1986

Nonparametric Recursive Estimation of a Multivariate, Marginal and Conditional DGP with an Application to Specification of Econometric Models

Radhey S. Singh

Aman Ullah

Follow this and additional works at: https://ir.lib.uwo.ca/economicscdse_tr

Part of the [Economics Commons](#)

Citation of this paper:

Singh, Radhey S., Aman Ullah. "Nonparametric Recursive Estimation of a Multivariate, Marginal and Conditional DGP with an Application to Specification of Econometric Models." Centre for Decision Sciences and Econometrics Technical Reports, 14. London, ON: Department of Economics, University of Western Ontario (1986).

THE CENTRE FOR DECISION SCIENCES AND ECONOMETRICS

Nonparametric Recursive Estimation of a Multivariate,
Marginal and Conditional DGP with an Application
to Specification of Econometric Models

Radhey S. Singh
Aman Ullah

**TECHNICAL REPORT NO. 14
OCTOBER 1986**

Centre For Decision Sciences And Econometrics
Social Science Centre
The University of Western Ontario
London, Ontario N6A 5C2

ISSN: 0831-4454
ISBN: 0-7714-0796-3

NONPARAMETRIC RECURSIVE ESTIMATION OF A MULTIVARIATE,
MARGINAL AND CONDITIONAL DGP WITH AN APPLICATION
TO SPECIFICATION OF ECONOMETRIC MODELS

Radhey S. Singh

Aman Ullah

Department of Mathematics
and Statistics
University of Guelph,
Guelph, Ontario, Canada

Department of Economics
University of
Western Ontario
London, Ontario, Canada

Key Words and Phrases: *nonparametric; multivariate; marginal
and conditional density (DGP);
regression curve; kernel; asymptotics;
normality.*

ABSTRACT

Nonparametric recursive kernel estimators of a multivariate data generating process (DGP) are presented and their asymptotic biases, variances and distributions are examined. Weak and strong consistency of these estimators are also proved. Remarks on the choice of the kernel function and the bandwidth function are made. Recursive estimates of the marginal and the conditional DGP are deduced from the estimates of the multivariate density. Finally, an application of these estimates to estimation and specification of econometric models is pointed out.

1. INTRODUCTION

In econometrics, as well as in many other areas of applied research, the data generating process (DGP)¹ is usually of a multivariate nature and quite often it becomes necessary to examine such processes. The true MDGP's (M stands for multivariate) are, however, rarely if ever known in practice, and in some situations the parametric families these DGP's belong to are also not known. Further, there are useful functionals in econometrics which need to be analyzed or estimated but they are heavily and sometimes exclusively dependent on the underlying unknown DGP's. Fisher information function, hazard function, regression function, entropy function are examples of some such functions. Singh (1977b), and more recently Singh, Ullah and Carter (1985) and Ullah and Singh (1984), have pointed out some useful applications of estimates of a multivariate density function. With this in view, we consider in this paper the estimation (nonparametrically) of a MDGP.

Among various methods of univariate DGP estimation, including the orthogonal series (e.g. Schwartz (1967), Kronmal and Tarter (1968) and Watson (1969), the maximum likelihood (e.g. Wegman (1970)), the histogram (Van Ryzin (1973)), the most widely used one is the so-called kernel method of estimation introduced by Rosenblatt (1956), and studied in detail first by Parzen (1962) and Bartlett (1963). Cacoullos (1966) extended techniques developed in Rosenblatt and Parzen to introduce kernel estimates of a MDGP and to study the consistency of such estimates. Cacoullos' work was extended to estimation of partial (or any mixed partial) derivatives of a MDGP by Singh (1978, 1981). In the latter paper of Singh, the speeds of various consistencies of the estimates are also examined. Recently, Prasad and Singh (1980) gave a good review of kernel estimates of a density function.

There has been a computational drawback in the univariate DGP estimates, viz. with respect to updating the additional data points. This drawback has been remedied by the recursive estimation technique of Wolverton and Wagner (1969), Yamato (1971), Deheuvels (1974) and Menon, Prasad and Singh (1983). However, we note that such results are not available for the MDGP estimates of Cacoullos (1966) and Singh (1978, 1981) which suffer from a similar computational drawback.

The aim of this paper is to present (kernel) estimates of a MDGP which not only have computational advantage over those estimates given by Cacoullos and Singh but are also more efficient than their estimators. In addition to this, we provide estimators of marginal DGP, conditional DGP and regression functions based on the estimator of MDGP.

The plan of the paper is as follows. In Section 2.1 we present a class of estimators of a given MDGP based on a random sample of size n . In Section 2.2 we state our main results on biasedness, variance, weak and strong consistencies and distribution of our estimator. In Section 2.3 we comment on how to choose kernel and window-width functions used in the construction of our estimator. The estimators of marginal and conditional DGP's are presented in Section 2.4. Estimation of a regression function is considered in Section 3 and its application to model specification problems in econometrics is pointed out. Proofs of the theorems in Section 2 are presented in Section 4.

2. NONPARAMETRIC ESTIMATORS OF A MDGP

2.1 Estimators of MDGP

Let x_t , $t=1, \dots, n$, be independent $m \times 1$ random vectors generated from an unknown m -variate DGP, f . Let $\mathcal{K}$ be a class of all Borel-measurable real value bounded functions K on the m -dimensional Euclidean space R^m such that

$$(2.1) \quad \int K(w)dw = 1, \quad \int |K(w)|dw < \infty,$$

$$\|w\|^m |K(w)| \rightarrow 0 \text{ as } \|w\| \rightarrow \infty$$

where $\|w\|$ is the usual Euclidean norm of w in R^m . Notice that m -variate normal density and the function $K(w_1, \dots, w_m) =$

$\prod_{j=1}^m \{ \frac{1}{2} I(-1 < w_j < 1) \}$ are among many others that belong to $\mathcal{K}$.

Further, notice that functions in $\mathcal{K}$ need not be non-negative.

Let h_t be a positive decreasing function of t such that $h_t \rightarrow 0$ as $t \rightarrow \infty$. Further, let x be a point in R^m at which the DGP f is to be estimated. For a fixed K belonging to $\mathcal{K}$, define random functions δ_t on R^m by

$$(2.2) \quad \delta_t(x) = h_t^{-m} K\left(\frac{x - x_t}{h_t}\right).$$

Then our estimator of f at x is

$$(2.3) \quad \hat{f}_n(x) = \hat{f}_n(x) = n^{-1} \sum_{t=1}^n \delta_t(x) = n^{-1} \sum_{t=1}^n h_t^{-m} K\left(\frac{x - x_t}{h_t}\right).$$

The estimators considered by Cacoullos (1966) and Singh (1978, 1981), however, are of the type

$$(2.4) \quad \hat{f}_n^*(x) = n^{-1} h_n^{-m} \sum_{t=1}^n K\left(\frac{x - x_t}{h_n}\right).$$

Notice that, if additional data become available, say at some later stage, then to update the estimator (2.4) incorporating all the $n+1$ data, one has to recompute all over again the values of all the $n+1$ functions $K((x_t - x)/h_{n+1})$, $t=1, \dots, n+1$, a rather tedious job even with a computer. In contrast to this, our proposed estimator in (2.3) is recursive in the sense that it can be updated recursively by the formula

$$\begin{aligned}\hat{f}_{n+1}(x) &= (n+1)^{-1} \sum_{t=1}^{n+1} \delta_t(x) \\ &= (n+1)^{-1} (n \hat{f}_n(x) + \delta_{n+1}(x)), \quad \hat{f}_n(x) = \hat{f}_n^*(x),\end{aligned}$$

as soon as the $(n+1)^{\text{th}}$ observation x_{n+1} from f becomes available. In this sense, the estimator $\hat{f}_n(x)$ has clear computational advantage over the estimator $\hat{f}_n^*(x)$ considered earlier in the literature. However, it may be noticed that the summands in (2.3) are, unlike in (2.4), no longer identically distributed and therefore we require some new techniques for proofs given in Section 4.

2.2 Main Results on the Estimates of MDGP

Here we state the main results on unbiasedness, variance, consistency and distribution of our estimate $\hat{f}$ of MDGP f . Proofs of the theorem are presented in Section 4. We show that $\hat{f}$ is asymptotically unbiased and is a weakly (in probability) as well as strongly (with probability one) consistent² estimator of the MDGP f at every continuity point x of f . Variance and distribution of $\hat{f}$ are examined. It turns out that the variance of $\hat{f}$ is strictly smaller than that of the nonrecursive estimator $\hat{f}_n^*$ (defined in (2.4)), and for any distinct points $x_1, \dots, x_2$ in $\mathbb{R}^m$, $f(x_1), \dots, f(x_2)$ are uncorrelated.

Throughout the remainder of this paper, all convergences, unless otherwise stated, are with respect to $n \rightarrow \infty$.

Theorem 2.1 (Asymptotic unbiasedness). Let x be a point in $\mathbb{R}^m$ at which f is continuous. Then the statistic $\hat{f}(x)$ is an asymptotically unbiased estimator of $f(x)$, that is

$$(2.5) \quad \lim_{n \rightarrow \infty} E(\hat{f}(x)) = f(x).$$

Theorem 2.2 (Variance). At every continuity point x of f ,

$$(2.6) \quad \text{var}(\hat{f}(x)) = n^{-2} \sum_{t=1}^n h_t^{-m} \{ (f(x))^2 K + o(1) \}$$

where $\int K^2 = \int K^2(w)dw$ is assumed to be finite.

Remark 2.1 If f is estimated by a nonrecursive estimator (2.4) (considered by Cacoullos (1966) and Singh (1978, 1981)), then (e.g. see Parzen (1962) for $m=1$, and Cacoullos (1966) for $m \geq 1$),

$$(2.6)* \quad \text{var}(f^*(x)) = (nh_n^m)^{-1} (f(x) \int K^2 + o(1))$$

Thus, ignoring the term $o(1)$ in (2.6) and (2.6)*, we have

$$\text{var}(\hat{f}(x)) < \text{var}(f^*(x))$$

which indicates that under similar assumption on f , K and $\langle h_n \rangle$, our estimator is more efficient than those proposed in Cacoullos (1966) or Singh (1978, 1981).

From Theorems 2.1 and 2.2, it follows that at every continuity point x of f ,

$$(2.7) \quad \text{MSE}(\hat{f}(x)) = E|\hat{f}(x) - f(x)|^2 = o(1)$$

provided

$$(2.8) \quad n^{-2} \sum_{t=1}^n h_t^{-m} = o(1).$$

Since h_t decreases in t , a sufficient condition for (2.8) is

$$(2.8)* \quad (nh_n^m)^{-1} = o(1).$$

Choices of the sequence $\langle h_t \rangle$ satisfying (2.8) and (2.8)* are given in the next subsection. The following corollary follows from (2.7) and an application of Chebyshev's inequality.

Corollary 2.1 (Weak Consistency) If (2.8) holds, then at every continuity point x of f , $\hat{f}$ is consistent (in prob.)³ estimator of f .

Remark 2.2. From the proofs of Theorems 2.1 and 2.2 it follows that if f is uniformly continuous, then (2.5) and (2.6) hold uniformly in x . Thus, if in addition, f is bounded, and h_t satisfies (2.8), then

$$(2.7)^* \quad \sup_x (\text{MSE } \hat{f}(x)) = o(1).$$

For (2.7)*, the boundedness of f is in fact not necessary if we make a stronger (than (2.8)) assumption on h_t , namely

$$n^{-2} \sum_{t=1}^n h_t^{-2m} = o(1), \text{ because } \sup_x \text{var}(\hat{f}(x)) \leq$$

$$n^{-2} \sum_{t=1}^n h_t^{-2m} \sup_x \text{EK} \left(\frac{x_t^2 - x^2}{h_t} \right) \text{ and } \text{EK} \left(\frac{x_t^2 - x^2}{h_t} \right) \text{ is bounded in } x$$

since K is bounded.

Remark 2.3. The consistency of Corollary 2.1 does not necessarily imply the convergence of $\hat{f}$ to f with probability one (see footnote 3). This latter property of an estimator, which is more powerful than the former one (the consistency in probability), e.g., see Rao (1973), p. 110, can, however, be achieved by our estimator if the sequence $\langle h_t \rangle$ satisfies a stronger (than (2.8), used in Corollary 2.1) condition, namely

$$(2.8)^* \quad \sum_{n=1}^{\infty} \exp(-\epsilon n h_n^m) < \infty \text{ for every fixed } \epsilon > 0.$$

Choices of sequences $\langle h_t \rangle$ satisfying (2.8)** are given in the next subsection.

Theorem 2.3 (Strong Consistency). Let the sequence $\langle h_t \rangle$ satisfy (2.8)**. Then at every continuity point of f , $\hat{f}$ is a consistent estimator of f with probability one.

The strong consistency of a recursive kernel estimate of a density is also proved in Deheuvels (1974), Isogal (1980) and Menon, Prasad and Singh (1983).

Next we study the distributional properties of $\hat{f}$. In order to build up any tests or any confidence intervals for the unknown density f (or for any of its functionals) based on

the estimator $\hat{f}$ of f , it is necessary to obtain the distribution of $\hat{f}$, at least for large n , if not for every n . In this section we will show that the distribution of $(\hat{f}(x) - E\hat{f}(x))/\{\text{var } \hat{f}(x)\}^{1/2}$ is asymptotically normal with mean zero and variance one. We further show that for any two distinct points x, x' in R^m , $\text{cov}(\hat{f}(x), \hat{f}(x')) = o(1)$, from which we conclude that, for any distinct points $x_1, x_2, \dots, x_l$ in R^m , the asymptotic distribution of the l -variate vector $(\hat{f}(x_1), \dots, \hat{f}(x_l))$ is a l -variate normal. More specifically, we conclude that for distinct $x_1, \dots, x_l$ continuity points (of f) in R^m , $(\hat{f}(x_1), \dots, \hat{f}(x_l))$ is asymptotically multi-variate normal with mean vector $(f(x_1), \dots, f(x_l))$ and variance-covariance matrix as a $l \times l$ diagonal matrix whose i^{th} element is asymptotically equivalent to $n^{-2} (\sum_{t=1}^n h_t^{-m}) f(x_i)^2 |K|^2$.

Theorem 2.4 (Asymptotic Normality) Let Φ be the (cumulative) probability distribution of the univariate standard normal random variable Z (i.e. $\Phi(\cdot) = P[Z \leq \cdot]$). Then at every continuity point x of f ,

$$(2.9) \quad \sup_{t \in R} 1/P\left[\frac{\hat{f}(x) - E\hat{f}}{\{\text{var } \hat{f}(x)\}^{1/2}} \leq t\right] - \Phi(t) = O\left(\frac{\epsilon}{n}\right)$$

where

$$\epsilon_n = \sum_{t=1}^n h_t^{-2m} / \left(\sum_{t=1}^n h_t^{-m}\right)^{3/2}.$$

Remark 2.4. Notice that for a positive α , if h_t is of the order $t^{-\alpha}$, then $\sum_{t=1}^n h_t^{-2m}$ and $\sum_{t=1}^n h_t^{-m}$ are of the order $n^{1+2\alpha m}$ and $n^{1+\alpha m}$ respectively, and hence, ϵ_n is of the order $n^{-(1-\alpha m)/2}$ which converges to zero as n tends to infinity for every $\alpha <$

$1/m$. Thus, in addition to proving the asymptotic normality of $\hat{f}$, Theorem 2.4 also provides the minimum speed (w.r.t. $n \rightarrow \infty$) with which the distribution of $\hat{f}$ approaches to the normal distribution.

$$\text{By Theorem 2.2, } \text{var}(\hat{f}_n(x)) = (n^{-2} \sum_{t=1}^n h_t^{-m}) [f(x)] K^2 + o(1).$$

Hence, Theorem 2.4 shows that

$$(2.10) \quad n \left(\sum_{t=1}^n h_t^{-m} \right)^{-1/2} (\hat{f}(x) - E\hat{f}(x)) \xrightarrow{\mathcal{D}} N(0, f(x) \int K^2)$$

where " $\mathcal{D}$ " stands for convergence in distribution. Thus if we can show that

$$(2.11) \quad n \left(\sum_{t=1}^n h_t^{-m} \right)^{-1/2} (E\hat{f}(x) - f(x)) \rightarrow 0,$$

then (by X(b), p. 122 Rao (1973))⁴

$$(2.12) \quad n \left(\sum_{t=1}^n h_t^{-m} \right)^{-1/2} (\hat{f}(x) - f(x)) \xrightarrow{\mathcal{D}} N(0, f(x) \int K^2)$$

In the next subsection we give sufficient conditions on the kernel K , f and the sequence h_t under which (2.11) (and hence (2.12)) holds.

Remark 2.5. (Asymptotic Covariance). Since δ_t 's are independent, for distinct points x and x' in $\mathbb{R}^m$,
 $\text{cov}(\hat{f}(x), \hat{f}(x')) = n^{-2} \sum_{t=1}^n h_t^{-2m} \text{cov}(\delta_t(x), \delta_t(x'))$. But if x and x'

are continuity points of f , then from the proof of Theorem 2.1,

$$h_t^{-m} (E\delta_t(x))(E\delta_t(x')) = h_t^m (f(x)f(x') + o(1)) = O(h_t^m) \text{ as } t \rightarrow \infty.$$

$$\text{Also, } h_t^{-m} E(\delta_t(x)\delta_t(x')) = \int K(y)K(y + \frac{x-x'}{h_t})f(x + h_t y)dy$$

which is in turn $o(1)$ since K is bounded and $h_t \rightarrow 0$ as $t \rightarrow \infty$

implies $\left| K(y + \frac{x-x'}{h_t}) \right| \rightarrow 0$ as $t \rightarrow \infty$ by (2.1). Hence

$$(2.13) \quad \text{Cov}(\hat{f}(x), \hat{f}(x')) = O(n^{-2} \sum_{j=1}^n h_j^{-m})$$

Remark 2.6. Since uncorrelated normal random variables are independently distributed, in view of (X1) of Rao (1973), p. 123, it follows from (2.12) and (2.13) that for $x_1, x_2, \dots, x_l$, distinct continuity points in R^m of f ,

$$n \left(\sum_{j=1}^n h_j^{-m} \right)^{-1/2} \{ (f(x_1) - f(x_1)), \dots, (f(x_l) - f(x_l)) \} \rightarrow$$

$MN(0, D(f(x_i)/K^2))$ where "MN" stands for multivariate

normal, and $D(f(x_i)/K^2)$ is the $l \times l$ diagonal matrix with $f(x_i)/K^2$ as its i^{th} diagonal element.

2.3 Comments on the Choice of the Kernel K and the Window-width Function h_t

To make any comment on the choice of K and h_t we have to examine the values of the bias and variance of f in terms of K and h_t . Because the speed of convergence in $\lim_{n \rightarrow \infty} E f(x) = f(x)$

(concluded in Theorem 2.1) is directly connected (as we will see) with the choice of K and h_t , we will take this opportunity to investigate the speed of convergence in Theorem 2.1 as well. Besides, we must establish some speed of asymptotic unbiasedness in order to have (2.11) ensuring (2.12).

If for some integer $r > 0$, all partial derivatives of f are continuous at x , the point in R^m at which f is to be estimated, then taking K belonging to in the definition of δ_t , such that

$$\int_{w_1}^{i_1} \dots \int_{w_m}^{i_m} K(w_1, \dots, w_m) = \begin{cases} 1 & \text{if } i_1 = \dots = i_m = 0 \\ 0 & \text{if } 0 < i_1 + \dots + i_m < r \end{cases}$$

and $\int_{w_1}^{i_1} \dots \int_{w_m}^{i_m} |K(w)| < \infty$ for $i_1 + \dots + i_m = r$, it can be

shown by the arguments used in the proofs of Theorem 2.1 here, Theorem of Singh (1977a) and in Theorem 1 of Singh (1981) that

$$(2.14) \quad (\hat{E}f(x) - f(x)) = (n^{-1} \sum_{t=1}^n h_t^r) (c(x, r, m) + o(1))$$

$$\text{where } c(x, r, m) = \sum_{i=1}^m \{ (\partial^r f(x) / \partial x_i^r) (\int_{w_1}^r K(w_1, \dots, w_m) dw / r!) \}.$$

An examination of (2.14) and (2.6) clearly indicates that

$|\text{bias}(\hat{f})|$ will be large whenever $|\int_{w_1}^r K(w_1, \dots, w_m) dw|$ for $i=1, \dots, m$ is large, and $\text{var}(f)$ will be large whenever $\int K^2$ is large. Hence attention must be paid in selection of a kernel so that neither $|\int_{w_1}^r K(w_1, \dots, w_m) dw|$ for $i=1, \dots, m$ nor $\int K^2$ is "too large" (in fact the smaller the better).

Having selected a kernel, we can now examine the criteria for selecting the window-width function h_t . An examination of (2.14) and (2.6) suggests that if h_t is chosen so that h_t is proportional to $t^{-1/(m+2r)}$, say,

$$(2.15) \quad h_t = \alpha_r(x, m) t^{-\gamma} \quad \gamma = 1/(m+2r)$$

then both the bias and the variance of $\hat{f}$ remains controlled simultaneously as $n \rightarrow \infty$.

Now it remains to examine the constant of proportionality in (2.15). For $m=1$, Singh (1979) has given the optimal choice of the constant of proportionality in (2.15). However, this optimal choice depends on the true values of the unknown DGP and its r^{th} order partial derivative, and therefore is not usable in

practice. Woodroffe's (1970) iterative method is also quite cumbersome to use in practice. We attempt to present some solution to this problem as follows. For the case $m=1$, let the kernel be K_1 (on R^1) and the DGP of interest be f_1 . Then the bias in (2.14) becomes $(n^{-1} \sum_{t=1}^n h_t^{-r} k_r)$ times $(\partial f_1^r(x_1)/\partial x_1^r)$, where

$$k_r = \int_{-r}^r w K_1(w) dw / r!). \text{ Hence, if } |k_r| \text{ is large we should}$$

choose h_t small so that $h_t^r k_r$ remains small, hence keeping the bias small. On the other hand, since variance is

$$(n^{-2} \sum_{t=1}^n h_t^{-1} \int K_1^2) f_1^2(x_1) \text{ if } \int K_1^2 \text{ is large, we should choose } h_t$$

large so that $h_t^{-1} \int K_1^2$ remains small, and hence keeping the

variance small. With these considerations in view, we see that $\alpha_r(x, m)$ in (2.15) for the case $m=1$ should be

$$(2.16) \alpha_r(x, 1) \text{ proportional to } [(\int K_1^2)^{2(1)} / \{2r | \int w K_1(w) | / r!\}]^{(1)} 1/(1+2r)$$

where $a^{(1)} = \max \{1, a\}$, and the constant of proportionality does not depend on K_1 .

Now consider the $\text{var}(\hat{f})$ for the $m=1$ case. Since $\text{var}(\hat{f}) =$

$$n^{-2} \sum_{t=1}^n h_t^{-2} \text{var} K_1\left(\frac{x_{1t} - x_1}{h_t}\right), \text{ it is inflated whenever the}$$

$\text{var} K_1\{x_{1t} - x_1\}/h_t$ does which, in turn, gets inflated whenever

$\text{var}(x_{1t}) = \sigma_1^2$ (say) does. The effect of σ_1^2 on the $\text{var}(K(x_{1t} - x_1/h_t))$ is eliminated to some extent if we take $h_t = \sigma_1^* h_t^*$, so that $K(x_{1t} - x_1)/h_t$ is now $K((x_{1t} - x_1)/\sigma_1^* h_t^*)$, and let h_t^* satisfy (2.15) and (2.16). Thus we conclude that one is expected to do better if he chooses, for the $m=1$ case,

$$(2.17) \quad h_t = \alpha_r t^{-\gamma}, \quad \gamma = 1/(1+2r)$$

where $\alpha_r \propto \sigma_{10}^2 [(JK_1^2)^{(1)} / \{2r \int w_1^r K_1(w_1) / (r!) \}^1]^{1/(1+2r)}$,

$a^{(1)} = \max\{1, a\}$, and σ_{10} is a "good" guess value of σ_1 . For the case $m \geq 1$, recommended choice of α_r in (2.15) is

$$\alpha_r \propto \eta_m^2 [(JK^2) / \{2r \sum_{i=1}^m \int w_i^r K(w_i, \dots, w_m) dw / r!\}^{(1)}]^{1/(1+2r)}$$

where η_m is $m^{-1} \sum_{i=1}^m \sigma_{i0}^{-1}$, $(m^{-1} \sum_{i=1}^m \sigma_{i0}^{-2})^{1/2}$, $(\prod_{i=1}^m \sigma_{i0})$ or some

likewise function, and σ_{i0}^2 is a "good" guess value of $\sigma_i^2 =$

$\text{var}(x_{it})$, x_{it} being the i^{th} component of x_t .

Remark 2.7. Instead of estimator (2.3), it may be appropriate to consider the estimator

$$(2.18) \quad f^*(x) = n^{-1} \sum_{t=1}^n \{ (\prod_{i=1}^m h_{it}^{-1}) K(\frac{x_{1t} - x_1}{h_{1t}}, \dots, \frac{x_{mt} - x_m}{h_{mt}}) \}$$

where $x_t = (x_{1t}, \dots, x_{mt})$, $x = (x_1, \dots, x_m)$ and for each $i=1, \dots, m$, $0 < h_{it} \rightarrow 0$ as $t \rightarrow \infty$. Then, results similar to those stated in Subsection 2.2 for the

estimator $\hat{f}$ remain true for f^* too under the same assumption on f . For example, at every continuity point x of f ,

$$Ef^*(x) = f(x) + o(1), \quad \text{var}(f^*(x)) = n^{-2} (\sum_{t=1}^n (\prod_{i=1}^m h_{it}^{-1})) f(x) JK^2 + o(1)$$

and

$$n (\sum_{t=1}^n (\prod_{i=1}^m h_{it}^{-1}))^{-1/2} (f^*(x) - Ef^*(x)) \rightarrow N(0, f(x) JK^2).$$

Further, if for a fixed integer $r > 0$, all partial derivatives of order r of f are continuous at x and K satisfies the moment conditions indicated in the beginning of this section, then

$$(Ef^*(x) - f(x)) = n^{-1} \sum_{t=1}^n \{ \sum_{i=1}^m \{ h_{it}^r (\partial^r f(x) / \partial x_i^r) \int w_i^r K(w) dw / r! \} + o(1) \}$$

The foregoing analysis shows that for f^* we will be better off if we take

$$(2.19) \quad h_{it} \propto [(fK)^2]^{(1)} / \{2r |f_w^r K(w)| / (r!)\}^{(1)} t^{1/(1+2r)} \quad \gamma$$

for $i=1, \dots, m$

Remark 2.8. The choice of h_t in (2.15) ensures that $MSE(\hat{f}) = O(n^{-2r/(1+2r)})$. A choice of $h_t \propto t^{-\gamma}$, where $\gamma = (m+2r)^{-1}$, ensures (2.11) in view of (2.14). Recall that (2.11) is needed for (2.12). These choices of h_t also ensure (2.8)*, needed the

strong consistency (Theorem 2.3) of $\hat{f}$.

Notice from Remark 2.8 that the speed for the MSE approaches to $O(n^{-1})$ as r gets large. However for large r , construction of a kernel satisfying the moment conditions indicated earlier in this subsection will not be an easy job, and the estimator based on such a kernel will be difficult to compute.

For $r=1$ or 2 , K can be taken any symmetric (about zero) multivariate density. For $r=3$ or 4 , a choice of K could be

$$K(y_1, \dots, y_m) = \prod_{j=1}^m K^*(y_j), \text{ where } K^*(t) =$$

$$(2\pi)^{-1/2} [2e^{-t^2/2} - \frac{1}{2}e^{-t^2/4}] I(-\infty < t < \infty) \text{ or } K^*(t) =$$

$$(2\pi)^{-1/2} \frac{1}{2}(3-t^2)e^{-t^2/2} I(-\infty < t < \infty).$$

2.4 Estimators of Marginal and Conditional DGP

Let us write the m components of the $m \times 1$ vector x_t of Subsection 2.1 as

$$(2.20) \quad x_t = [y_t', z_t']', \quad t=1, \dots, n$$

where y_t is a $p \times 1$ vector and z_t is a $q \times 1$ vector such that $p+q = m$. Similarly, as in Section 2, let K_1 and K_2 be p and q variate functions obtained after integrating out $K(w_1, \dots, w_p)$ with

respect to $(w_{p+1}, \dots, w_m)$ and with respect to $(w_1, \dots, w_p)$ respectively. Further, consider $x = [y', z']'$ to be a point in R^m at which the DGP is to be estimated.

Now the estimator of the joint DGP f at x from (2.3) is

$$(2.21) \quad \hat{f}(x) = n^{-1} \sum_{t=1}^n h_t^{-m} K\left(\frac{x - x_t}{h_t}\right) = \hat{f}(y, z).$$

Using this we can write the estimate of the marginal MDGP of z_t at z as

$$(2.22) \quad \begin{aligned} \hat{f}(z) &= \int \hat{f}(y, z) dy = n^{-1} \sum_{t=1}^n h_t^{-m} \int K\left(\frac{x - x_t}{h_t}\right) dy \\ &= n^{-1} \sum_{t=1}^n h_t^{-q} K_2\left(\frac{z - z_t}{h_t}\right). \end{aligned}$$

The estimate of the marginal MDGP of y_t can be similarly written.

Next, the estimator of the conditional MDGP of y_t given $z_t = z$ can be obtained as

$$(2.23) \quad \hat{f}(y|z) = \frac{\hat{f}(y, z)}{\hat{f}(z)} = \frac{\sum_{t=1}^n h_t^{-m} K\left(\frac{y - y_t}{h_t}, \frac{z - z_t}{h_t}\right)}{\sum_{t=1}^n h_t^{-q} K_2\left(\frac{z - z_t}{h_t}\right)}$$

where $\hat{f}(y, z)$ and $\hat{f}(z)$ are as given in (2.3) and (2.22) respectively.

In a special case when the joint kernel function is

$$(2.24) \quad \begin{aligned} K(w_1, \dots, w_m) &= k_1(w_1), \dots, k_p(w_p) k_{p+1}(w_{p+1}), \dots, k_m(w_m) \\ &= \prod_{i=1}^p k_i(w_i) \prod_{j=1}^q k_{p+j}(w_{p+j}), \end{aligned}$$

where k_j are univariate symmetric (around zero) Borel-measurable functions, satisfying (2.1) for the $m=1$ case. Notice that k_j , for $j=1, \dots, m$, can be identical. Taking $k_1 = \dots = k_m = k_1$ (say) we can write (2.23) as

$$(2.25) \quad \hat{f}(y|z) = \frac{\sum_{t=1}^n h_t^{-m} \left\{ \prod_{i=1}^p k\left(\frac{y_{ti} - y_i}{h_i}\right) \prod_{j=1}^q k\left(\frac{z_{tj} - z_j}{h_j}\right) \right\}}{\sum_{t=1}^n h_t^{-q} \prod_{j=1}^q k\left(\frac{z_{tj} - z_j}{h_j}\right)}$$

where y_{ti} , $i=1, \dots, p$ are p elements of y_t and z_{tj} , $j=1, \dots, q$, are q components of z_t .

Examination of the distributional and other properties of our conditional MDGP is beyond the scope of this paper, and will be done in our forthcoming paper.

3. ESTIMATION AND SPECIFICATION OF ECONOMETRIC MODELS

In this section we point out an application of estimators of MDGP and conditional MDGP to the specification of econometric models. Let us rewrite the partition of $m \times 1$ vector x_t as $x_t = [y_t', z_t']'$ where y_t and z_t are $p \times 1$ and $q \times 1$ respectively. The conditional expectation of y given the value of the vector z , that is the regression of y on z , is then given by

(3.1) $E(y|z) = g(z)$, say.

The function $g(z)$ implies how the average values of the vector of p variables y change with a change in the value of z . The problem of specifying $g(z)$ is essentially the problem of specifying multivariate econometric model. Thus the function $g(z)$ plays a significant role in econometrics for the purposes of prediction and testing economic theories. However, we note that g is known only if either the true MDGP is known or the true conditional MDGP is known. Since these are rarely, if ever, known, econometricians have used the following specification, called a multivariate regression function, as

$$(3.2) \quad E(y|z) = g(z) = z\beta,$$

where β is a $q \times 1$ parameter vector. For $p=1$, (3.2) is a univariate regression function. It is well known that the approximation of the mean of the (unknown) conditional MDGP by $g(z) = z\beta$ introduces a serious inference problem due to various misspecifications. In addition, most of the econometric studies tend to assume z_t to be fixed (nonstochastic) while it is not, due to the nonexperimental nature of the DGP of the economic variables.

The alternative proposed here is to directly estimate $E(y|z)$ by the nonparametric method. Using (2.23), this can be obtained as

$$(3.3) \quad \hat{E}(y|z) = \int y \hat{f}(y|z) dy \\ = \sum_{t=1}^n y_t r_t(z)$$

where $r_t(z)$ is a normalized ratio depending on z given by

$$(3.4) \quad r_t(z) = \frac{h_t^{-q} K_2 \left(\frac{z - z_t}{h_t} \right)}{\sum_{t=1}^n h_t^{-q} K_2 \left(\frac{z - z_t}{h_t} \right)}.$$

The expression (3.3) corresponding to (2.25) can similarly be written. Note that for fixed sample size n , when $h_t = h_n = h$ for $t=1, \dots, n$, then

$$(3.5) \quad r_t(z) = \frac{K_2 \left(\frac{z - z_t}{h} \right)}{\sum_{t=1}^n K_2 \left(\frac{z - z_t}{h} \right)}.$$

Remark 3.1. It is clear from (3.3) that the nonparametric estimator $\hat{E}(y|z)$ is a weighted sum of y where the weights are some nonlinear function of z depending on K_2 and h . However, if

we had used the approximate parametric specification $E(y|z) = z\beta$ of the econometrician in (3.2) we would get (for $p=1$) the least squares (LS) estimator as

$$(3.6) \quad \tilde{E}(y|z) = z\tilde{\beta} = z(Z'Z)^{-1}Z'y = a(z)y = \sum_{t=1}^n y_t a_t(z)$$

where $a(z) = z(Z'Z)^{-1}Z'y$ is a $1 \times n$ vector and Z is an $n \times q$ matrix of observations on q variable vector z .

Remark 3.2. Note that the nonparametric estimator $\hat{E}(y|z)$ in (3.3) is obtained without making assumptions about the functional form and the DGP of y and z . We hope that the nonparametric residuals

$$(3.7) \quad y_t - \hat{E}(y_t|z_t) = \hat{u}_t, \quad t=1, \dots, n$$

are robust and they can be used to perform meaningful diagnostic checkups for serial correlation, heteroscedasticity and simultaneity among others. In addition, we also hope that the comparison of the nonparametric residual sum of squares $\sum \hat{u}_t^2$ with the parametric residual sum of squares $\sum u_t^2$, $u_t = y_t - \tilde{E}(y_t|z_t)$ from (3.6), will indicate whether or not the parametric model specified by the econometrician is well fitted. Some work on this has been attempted in Ullah (1985).

Remark 3.3. The expression in (3.6) can be used to estimate econometric parameters of interest. For example, the response coefficient of y with respect to z (suppose $p=1$) can be obtained by calculating $\frac{\partial}{\partial z} \tilde{E}(y|z)$. This nonparametric estimate is useful

since it does not use any assumption about the functional form, see e.g., Ullah and Vinod (1985).

4. PROOFS OF THEOREMS IN SECTION 2

Proof of Theorem 2.1: With $B_n = (\hat{E}f - f)$, our aim is to prove that

$$(4.1) \quad \lim_n B_n(x) = 0$$

at every continuity point of f .

From (2.3), $E\hat{f} = n^{-1} \sum_{t=1}^n E\delta_t$. We first consider the behavior

of $E\delta_t(x)$ as $t \rightarrow \infty$. Since $\int K=1$, $a_t(x) = (E\delta_t(x) - f(x)) = \int K(w)(f(x+h_t w) - f(x))dw$. Splitting the region of integration into two parts $\|h_t w\| \leq \eta$ and $\|h_t w\| > \eta$ for an $\eta > 0$, we have

$$(4.2) \quad |a_t(x)| \leq \sup_{\|h_t w\| < \eta} |f(x+h_t w) - f(x)| \int |K(w)| dw \\ + \int_{\|u\| > \eta} h_t^{-m} |K(h_t^{-1} u) f(x+u) du + f(x) \int_{\|w\| > \eta h_t^{-1}} |K(w)| dw$$

Since x is a point of continuity of f , for any $\epsilon > 0$, we can choose $\eta = \eta(\epsilon) > 0$ so that the first term on the right-hand side of (4.2) is no more than $\epsilon/3$. Since $\int f(x+u)du = \int f=1$, the second term is no more than $\eta^{-m} \sup_{\|u\| > \eta} \{(h_t^{-1} \|u\|)^m |K(h_t^{-1} u)|\}$ which, in turn, is bounded by $\eta^{-m} \sup_{\|u\| > \eta h_t^{-1}} \|u\|^m |K(u)|$,

but this tends to zero as $t \rightarrow \infty$ by (2.1). The third term also tends to zero as $h_t^{-1} \rightarrow \infty$ or $t \rightarrow \infty$ since the region $\|w\| > \eta h_t^{-1}$ tends to empty set as $t \rightarrow \infty$. We thus conclude that $a_t(x) \rightarrow 0$ as

$t \rightarrow \infty$, which in turn implies $B_n(x) = n^{-1} \sum_{t=1}^n a_t(x) \rightarrow 0$ as $n \rightarrow \infty$.

Proof of Theorem 2.2: Since δ_j are independent,

$$\text{var}(\hat{f}_n(x)) = n^{-2} \sum_{t=1}^n \{E\delta_t^2(x) - (E\delta_t(x))^2\}. \text{ But from the proof}$$

$$\text{of Theorem 2.1, } E\delta_t^2(x) = h_t^{-m} \int K^2(y) f(x+h_t y) dy =$$

$h_t^{-m} f(x) \int K^2(y) dy + o(1)$ and $E\delta_t(x) = f(x) + o(1)$ as $t \rightarrow \infty$. Hence (2.6) follows.

Proof of Theorem 2.3: Since $\hat{f} = n^{-1} \sum_{t=1}^n \delta_t$ and $\delta_t \leq h_t^{-m} M$

where $M = \sup_w |K(w)|$, Bernstein's⁵ inequality given in

(2.13) of Hoeffding (1963) gives, for any $\epsilon > 0$,

$$(4.3) \quad P[|\hat{f} - Ef| > \epsilon] \leq 2 \exp \left[-\frac{3}{2} n \epsilon^2 / (3n \text{var}(\hat{f}) + h_n^{-m} \epsilon M) \right]$$

But from Theorem 2.2, $\text{var}(\hat{f}(x)) = (n^{-2} \sum_{t=1}^n h_t^{-m})(f(x) \int K^2 + o(1))$

$\leq (n h_n^m)^{-1} (f(x) \int K^2 + o(1))$. Therefore, for some positive constant c , the rhs of (4.3) is no more than $2 \exp(-c \epsilon n h_n^m)$.

Thus for every $\epsilon > 0$, $\sum_{n=1}^{\infty} P[|\hat{f}(x) - Ef(x)| > \epsilon] < \infty$ by (2.8)**.

Hence by Borel-Cantelli lemma⁶ $|\hat{f}(x) - Ef(x)| = o(1)$ with probability one. Now the proof is complete in view of Theorem 2.1.

Proof of Theorem 2.4: Denoting $n^{-1}(\delta_t(x) - E\delta_t(x)) / (\text{var} \hat{f}(x))^{1/2}$ by L_{nt} , we see that $(\hat{f}(x) - Ef(x)) / (\text{var} \hat{f}(x))^{1/2}$ is $S = \sum_{t=1}^n L_{nt}$, a sum of n independent

centered random variables L_{nt} with $\text{var}(S) = \sum_{t=1}^n \text{var}(L_{nt}) = 1$.

Therefore, by Berry-Essen theorem (see Theorem 7.4.1 of Chung (1968)), the lhs of (2.9) is

$$(4.4) \quad \sup_{\xi \in \mathbb{R}} \frac{1}{n} |P[S \leq \xi] - \Phi(\xi)| \leq c \sum_{t=1}^n E|L_{nt}|^3$$

where c is a constant. To complete the proof we have to examine the behavior of $E|L_{nt}|^3$ for each t .

By C_Y -inequality (Loeve (1963), p. 155), $E|n^{-1}(\delta_t(x) - E\delta_t(x))|^3 \leq 4n^{-3} E|\delta_t(x)|^3$ which is by (2.1),
 $4(n^3 h_t^{2m})^{-1} \int |K(w)|^3 f(x+h_t w) dw = 4(n^3 h_t^{2m})^{-1} [f(x) \int |K|^3 + o(1)]$ as $t \rightarrow \infty$, by the arguments used in the proof of Theorem 2.1. Since by Theorem 2.2,

$$\text{var}(f_n(x)) = (n^{-2} \sum_{t=1}^n h_t^{-m}) [f(x) \int K^2 + o(1)], \text{ we get } \sum_{t=1}^n E |L_{nt}|^3 =$$

$$\epsilon_n (f(x))^{-1/2} \{ \int |K|^3 / (\int |K|^2)^{3/2} \} + o(\epsilon_n) = o(\epsilon_n) \text{ where } \epsilon_n \text{ is as}$$

defined in (2.9). The proof of the theorem is now complete.

FOOTNOTES

¹Throughout the paper the multivariate density function will be referred to as the multivariate data generating process.

²A density function f (of a random variable X) at a point x is defined by $f(x) = \lim_{h \rightarrow 0} h^{-1} P[x - (h/2) < X < x + (h/2)] =$

$$\lim_{h \rightarrow 0} h^{-1} EI(-\frac{1}{2} < \frac{X-x}{h} < \frac{1}{2}) \quad I(a < X < b) \text{ is 1 or 0 according as } X \text{ is}$$

in the interval (a, b) or outside this interval. If $X_1, \dots, X_n$

are i.i.d. observations, then $n^{-1} \sum_{j=1}^n I(-\frac{1}{2} < \frac{X_j - x}{h} < \frac{1}{2})$ is a

consistent (as $n \rightarrow \infty$) estimator of $EI(-\frac{1}{2} < \frac{X-x}{h} < \frac{1}{2})$. Hence if h

is such that $h \rightarrow 0$ whenever $n \rightarrow \infty$, $f(x) = \lim_{h \rightarrow 0} h^{-1} EI(-\frac{1}{2} < \frac{X-x}{h} < \frac{1}{2})$ is

consistently estimated by $\alpha_n(x) = h^{-1} n^{-1} \sum_{j=1}^n I(-\frac{1}{2} < \frac{X_j - x}{h} < \frac{1}{2})$.

Rosenblatt (1956), while proving various properties of α_n , noted that if the indicator function $I(\cdot) = I(-\frac{1}{2} < \cdot < \frac{1}{2})$ in definition of α_n is replaced by any non-negative function $K(\cdot)$ such that $\int K(t) dt = 1$, the consistency property of α_n remains invariant. For this reason he considered estimating $f(x)$ by a

(more general) estimator $h^{-1} n^{-1} \sum_{j=1}^n K(\frac{X_j - x}{h})$, and called it a

kernel estimator as for each K we have one estimator. Parzen (1962) considered even more general estimator by dropping Rosenblatt's "non-negativity" condition on K .

³ A sequence of r.v.'s $\{L_n\}$ is said to converge to l in probability if for every $\epsilon > 0$, $\lim_{n \rightarrow \infty} P[|L_n - l| > \epsilon] = 0$, it is

said to converge l with probability one if $P[\lim_{n \rightarrow \infty} L_n = l] = 1$.

The latter is stronger than former one (see Rao (1973), p. 110).

⁴ For $\{L_n, T_n\}$, $n=1,2,\dots$ of pairs of random variables, if

$L_n \xrightarrow{P} l$, a constant and $T_n \xrightarrow{D} T$, a r.v., where "P" and "D" stand

for convergence in probability and in distribution, respectively, then $(T_n + L_n) \xrightarrow{D} (T + l)$ (Rao (1973), p. 122).

⁵ (Bernstein's inequality, Hoeffding (1963).) Let w_t be independent random variables and for positive b , $w_j \leq b$.

Then, with $w = \sum_{t=1}^n w_t$, for every $\xi > 0$, $P[|w - Ew| > \xi] \leq$

$2 \exp[-3n\xi^2 / \{2(3\sigma^2 + b\xi)\}]$, where $\sigma^2 = \text{var}(w)$.

⁶ Borel-Cantelli lemma. For a sequence of events $\langle A_n \rangle$, if $\sum_{n=1}^{\infty} P(A_n) < \infty$ then with probability one, the event A_n happens

only for finitely many n .

ACKNOWLEDGEMENT

Major work of this research was conducted during the first author's one month (March-April, 1984) visit to the University of Western Ontario. The authors are grateful to NSERC for the research supports, and to Virginia Ho for the research assistance.

REFERENCES

- Bartlett, M.S. (1963), "Statistical Estimation of Density Function," *Sankhya A*, 25: 245-254.
- Billingsley, P. (1968), *Convergence of Probability Measures* [New York: John Wiley].
- Cacoullos (1966), "Estimation of a Multivariate Density," *Ann. Ins. Statist. Math.* 18: 179-189.
- Chung, K.L. (1968), *A Course in Probability Theory* [Brace and World, Inc. Harcourt].
- Deheuvels, P. (1974), "Estimation Sequentielle de la densite, Universite de Paris VI, 1974), "Conditions necessaires et suffisantes de convergence ponctuelle presque sure et uniforme presque sure des estimateurs de la densite," *Comptes Rendus de L'Academic des Sciences de Paris*, Vol. 278, pp. 1217-1220.
- Hoeffding, W. (1963), "Probability Inequalities for Sums of Bonded Random Variables," *J. Amer. Statist. Assoc.* 58: 13-30.
- Isogal, E. (1980), "Strong consistency and optimality of a sequential density estimator," *Bull. Math. Statist.* 19: 55-69.
- Kronmal R. and M. Tarter (1968), "The Estimation of Probability Densities and Cumulatives by Fourier Series Methods," *J. Amer. Statist. Assoc.* 38: 482-93.
- Loeve, M. (1963), *Probability Theory*, 3rd edition [Princeton, N.J.: Van Nostrand].
- Menon, V.V., V. Prasad, and R.S. Singh (1984), "Nonparametric Recursive Estimates of A Probability Density Function and Its Derivatives," *J.S.P.E.*, pp. 73-82.
- Moore, D.S. and J.S. Yackel (1977), "Consistency Properties of Nearest Neighbor Density Estimates," *Ann. Statist.* 5: 143-154.
- Parzen, E. (1962), "On the Estimation of Probability Density and Mode," *Ann. Math. Statist.* 33: 1065-1076.
- Prasad, B. and R.S. Singh, "Nonparametric Kernel Estimates of a Density Function Along With Its Derivatives," *Colloquia Mathematica Societatis, Janos, Bolyai, Nonparametric Statistical Inference Budapest*, 1980.

- Rao, C.R. (1973), *Linear Statistical Inference and Its Applications* [New York: John Wiley].
- Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of Density Function," *Ann. Math. Statist.* 27: 832-837.
- Schwartz, S.C. (1967), "Estimation of a Probability Density By An Orthogonal Series," *Ann. Math. Statist.* 38: 1261-1265.
- Singh, R.S. (1977a), "Improvement on Some Known Nonparametric Uniformly Consistent Estimators of Derivatives of a Density," *Ann. Statist.* 5: 394-399.
- Singh, R.S. (1977b), "Applications of Estimators of a Density and Its Derivatives to Certain Statistical Problems," *J. Roy Statist. Soc., Series B*, 39: 357-363.
- Singh, R.S. (1978), "Nonparametric Estimation of Mixed Partial Derivatives of a Multivariate Density," *J. Multiv. Anal.* 6: 555-567.
- Singh, R.S. (1979), "Mean Squared Errors of Estimates of a Density and Its Derivatives," *Biometrika* 66: 177-180.
- Singh, R.S. (1980), "Speed of Convergence in Nonparametric Estimation of a Multivariate μ -Density and Its Mixed Partial Derivatives," *J. Statist. Planning Inf.* 5: 287-298.
- Singh, R.S., A. Ullah and R.A.L. Carter (1985), "Nonparametric Inference in Econometrics: New Applications" to appear in *Symposia in Statistics*, Festschrift in honor of V.M. Joshi, Reidel Press.
- Singh, R.S. and A. Ullah (1984), "Estimation of Multivariate DGP, University of Western Ontario manuscript.
- Ullah, A. (1985), "Specification Analysis of Econometric Models," *Journal of Quantitative Economics*, 1: 187-209.
- Ullah A. and R.S. Singh (1984), "Estimation of Density and Its Applications in Econometrics, Technical Report, C.D.S.E., University of Western Ontario.
- Ullah, A. and H.D. Vinod (1985), "Nonparametric Kernel Estimation of Econometric Parameters," University of Western Ontario manuscript.
- Van Ryzin, J. (1973), "On Histogram Method of Density Estimation," *Comm. Statist.* 2: 493-506.

- Watson, G.S. (1969), "Density estimation by orthogonal series," *Ann. Math. Statist.* 40: 1496-1498.
- Wegman, E.J. (1970, "Maximum Likelihood estimation of a unimodel density function," *Ann. Math. Statist.* 41: 143-161.
- Wolverton, C.T. and T.J. Wagner (1969), "Asymptotically Optimal Discriminant Functions for Pattern Classification," *IEEE Trans. Information Theory*, IT-15: 258-265.
- Woodrooffe, M. (1970), Discussion on Rosenblatt's Paper, *Nonparametric Techniques in Statistical Inference* (M.L. Puri, ed.), pp. 199-210 [Cambridge University Press].
- Yamato, H. (1971), "Sequential Estimation of a Continuous Probability Density Function and Mode," *Bull. Math. Statist.* 14: 1-23.