The Annals of Probability
1992, Vol. 20, No. 4, 1968–1982

FRÉCHET DIFFERENTIABILITY, p-VARIATION AND UNIFORM DONSKER CLASSES[1]

BY R. M. DUDLEY

Massachusetts Institute of Technology

Differentiability of functionals of the empirical distribution function is extended. The supremum norm is replaced by p-variation seminorms, which are the pth roots of suprema of sums of pth powers of absolute increments of a function over nonoverlapping intervals. Fréchet derivatives often exist for such norms when they do not for the supremum norm. For $1 < q < 2$, classes of functions uniformly bounded in q-variation are universal and uniform Donsker classes: The central limit theorem for empirical measures holds with respect to uniform convergence over such a class, also uniformly over all probability laws on the line. The integral $\int F\,dG$ was defined by L. C. Young if F and G are of bounded p- and q-variation respectively, where $p^{-1} + q^{-1} > 1$. Thus the normalized empirical distribution function $n^{1/2}(F_n - F)$ is with high probability in sets of uniformly bounded p-variation for any $p > 2$, uniformly in n.

1. Introduction. Classically, a statistical functional is defined on a space of distribution functions F on the real line with the supremum norm. The values may be real or themselves functions such as the quantile function F^{-1}. Nonlinear functionals are studied via their derivatives. This paper and a related one [Dudley (1991a)] will show that the differentiability properties originally proved by Reeds (1976), also treated in Fernholz (1983), can be improved substantially.

Differentiability of functionals is defined in general as follows. Let $(X, \|\cdot\|)$ and $(Y, |\cdot|)$ be two Banach spaces. Let U be an open set in X and u a point of U. A function T defined on U with values in Y is said to be *Fréchet differentiable* at u if there is a bounded linear operator A from X into Y such that

$$|T(x) - T(u) - A(x-u)| = o(\|x-u\|)$$

as $x \to u$. Equivalently, for every bounded set B in X,

$$\sup\{|T(u+tv) - T(u) - tA(v)|: v \in B\}/t \to 0 \quad \text{as } t \to 0,$$

where t is a real variable. If $\mathscr{C}$ is a collection of subsets of X, T will be called *$\mathscr{C}$-differentiable* at u if "every bounded set" is replaced by "each set in $\mathscr{C}$." Then if $\mathscr{C}$ is the class of all compact sets for a topology, T is said to be *compactly* or *Hadamard differentiable* at u for the given topology, which in past work has usually been, but need not be, that of a norm.

Received February 1991; revised October 1991.
[1]Research partially supported by an NSF grant.
AMS 1980 *subject classifications*. Primary 60F17, 62G30; secondary 26A42, 26A45.
Key words and phrases. Wilcoxon statistics, Riemann–Stieltjes integral, L. C. Young integral.

Reeds (1976) and Fernholz (1983) showed that once the domain and range of functionals are suitably specified, in several cases functionals of statistical interest are compactly but not Fréchet differentiable at some points. To my knowledge, most if not all such examples were for the sup norm $\|\cdot\|_\infty$ in $\mathbb{R}$, where $\|F\|_\infty := \sup_t |F(t)|$. Although Fréchet differentiability holds in some other useful cases [e.g., Fernholz (1983), Proposition 6.1.3, page 72, Corollary 6.1.4, page 75, and Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 54], many statisticians have been convinced that in general it is too strong [e.g., Huber (1981), page 37, Gill (1989), page 101, and Serfling (1980), page 220]. But Serfling (1980), page 218, mentions the "choice of norm": A larger norm increases the chances for Fréchet differentiability, and will work well if probability limit theorems still hold for the larger norm.

This paper expands on the "choice of norm." It will be shown that in p-variation norms for suitable values of p, defined in Sections 2 and 3, uniform central limit theorems still hold. In this paper, the result will be applied to the bilinear functional $(F, G) \mapsto \int F\,dG$ (Section 4). Another paper [Dudley (1991a)] will treat the inverse operator $F \mapsto F^{-1}$ and composition $(F, G) \mapsto F \circ G$.

In a nonseparable normed space, specifically $D[0,1]$ with $\|\cdot\|_\infty$, compact sets are very small compared to open, bounded sets. If F is a continuous distribution function and K is a compact subset of $D[0,1]$ for the sup norm, the probability that the empirical distribution function $F_n \in K$, or $n^{1/2}(F_n - F) \in K$, is 0! So there are technical problems in applying norm-compact differentiability. The problems have been treated by modifying empirical distribution functions to be continuous [e.g., Fernholz (1983), pages 32–42]. Or, the definition of compact differentiation can (also for very general sample spaces and norms) be modified to hold "tangentially to a subspace" [Gill (1989), page 102, and Pons and Turckheim (1989)]. Compact differentiation can also be applied to almost surely convergent realizations. At any rate, no such modification is needed when we have $\mathscr{C}$-differentiability for a class $\mathscr{C}$ of sets of functions such that $n^{1/2}(F_n - F)$ satisfies a tightness condition for sets in $\mathscr{C}$, as will be shown in Proposition 3.7 and Corollary 3.8 for classes of bounded p-variation for $p > 2$.

For a probability measure P and observations $X_1, X_2, \ldots,$ i.i.d. with law P, we form the empirical measures $P_n := n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n})$ and the empirical process $\nu_n := n^{1/2}(P_n - P)$. If $\mathscr{F}$ is a class of measurable functions and ν is a signed measure, let $\|\nu\|_{\mathscr{F}} := \sup\{|\int f\,d\nu|: f \in \mathscr{F}\}$. Here a class of measurable sets can be identified with the class of indicator functions of the sets. The "sup norm" or "Kolmogorov norm" of the empirical process is the supremum of its absolute value over the family $\mathscr{H}$ of all sets (half-lines) $]-\infty, x]$, so $\|\cdot\|_\infty \equiv \|\cdot\|_{\mathscr{H}}$. $\mathscr{H}$ is taken into itself by all nondecreasing transformations of $\mathbb{R}$ and is taken one to one and onto itself by all strictly increasing, continuous transformations of $\mathbb{R}$ onto itself. It is known that a central limit theorem holds in the sup norm for all probability laws P on $\mathbb{R}$, at a rate which is uniform in P. On extensions to general sample spaces and classes $\mathscr{F}$, see Dudley (1987) for the "universal Donsker" property and Giné and Zinn (1991) for the uniformity in P.

The sup norm is defined by a relatively small and quite tractable class $\mathscr{H}$ of sets, while it also encompasses larger classes of functions: Let BV1 be the class of all functions f from $\mathbb{R}$ into itself of total variation at most 1, with $f(x) \to 0$ as $x \to +\infty$. Let DE1 be the class of nonincreasing functions in BV1. Then

$$\|\cdot\|_\infty = \|\cdot\|_{\mathrm{DE1}} \le \|\cdot\|_{\mathrm{BV1}} \le 2\|\cdot\|_{\mathrm{DE1}}$$

[e.g., Dudley (1987), proof of Theorem 2.1i], although BV1 and DE1 are in many ways much larger than $\mathscr{H}$. So it is not surprising that the sup norm has long been considered as the main norm for empirical processes and differentiability on $\mathbb{R}$. But p-variation norms, besides their other advantages, are also preserved by any continuous increasing function from $\mathbb{R}$ onto itself.

I suggested [Dudley (1989b, 1990)] that statistical functionals might be Fréchet differentiable for norms $\|\cdot\|_{\mathscr{F}}$ for universal Donsker classes $\mathscr{F}$ other than $\mathscr{H}$. It will be seen that classes of functions of uniformly bounded q-variation for $1 < q < 2$ are universal (and uniform) Donsker classes $\mathscr{F}$ on $\mathbb{R}$. Via the duality theory in Section 3, this is equivalent to $n^{1/2}(F_n - F)$ being with high probability in sets of uniformly bounded p-variation, $p > 2$.

Compact differentiability for $\|\cdot\|_\infty$ implies Fréchet differentiability for $\|\cdot\|_{\mathscr{F}}$ directly if all bounded sets for $\|\cdot\|_{\mathscr{F}}$ are compact for $\|\cdot\|_\infty$. But matters are not quite so easy, since no such $\mathscr{F}$ is a universal Donsker class.

PROPOSITION 1.1. *There is no universal Donsker class $\mathscr{F}$ of Borel-measurable functions on $\mathbb{R}$ such that all bounded sets of finite signed measures for $\|\cdot\|_{\mathscr{F}}$ are compact or even separable for $\|\cdot\|_\infty$.*

PROOF. Any universal Donsker class $\mathscr{F}$ is uniformly bounded up to additive constants [Dudley (1987), Proposition 1.1]. Thus the set B of all differences $\delta_x - \delta_y$ for x and y in $\mathbb{R}$ is bounded for $\|\cdot\|_{\mathscr{F}}$. But if $x \ne z$, $\|\delta_x - \delta_y - (\delta_z - \delta_y)\|_{\mathscr{H}} = 1$, so B is not separable for the supremum norm $\|\cdot\|_\infty = \|\cdot\|_{\mathscr{H}}$. □

This paper does not treat inverse or implicit function theorems and M-estimators, on which I hope to complete a separate paper [Dudley (1991b)].

The present paper is organized as follows: Section 2 reviews p-variation and proves a uniform central limit theorem (Donsker property) for classes of functions of bounded p-variation, $p < 2$. Section 3 treats Young's duality theory of r-variation spaces W_r, W_s via $(F, G) \mapsto \int F\,dG$, $F \in W_r$, $G \in W_s$, $1/r + 1/s > 1$. Section 4 treats differentiability of the functional $(F, G) \mapsto \int F\,dG$ and compares it with the results of Gill (1989).

2. Functions of bounded p-variation and Donsker classes. A function f from an interval $J \subset \mathbb{R}$ into $\mathbb{R}$ has *p-variation* defined by

$$v(f, p) := v(f, p, J) := \sup\left\{ \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|^p : x_0 < x_1 < \cdots < x_n \in J, \right.$$
$$\left. x_0 \in J,\ n = 1, 2, \dots \right\}.$$

For $p = 1$, this is the usual total variation. Apparently the notion of p-variation was first defined by Wiener (1924), who treated mainly the case $p = 2$, "quadratic variation." For $p \neq 2$, the first major work, including the duality theory (Section 3), was done in the late 1930s by L. C. Young, partly with E. R. Love; see Young (1936) and Love and Young (1937). More recently, Bruneau (1974) treats p-variation, mainly in a different direction ("fine" variation). There are recent applications in probability theory [e.g., Bertoin (1989), Xu (1986, 1988) and Pisier and Xu (1987, 1988)]. Apparently p-variation for empirical processes was not previously addressed explicitly, but see the discussion after Theorem 2.2 below. The following will be proved.

THEOREM 2.1. *For any (bounded or unbounded) interval* $J \subset \mathbb{R}$, *any* p *with* $0 < p < 2$ *and any* $M < \infty$, $\mathscr{F}_{p,M} := \{f\colon J \mapsto \mathbb{R},\ v(f,p) \le M\}$ *is a universal Donsker class.*

PROOF. Recall the notion of Kolchinskii–Pollard entropy $\log D^{(2)}$ for the L^2 norm over finite sets, where $\int_0^1 (\log D^{(2)}(\varepsilon, \mathscr{F}))^{1/2}\, d\varepsilon < \infty$ and some measurability conditions imply that $\mathscr{F}$ is a universal Donsker class by a theorem of Pollard (1982), stated with more general measurability conditions in Dudley (1987), Theorem 2.1h.

We can assume without loss of generality that $M = 1$.

For any $f \in \mathscr{F} := \mathscr{F}_{p,1}$, we have $\operatorname{diam} f := \sup f - \inf f$ satisfying $(\operatorname{diam} f)^p \le 1$, so $\operatorname{diam} f \le 1$. So $\mathscr{F}$ is uniformly bounded up to additive constants and in proving the universal Donsker property we can replace the functions f in $\mathscr{F}$ by the functions $f - \inf f$, so that we can assume $0 \le f \le 1$, $f \in \mathscr{F}$ [cf. Dudley (1987)].

We can assume $p \ge 1$ and $J = \mathbb{R}$. Given $f \in \mathscr{F}$, let $h(x) := v(f, p,]-\infty, x])$, the total p-variation of f up to x. Then h is a nondecreasing function. For any $x < y$, $|f(y) - f(x)|^p \le h(y) - h(x)$, so $|f(y) - f(x)| \le (h(y) - h(x))^{1/p}$. Thus f is a function of h, $f(x) \equiv g(h(x))$, for a function g which satisfies a Hölder condition $|g(u) - g(v)| \le |u - v|^{1/p}$ for all u and v in the range of h, as was known [Love and Young (1937), page 244, and Bruneau (1974), page 3]. Since $0 < 1/p \le 1$, $e(u, v) := |u - v|^{1/p}$ is a metric, and the ranges of g and h are included in $[0, 1]$, so we can assume g is defined and satisfies the same Hölder condition on all of $[0, 1]$ into $[0, 1]$ by the Kirszbraun–McShane extension theorem [e.g., Dudley (1989a), Theorem 6.1.1].

Let G be the set of all functions from $[0, 1]$ into itself satisfying the given Hölder condition. Then there are constants C_1 and C_2 such that for $0 < \varepsilon \le 1$, there is a set $G_\varepsilon \subset G$, dense within ε for the supremum norm and containing at most $C_1 \exp(C_2 \varepsilon^{-p})$ functions [Kolmogorov and Tihomirov (1959), Theorem XIII].

Let H be the set of nondecreasing functions from $\mathbb{R}$ into $[0, 1]$. Then H is included in the set of sequential pointwise limits of convex combinations of members of the set H_1 of all indicators of half-lines, $1_{[a,\infty[}$ [e.g., Dudley (1987), proof of Theorem 2.1(a), and since an open half-line is a countable union of closed half-lines). For any probability measure Q on $\mathbb{R}$ (such as one with finite support), for $0 < \varepsilon \le 1$ there is a set of at most $2/\varepsilon^2$ members of H_1, dense within ε^2 in H_1 in the $L^1(Q)$ norm and so dense within ε in H_1 in the $L^2(Q)$ norm. Thus, for any $t > 1$, there are constants C_3 and C_4 such that for $0 < \varepsilon \le 1$ there is a set $H_\varepsilon \subset H$, dense in H within ε^p in the $L^2(Q)$ norm and containing at most $C_3 \exp(C_4 \varepsilon^{-pt})$ functions [Dudley (1987), Theorem 5.1]. Choose t so that $pt < 2$.

Now take any $f \in \mathscr{F}$. Then $f \equiv g \circ h$ for some $g \in G$ and $h \in H$. Given $0 < \varepsilon \le 1$, take $\gamma \in G_\varepsilon$ with $\sup|g - \gamma| \le \varepsilon$ and $\theta \in H_\varepsilon$ with $\|h - \theta\|_2 \le \varepsilon^p$ for Q. Then

$$\| f - (\gamma \circ \theta) \|_2 \le \|g \circ h - \gamma \circ h\|_2 + \|\gamma \circ h - \gamma \circ \theta\|_2$$
$$\le \varepsilon + \|\gamma \circ h - \gamma \circ \theta\|_2 .$$

Next,

$$\int (\gamma \circ h - \gamma \circ \theta)^2 \, dQ(x) \le \int |(h - \theta)(x)|^{2/p} \, dQ(x)$$
$$\le \left(\int |(h - \theta)(x)|^2 \, dQ(x) \right)^{1/p} \le \varepsilon^{2p/p} = \varepsilon^2 .$$

So $\| f - (\gamma \circ \theta)\|_2 \le 2\varepsilon$. So we have found a set $\mathscr{F}_\varepsilon$ dense within 2ε in $\mathscr{F}$ for the $L^2(Q)$ norm, where for some constants C and D, the number of functions in $\mathscr{F}_\varepsilon$ is at most $C \exp(D\varepsilon^{-pt})$. This implies Pollard's entropy condition. So it will be enough to show that the image admissible Suslin measurability condition holds.

Any function f of bounded p-variation is in the set $E(\mathbb{R})$ of real functions on $\mathbb{R}$ having right limits on $[-\infty, \infty[$ and left limits on $]-\infty, \infty]$. Let $D(\mathbb{R})$ be the set of all right-continuous functions in $E(\mathbb{R})$. Let $c_0(\mathbb{R})$ be the set of all functions h from $\mathbb{R}$ into $\mathbb{R}$ such that for each $\varepsilon > 0$, $\{x: |h(x)| > \varepsilon\}$ is finite. Then h is 0 except on a countable set.

Given $f \in E(\mathbb{R})$, let $g(x) := f(x+) := \lim\{f(y): y \downarrow x\}$ for all x. Then since $f \in E(\mathbb{R})$, we have $g \in D(\mathbb{R})$, and f is continuous except at most on a countable set C, so $g(x) = f(x)$ for $x \notin C$. Let $h := f - g$. For any $\varepsilon > 0$, f can only have finitely many jumps of heights larger than ε, so $h \in c_0(\mathbb{R})$.

Now $h \in c_0(\mathbb{R})$ if and only if there is a sequence $\{x_i\}$ of distinct real numbers and a sequence $y_i \to 0$ such that $h(x) = \sum_i y_i 1\{x = x_i\}$. The set of such sequences $\{x_i\}$ is a countable intersection of open sets $\{x_i \ne x_j\}$, $i \ne j$, in the product topology, so it is Polish [e.g., Dudley (1989a), Theorem 2.5.4]. The set of sequences $\{y_i\}$, $y_i \to 0$, is a Banach space with supremum norm. Then evaluation of h is jointly measurable, so $c_0(\mathbb{R})$ is image admissible Suslin.

Let $\mathscr{S}(T)$ be the smallest σ-algebra on a set of real functions with domain including T such that evaluation at each point of T is measurable. For $f \in D[0,1]$, let $f_n(x) := f(k/n)$ for $(k-1)/n \le x < k/n$, $k = 1, \dots, n$, $f_n(1) := f(1)$. Then $f_n(x) \to f(x)$ for all x, and $(x, f) \mapsto f_n(x)$ is jointly measurable for the Borel σ-algebra in $x \in \mathbb{R}$ and the σ-algebra $\mathscr{S}(T)$, where T is the set of rational numbers. Since the given σ-algebras are Borel σ-algebras of Polish spaces, the same holds for $(x, f) \mapsto f(x)$, so $D(\mathbb{R})$ is image admissible

Suslin, and so by addition $E(\mathbb{R})$ is image admissible Suslin. In this class, the set of functions of p-variation bounded by 1 is the image of a Borel set, since we can consider the p-variation on finite sets of rationals or points x_i for h in $c_0(\mathbb{R})$. □

Now it will be shown that $\mathscr{F}_{p,M}$ is not only a universal but a *uniform* Donsker class, in the sense that the central limit theorem for empirical measures for uniform convergence over $\mathscr{F}_{p,M}$ also holds uniformly in P. Giné and Zinn (1991) gave a general, natural definition of the uniform Donsker property in terms of dual-bounded-Lipschitz metrics, and found a striking Gaussian characterization. Here, a different property will be defined, closer to the traditional uniformity over distribution functions on $\mathbb{R}$.

DEFINITION. Let $\mathscr{F}$ be a class of real-valued measurable functions on a measurable space $(X, \mathscr{A})$. Then $\mathscr{F}$ is a *dominated uniform Donsker class* if there is a law λ on $(X, \mathscr{A})$ for which $\mathscr{F}$ is a Donsker class, and for every law P on $(X, \mathscr{A})$ there is a measurable function f_P from X into itself such that the image measure $\lambda \circ f_P^{-1} = P$, and such that the class $\mathscr{F} \circ f_p$ of all compositions $f \circ f_p$, $f \in \mathscr{F}$, is included in $\mathscr{F}$.

When the condition just defined holds, other definitions of uniform Donsker class could be verified, but the details will just be sketched here. If λ_n are empirical measures for λ, then $\lambda_n \circ f_P^{-1}$ have all the properties of the empirical measures P_n, so $n^{1/2}(P_n - P)$ can be written as $n^{1/2}(\lambda_n - \lambda) \circ f_P^{-1}$. For a class $\mathscr{F}$ of functions, we then have by the image measure theorem:

$$\begin{aligned}\|n^{1/2}(P_n - P)\|_{\mathscr{F}} &= n^{1/2} \sup\left\{\left|\int f\, d(\lambda_n - \lambda) \circ f_P^{-1}\right| : f \in \mathscr{F}\right\} \\ &= n^{1/2} \sup\left\{\left|\int f \circ f_P \, d(\lambda_n - \lambda)\right| : f \in \mathscr{F}\right\} \\ &= \|n^{1/2}(\lambda_n - \lambda)\|_{\mathscr{F} \circ f_P}.\end{aligned}$$

Moreover, for the limiting Gaussian processes G_P and G_λ, for any $\mathscr{F} \subset \mathscr{L}^2(P)$, $f \mapsto G_\lambda(f \circ f_P)$ has the same distributions as G_P on $\mathscr{F}$ since it is Gaussian with mean 0 and the same variances and covariances. It then follows, for example by existence of almost surely convergent realizations [Dudley (1985)] that the convergence in limit theorems for P_n for any P is at least as fast as for λ_n.

For the classes in Theorem 2.1 we have the following theorem.

THEOREM 2.2. *Each $\mathscr{F}_{p,M}$ on $\mathbb{R}$ for $0 < p < 2$ and $M < \infty$ is a dominated uniform Donsker class.*

PROOF. Let λ be Lebesgue measure on $[0, 1]$. Let P be any probability measure on $\mathbb{R}$ with distribution function F. For $0 < y < 1$, let $F^{-1}(y) :=$

$\inf\{x\colon F(x) \ge y\}$. Take f_P as the function F^{-1}. Then the image measure $\lambda \circ (F^{-1})^{-1} = P$ as desired. If $f \in \mathscr{F}_{p,M}$, then for any nondecreasing function G, such as $G = F^{-1}$, we have $f \circ G \in \mathscr{F}_{p,M}$ since for any $y_1 < \cdots < y_n$, letting $x_i := G(y_i)$ gives

$$\sum_{i=2}^{n} |f(G(y_i)) - f(G(y_{i-1}))|^p = \sum_{i=2}^{n} |f(x_i) - f(x_{i-1})|^p,$$

which (with Theorem 2.1 for λ) completes the proof. □

Gilles Pisier, in a letter dated September 3, 1991, has kindly pointed out the following: The work of Pisier and Xu (1987) and Xu (1986, 1988) easily implies at least a fact close to Theorem 2.2, which I will call Theorem 2.2′, where convergence in distribution of ν_n is replaced by boundedness in L^2: $\sup_P \sup_n E\|\nu_n\|^2_{\mathscr{F}} < \infty$ for $\mathscr{F} = \mathscr{F}_{p,M}$.

3. Duality inequalities for r-variation spaces. For $1 \le r < \infty$ and $-\infty \le a < b \le +\infty$, let $W_r := W_r[a,b]$ be the class of all real-valued functions f on $[a,b]$ with finite r-variation $v(f,r) = v(f,r,[a,b]) < \infty$. For any $f \in W_r$, we have the seminorm $\|f\|_{(r)} := v(f,r)^{1/r}$, which is 0 only for constants. If f has finite r-variation on an open interval $]a,b[$, then it has a limit as $x \to a$ or b respectively, even if $a = -\infty$ or $b = +\infty$, in which case $f(-\infty)$ or $f(+\infty)$ is defined as the corresponding limit.

For $f \in W_r[a,b]$, let $\|f\|_{[r]} := \|f\|_\infty + \|f\|_{(r)}$, where $\|f\|_\infty := \sup_x |f(x)|$. Then $\|\cdot\|_{[r]}$ is a norm on W_r. It is not hard to show that W_r is complete for $\|\cdot\|_{[r]}$ and so W_r is a Banach space.

To define integrals $\int F\,dG$, first consider the Riemann–Stieltjes integral defined as follows: Given an interval $[a,b]$, a *partition* will be a finite sequence $x_0 = a < x_1 < \cdots < x_n = b$. A *Riemann–Stieltjes sum* for F, G and the given partition will be any sum $\sum_{i=1}^{n} F(t_i)(G(x_i) - G(x_{i-1}))$, where $x_{i-1} \le t_i \le x_i$ for $i = 1,\dots,n$. Then the *Riemann–Stieltjes integral* $\int_a^b F\,dG$ exists and equals c if for every $\varepsilon > 0$ there exists a partition $\pi = \{x_0, x_1, \dots, x_n\}$ as above such that for all partitions τ including π and all Riemann–Stieltjes sums S based on τ, $|S - c| < \varepsilon$. [This is one of two definitions most often given in the literature, and is sometimes called the Moore–Pollard definition. The other definition requires convergence of sums S to c as the *mesh* $\max_i (x_i - x_{i-1}) \to 0$.] When the integral exists, we say $F \in \mathscr{R}(G)$. Hildebrandt (1938) is a survey on Riemann–Stieltjes integration, with references to earlier sources for most results.

Riemann–Stieltjes integrals have been considered mainly when one of F and G is continuous and the other is of bounded variation. The theory of p-variation provides a class of cases, to be given below, where neither F nor G is of bounded variation.

The integral $\int F\,dG$ will not be defined as a Riemann–Stieltjes integral, even if both F and G are of bounded variation, if they are discontinuous on the same side of the same point [Hildebrandt (1938), 4.13, "σ" case; for one example, see Rudin (1976), page 138, Example 3]. So the integral has to be defined otherwise. For a function h and point x, let $h(x-) = \lim_{y \uparrow x} h(y)$ and $h(x+) = \lim_{y \downarrow x} h(y)$. If $h \in W_r[a, b]$ for any $r < \infty$, then $h(x+)$ exists for $a \le x < b$ and $h(x-)$ exists for $a < x \le b$. The following is known [Hildebrandt (1938), Theorem 5.32].

LEMMA 3.1. *The Riemann–Stieltjes integral $\int_a^b F\,dG$ exists if G is of bounded variation and right continuous $[G(x+) = G(x),\ a \le x < b]$ while F is left continuous $[F(x) = F(x-),\ a < x \le b]$ and has right limits $F(x+)$ for $a \le x < b$.*

The *Young's* or (Y_1) *integral*

$$\begin{aligned}(Y_1)\int_a^b F\,dG := \int_a^b & F(x+)\,dG(x-) \\ &+ \sum_{a<x<b} (F(x) - F(x+))(G(x+) - G(x-)) \\ &+ (F(a) - F(a+))(G(a+) - G(a)) \\ &+ F(b)(G(b) - G(b-))\end{aligned}$$

will be so defined if $\int_a^b F(x+)\,dG(x-)$ exists as a Riemann–Stieltjes integral, and in the sum, the summands are 0 except for at most countably many values of x and the series is absolutely convergent. Here in the integral, $G(a-)$ is replaced by $G(a)$ [and $F(b+)$ by $F(b)$, which does not matter since $x \mapsto G(x-)$ is left continuous at b]. Note that if F is continuous from the right, $F(x) \equiv F(x+)$, the (Y_1) integral becomes

$$\int_a^b F(x)\,dG(x-) + F(b)(G(b) - G(b-)).$$

W. H. Young (1914) contributed to defining extended Riemann–Stieltjes integrals. L. C. Young (1936), pages 263–265, shows that $(Y_1)\int_a^b F\,dG$ exists if $F \in W_p[a, b]$ and $G \in W_q[a, b]$, where $p^{-1} + q^{-1} > 1$. [Hildebrandt (1938) distinguishes different integrals by putting symbols such as N, Y and/or σ before the integral sign.

An alternate definition of integral is obtained by applying the definition of the (Y_1) integral to the interval in the reverse order, taking $-(Y_1)\int_{-b}^{-a} F(-x)\,dG(-x)$. The resulting integral, which will also be called a Young integral, is here defined as

$$\begin{aligned}(Y_2)\int_a^b F\,dG := \int_a^b & F(x-)\,dG(x+) \\ &+ \sum_{a<x<b} (F(x) - F(x-))(G(x+) - G(x-)) \\ &+ F(a)(G(a+) - G(a)) \\ &+ (F(b) - F(b-))(G(b) - G(b-)),\end{aligned}$$

which will, as with the (Y_1) integral, be said to exist if the integral on the right exists as a Riemann–Stieltjes integral and the sum is absolutely convergent, and here $G(b+)$ is replaced by $G(b)$ in the integral [and $F(a-)$ by $F(a)$, which does not matter since $x \mapsto G(x+)$ is right continuous at a]. The inversion $x \mapsto -x$ shows that the (Y_2) integral likewise exists for $F \in W_p$ and $G \in W_q$ with $p^{-1} + q^{-1} > 1$. The two integrals actually agree when defined, as follows.

THEOREM 3.2. *Suppose that F and G are two bounded real functions on $[a, b]$ having right limits on $[a, b[$ and left limits on $]a, b]$. Then if both of the integrals $(Y_1)\int_a^b F\,dG$ and $(Y_2)\int_a^b F\,dG$ exist, they are equal.*

PROOF. Since the Riemann–Stieltjes integral $I_1 := \int_a^b F(x+)\,dG(x-)$ exists, for any $\varepsilon > 0$ there is a partition τ given by $a = x_0 < x_1 < \cdots < x_n = b$ of $[a, b]$ such that any Riemann–Stieltjes sum S for I_1 based on a partition which is a refinement of τ differs from the integral by at most ε. Here $S = \sum_{i=1}^n F(y_i+)[G(x_i-) - G(x_{i-1}-)]$ for any y_i in the interval $[x_{i-1}, x_i]$, replacing $a-$ by a and $b+$ by b. The same holds for $I_2 := \int_a^b F(x-)\,dG(x+)$, so taking a common refinement of the two partitions, we can assume τ is such that it holds for both Riemann–Stieltjes integrals. Also, τ can be chosen to contain enough of the points of discontinuity of F and G such that the sums in the definitions of the (Y_1) and (Y_2) integrals over all other points are at most ε each.

For the given partition τ, consider refinements formed by adjoining points u_i at which G is continuous with $x_{i-1} < u_i < x_i$ for $i = 1, \ldots, n$. [Since $G(x-)$ and $G(x+)$ exist for $a < x < b$ by assumption, G is continuous except at most on a countable set.] We can then form Riemann–Stieltjes sums for I_1 by evaluating F at $x_{i-1}+$ for $[x_{i-1}, u_i]$ and at x_i- (a limit of points $x+$ as $x \uparrow x_i$) for $[u_i, x_i]$. In forming sums for I_2, we can evaluate F at the same places, interchanging the reasons for the lower and higher of the two intervals. Then $I_1 - I_2$ differs by at most 2ε from

$$\sum_{i=1}^{n} F(x_i-)[G(x_i-) - G(x_i+)] + F(x_{i-1}+)[G(x_{i-1}+) - G(x_{i-1}-)]$$

$$= R(a, b) + \sum_{i=1}^{n-1} [G(x_i+) - G(x_i-)][F(x_i+) - F(x_i-)],$$

where

$$R(a, b) := F(a+)[G(a+) - G(a)] - F(b-)[G(b) - G(b-)].$$

So if we subtract the (Y_2) from the (Y_1) integral, all the terms in the approximating sums cancel so the integrals differ at most by 4ε. Let $\varepsilon \downarrow 0$ to complete the proof. □

When G has bounded variation, Young's integrals are equal to Lebesgue–Stieltjes integrals, as has been known at least under some conditions [e.g., Hildebrandt (1938), page 275; Young (1936), page 266].

LEMMA 3.3. *Let F be a function on $[a, b]$ which has left and right limits at all points. Let ν be a finite signed measure on $]a, b]$ with $G(x) := \nu([a, x])$, $a \le x \le b$. Then* $\int_a^b F\,d\nu = (Y_2)\int_a^b F\,dG$.

PROOF. Since G is right continuous, $G(x+) \equiv G(x)$, and $G(x) - G(x-) = \nu(\{x\})$, which is 0 by assumption if $x = a$. We can write $G = G_c + G_{\text{at}}$, where G_c is continuous and G_{at} is the distribution function of the purely atomic part ν_{at} of ν, $G_{\text{at}}(x) \equiv \nu_{\text{at}}(]a, x]) \equiv \sum_{a<y\le x}\nu(\{y\})$. Then the sum in the definition of (Y_2) integral (with the $x = a, b$ terms) is $\int_a^b F(x) - F(x-)\,d\nu_{\text{at}}(x)$.

The next fact follows directly from Billingsley (1968), page 110, Lemma 1, interchanging the sides on which intervals are closed and functions are continuous.

LEMMA 3.4. *A function F from a closed interval $[a, b]$ into $\mathbb{R}$ is left continuous with right limits if and only if F is a uniform limit of step functions which are finite sums of the form $\sum_i c_i 1_{]a_i, b_i]}$ for some real c_i and left open, right closed intervals $]a_i, b_i]$.*

Each step function as in Lemma 3.4 is left continuous with right limits and so in $\mathscr{R}(G)$ with $\int 1_{]c,d]}\,dG \equiv G(d) - G(c)$, $a \le c < d \le b$. Also, clearly, a uniform limit of functions in $\mathscr{R}(G)$ is in $\mathscr{R}(G)$ and the integrals converge [e.g., Hildebrandt (1938), 5.41]. So under the conditions of Lemma 3.3, since $x \mapsto F(x-)$ is left continuous with right limits, we have $\int_a^b F(x-)\,d\nu(x) = (Y_2)\int_a^b F(x-)\,dG(x)$. For any $\varepsilon > 0$, $|F(x) - F(x-)| > \varepsilon$ for at most finitely many values of x, so $F(x) - F(x-)$ is a uniform limit of functions with finite support, and its integrals for dG and $d\nu$ equal those for dG_{at} and $d\nu_{\text{at}}$ respectively. Lemma 3.3 then follows. □

If $G(a-)$ is defined, let

$$(Y_2)\int_{a-}^{b} F\,dG := (Y_2)\int_a^b F\,dG + F(a)(G(a) - G(a-)).$$

For this integral, Lemma 3.3 extends to the case where ν has an atom at a, with $G(a-) = 0$ and $G(a) = \nu(\{a\})$.

THEOREM 3.5 (L. C. Young). *Suppose on an interval $[a, b]$ we have $f \in W_p$ and $g \in W_q$, where $s := 1/p + 1/q > 1$. Then $(Y_i)\int_a^b f\,dg$ is well defined for $i = 1, 2$, and for any ξ in $[a, b]$,*

$$\left|(Y_i)\int_a^b (f(x) - f(\xi))\,dg(x)\right| \le (1 + \zeta(s))\|f\|_{(p)}\|g\|_{(q)}, \qquad i = 1, 2,$$

where ζ is the Riemann zeta function $\zeta(s) := \sum_{n\ge1} n^{-s}$. Also,

$$\left|(Y_i)\int_a^b f(x)\,dg(x)\right| \le (1 + \zeta(s))\|f\|_{[p]}\|g\|_{(q)}, \qquad i = 1, 2.$$

PROOF. The first inequality is given in Young (1936), pages 264–266, for the (Y_1) integral, and follows for the (Y_2) integral by the transformation $x \mapsto -x$. The second inequality then holds, since for any ξ,

$$\left|\int_a^b f(\xi)\, dg(x)\right| \le |f(\xi)|(\sup - \inf)(g) \le \|f\|_\infty \|g\|_{(q)}. \qquad \square$$

As $s \downarrow 1$, the constant $1 + \zeta(s)$ goes to $+\infty$. Young (1936), Section 7, shows that this is inevitable: For $s = 1$, $1 + \zeta(1)$ cannot be replaced in Theorem 3.5 by any finite constant. But there is a converse inequality, as follows. For $\delta > 0$, let $v(f, r, [a, b], \delta)$ be the supremum of all sums $\sum_{i=1}^n |f(x_i) - f(x_{i-1})|^r$, where $a \le x_0 < x_1 < \cdots < x_n \le b$, and where $x_i - x_{i-1} \le \delta$ for all $i = 1, \dots, n$. For $1 < r < \infty$, let $C_r^* := C_r^*[a, b]$ be the set of all functions f in W_r such that $v(f, r, [a, b], \delta) \to 0$ as $\delta \downarrow 0$. Clearly, each function in C_r^* is continuous. [Young (1936), page 261, defines the class $W_r^*[a, b]$ of functions in W_r such that $\lim_{\delta \downarrow 0} v(f, r, [a, b], \delta)$ equals the r-variation coming from the jumps of f. Then C_r^* is the set of continuous functions in W_r^*.] A dual norm will be defined with respect to C_r^*: For any linear functional h on C_r^*, let

$$\|h\|'_{[r]} := \sup\{|h(f)|: f \in C_r^*, \|f\|_{[r]} \le 1\}.$$

THEOREM 3.6 (Love and Young). *Let $1 < s < \infty$ and let $L \in (C_s^*)'$. Then there exists a function $g \in W_r$, where $r^{-1} + s^{-1} = 1$ such that $L(f) = \int_a^b f\,dg$, the Riemann–Stieltjes integral existing for all $f \in C_s^*$, and where $\|g\|_{(r)} \le 2^{2+1/r}\|L\|'_{[s]}$.*

PROOF. Love and Young (1937), page 248, give this fact except for using the norm $\|f\|_{[p,a]} := |f(a)| + \|f\|_{(p)}$ in place of $\|\cdot\|_{[p]}$ and having a factor of $2^{1+1/r}$. The conclusion follows since $|f(a)| \le \|f\|_\infty \le \|f\|_{[p,a]}$, so $\|f\|_{[p,a]} \le \|f\|_{[p]} \le 2\|f\|_{[p,a]}$ for all f. $\square$

Now given any finite signed measure ν on the Borel sets of $[a, b]$, let G be the distribution function $G(x) := \nu([a, x])$. Then $\int f\,d\nu = \int f\,dG$, where the Riemann–Stieltjes integral on the right exists for all continuous f [by Lemma 3.1, or, e.g., Kolmogorov and Fomin (1970), page 368]. Since G is of bounded variation, we also have $\|G\|_{(t)} < \infty$ for $1 \le t < \infty$.

Suppose $\int f\,d\nu = \int f\,dg$ for all $f \in C_s^*$ where $g \in W_r$, $r^{-1} + s^{-1} = 1$. For $a < c \le b$, consider the functions $f_n(x) := \max(0, \min(1, n(c - x)))$ on $[a, \mathrm{b}]$. Then $f_n \uparrow f := 1_{[a,c[}$ and f_n are uniformly of bounded variation. Let $c(n) := c - 1/n$ and $d(n) := c - 1/n^2$. Then as $n \to \infty$,

$$\int_a^b f_n\,dg = \left(\int_a^{c(n)} + \int_{c(n)}^{d(n)} + \int_{d(n)}^c\right) f_n\,dg \to g(c-) - g(a),$$

because the first integral on the right equals $g(c_n) - g(a)$ which converges to the given limit, the third integral goes to 0 since $0 \le f_n \le 1/n$ on $[d(n), c]$, and the second integral goes to 0 since the r-variation of g on the half-open interval $[c(n), c[$ goes to 0 while the 1-variation of f_n remains bounded by 1, so Theorem 3.5 applies. It follows that $g(c-)-g(a)$ is uniquely determined by ν. Likewise, taking $h_n \downarrow 1_{[a,c]}$, so is $g(c+)-g(a)$, and clearly $g(b)-g(a)=\int_a^b 1\,dg$.

By adding a constant to g, we can take $g(a)=0$. Then on the set of functions $g \in W_r[a,b]$ with the same $g(x-)$ for $a<x\le b$, the same $g(x+)$ for $a \le x < b$, the same $g(b)$, and the same $\int f\,dg$ for all $f \in C_s^*$, $\|g\|_{(r)}$ is minimized when g is right continuous at x for $a<x<b$, because for such a g and x, in any r-variation sum, if x appears, it can be replaced by $x+\delta$ for $\delta \downarrow 0$ such that g is continuous at $x+\delta$, so that the supremum of such sums is the same as the supremum of sums in which x does not appear. A different value of $g(x)$ could only make the r-variation larger.

Also considering f_n in C_s^* with $f_n \downarrow 1_{\{a\}}$ and $\|f_n\|_{[s]} \equiv 2$, we get $|\nu(\{a\})| = |\lim_{n\to\infty} \int f_n\,d\nu| \le 2\|\nu\|'_{[s]}$. So we obtain the following result.

PROPOSITION 3.7. *If ν is a finite signed measure on $[a,b]$, $G(x) \equiv \nu([a,x])$, $1<s<\infty$ and $s^{-1}+r^{-1}=1$, then*

$$\|G\|_{(r)} \le 2^{2+1/r}\|\nu\|'_{[s]} \quad \textit{and} \quad \|G\|_{[r]} \le (2^{3+1/r}+2)\|\nu\|'_{[s]}.$$

Note that conversely, by Theorem 3.5, for any $\varepsilon>0$,

$$\|\nu\|'_{[s]} \le \left(1+\zeta\left(\frac{1}{s}+\frac{1}{r-\varepsilon}\right)\right)\|G\|_{(r-\varepsilon)}.$$

COROLLARY 3.8. *For any r with $2<r<\infty$ and $\varepsilon>0$, there is an $M<\infty$ such that for any distribution function F on $\mathbb{R}$ and its empirical distribution functions F_n, $\Pr\{\|n^{1/2}(F_n-F)\|_{[r]}>M\}<\varepsilon$ for all n.*

PROOF. $\|F_n-F\|_{[r]}$ is a measurable random variable since in both $\|F_n-F\|_{(r)}$ and $\|F_n-F\|_\infty$ one can restrict to the rational numbers. Let P be the law with distribution function F and P_n its empirical measures. Apply Proposition 3.7 to $\nu := \nu_n := n^{1/2}(P_n-P)$, $a := -\infty$, $b := +\infty$ and $s := r/(r-1)$. Then $1<s<2$, and we can treat $\|\nu_n\|'_{[s]}$ in place of $\|n^{1/2}(F_n-F)\|_{[r]}$. The unit ball of C_s^*, being a subset of that of W_s, is a dominated uniform Donsker class by Theorem 2.2, and the result follows. □

COROLLARY 3.9. *Let $T(\cdot)$ be a functional on some open set U in W_r, $r>2$, such that T is Fréchet differentiable with derivative $L(\cdot)$ at some distribution function $F \in U$ with respect to $\|\cdot\|_{[r]}$. Then for the empirical distribution functions F_n,*

$$T(F_n)=T(F)+L(F_n-F)+o_p(n^{-1/2}) \quad \textit{as } n\to\infty,$$

where the $o_p(n^{-1/2})$ is uniform in F.

PROOF. By definition of Fréchet differentiability, the remainder is $o(\|F_n-F\|_{[r]})$, which is uniformly $o_p(n^{-1/2})$ by Corollary 3.8. □

Some functionals with the differentiability stated in Corollary 3.9 are the inverse operator $F \mapsto F^{-1}$ and the composition operator $(F, G) \mapsto F \circ G$ with respect to F, while G varies in L^p, under suitable conditions [Dudley (1991a)].

Corollary 3.8 also follows directly from:

(a) the results of Pisier and Xu, discussed after Theorem 2.2 above, and
(b) again, the Love and Young (1937) duality.

There is still another proof of Corollary 3.8, by martingales [Dudley (1991c)].

4. Bilinear operations and Young integrals. Let X, Y and Z be three Banach spaces and let B be a function from $X \times Y$ into Z which is *bilinear*, so that $B(\cdot, y)$ is linear on X for each $y \in Y$ and $B(x, \cdot)$ is linear on Y for each $x \in X$. Then B provides its own partial derivatives with respect to each variable, since

$$B(x+u, y) - B(x, y) \equiv B(u, y) \quad \text{and} \quad B(x, y+v) - B(x, y) \equiv B(x, v).$$

The remainder in differentiating with respect to both variables is

$$B(x+u, y+v) - B(x, y) - B(u, y) - B(x, v) \equiv B(u, v).$$

So B is jointly Fréchet differentiable from $X \times Y$ into Z if $\|B(u, v)\| = o(\|u\| + \|v\|)$ as $\|u\| \to 0$ and $\|v\| \to 0$. If B is a *bounded* bilinear form in the sense that for some $K < \infty$, $\|B(u, v)\| \le K\|u\|\,\|v\|$ for all $u \in X$ and $v \in Y$, then B is jointly Fréchet differentiable. If X, Y and Z are Banach spaces, boundedness is equivalent to joint continuity, and even to separate continuity, where $B(x, y)$ is continuous in x for each fixed y and in y for each fixed x [e.g., Schaefer (1966), page 88].

Functionals of the form $B(F, G) := \int F\,dG$ are bilinear. Such functionals are basic in the duality theory of functions of bounded p-variation as in Section 3. On the other hand, for empirical distribution functions F_m and G_n, $\int F_m\,dG_n$ is (up to normalization) a two-sample Wilcoxon statistic. Gill (1989), Lemma 3, page 110 shows that B is compactly differentiable for supremum norms at F_0, G_0 on $D[-\infty, \infty] \times E_1$ if E_1 is the set of functions in $D[-\infty, \infty]$ of total variation at most C for some $C < \infty$, and F_0 has bounded variation. Gill (1989), Lemma 1, page 105, then gives an extension, compactly differentiable for sup norms, of B to $D[-\infty, \infty] \times D[-\infty, \infty]$, at (F_0, G_0) of bounded variation. But this extension simply deletes the remainder term $B(f, g)$ in

$$B(F_0 + f, G_0 + g) \equiv B(F_0, G_0) + B(F_0, g) + B(f, G_0) + B(f, g)$$

if $G_0 + g \notin E_1$, in other words, if $G_0 + g$ has total variation greater than C. The resulting function is no longer bilinear, and is discontinuous even along lines. As Gill notes, for any probability distribution functions G and $G + g$ such as $G + g = G_n$, if $C \ge 1$, then G and $G + g$ are both in E_1, while if $C \ge 2$, then also $g \in E_1$, so the extension makes no difference.

Young's p-variation duality theory (Section 3) provides an alternative approach.

THEOREM 4.1. *The integrals* $(F, G) \mapsto (Y_i)\int F\,dG$, $i = 1, 2$, *are Fréchet differentiable, on* $W_r \times W_s$ *for* $\|\cdot\|_{[r]}, \|\cdot\|_{(s)}$ *whenever* $r^{-1} + s^{-1} > 1$.

PROOF. By a theorem of Young (the latter part of Theorem 3.5 above), the integrals are bounded bilinear forms for the given norm and seminorm, so they are Fréchet differentiable. □

Corollary 3.8 and Theorem 4.1 cannot be applied directly if F and G are both replaced by empirical processes, since we cannot take both $r > 2$ and $s > 2$. On the other hand, it may sometimes be useful that empirical processes do have sample functions in r-variation spaces also for $1 \le r \le 2$, without the uniformity of Corollary 3.8.

In this case [unlike the composition and inverse operators treated by Reeds (1976), Fernholz (1983) and Dudley (1991a)], the derivative need not be taken at F_0 or G_0 having stronger smoothness properties; neither of them has to be of bounded variation. While the $\|\cdot\|_{[r]}$ norm for f is stronger than the sup norm, recall that compact sets for the sup norm are small while bounded sets for $\|\cdot\|_{[r]}$ can be quite large (nonseparable) for the sup norm as well as for the $\|\cdot\|_{[r]}$ norm itself.

Acknowledgments. I thank Richard Gill, Evarist Giné, Jose Gonzalez-Barrios, Gilles Pisier, Aad van der Vaart, Jon Wellner and the referee for helpful comments on this paper and the subject generally.

REFERENCES

BERTOIN, J. (1989). Sur une intégrale pour les processus à α-variation bornée. *Ann. Probab.* **17** 1521–1535.

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

BRUNEAU, M. (1974). *Variation totale d'une fonction. Lecture Notes in Math.* **413**. Springer, New York.

DUDLEY, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. *Probability in Banach Spaces V. Lecture Notes in Math.* **1153** 141–178. Springer, New York.

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

DUDLEY, R. M. (1989a). *Real Analysis and Probability.* Wadsworth and Brooks/Cole, Pacific Grove, Calif.

DUDLEY, R. M. (1989b). Comment on "Asymptotics via empirical processes" by D. Pollard. *Statist. Sci.* **4** 354.

DUDLEY, R. M. (1990). Nonlinear functionals of empirical measures and the bootstrap. In *Probability in Banach Spaces VII* (E. Eberlein, J. Kuelbs and M. B. Marcus, eds.). *Progress in Probability* **21** 63–82. Birkhäuser, Boston.

DUDLEY, R. M. (1991a). Differentiability of the composition and inverse operators for regulated and a.e. continuous functions. Preprint.

DUDLEY, R. M. (1991b). Implicit functions of empirical measures and robust statistics. Unpublished manuscript.

DUDLEY, R. M. (1991c). The order of the remainder in derivatives of composition and inverse operators for Φ-variation norms. Preprint.

FERNHOLZ, L. T. (1983). *von Mises Calculus for Statistical Functionals. Lecture Notes in Statist.* **19**. Springer, New York.

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

GINÉ, E. and ZINN, J. (1991). Gaussian characterization of uniform Donsker classes of functions. *Ann. Probab.* **19** 758–782.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

HILDEBRANDT, T. H. (1938). Definitions of Stieltjes integrals of the Riemann type. *Amer. Math. Monthly* **45** 265–278.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

KOLMOGOROV, A. N. and FOMIN, S. V. (1970). *Introductory Real Analysis.* Translated and edited by R. A. Silverman. Prentice-Hall, Englewood Cliffs, NJ.

KOLMOGOROV, A. N. and TIHOMIROV, V. M. (1961). ε-entropy and ε-capacity of sets in functional spaces. *Amer. Math. Soc. Transl.* (*Ser.* 2) **17** 277–364. [*Uspehi Mat. Nauk* **14** no. 2 (1959) 3–86.]

LOVE, E. R. and YOUNG, L. C. (1937). Sur une classe de fonctionelles linéaires. *Fund. Math.* **28** 243–257.

PISIER, G. and XU, Q. (1987). Random series in the real interpolation spaces between the spaces v_p. In *Geometrical Aspects of Functional Analysis* (J. Lindenstrauss and V. Milman, eds.). *Lecture Notes in Math.* **1267** 185–209. Springer, New York.

PISIER, G. and XU, Q. (1988). The strong p-variation of martingales and orthogonal series. *Probab. Theory Related Fields* **77** 497–514.

POLLARD, D. B. (1982). A central limit theorem for empirical processes. *J. Austral. Math. Soc. Ser. A* **33** 235–248.

POLLARD, D. (1989). Asymptotics via empirical processes (with discussion). *Statist. Sci.* **4** 341–366.

PONS, O. and DE TURCKHEIM, E. (1989). Méthode de von Mises, Hadamard-différentiabilité et bootstrap dans un modèle non paramétrique sur un espace métrique. *C. R. Acad. Sci. Paris Sér. I* **308** 369–372.

REEDS, J. A., III (1976). On the definition of von Mises functionals. Ph.D. dissertation, Dept. Statistics, Harvard Univ.

RUDIN, W. (1976). *Principles of Mathematical Analysis*, 3d ed. McGraw-Hill, New York.

SCHAEFER, H. H. (1966). *Topological Vector Spaces.* MacMillan, New York.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

WIENER, N. (1924). The quadratic variation of a function and its Fourier coefficients. *J. Math. Phys.* **3** 72–94.

XU, Q. (1986). Espaces d'interpolation réels entre les espaces V_p: propriétés géométriques et applications probabilistes. *Séminaire d'Analyse Fonctionelle 1985-86-87* 77–123. Univ. Paris VI–VII; also in Xu (1988), II, with "Remarques sur l'exposé ..." (3 pages).

XU, Q. (1988). Thèse de doctorat, math., Univ. Paris VI.

YOUNG, L. C. (1936). An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.* **67** 251–282.

YOUNG, W. H. (1914). On integration with respect to a function of bounded variation. *Proc. London Math. Soc.* (2) **13** 109–150.

DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139