $n^{\frac{1}{2}}$. By systematic use of empirical process methods, they redevelop known results for the shorth estimator of location and the Grenander estimator of a monotone density studied by Groeneboom (1985), and go on to obtain new results concerning mode estimation in higher dimensions and least median of squares estimates in econometrics. It would be interesting to see their methods applied to the interval censoring models studied by Groeneboom (1991)—which involve $n^{1/3}$ and $(n \log n)^{1/3}$ normalizations. For an important alternative way of *distinguishing* standard $n^{\frac{1}{2}}$-situations from nonstandard situations, see Van der Vaart (1991a).

Some further developments using similar tools are due to Nolan: Nolan (1989a) applies empirical process tools to multivariate trimming; Nolan (1989b) studies analogs of the shorth on the spheres S^p; and Nolan (1989c) examines a particular multivariate analog of the median: the point of 'greatest depth' with respect to the empirical measure.

Work on the asymptotic theory of U-statistics and U-statistics processes has proceeded rapidly, with attendant applications: see for example Nolan & Pollard (1987), (1988), and Arcones & Giné (1991). The latter give an interesting application to Liu's (1990) 'simplicial depth process'.

Problems involving increasing dimension of the underlying sample space or increasing numbers of parameters with sample size have been arising more frequently in statistics in recent years: see e.g. Diaconis & Freedman (1984) for an example of the former, and Portnoy (1988), Sauermann (1989), and the references contained therein for the latter. As mentioned in Section III.1, Diaconis & Freedman (1984) made good use of one of the Vapnik–Červonenkis (1971) inequality to obtain consistency results. Several of the suggestions of Pyke (1991) concerning exchangeable models seem closely related. The literature concerned with increasing dimensionality of the parameter space has not yet interacted substantially with empirical process theory, but with a few exceptions: see e.g. Donoho & Liu (1991) (who make good use of 'chaining arguments'), Birgé (1983), and Le Cam (1973). I predict that a great unification and simplification of these problems will result from the application of empirical process ideas and methods in the next few years.

References

Adler, R.J. (1990). An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. *IMS Lecture Notes—Monograph Series*, **12**, Institute of Mathematical Statistics, Hayward.

Adler, R.J. & Brown, L.D. (1986). Tail behaviour for suprema of empirical processes. *Ann. Probability* **14**, 1–30.

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

Alexander, K.S. (1987a). Rates of growth and sample moduli for weighted empirical processes indexed by sets. *Probab. Th. Rel. Fields* **75**, 379–423.

Alexander, K.S. (1987b). The central limit theorem for empirical processes on Vapnik–Červonenkis classes. *Ann. Probability* **15**, 178–203.

Andersen, N.T., Giné, E., Ossiander, M. & Zinn, J. (1988). The central limit theorem and the law of the iterated logarithm for empirical processes under local conditions. *Probab. Th. Rel. Fields* **77**, 271–306.

Arcones, M.A. & Giné, E. (1990). On the Bootstrap of M-estimators and other statistical functionals. *Preprint*, Department of Statistics, University of Connecticut, Storrs.

Arcones, M.A. & Giné, E. (1991). Limit theorems for U-processes. *Preprint*.

Barron, A. (1989). Statistical properties of artificial neural networks, in *Proceedings of the 28th IEEE Conference on Decision and Control*, **1**, pp. 280–285, IEEE Press, New York.

Baum, E.B. (1988). On the capabilities of multilayer perceptrons. *J. of Complexity* **4**, 193–215.

Beran, R. & Millar, P.W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.* **14**, 431–443.

Beran, R. & Millar, P.W. (1987). Stochastic estimation and testing. *Ann. Statist.* **15**, 1131–1154.

Beran, R. & Millar, P.W. (1989). A stochastic minimum distance test for multivariate parameteric models. *Ann. Statist.* **17**, 125–140.

Beran, R.J., Le Cam, L. & Millar, P.W. (1987). Convergence of stochastic empirical measures. *J. Mult. Anal.* **23**, 159–168.

Bickel, P.J. (1969). A distribution free version of the Smirnov two-sample test in the p-variate case. *Ann. Math. Statist.* **40**, 1–23.

Bickel, P.J. & Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196–1217.

Birgé, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. verw. Gebiete* **65**, 181–237.

Blum, J.R. (1955). On the convergence of empiric distribution functions. *Ann. Math. Statist.* **26**, 527–529.

Blumer, A., Ehrenfeucht, A., Haussler, D. & Warmuth, M. (1986). Classifying learnable geometric concepts with the Vapnik–Červonenkis dimension, in *Proceedings 18th ACM Symp. on Theory of Computing*, pp. 273–282, Assoc. Comp. Mach., New York.

Bolthausen, E. (1977). Convergence in distribution of minimum distance estimators. *Metrika* **24**, 215–227.

Bolthausen, E. (1978). On weak convergence of an empirical process indexed by the closed convex subsets of I-2. *Z. Wahrsch. verw. Gebiete* **43**, 173–181.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984). *Classification and regression trees*. Wadsworth, Belmont.

Breslow, N. & Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2**, 437–453.

Chang, M. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18**, 391–404.

Cramér, H. (1928). On the composition of elementary errors, II. *Skand. Aktuar.* **11**, 141–180.

Csörgő, M., Csörgő, S., Horváth, L. & Mason, D.M. (1986). Weighted empirical and quantile processes. *Ann. Probability* **14**, 31–85.

Csörgő, S. & Mason, D.M. (1989). Bootstrapping empirical functions. *Ann. Statist.* **17**, 1447–1471.

Dabrowska, D.M. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16**, 1475–1489.

Dabrowska, D.M. (1989). Kaplain–Meier estimate on the plane: weak convergence, LIL, and the bootstrap. *J. Mult. Anal.* **29**, 308–325.

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

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

Donoho, D.L. & Liu, R.C. (1991). Geometrizing rates of convergence, II. *Ann. Statist.* **19**, 633–667.

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

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

Dudley, R.M. (1978). Central limit theorems for empirical measures. *Ann. Probability* **6**, 899–929. Correction (1979) **7**, 909–911.

Dudley, R.M. (1981). Some recent results on empirical processes. *Lect. Notes in Math.* **806**, 107–123, Springer-Verlag, New York.

Dudley, R.M. (1984). A course on empirical processes; École d'Été de Probabilités de St. Flour. *Lecture Notes in Math.* **1097**, 2–142, Springer Verlag, New York.

Dudley, R.M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. *Lecture Notes in Math.* **1153**, 141–178, Springer-Verlag, New York.

Dudley, R.M. (1987). Universal Donsker classes and metric entropy. *Ann. Probability* **15**, 1306–1326.

Dudley, R.M. (1990a). Nonlinear functionals of empirical measures and the bootstrap. *Probability in Banach Spaces, Progress in Probability* **21**: 7, 63–82, Birkhauser, Boston.

Dudley, R.M. (1990b). Fréchet differentiability, p-variation, and universal Donsker classes. *Preprint.*

Dudley, R.M., Giné, E. & Zinn, J. (1991). Uniform and universal Glivenko–Cantelli classes. *J. Theor. Prob.* **4**, 485–510.

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

Dvoretzky, A., Kiefer, J. & Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27**, 642–669.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1–16.

Efron, B. (1982). The Jackknife, the Bootstrap, and Other Resampling Plans. *CBMS-NSF Regional Conference Series in Applied Mathematics,* **38**, Society for Industrial and Applied Mathematics, Philadelphia.

Fortet, R. and Mourier, E. (1953). Convergence de la répartition empirique vers la repartition théorique. *Ann. Scient. École Norm. Sup.* **70**, 266–285.

Gaenssler, P. (1983). Empirical Processes. *Lecture Notes—Monograph Series,* **3**, Institute of Mathematical Statistics, Hayward.

Gaenssler, P. (1986). Bootstrapping empirical measures indexed by Vapnik–Chervonenkis classes of sets, in *Proceedings IV Vilnius Conference* (1985) *Prob. Theory and Math. Statist.* **1**, 467–481.

Gill, R.D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 49–58.

Gill, R.D. (1989). Non- and semiparametric maximum likelihood estimators and the von-Mises method (part I). *Scand. J. Statist.* **16**, 97–128.

Gill, R.D. (1990). Multivariate survival analysis, in *Proceedings of the 2nd World Congress of the Bernoulli Society, 53rd annual meeting of the IMS; Uppsala,* 1990, Department of Mathematics, University of Utrecht.

Gill, R.D., Van der Laan, M. & Wellner, J.A. (1992). Asymptotic theory for bivariate estimators in three models. *Manuscript in preparation.*

Gill, R.D. & Johansen, S. (1990). A survey of product integration with a view toward application in survival analysis. *Ann. Statist.* **18**, 1501–1555.

Gill, R., Vardi, Y. & Wellner, J.A. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16**, 1069–1112.

Giné, E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.* **3**, 1243–1266.

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

Giné, E. & Zinn, J. (1986). Lectures on the central limit theorem for empirical processes. *Lect. Notes in Math.* **1221**, 50–113, Springer, Berlin.

Giné, E. & Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Probability* **18**, 851–869.

Giné, E. & Zinn, J. (1991). Gaussian characterization of uniform Donsker classes of functions. *Ann. Probability* **19**, 758–782.

Grenander, U. (1956). On the theory of mortality measurement, Part II, *Skand. Akt.* **39**, 125–153.

Groeneboom, P. (1985). Estimating a monotone density, in *Proceedings of the Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II,* ed. L.M. Le Cam and R.A. Olshen, Wadsworth, Belmont.

Groeneboom, P. (1988). Asymptotics for incomplete censored observations, *Technical Report,* **87-18**, Department of Mathematics, University of Amsterdam.

Groeneboom, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution, *Technical Report,* **91-53**, Faculty of Technical Mathematics and Informatics, Delft University of Technology.

Groeneboom, P. & Wellner, J.A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation,* DMV Seminar, Band 19. Birkhauser, New York.

Grübel, R. (1988). The length of the shorth, *Ann. Statist.* **16,** 619–628.

Hájek, J. & Šidák, Z. (1967). *Theory of Rank Tests,* Academia, Prague.

Hoffmann-Jørgensen, J. (1984). *Stochastic Processes on Polish Spaces,* Unpublished manuscript.

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

Jupp, P.E. & Spurr, B.D. (1983). Sobolev tests for symmetry of directional data, *Ann. Statist.* **11,** 1225–1231.

Jupp, P.E. & Spurr, B.D. (1985). Sobolev tests for independence of directions, *Ann. Statist.* **13,** 1140–1155.

Kaplan, E.L. & Meier, P. (1958). Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.* **53,** 457–481.

Keiding, N. & Gill, R.D. (1990). Random truncation models and Markov processes, *Ann. Statist.* **18,** 582–602.

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

Kiefer, J. & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters, *Ann. Math. Statist.* **27,** 887–906.

Kim, J. & Pollard, D. (1990). Cube root asymptotics, *Ann. Statist.* **18,** 191–219.

Klaassen, C.A.J. & Wellner, J.A. (1992). Kac empirical processes and the bootstrap, in *Proceedings of the Eighth International Conference on Probability in Banach Spaces,* p. 411–429, Birkhauser, New York, M. Hahn and J. Kuelbs, editors.

Kolčinskii, V.I. (1981). On the central limit theorem for empirical measures, *Theor. Probability Math. Statist.* **24,** 71–82.

Lai, T.L. & Ying, Z. (1991). Estimating a distribution with truncated and censored data, *Ann. Statist.* **19,** 417–442.

Le Cam, L. (1973). Convergence of estimates under dimensionality restrictions, *Ann. Statist.* **1,** 38–53.

Ledoux, M. & Talagrand, M. (1986). Conditions d'integrabilite pour les multiplicateurs dans le TLC Banachique, *Ann. Probability* **14,** 916–921.

Ledoux, M. & Talagrand, M. (1988). Un critère sur les petites boules dans le théorème limite central, *Probab. Th. Rel. Fields* **77,** 29–47.

Ledoux, M. & Talagrand, M. (1991). *Probability in Banach Spaces,* Springer-Verlag, New York.

Liu, R.V. (1990). On a notion of data depth based on random simplices, *Ann. Statist.* **18,** 405–414.

Lo, A.Y. (1987). A large sample study of the Bayesian bootstrap, *Ann. Statist.* **15,** 360–375.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars, *Mon. Not. R. Astr. Soc.* **166,** 281–295.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations, *Proceedings Fifth Berkeley Symp. Math. Statist. Prob.* **1,** 281–297, University of California Press, Berkeley. L. Le Cam and J. Neyman, editors.

Massart, P. (1986). Rates of convergence in the central limit theorem for empirical measures, *Ann. Inst. Henri Poincaré* **22,** 381–423.

Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via KMT constructions, *Ann. Probability* **17,** 266–291.

Massart, P. (1990). The tight constant in the Dvoretzky–Kiefer Wolfowitz inequality, *Ann. Probability* **18,** 1269–1283.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration, *Econometrica* **57,** 995–1026.

Nolan, D. (1989a). Asymptotics for multivariate trimming, *Technical Report* **151,** Department of Statistics, University of California, Berkeley.

Nolan, D. (1989b). Estimation on the sphere: a geometric approach, *Preprint,* Department of Statistics, University of California, Berkeley.

Nolan, D. (1989c). On min-max majority and deepest points, *Preprint,* Department of Statistics, University of California, Berkeley.

Nolan, D. & Marron, J.S. (1989). Uniform consistency of automatic and location-adaptive delta-sequence estimators, *Probab. Th. Rel. Fields* **80,** 619–632.

Nolan, D. & Pollard, D. (1987). U-processes: rates of convergence, *Ann. Probability* **15,** 780–799.

Nolan, D. & Pollard, D. (1988). Functional limit theorems for U-processes, *Ann. Probability* **16,** 1291–1298.

Olshen, R.A., Biden, E.N., Wyatt, M.P. & Sutherland, D.H. (1989). Gait analysis and the bootstrap, *Ann. Statist.* **17,** 1419–1440.

Ossiander, M. (1987). A central limit theorem under metric entropy with L_2 bracketing, *Ann. Probab.* **15,** 897–919.

Pakes, A. (1986). Patents as options: some estimates of the value of holding European patent stocks, *Econometrica* **54,** 755–785.

Pakes, A. & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators, *Econometrica* **57,** 1027–1057.

Pisier, G. (1986). Probabilistic methods in the geometry of Banach spaces, *Lecture Notes in Math.* **1206,** 167–241, Springer, Berlin.

Pisier, G. (1989). *The Volume of Convex Bodies and Banach Space Geometry,* Cambridge University Press, Cambridge.

Pollard, D. (1979). General chi-square goodness-of-fit tests with data dependent cells, *Z. Wahrsch. verw. Geb.* **50**, 317–331.

Pollard, D. (1980). The minimum distance method of testing, *Metrika* **27**, 43–70.

Pollard, D. (1982a). A central limit theorem for k-means clustering, *Ann. Probability* **10**, 919–926.

Pollard, D. (1982b). A central limit theorem for empirical processes, *J. Austral. Math. Soc.* **A33**, 235–248.

Pollard, D. (1985). New ways to prove central limit theorems, *Econometric Theory* **1**, 295–314.

Pollard, D. (1987). Rates of uniform almost-sure convergence for empirical processes indexed by unbounded classes of functions, *Preprint*.

Pollard, D. (1989). Asymptotics via empirical processes, *Statistical Science* **4**, 341–366.

Pollard, D. (1990). *Empirical Processes: Theory and Applications. CBMS/NSF Regional Conference Series in Applied Mathematics*; Lectures at the University of Iowa, 20–24 June 1988, SIAM/IMS, Philadelphia.

Pons, O. & Turckheim, E. de (1989). Méthode de von Mises, Hadamard—différentiabilité, et bootstrap dans un modèle non paramétrique sur un espace métrique, *C.R. Acad. Sci. Paris* **308**, 369–372.

Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity, *Ann. Statist.* **16**, 356–366.

Powell, J.L. (1984). Least absolute deviations estimation for the censored regression model, *J. Econometrics* **25**, 303–325.

Praestgaard, J. (1990). Bootstrap with general weights and multiplier central limit theorems, *Technical Report* **195**, Department of Statistics, University of Washington.

Praestgaard, J. (1991). General Weights Bootstrap of the Empirical Process, *Unpublished Ph.D. dissertation*, Department of Statistics, University of Washington.

Praestgaard, J. & Wellner, J. A. (1992). Exchangeably weighted bootstraps of the general empirical process, *Technical Report*, Department of Statistics, University of Washington, Seattle.

Pyke, R. (1984). Asymptotic results for empirical and partial sum processes: a review, *Canad. J. Statist.* **12**, 241–264.

Pyke, R. (1992). Probability in Mathematics and Statistics: A century's predictor of future directions, Jber.d.D.Math.-Verein., Jubiläumstagung 1990, 239–264.

Pyke, R. & Wilbour, D.C. (1988). New approaches for goodness-of-fit tests for multidimensional data, in *Statistical Theory and Data Analysis II*, 139–154, Elsevier Science Publishers (North Holland), Amsterdam.

Prakasa Rao, B.L.S. (1983). Estimation of a unimodal density, *Sankhya Ser. A* **31**, 23–36.

Reeds, J. (1976). On the definition of von Mises functionals, *Research Report* **S-44**, Department of Statistics, Harvard University.

Rhee, W.T. & Talagrand, M. (1988). Exact bounds for the stochastic upward matching problem, *Trans. Amer. Math. Soc.* **307**, 109–125.

Romano, J.P. (1988). A bootstrap revival of some nonparametric distance tests, *J. Amer. Stat. Assoc.* **83**, 698–708.

Romano, J.P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses, *Ann. Statist.* **17**, 141–159.

Rubin, D. (1981). The Bayesian bootstrap, *Ann. Statist.* **9**, 130–134.

Sauermann, W. (1989). Bootstrapping the maximum likelihood estimator in high-dimensional log-linear models, *Ann. Statist.* **17**, 1198–1216.

Scholz, F.W. (1980). Towards a unified definition of maximum likelihood, *Canad. J. Statist.* **8**, 193–203.

Sheehy, A. (1988). Maximal Kullback–Leibler divergence cluster analysis, *Technical Report*, Stanford University.

Sheehy, A. & Wellner, J.A. (1988). Uniformity in P of some limit theorems for empirical measures and processes, *Technical Report* **134**, Department of Statistics, University of Washington, Seattle.

Sheehy, A. & Wellner, J.A. (1992). Uniform Donsker classes of functions, *Ann. Probability* **20**, to appear.

Shor, P.W. & Yukich, J.E. (1991). Minimax grid matching and empirical measures, *Ann. Probability* **19**, 1338–1348.

Shorack, G.R. & Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

Stengle, G. & Yukich, J.E. (1989). Some new Vapnik–Chervonenkis classes, *Ann. Statist.* **17**, 1441–1446.

Turnbull, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data, *J. Amer. Statist. Assoc.* **69**, 169–173.

Van de Geer, S. (1988). Regression Analysis and Empirical Processes, *CWI Tract* **45**, Centre for Mathematics and Computer Science, Amsterdam.

Van de Geer, S. (1990). Estimating a regression function, *Ann. Statist.* **18**, 907–924.

Van der Vaart, A.W. (1988). Statistical estimation in large parameter spaces, *CWI Tract* **44**, Centrum voor Wiskunde en Informatica, Amsterdam.

Van der Vaart, A.W. (1991a). On differentiable functionals, *Ann. Statist.* **19**, 178–204.

Van der Vaart, A.W. (1991b). Efficiency and Hadamard differentiability, *Scand. J. Statist.* **18**, 63–75.

Van der Vaart, A.W. & Wellner, J.A. (1990). Prohorov and continuous mappings theorems in the Hoffmann–Jørgensen weak convergence theory with applications to convolution and asymptotic minimax theorems, *Technical Report* **157**, Department of Statistics, University of Washington, Seattle.

Vapnik, V. (1982). *Estimation of Dependences*, Springer-Verlag, New York.

Vapnik, V.N. & Chervonenkis, A.Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities, *Theor. Prob. Appl.* **16**, 264–280.

Vapnik, V. & Čhervonenkis, A.Ya. (1974). *The Theory of Pattern Recognition*, Nauka, Moscow.

Vardi, Y. (1985). Empirical distributions in selection bias models, *Ann. Statist.* **13**, 178–203.

Wang, M.-C., Jewell, N.P. & Tsai, W.-Y. (1986). Asymptotic properties of the product limit estimate under random truncation, *Ann. Statist.* **14**, 1597–1605.

Wellner, J.A. (1979). Permutation tests for directional data, *Ann. Statist.* **7**, 929–943.

Wellner, J.A. (1989). Discussion of: "Non- and semiparametric maximum likelihood estimators and the von Mises method (Part I)" by Richard Gill, *Scand. J. Statist.* **16**, 124–127.

White, H. (1989). Learning in artificial neural networks: a statistical perspective, *Neural Computation* **1**, 425–464.

White, H. (1990). Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings, *Neural Networks* **3**, 535–549.

Woodroofe, M. (1985). Estimating a distribution function with truncated data, *Ann. Statist.* **13**, 163–177. Correction: *Ann. Statist.* (1987) **15**, 883.

Yukich, J.E. (1989). Optimal matching and empirical measures, *Proc. Amer. Math. Soc.* **102**, 1051–1059.

Yukich, J.E. (1989). A note on limit theorems for perturbed empirical processes, *Stoch. Processes and Their Applications* **33**, 163–174.

Résumé

Nous résumons les applications récentes de la théorie des processus empiriques, en particulier les processus empiriques indexés par des ensembles et des fonctions. Nous commende la théorie des processus empiriques. Ensuite, nous décrivons les applications de cette théorie à l'estimation (données censurées et tronquées, échantillonage biasé, régression et estimation de la densité, méthodes de distance minimale), aux test d'hypothèses (tests classiques d'ajustement, tests de distance minimale, tests de permutations et tests bootstrap, alternatives locales et puissance), reconnaissance des formes, classification, analyse discrimminante, bootstrap des mesures empiriques, et la méthode delta. Nous présentons un nouveau théorème sur le comportement asymptotique des processus empiriques sous des alternatives locales.