Anil K. Jain (S'70–M'72) was born in Basti, India, on August 5, 1948. He received the B.Tech. degree with distinction from the Indian Institute of Technology, Kanpur, India, in 1969, and the M.S. and Ph.D. degrees in electrical engineering from Ohio State University, Columbus, in 1970 and 1973, respectively.

From 1971 to 1972 he was a Research Associate in the Communications and Control Systems Laboratory, Ohio State University. Then, from 1972 to 1974, he was an Assistant Professor in the Department of Computer Science, Wayne State University, Detroit, MI. In 1974, he joined the Department of Computer Science, Michigan State University, where he is currently an Associate Professor. His research interests are in the areas of pattern recognition and image processing. He was recipient of the National Merit Scholarship in India.

Dr. Jain is a member of the Association for Computing Machinery, the Pattern Recognition Society, and Sigma Xi.

On the Inequality of Cover and Hart in Nearest Neighbor Discrimination

LUC DEVROYE

Abstract—When $(X_1, \theta_1), \cdots, (X_n, \theta_n)$ are independent identically distributed random vectors from $\mathbb{R}^d \times \{0, 1\}$ distributed as (X, θ), and when θ is estimated by its nearest neighbor estimate $\theta_{(1)}$, then Cover and Hart have shown that $P\{\theta_{(1)} \neq \theta\} \xrightarrow{n \to \infty} 2E\{\eta(X)(1 - \eta(X))\} \leqslant 2R^*(1 - R^*)$ where R^* is the Bayes probability of error and $\eta(x) = P\{\theta = 1 | X = x\}$. They have conditions on the distribution of (X, θ). We give two proofs, one due to Stone and a short original one, of the same result for *all* distributions of (X, θ).

If ties are carefully taken care of, we also show that $P\{\theta_{(1)} \neq \theta | X_1, \theta_1, \cdots, X_n, \theta_n\}$ converges in probability to a constant for *all* distributions of (X, θ), thereby strengthening results of Wagner and Fritz.

Index Terms—Bayes' risk, inequality of Cover and Hart, nearest neighbor rule, nonparametric discrimination, probability of error.

I. Introduction

Let $(X, \theta), (X_1, \theta_1), \cdots, (X_n, \theta_n)$ be independent identically distributed $\mathbb{R}^d \times \{0, 1\}$-valued random vectors and estimate θ from X and the (X_i, θ_i)'s by $\theta_{(1)}$, the nearest neighbor estimate that is obtained by reordering the (X_i, θ_i) according to increasing values for $||X_i - X||$ and taking $\theta_{(1)}$ from the nearest neighbor $X_{(1)}$ (ties are broken by comparing original indexes).

Cover and Hart [1] have shown the following inequality. When

$$L_{n1} = P\{\theta_{(1)} \neq \theta | X_1, \theta_1, \cdots, X_n, \theta_n\}$$

and

$$\eta(x) = P\{\theta = 1 | X = x\},$$

then

$$E\{L_{n1}\} = P\{\theta_{(1)} \neq \theta\} \xrightarrow{n \to \infty} 2E\{\eta(X)(1 - \eta(X))\} \leqslant 2R^*(1 - R^*) \tag{1}$$

where

$$R^* = E\{\min(\eta(X), 1 - \eta(X))\}$$

is the Bayes probability of error. They require, however, that X have a density f and that f and η are almost everywhere continuous. It should be noted, however, that the proof in Cover and Hart holds for X taking values in a separable metric space. Stone [4] has implicitly shown that (1) is true for *all* distributions of (X, θ). The purpose of this note is to give two short proofs of (1) and to obtain additional results on the convergence of L_{n1}.

II. The Basic Theorem

Theorem 1:

$$E\{L_{n1}\} \xrightarrow{n \to \infty} 2E\{\eta(X)(1 - \eta(X))\} \leqslant 2R^*(1 - R^*).$$

Proof: Fix a version η of $P\{\theta = 1 | X = x\}$, and let $X_{(1)}^x$ be the nearest neighbor of x while $X_{(1)}^X$ is the nearest neighbor of the random variable X. Further, let

$$\xi(x) = E\{\eta(X_{(1)}^x)\}$$

and

$$r_n(x) = \xi(x)(1 - \eta(x)) + (1 - \xi(x))\eta(x).$$

The inequality in Theorem 1 follows from $\eta(x)(1 - \eta(x)) =$

Manuscript received August 20, 1979; revised January 17, 1980. This work was supported by AFOSR Grant 77-3385 while the author was at the University of Texas at Austin during the Summer of 1979.

The author is with the School of Computer Science, McGill University, Montreal, P.Q., Canada.

0162-8828/81/0100-0075$00.75 © 1981 IEEE

$\min(\eta(x), 1-\eta(x)) \times (1-\min(\eta(x), 1-\eta(x)))$ and Jensen's inequality.

Next,

$$\begin{aligned}|r_n(x) - 2\eta(x)(1-\eta(x))| &\leq |\xi(x) - \eta(x)| \\ &\leq E\{|\eta(X^x_{(1)}) - \eta(x)|\}.\end{aligned} \tag{2}$$

We will show for almost all $x(\mu)$ (μ is the probability measure for X), that (2) tends to 0 as $n \to \infty$. By the dominated convergence theorem, we may then certainly conclude that $E\{|r_n(X) - 2\eta(X)(1-\eta(X))|\} \xrightarrow{n\to\infty} 0$. The theorem then follows because

$$\begin{aligned}E\{r_n(X)\} &= E\{\xi(X)(1-\eta(X)) + (1-\xi(X))\eta(X)\} \\ &= E\{\eta(X^X_{(1)})(1-\eta(X)) + (1-\eta(X^X_{(1)}))\eta(X)\} \\ &= E\{L_{n1}\}.\end{aligned} \tag{3}$$

When I is the indicator function and $a > 0$ is a constant, we have

$$\begin{aligned}E\{|\eta(X^x_{(1)} - \eta(x)|\} &\leq P\{\|X^x_{(2)} - x\| > a\} \\ &+ \sup_{0<b\leq a} \frac{\int_{S_{x,b}} |\eta(y) - \eta(x)|\,\mu(dy)}{\mu(S_{x,b})}\end{aligned} \tag{4}$$

where $a > 0$ is arbitrary and $S_{x,r}$ is the closed sphere centered at x and radius r. The last term in (4) tends to 0 as $a \to 0$ for almost all $x(\mu)$ by a theorem on the relative differentiation of measures (Wheeden and Zygmund [6, pp. 185-190]). The first term on the right-hand side of (4) tends to 0 for all $a > 0$ whenever $x \in \text{support}(\mu)$. But $\mu(\text{support}(\mu)) = 1$ (see [1]) and the theorem is proved.

We will sketch a second proof that is essentially due to Stone [4]. Again, we will show that

$$E\{|\eta(X^X_{(1)}) - \eta(X)|\} \xrightarrow{n\to\infty} 0. \tag{5}$$

For fixed $\epsilon > 0$, find g: $\mathbb{R}^d \to [0, 1]$, g continuous, such that $E\{|g(X) - \eta(X)|\} < \epsilon$ (see [6, p. 149]). Estimate (5) from above by

$$\begin{aligned}&E\{|\eta(X^X_{(1)}) - g(X^X_{(1)})|\} \\ &+ E\{|g(X^X_{(1)}) - g(X)|\} \\ &+ E\{|g(X) - \eta(X)|\}.\end{aligned} \tag{6}$$

Stone [4, p. 613] has shown that for any function $f \in L^1(\mu)$,

$$E\{|f(X^X_{(1)})|\} \leq \alpha(d)\,E\{|f(X)|\} \tag{7}$$

where $\alpha(d) > 0$ is a constant depending upon d only. Thus, the first and third terms of (6), summed together, are not greater than $(\alpha(d) + 1)\epsilon$.

For all $x \in \text{support}(\mu)$, we have $X^x_{(1)} \xrightarrow{n\to\infty} x$ a.s., and thus $g(X^x_{(1)}) \xrightarrow{n\to\infty} g(x)$ a.s., so that the second term of (6) tends to 0 by the dominated convergence theorem. Since $\epsilon > 0$ was arbitrary, we may conclude the proof of the theorem.

Remark 1: We have in fact shown that

$$r_n(x) \to 2\eta(x)(1-\eta(x)) \tag{8}$$

for almost all $x(\mu)$.

Remark 2: If g and h_n are uniformly bounded Borel measurable functions of their arguments, then it is true that

$$E\{|g(X^X_{(1)}) - g(x)|\,h_n(X_1, \cdots, X_n, x)\} \xrightarrow{n\to\infty} 0 \tag{9}$$

for almost all $x(\mu)$, and

$$E\{|g(X^X_{(1)}) - g(X)|\,h_n(X_1, \cdots, X_n, X)\} \xrightarrow{n\to\infty} 0, \tag{10}$$

for *all* distributions of X.

Remark 3: The proof given above work for $\mathbb{R}^d$, but it is not clear how they can be generalized to separable metric spaces.

III. The Conditional Probability of Error

For general μ, L_{n1} does not converge to a constant in probability. For example, take $\mu(\{0\}) = 1$, $\eta(0) = \frac{1}{3}$. Clearly,

$$L_{n1} = \tfrac{1}{3} I_{[\theta_1=0]} + \tfrac{2}{3} I_{[\theta_1=1]}$$

and convergence to a constant is thus excluded. Nevertheless, we have the following.

Theorem 2: If μ has no atoms, then

$$L_{n1} \xrightarrow{n\to\infty} 2E\{\eta(X)(1-\eta(X))\}$$

in probability.

Note: Wagner [5] has shown Theorem 2 for the special case that μ has a density f and that η and f are almost everywhere continuous. For $d = 1$, he has shown that

$$L_{n1} \xrightarrow{n\to\infty} 2E\{\eta(X)(1-\eta(X))\} \text{ a.s.,}$$

under the same assumptions. Fritz [2] proved the a.s. convergence of L_{n1} to $2E\{\eta(X)(1-\eta(X))\}$ when μ has no atoms and η is almost everywhere continuous (μ). Our Theorem 2, in contrast, holds for *all* nonatomic measures μ and *all* η.

The proof of Theorem 2 will be postponed until Theorem 3. To take care of the atomic part of μ, Stone [4] proposed replacing $\theta_{(1)}$ by $\hat{\theta}$, where $\hat{\theta}$ is defined as follows.

Reorder the (X_i, θ_i)'s to obtain $(X^X_{(i)}, \theta_{(i)})$, $1 \leq i \leq n$. If

$$\begin{aligned}\|X^X_{(1)} - X\| = \cdots = \|X^X_{(k)} - X\| \\ < \|X^X_{(k+1)} - X\| \leq \cdots \leq \|X^X_{(n)} - X\|,\end{aligned}$$

then let $\hat{\theta}$ be the integer most frequently occurring among $\theta_{(1)}, \cdots, \theta_{(k)}$ (in case of a tie, $\hat{\theta}$ is taken arbitrarily among the integers involved in the tie). Define

$$L_n = P\{\hat{\theta} \neq \theta \mid X_1, \theta_1, \cdots, X_n, \theta_n\},$$

$$A = \text{set of atoms of } \mu$$

and for general Borel sets B from $\mathbb{R}^d$,

$$R^*(B) = E\{I_{[X\in B]} \min(\eta(X), 1-\eta(X))\},$$

$$L(B) = E\{I_{[X\in B]}\, 2\eta(X)(1-\eta(X))\}.$$

Thus, Theorem 1 can be rewritten as

$$E\{L_{n1}\} \xrightarrow{n\to\infty} L(\mathbb{R}^d)$$

and the Bayes probability of error is $R^* = R^*(\mathbb{R}^d)$. In fact, $R^*(\cdot)$ and $L(\cdot)$ can be considered as finite measures on the

Borel sets of $\mathbb{R}^d$, but this matter will not be pursued any further.

Theorem 3:

$$L_n \xrightarrow{n\to\infty} R^*(A) + L(A^c)$$

in probability and

$$E\{L_n\} \xrightarrow{n\to\infty} R^*(A) + L(A^c)$$

where A^c is the complement of A.

Remark 4: $R^*(A)$ is the portion of the Bayes probability of error due to the atomic part of μ; $L(A^c)$ is the portion of the asymptotic nearest neighbor probability of error due to the nonatomic part of μ. Clearly,

$$R^*(\mathbb{R}^d) \leq R^*(A) + L(A^c) \leq L(\mathbb{R}^d) = E\{2\eta(X)(1-\eta(X))\}$$

(the asymptotic nearest neighbor probability of error for $\theta_{(1)}$) so that, in a sense, $\hat{\theta}$ is always better than $\theta_{(1)}$.

Remark 5: If μ is nonatomic, then $L_n = L_{n1}$ a.s. because $\hat{\theta} = \theta_{(1)}$ a.s. Therefore, Theorem 2 is a corollary of Theorem 3.

Remark 6: If μ is atomic, then $L_n \to R^*$ a.s. by Lemma 4 below.

IV. Lemmas Needed to Prove Theorem 3

In this section, we give some lemmas, all of a measure-theoretical nature, that will be used further on. The proofs can be found in the Appendix.

Lemma 1 is an extension of the dominated convergence theorem.

Lemma 1 [3]: Let $|f_n| \leq c < \infty$ be a sequence of Borel measurable functions of $x, X_1, \theta_1, \cdots, X_n, \theta_n$, and let

$$f_n(x, X_1, \theta_1, \cdots, X_n, \theta_n) \xrightarrow{n\to\infty} f(x) \text{ a.s.}$$

for almost all $x(\mu)$, then

$$E\{|f_n(X, X_1, \theta_1, \cdots, X_n, \theta_n) - f(X)| \,|X_1, \theta_1, \cdots, X_n, \theta_n\} \xrightarrow{n\to\infty} 0 \text{ a.s.}$$

The sample $X_1, \cdots, X_n$ partitions $\mathbb{R}^d$ up into at most n sets $A_{1n}, \cdots, A_{nn}$, where A_{in} is the collection of all $x \in \mathbb{R}^d$ for which X_i is the nearest neighbor among $X_1, \cdots, X_n$. Lemma 2 below states that for all nonatomic measures, the μ-measure of these sets tends to 0 a.s. uniformly in i as $n \to \infty$.

Lemma 2 [5]: If μ is a nonatomic finite measure, then

$$\sup_{1\leq i\leq n} \mu(A_{in}) \xrightarrow{n\to\infty} 0 \text{ a.s.}$$

and

$$E\{\sup_{1\leq i\leq n} \mu(A_{in})\} \xrightarrow{n\to\infty} 0.$$

Remark 7: For *any* finite measure μ we thus have

$$\sup_{1\leq i\leq n} \mu(A_{in}\cap A^c) \xrightarrow{n\to\infty} 0 \text{ a.s.}$$

and

$$E\{\sup_{1\leq i\leq n} \mu(A_{in}\cap A^c)\} \xrightarrow{n\to\infty} 0.$$

We will also need some result on the "separation" of the atomic and the nonatomic parts of μ. Lemma 3 below to some degree qualifies the statement that almost all "nonatomic" x's have "nonatomic" nearest neighbors $X_{(1)}^x$ with probability tending to 1 as $n \to \infty$.

Lemma 3:

$$P\{\|X_{(1)}^X - X\| = \|X_{(2)}^X - X\|,\ X \in A^c\} \xrightarrow{n\to\infty} 0.$$

Lemma 4:

$$P\{\hat{\theta} \neq \theta, X \in A \,|\, X_1, \theta_1, \cdots, X_n, \theta_n\} \xrightarrow{n\to\infty} R^*(A) \text{ a.s.}$$

Lemma 5:

$$P\{\hat{\theta} \neq \theta, X \in A^c \,|\, X_1, \theta_1, \cdots, X_n, \theta_n\} \xrightarrow{n\to\infty} L(A^c)$$

in probability when

$$P\{\theta_{(1)} \neq \theta, X \in A^c \,|\, X_1, \theta_1, \cdots, X_n, \theta_n\} \xrightarrow{n\to\infty} L(A^c)$$

in probability.

We can now handle the atomic and nonatomic parts of the probability of error separately. The basic results are that for $X \in A$, $\hat{\theta}$ is asymptotically Bayes (Lemma 4), and that for $X \notin A$, $\hat{\theta}$ and $\hat{\theta}_{(1)}$ are asymptotically equivalent (Lemma 5).

Appendix

Proof of Lemma 2: We first recall that for any compact set $K \subseteq \mathbb{R}^d$,

$$V_n(K) = \sup_{K\cap \text{support}(\mu)} \|X_{(1)}^x - x\| \xrightarrow{n\to\infty} 0 \text{ a.s. [5].}$$

Let $W_n = \sup_{1\leq i\leq n} \mu(A_{in})$. Arguing again as in [5], we have for arbitrary $\epsilon > 0$, for all $n \geq 1$ and for arbitrary compact K,

$$[V_n(K) < \epsilon] \subseteq [W_n < \mu(K^c) + \sup_K \mu(S_{x,\epsilon})]$$

where $S_{x,\epsilon}$ is a closed sphere centered at x with radius ϵ. Choose K such that $\mu(K^c) < \delta/2$ where $\delta > 0$ is given. Since $\sup_K \mu(S_{x,\epsilon}) \to 0$ as $\epsilon \to 0$, it is clear that by choosing ϵ sufficiently small we can ensure

$$[V_n(K) < \epsilon] \subseteq [W_n < \delta].$$

Hence, $W_n \xrightarrow{n\to\infty} 0$ a.s. and $E\{W_n\} \xrightarrow{n\to\infty} 0$.

Proof of Lemma 3: We will show that for almost all $x \in A^c$, $P\{\|X_{(1)}^x - x\| = \|X_{(2)}^x - x\|\} \xrightarrow{n\to\infty} 0$, and Lemma 3 then follows by the dominated convergence theorem.

Without loss of generality we can assume that $x \in \text{support}(\mu)$ because $\mu(\text{support}(\mu)) = 1$. Let ν be the measure on $[0,\infty)$ corresponding to $\|X_1 - x\|$, and let A^* be the set of atoms of ν. Clearly,

$$\begin{aligned} P\{\|X_{(1)}^x - x\| = \|X_{(2)}^x - x\|\} &\leq P\{\|X_{(1)}^x - x\| \in A^*\} \\ &\leq P\{\|X_{(2)}^x - x\| > a\} \\ &\quad + \sup_{0<b\leq a} \nu(A\cap[0,b])/\nu([0,b]) \end{aligned}$$

for arbitrary a. The last term in this expression is arbitrarily small by choice of a [6, Corollary 10.50] and the first term tends to 0 as $n \to \infty$ because $x \in \text{support}(\mu)$. This concludes the proof of Lemma 3.

Proof of Lemma 4: When $\eta(x) < 1 - \eta(x)$, $\mu(\{x\}) > 0$, we have $P\{\hat{\theta} = 0, X = x | X_1, \theta_1, \cdots, X_n, \theta_n\} \xrightarrow{n\to\infty} \mu(\{x\})$ a.s.

by the strong law of large numbers. If $\eta(x) > 1 - \eta(x)$, then it tends to 0 a.s.

Thus,

$$\begin{aligned} &P\{\hat{\theta} \neq \theta, X \in A \mid X_1, \theta_1, \cdots, X_n, \theta_n\} \\ &\quad = \sum_{x \in A} (\eta(x) P\{\hat{\theta} = 0, X = x \mid X_1, \theta_1, \cdots, X_n, \theta_n\} \\ &\qquad + (1 - \eta(x)) P\{\hat{\theta} = 1, X = x \mid X_1, \theta_1, \cdots, X_n, \theta_n\}) \\ &\quad \xrightarrow{n \to \infty} \sum_{x \in A} \mu(\{x\}) \min(\eta(x), 1 - \eta(x)) \text{ a.s.} \end{aligned}$$

by Lemma 1.

Proof of Lemma 5:

$$\begin{aligned} &|P\{\hat{\theta} \neq \theta, X \in A^c \mid X_1, \theta_1, \cdots, X_n, \theta_n\} \\ &\quad - P\{\theta_{(1)} \neq \theta, X \in A^c \mid X_1, \theta_1, \cdots, X_n, \theta_n\}| \\ &\quad \leqslant P\{\|X_{(1)}^X - X\| = \|X_{(2)}^X - X\|, \\ &\qquad X \in A^c \mid X_1, \theta_1, \cdots, X_n, \theta_n\} \\ &\quad \xrightarrow{n \to \infty} 0 \end{aligned}$$

in probability (Lemma 3).

Proof of Theorem 3: Let us define $L_n(1)$, $L_n(2)$, $L_n^*(2)$ as follows:

$$\begin{aligned} L_n &= P\{\hat{\theta} \neq \theta, X \in A \mid X_1, \theta_1, \cdots, X_n, \theta_n\} \\ &\quad + P\{\hat{\theta} \neq \theta, X \in A^c \mid X_1, \theta_1, \cdots, X_n, \theta_n\} \\ &= L_n(1) + L_n^*(2) \end{aligned} \tag{11}$$

and $L_n(2) = P\{\theta_{(1)} \neq \theta, X \in A^c \mid X_1, \theta_1, \cdots, X_n, \theta_n\}$. We have seen that $L_n(1) \to R^*(A)$ a.s. (Lemma 4). $L_n(2)$ can be rewritten as

$$\begin{aligned} &E\{I_{[X \in A^c]} \eta(X) I_{[\theta_{(1)} = 0]} \mid X_1, \theta_1, \cdots, X_n, \theta_n\} \\ &\quad + E\{I_{[X \in A^c]} (1 - \eta(X)) I_{[\theta_{(1)} = 1]} \mid X_1, \theta_1, \cdots, X_n, \theta_n\} \end{aligned} \tag{12}$$

and each term of (12) converges in probability to $1/2\, L(A^c)$ as we will see below. Because $|L_n(2) - L_n^*(2)| \xrightarrow{n \to \infty} 0$ in probability (Lemma 5), we will have shown Theorem 3.

Consider the first term of (12):

$$\begin{aligned} &|E\{I_{[X \in A^c]} \eta(X) I_{[\theta_{(1)} = 0]} \mid X_1, \theta_1, \cdots, X_n, \theta_n\} \\ &\quad - E\{I_{[X \in A^c]} \eta(X) (1 - \eta(X))\}| \\ &\quad \leqslant E\{|\eta(X_{(1)}^X) - \eta(X)| \mid X_1, \theta_1, \cdots, X_n, \theta_n\} \\ &\quad + |E\{I_{[X \in A^c]} \eta(X) (I_{[\theta_{(1)} = 1]} \\ &\quad - \eta(X_{(1)})) \mid X_1, \theta_1, \cdots, X_n, \theta_n\}| \\ &\quad = U_n(1) + U_n(2). \end{aligned} \tag{13}$$

Clearly, $E\{U_n(1)\} \xrightarrow{n \to \infty} 0$ by Remark 2. If A_{in} is defined as in Lemma 2, and

$$\nu(B) = \int_{B \cap A^c} \eta(x) \mu(dx), \quad \text{for } B \text{ a Borel set of } \mathbb{R}^d,$$

then

$$U_n(2) = \left| \sum_{i=1}^n (I_{[\theta_i = 1]} - \eta(X_i)) \nu(A_{in}) \right|$$

and

$$\begin{aligned} E\{U_n^2(2)\} &= E\{E\{U_n^2(2) \mid X_1, X_2, \cdots, X_n\}\} \\ &= E\left\{\sum_{i=1}^n \eta(X_i)(1 - \eta(X_i)) \nu^2(A_{in})\right\} \\ &\leqslant E\{\sup_{1 \leqslant i \leqslant n} \nu(A_{in})\} \\ &\xrightarrow{n \to \infty} 0 \quad \text{(Lemma 2).} \end{aligned}$$

Thus, $L_n(2) \xrightarrow{n \to \infty} L(A^c)$ in probability, thereby completing the proof of Theorem 3.

References

[1] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 21-27, 1967.

[2] J. Fritz, "Distribution-free exponential error bound for nearest neighbor pattern classification," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 552-557, 1975.

[3] N. Glick, "Consistency conditions for probability estimators and integrals of density estimators," *Utilitas Mathematica*, vol. 6, pp. 61-74, 1974.

[4] C. J. Stone, "Consistent nonparametric regression," *Ann. Statist.*, vol. 5, pp. 595-645, 1977.

[5] T. J. Wagner, "Convergence of the nearest neighbor rule," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 566-571, 1971.

[6] R. L. Wheeden and A. Zygmund, *Measure and Integral.* New York: Marcel Dekker, 1977.

Luc Devroye was born in Tienen, Belgium, on August 6, 1948. He received the Ph.D. degree from the University of Texas at Austin in 1976.

In 1977 he became an Assistant Professor at the School of Computer Science, McGill University, Montreal, P.Q., Canada. He is interested in various applications of probability theory and mathematical statistics such as nonparametric estimation, probabilistic algorithms, the computer generation of random numbers, and the strong convergence of random processes.