The Annals of Statistics
1982, Vol. 10, No. 4, 1125–1136

NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION OF SPATIAL PATTERNS[1]

BY THOMAS W. SAGER

University of Texas at Austin

Let $\mathbf{X}$ be an absolutely continuous random variable in $\mathbb{R}^k$ with distribution function $F(\mathbf{x})$ and density $f(\mathbf{x})$. Let $\mathbf{X}_1, \cdots, \mathbf{X}_n$ be independent random variables distributed according to F. Mapping the spatial distribution of $\mathbf{X}$ normally entails drawing a map of the isopleths, or level curves, of f. In this paper, it is shown how to map the isopleths of f nonparametrically according to the criterion of maximum likelihood. The procedure involves specification of a class $\mathscr{L}$ of sets whose boundaries constitute admissible isopleths and then maximizing the likelihood $\prod_{i=1}^n g(\mathbf{x}_i)$ over all g whose isopleths are boundaries of $\mathscr{L}$-sets. The only restrictions on $\mathscr{L}$ are that it be a σ-lattice and an F-uniformity class. The computation of the estimate is normally straightforward and easy. Extension is made to the important case where $\mathscr{L}$ may be data-dependent up to locational and/or rotational translations. Strong consistency of the estimator is shown in the most general case.

1. Introduction. Let $F(\mathbf{x})$ be the distribution function of an absolutely continuous random variable $\mathbf{X}$ on $\mathbb{R}^k$, $k = 1, 2, 3, \cdots$, with density $f(\mathbf{x})$. Let $\mathbf{X}_1, \cdots, \mathbf{X}_n$ be independent random variables distributed according to F. It is often of interest, particularly when $k = 2$, to map the distribution of $\mathbf{X}$ by drawing a series of isopleths (level curves) of f. A well-constructed map will convey at a glance the pattern of geographic variation in concentrations of the phenomenon $\mathbf{X}$ under study. But few maps are drawn from complete knowledge of the true density f and would be needlessly detailed for many purposes, if they were. Instead, the pattern of geographic variation is sampled. We may distinguish two different sampling schemes at opposite ends of a spectrum of possibilities. On the one hand, the data may consist of a fixed and small number of direct samples $f_1, \cdots, f_n$ of the density taken with error at corresponding fixed sampling points $\mathbf{x}_1, \cdots, \mathbf{x}_n$. This kind of problem is called the *interpolation problem* because the goal is to estimate f at nonsampling points. An example is the monitoring of oxidant pollution from 30-odd fixed site monitoring stations in the Bay Area counties of California. The pollutant values f, rather than the monitoring sites $\mathbf{x}$, exhibit sampling variation. The theoretical framework of this paper does not apply to the interpolation problem. On the other hand, the data may consist of a large number of observations $\mathbf{x}_1, \cdots, \mathbf{x}_n$ on random geographic coordinates $\mathbf{X}$. In this case, the $\mathbf{x}$-data are the geographic locations where random phenomena were observed to occur. The geographic distribution of people stricken with lung cancer is an example of this kind of problem, which we call the *density estimation problem* and which is the subject of this paper. In the applications, f need not be restricted to densities but may encompass other kinds of surfaces which can be suitably normalized to densities. Figures 1 and 2 show estimates of a bivariate density computed by the methods of this paper.

In the old days, after the data had been collected, the statistician or cartographer would prepare a map by hand-drawing a few isopleths. More recently, the statistician-cartographer might use one of the available computer graphics packages to draw his map. However, few of these packages have paid much attention to traditional statistical criteria and many

Received October 1979; revised May 1982.

[1] Prepared under support from the Department of Energy (DOE), Rockefeller Foundation, Sloan Foundation, Environmental Protection Agency (EPA), and the National Science Foundation (NSF).

AMS 1980 *subject classification*. Primary 62G05, secondary 60F15, 62H99.

Key words and phrases. Density estimation, isopleth, nonparametric estimation, consistency, isotonic regression, maximum likelihood estimation, σ-lattice, P-uniformity.

use ad hoc fitting methods with unknown statistical properties. Without theoretical analysis of the methods, only tentative judgments can be rendered on their adequacy even though they may produce nice-looking maps in many instances.

However, in recent years more statisticians have turned their attention to redressing the situation. Working in the interface between theory and applications and well-qualified in each, these statisticians are starting to provide effective and efficient mapping techniques which are also firmly rooted in theory. Much of the serious effort to automate cartography is found in the earth sciences, geology, and geography. Davis and Cullagh (1975) collect a series of papers with extensive bibliographies where one may find early papers on theoretical aspects of spatial analysis. Mathéron (1973) provides a theoretical foundation for Kriging, a technique used in oil exploration and gold mining. Much of this literature treats what we have called the interpolation problem. Turning to the density estimation problem, we find a number of instances of effective and theoretically rooted techniques. Boneva, Kendall, and Stefanov (1971) and a long series of papers by Wahba on spline techniques, Tarter and Kronmal (1976) on Fourier series methods, and Silverman (1978) on kernel estimates are examples of contributions to their respective density estimation methods in which practical exigencies and theory blend felicitously. Tapia and Thompson (1978) perform the same service for maximum penalized likelihood estimation and, in addition, provide the most recent and most thorough review yet in print of density estimation methods.

Before collecting any data, the statistician often has in mind a family of curves from which he would be willing to select his isopleths. We assume that to be the case in this paper. However, instead of focusing our attention on the curves, we shall concentrate equivalently on the sets bounded by those curves. Let $\mathscr{L}$ denote this class of sets. Since an isopleth is a level curve $\{\mathbf{x}, f(\mathbf{x}) = c\}$ which locally divides the region $\{f > c\}$ from the region $\{f < c\}$, we call sets of the form $\{\mathbf{x}; f(\mathbf{x}) \geq c\}$ modal regions of f. Restricting the statistician's choice of modal regions [isopleths] to $\mathscr{L}$ [boundaries of $\mathscr{L}$-sets] means that any estimate g of f should satisfy $\{\mathbf{x}; g(\mathbf{x}) \geq c\} \in \mathscr{L}$ for all c. We say that g is $\mathscr{L}$-measurable if and only if this condition is satisfied. The fitting criterion we propose is maximum likelihood: Given $\mathscr{L}$ and observations $\mathbf{X}_1 = \mathbf{x}_1, \cdots, \mathbf{X}_n = \mathbf{x}_n$, we say that an $\mathscr{L}$-measurable function $\hat{f}_n$ is an $\mathscr{L}$-maximum likelihood estimate ($\mathscr{L}$-mle) if and only if

$$\prod_{i=1}^{n} \hat{f}_n(\mathbf{x}_i) = \max \{\prod_{i=1}^{n} g(\mathbf{x}_i); g \text{ is } \mathscr{L}\text{-measurable}\} < \infty.$$

$\mathscr{L}$ plays the same role in this conceptualization of maximum likelihood estimation as the more usual specification of a (parameterized) family of densities $\mathscr{F}$ over which $\prod_{i=1}^{n} g(\mathbf{x}_i)$, $g \in \mathscr{F}$, is to be maximized. Sager (1979) considers a different but somewhat related criterion for fitting convex isopleths.

In this paper we shall study the statistical properties of $\hat{f}_n$ under two restrictions on $\mathscr{L}$. The first is that $\mathscr{L}$ be a complete σ-lattice of Borel or Lebesgue sets and the second is that $\mathscr{L}$ be an F-uniformity class. The first restriction enables us to use the tools of isotonic regression and the second facilitates consistency arguments.

DEFINITION 1. $\mathscr{L}$ is a complete σ-lattice if and only if $\mathscr{L}$ contains the empty set and the whole space and is closed under arbitrary union and intersection. (We also require measurability for the members of $\mathscr{L}$.)

DEFINITION 2. A class $\mathscr{A}$ of measurable sets is a P-uniformity class if and only if $\sup_{P_n \Rightarrow P} \lim \sup_{n\to\infty} \sup_{A \in \mathscr{A}} | P_n(A) - P(A) | = 0$, where the first sup is over all measures P_n converging weakly to P.

The framework presented here lends itself readily to nonparametric maximum likelihood estimation of a unimodal univariate density (see Examples 4.1 and 4.2). The univariate mle for the latter problem was obtained and consistency shown by Robertson (1967) in the case of known mode and by Wegman (1969) in the case of unknown mode.

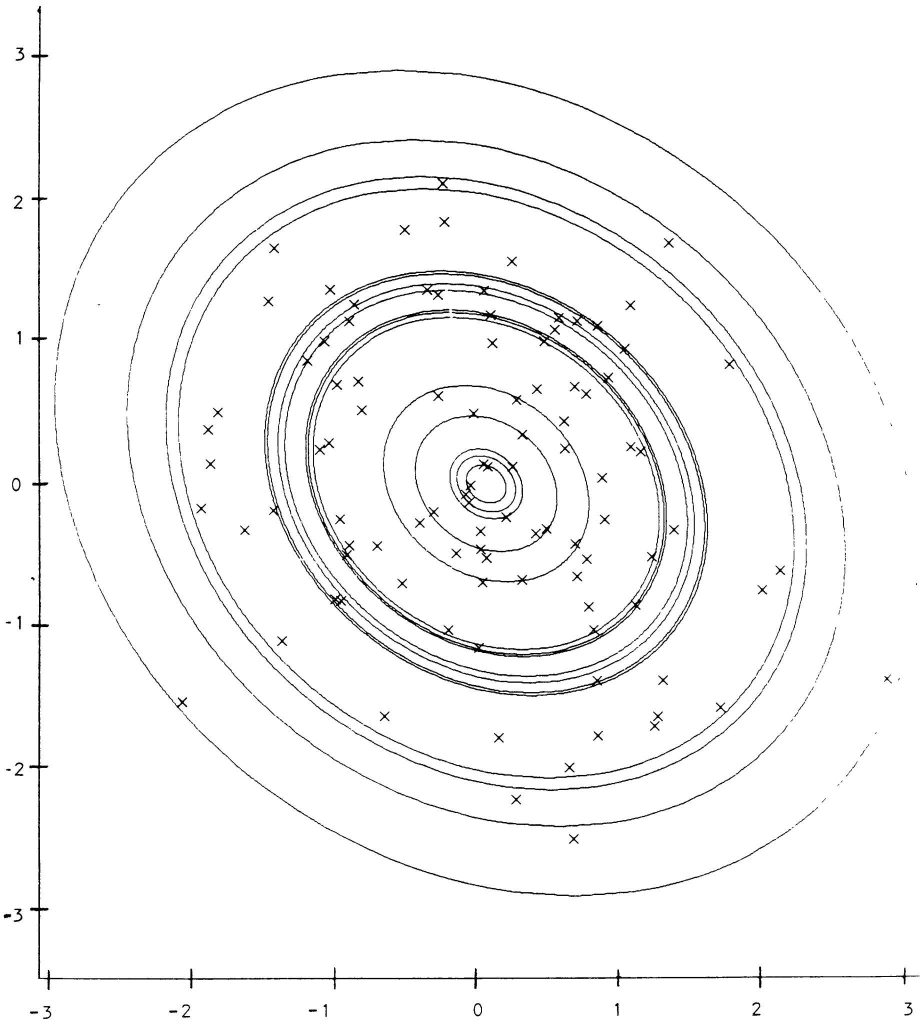

FIG. 1. *Estimated density, using elliptical contours, for a sample of* $n = 100$ *independent* N (0, 1) *pairs. Estimated values on the* 16 *contours*: .510, .289, .288, .128, .110, .090, .078, .077, .077, .072, .061, .054, .038, .009, .005, .002.

Additionally, Robertson (1967) provided a theoretical representation for the mle with respect to a σ-lattice measurable density in general measure spaces. However, he did not provide consistency arguments except in two special cases, nor did he consider the problem of an unknown mode (i.e., data-dependent $\mathscr{L}$). This paper addresses these issues, strengthens and generalizes Robertson's and Wegman's results.

In practice, the class of isopleth shapes available to the cartographer-statistician is more likely to be limited by the requirement that $\mathscr{L}$ be a σ-lattice than by the F-uniformity stricture. Topsøe (1970) presents very powerful methods for ascertaining when a given class is an F-uniformity class. The interested reader is referred to Topsøe's article for details and examples. As noted in Section 3, the full power of F-uniformity is not required, but for absolutely continuous F, not much is gained by relaxing this restriction.

2. The general framework. Let $(\Omega, \mathscr{B}, \mu)$ be a totally finite measure space and let $\mathscr{L}$ be a complete σ-lattice of subsets of Ω. Let $\omega_1, \cdots, \omega_n$ be a fixed set of points in Ω. Let $\sigma(\mathscr{S})$ be the smallest σ-lattice containing $\mathscr{S}$, for any class of sets $\mathscr{S}$. In the usual application

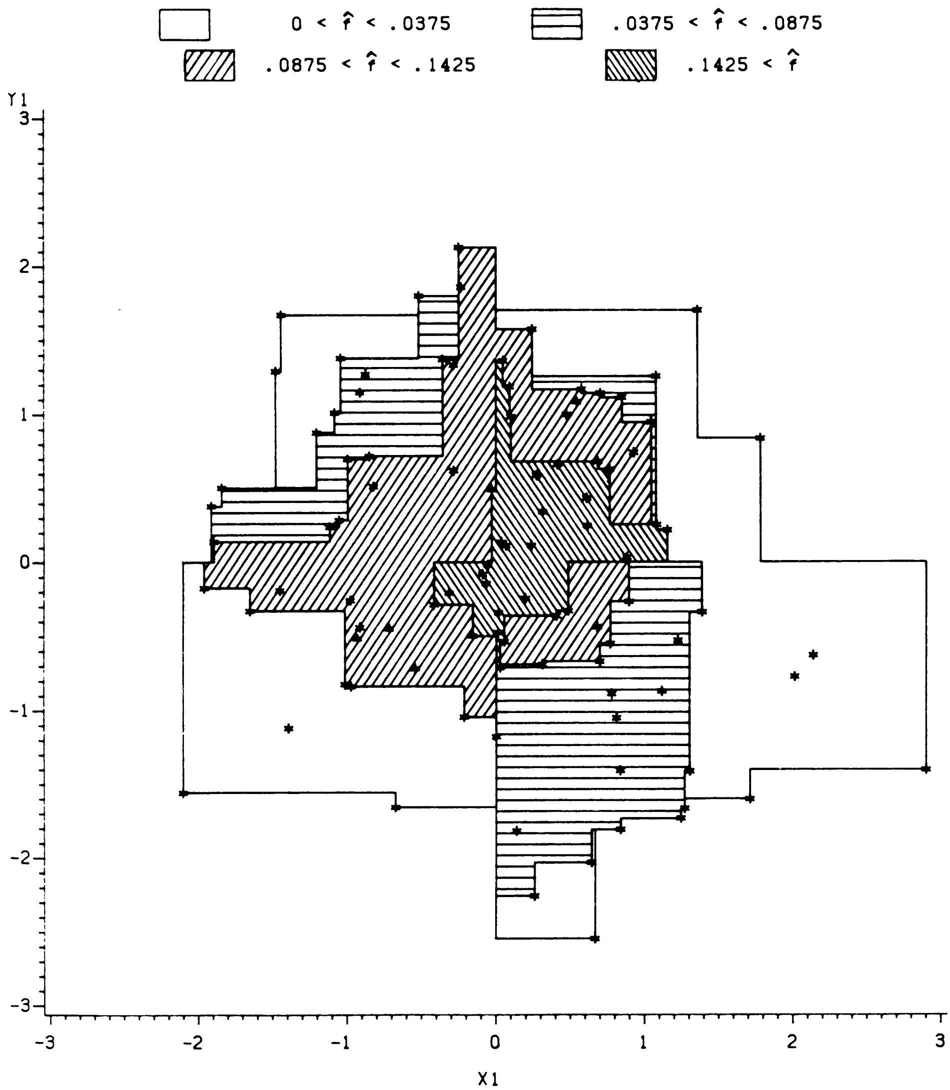

FIG. 2. *Estimated density, using sigma-lattice of Example* 4.4, *for a sample of* $n = 100$ *independent* N (0, 1) *pairs. Note*: *the same sample was used in Figure* 1. *Level sets have been consolidated into four.*

of the general framework, Ω will become $\mathbb{R}^k$, μ will denote Lebesgue measure, and $\omega_1, \cdots, \omega_n$ will be realizations $\mathbf{x}_1, \cdots, \mathbf{x}_n$ of $\mathbf{X}_1, \cdots, \mathbf{X}_n$. Figure 1 illustrates the setting in the plane. There, the σ-lattice is the class of ellipses centered and oriented by estimated values (see Example 4.5). The solution shown maximizes the likelihood of the data over the class of all densities with elliptical contours. The general $\mathscr{L}$-mle may be obtained as a solution to the following maximization problem:

(2.1) PROBLEM. Maximize $\prod_{i=1}^{n} g(\omega_i)$ subject to $g \geq 0$, $\int g \, d\mu = 1$, and g being $\mathscr{L}$-measurable.

Calculation of the solution $\hat{f}_n$ to this problem may be simplified by essentially replacing the class $\mathscr{L}$ by the readily identifiable σ-lattice $\sigma\{L(\omega_1), \cdots, L(\omega_n)\}$, where $L(\omega_i) = \cap\{L;\ \omega_i \in L \in \mathscr{L}\}$. This σ-lattice may then be iteratively modified and reduced by the maximum upper sets algorithm (or minimum lower sets algorithm), which is the most general of the available isotonic regression algorithms (see Barlow, et al, 1972):

1. Enter $\mathscr{L}^* = \sigma\{L(\omega_1), \cdots, L(\omega_n)\}$
2. Select $L' \in \mathscr{L}^*$ such that $F_n(L')/\mu(L') = \max_{L\in\mathscr{L}^*} F_n(L)/\mu(L)$
3. Set $\hat{f}_n(x) = F_n(L')/\mu(L')$ for all $\mathbf{x} \in L'$
4. Set $\mathscr{L}^* = \sigma\{L(\omega_1) - L', \cdots, L(\omega_n) - L'\}$
5. If $\mathscr{L}^*$ is empty, exit; otherwise, return to step 2. □

However, it is usually better to employ a specialized algorithm in a concrete problem in order to speed computer execution time. For example, the pool-adjacent-violators algorithm (see Barlow, et al, 1972) was used to produce Figure 1 because preliminary processing can reduce the σ-lattice to a linear order, for which the pool-adjacent-violators algorithm is much more efficient than the minimum lower sets algorithm. The σ-lattice for Figure 2 was not based on a linear order; a matrix-smoothing algorithm suggested by Professor Richard Dykstra (personal communication) was adapted. Although computation times will depend on the data and on the computer, a user could anticipate that estimates like Figure 1 could be obtained in a few seconds (or less) and in a few minutes for Figure 2 on a modern main-frame computer. Fortran programs for these algorithms are available from the author.

Solution. Let $L(\omega_i) = \cap\{L; \omega_i \in L \in \mathscr{L}\}$. First consider the above problem with the additional restriction that g be a simple function on $\mathscr{S} = \{A_1, \cdots, A_m\}$, the maximal disjoint base of nonempty sets for $\sigma\{L(\omega_1), \cdots, L(\omega_n)\}$ of which every such disjoint base is a refinement. Now a simple function on $\mathscr{S}$ is $\mathscr{L}$-measurable if and only if it is isotonic with respect to the quasi order $\lesssim$ on $\mathscr{S}$ defined by

$$A_i \lesssim A_j \Leftrightarrow \forall L \in \mathscr{L} \cap \sigma(\mathscr{S}) \text{ such that } A_i \subset L, \text{ also } A_j \subset L.$$

For if the simple g is $\mathscr{L}$-measurable and $A_i \lesssim A_j$, then

$$A_j \subset \{\omega; g(\omega) \geq g(A_i)\} \in \mathscr{L} \cap \sigma(\mathscr{S}) \Rightarrow g(A_i) \leq g(A_j).$$

And if the simple g is isotonic, then observing that $g \geq g(A_i)$ on $\cap\{L \in \mathscr{L} \cap \sigma(\mathscr{S}); A_i \subset L\}$, we see that

$$\{\omega; g(\omega) \geq c\} = \cup_{\{A_i; g(A_i) \geq c\}} \cap\{L \in \mathscr{L} \cap \sigma(\mathscr{S}); A_i \subset L\} \in \mathscr{L}.$$

Then when $\mu(A_i) > 0$ for all i, the modified problem may be posed as:

$$\text{minimize} \quad \textstyle\sum_{i=1}^{n} \dfrac{F_n(A_i)}{\mu(A_i)} \cdot \mu(A_i) \log g_i$$

subject to $g_i \geq 0$, $\sum_{i=1}^{m} g_i \cdot \mu(A_i) = 1$ and $g_i \lesssim g_j$ iff $A_i \lesssim A_j$.

(Here, F_n denotes empirical measure.) By Example 1.10, page 45–46 of Barlow, et al (1972), the solution to this modified problem is given by the isotonic regression of $\{F_n(A_i)/n\mu(A_i);\ i = 1, \cdots, n\}$ with weights $\mu(A_i)$, $i = 1, \cdots, n$ with respect to the quasi order $\lesssim$. The solution may be implemented by the minimum lower sets algorithm or other algorithms mentioned above. To show that this also solves the general problem, let h be any $\mathscr{L}$-measurable function and let $\xi_1, \cdots, \xi_n$ be the rearranged values of $\omega_1, \cdots, \omega_n$ such that $h(\xi_1) \leq h(\xi_2) \leq \cdots \leq h(\xi_n)$. If we define

$$h^*(\omega) = \begin{cases} c \cdot h(\xi_i) & \text{if} \quad \omega \in L(\xi_i) - \cup_{j=i+1}^{n} L(\xi_j), \qquad i = 1, \cdots, n-1 \\ c \cdot h(\xi_n) & \text{if} \quad \omega \in L(\xi_n) \\ 0 & \text{otherwise} \end{cases}$$

where c is a normalization constant, then h^* is $\mathscr{L}$-measurable, simple on $\mathscr{S}$ and

$$\textstyle\prod_{i=1}^{n} h^*(\omega_i) = c^n \prod_{i=1}^{n} h(\omega_i) \geq \prod_{i=1}^{n} h(\omega_i)$$

since

$$1 = c \sum_{i=1}^{n-1} h(\xi_i) \cdot \mu(L(\xi_i) - \cup_{j=i+1}^{n} L(\xi_j)) + c \cdot h(\xi_n) \cdot \mu(L(\xi_n))$$
$$\leq c \int_\Omega h(\omega)\, d\mu = c. \qquad \square$$

Note that the argument depends on $0 < \mu(A_i) < \infty$, $\forall i$. Whenever $\mu(A_i) = 0$ and $F_n(A_i) = 0$, delete the set A_i from the space Ω for the purposes of calculating the mle. The mle may then be extended to A_i without affecting the likelihood. The case $\mu(A_i) = 0$ and $F_n(A_i) > 0$ is discussed in Section 3. In case $(\Omega, \mathscr{B}, \mu)$ is not totally finite (e.g., $\mathbb{R}^k$), restrict the space to $\cup_{i=1}^{n} L(\omega_i)$, as in Section 5 of Robertson (1967).

The solution given here, although based on ideas of Robertson (1967), is somewhat more general in that it does not use his assumption (ii), page 484. In fact, $\mathscr{L}$ is freed from any connection with any quasi order on Ω (although such an order may always be defined after being given $\mathscr{L}$). This allows us to choose the σ-lattice of shapes for our isopleths without need for generating them from a more or less arbitrary ordering.

3. Existence and consistency of the MLE. Suppose f is a bivariate density and the statistician believes that the isopleths of f are nested ellipses of a given eccentricity with common center and major and minor axes. If the center and orientation of the axes are known, the method of Section 2 may be used to calculate the mle. This knowledge may often be available, particularly if the center is an identifiable source (like a factory smokestack) which distributes a variable in a topographically channeled manner. However, the knowledge is frequently not available. In the latter case, the statistician may still use the method of Section 2 provided he is willing to estimate the center and orientation of the nested ellipses (see Example 4.5 which also discusses estimating the eccentricity).

In general, the statistician may have in mind a basic class of sets $\mathscr{L}$ which he can specify up to a location and/or rotation parameter. We allow the statistician to estimate these parameters from the data before calculating the mle based on the translation of $\mathscr{L}$ according to those estimates. Any such location/rotation translation of members of $\mathscr{L}$ is rigid and preserves their Lebesgue measure. In the sequel, we shall suppose without loss of generality that the location and orientation of $\mathscr{L}$ are correct ones so that f is $\mathscr{L}$-measurable. The error made by the statistician in estimating the vector of location/rotation parameters is $\mathbf{e}_n$. Thus $\hat{f}_n$ is calculated with respect to the $\mathbf{e}_n$-translation of $\mathscr{L}$. Other notation used in this section is as follows:

μ	Lebesgue measure
$\mathscr{L}^{\mathbf{e}_n}$	the $\mathbf{e}_n$-location/rotation translation of $\mathscr{L}$
$A^{\mathbf{e}_n}$	the $\mathbf{e}_n$-translation of the set A
$F^{\mathbf{e}_n}$	the measure $F^{\mathbf{e}_n}(A) = F(A^{-\mathbf{e}_n})$
$f^{\mathbf{e}_n}$	the density of $F^{\mathbf{e}_n}$
$F_n^{\mathbf{e}_n}$	the empiric measure $F_n^{\mathbf{e}_n}(A) = F_n(A^{-\mathbf{e}_n})$
f_n	the formal density of $F_n: f_n(\mathbf{x}) = n^{-1}/\mu(A_i)$ if $\mathbf{x} \in A_i$ where $\mathscr{S} = \{A_1, \cdots, A_m\}$ is given in Section 2
$\hat{F}_n$	the measure $\hat{F}_n(A) = \int_A \hat{f}_n\, d\mu$
$T_a g$	the set $\{\mathbf{x}; g(\mathbf{x}) \geq a\}$ for any function g
$P_a g$	the set $\{\mathbf{x}; g(\mathbf{x}) > a\}$ for any function g

The large sample properties of $\hat{f}_n$ will depend upon the accuracy of the estimated translation parameters. We shall suppose that those estimates are consistent so that the error $\mathbf{e}_n \to 0$ almost surely. However, it is of interest to know what happens when the estimates converge to the wrong value, so the main theorem will be presented in the more general setting $\mathbf{e}_n \to \mathbf{e}$. Let $h_{\mathbf{e}} = E[f \mid \mathscr{L}^{\mathbf{e}}]$, the conditional expectation of f with respect to the σ-lattice $\mathscr{L}^{\mathbf{e}}$ (see Barlow, et al, 1972, Chapter 7). Also, let $H_e(A) = \int_A h_e \, d\mu$. Note that H_e and $F^{\mathbf{e}}$ are not the same, in general, unless $\mathbf{e} = \mathbf{0}$ or f is $\mathscr{L}^{\mathbf{e}}$-measurable. Throughout this section we suppose that $\int f^2 \, d\mu < \infty$.

We begin with some simple lemmas, whose proofs are in the Appendix.

LEMMA 3.1. *Let $\mathscr{A}$ be a class of measurable sets in $\mathbb{R}^k$. $\mathscr{A}$ is an F-uniformity class if and only if $\mathscr{A}$ is an $H_{\mathbf{e}}$-uniformity class.*

LEMMA 3.2. *Let $\mathscr{L}$ be an $F^{-\mathbf{e}}$-uniformity class and $\mathbf{e}_n \to \mathbf{e}$ almost surely. Then*

(a) $\sup_{L\in\mathscr{L}^{\mathbf{e}}} | F_n(L) - F(L) | \to 0$ a.s.,
(b) $\sup_{L\in\mathscr{L}^{\mathbf{e}}} | F_n^{\mathbf{e}-\mathbf{e}_n}(L) - F(L) | \to 0$ a.s.,
(c) $\sup_{L\in\mathscr{L}^{\mathbf{e}}} | H_{\mathbf{e}}^{\mathbf{e}_n-\mathbf{e}}(L) - H_{\mathbf{e}}(L) | \to 0$.

It is rather remarkable that F-uniformity of the σ-lattice $\mathscr{L}$ suffices to establish not only the existence of $\hat{f}_n$ but also its consistency. F-uniformity seems to entail that ratios of the form $F_n(L \cap U)/\mu(L \cap U)$ used in defining $\hat{f}_n(\mathbf{x})$ will be close to $F(L \cap U)/\mu(L \cap U)$ which will be close to $f(x)$. However, there are interesting exceptions to this behavior, which may arise when the modal regions of f can have "tunnels" in or "spikes" on their boundaries. An example is briefly mentioned in the Appendix as a pathology. Fortunately, we can identify a set $g_{\mathscr{L}}$ of F-measure zero, which contains all the exceptions.

DEFINITION 3.4. Let $\mathscr{L}$ be a class of sets and g be $\mathscr{L}$-measurable. Let $g_{\mathscr{L}}$ denote the set of all $\mathbf{x}$ for which at least one of the following holds:

(i) $$\exists a < g(\mathbf{x}) \quad \text{with} \quad \lim_{\mathbf{d}\to\mathbf{0}} \inf_{\{L\in\mathscr{L}^{\mathbf{d}};\mathbf{x}\notin L\}} \mu(T_a g - L) = 0,$$

(ii) $$\exists b > g(\mathbf{x}) \quad \text{with} \quad \lim_{\mathbf{d}\to\mathbf{0}} \inf_{\{L\in\mathscr{L}^{\mathbf{d}};\mathbf{x}\in L\}} \mu(L - P_b g) = 0.$$

THEOREM 3.5. *If $\mathscr{L}$ is a σ-lattice and an F-uniformity class, g is $\mathscr{L}$-measurable and $g_{\mathscr{L}}$ is measurable, then $F(g_{\mathscr{L}}) = 0$.*

PROOF. (See Appendix).

Existence of the mle. For the sake of notational convenience, we shall prove the existence of the mle only for the case $\mathbf{e}_n = \mathbf{0}$. There is little loss of generality in this, as noted below, provided we take care that the conditioning event, $0 < \mu(L(\mathbf{X}_i)) < \infty$, in Theorem 3.6, always be satisfied for $\mathbf{e}_n$ different from $\mathbf{0}$. For example, $\mathscr{L}$ may be a nested class of sets decreasing to a point. If $\mathscr{L}^{\mathbf{e}_n}$ is centered on an estimate of that point and the estimate is an observation $\mathbf{x}_i$, then $\mu(L(\mathbf{x}_i)) = 0$. Whenever $\mu(L(\mathbf{x}_i)) = 0$ [or ∞] for some i, then the mle does not exist, for the likelihood can be made as large [or small] as we please.

THEOREM 3.6. P[mle exists $| \; 0 < \mu(L(\mathbf{X}_i)) < \infty, i = 1, \cdots, n] = 1$.

PROOF. By induction on n. The case $n = 1$ is trivial, so suppose the theorem true for n. Within the event $[\hat{f}_n$ exists$] \cap [0 < \mu(L(\mathbf{X}_i)) < \infty, i = 1, \cdots, n + 1]$, condition on $\mathbf{X}_1 = \mathbf{x}_1, \cdots, \mathbf{X}_n = \mathbf{x}_n$. It suffices to show

$$F\{\mathbf{x}_{n+1}; L(\mathbf{x}_i)\Delta L(\mathbf{x}_{n+1}) \neq \phi \quad \text{and} \quad \mu(L(\mathbf{x}_i)\Delta L(\mathbf{x}_{n+1})) = 0\} = 0 \tag{3.1}$$

for $i = 1, \cdots, n$, where Δ is the symmetric difference operation. For if $L(\mathbf{x}_i) = L(\mathbf{x}_{n+1})$ then $\mathscr{S} = \{A_1, \cdots, A_m\}$, the maximal disjoint base for the σ-lattice $\sigma\{L(\mathbf{x}_1), \cdots, L(\mathbf{x}_{n+1})\}$ (see Section 2), is the same as the base for $\sigma\{L(\mathbf{x}_1), \cdots, L(\mathbf{x}_{n+1})\}$ and the induction hypothesis applies. Or if $\mu(L(\mathbf{x}_i)\Delta L(\mathbf{x}_{n+1})) > 0$, $i = 1, \cdots, n$, then by dividing the A_i into two groups: incumbents from the case of n, and new A_i, we see that $F_{n+1}(A_j) > 0 \Rightarrow \mu(A_j) > 0$, so the mle exists. The argument for (3.1) is similar to the proof of Theorem 3.5, but with $L(\mathbf{x}_i)$ replacing $T_r g$ and other obvious but somewhat tedious modifications and therefore is omitted.

The proof just given is actually valid for the case $\mathbf{e}_n \equiv \mathbf{e}$ (free of n). It does not strictly apply to general $\mathbf{e}_n$ because the addition of $\mathbf{x}_{n+1}$, by changing the σ-lattice from $\mathscr{L}^{\mathbf{e}_n}$ to $\mathscr{L}^{\mathbf{e}_{n+1}}$, may change the sets $L(\mathbf{x}_1), \cdots, L(\mathbf{x}_n)$. However, this difficulty may be overcome by insisting that $\mathbf{e}_{n+1}$ be chosen so that, for each n, the condition $0 < \mu(L(\mathbf{x}_i)) < \infty$, $i = 1, \cdots, n$ would have been met if $\mathbf{e}_{n+1}$ had been used at stage n in lieu of the translation $\mathbf{e}_n$. This is a relatively modest restriction which allows the induction argument to work in the case of general $\mathbf{e}_n$. □

The principal mathematical property of the estimate which can be proved is its consistency. If the method of estimating location and orientation of the σ-lattice is convergent, then so is the estimate $\hat{f}_n$, and when the location and orientation converge to the correct values then so does $\hat{f}_n$.

THEOREM 3.7. *Let $\mathscr{L}$ be a σ-lattice and an F- and $F^{-\mathbf{e}}$-uniformity class. If $\mathbf{e}_n \to \mathbf{e}$, then $\hat{f}_n(\mathbf{x}) \to h_{\mathbf{e}}(\mathbf{x})$* a.s. *$\forall\, \mathbf{x} \notin h_{\mathbf{e}\mathscr{L}}$, hence $\forall\, \mathbf{x}$ in a set of $H_{\mathbf{e}}$-measure one.*

PROOF. First we note an immediate consequence of Brunk's (1963) characterization of isotonic regression. If g is a density, $\mathscr{U}$ is a σ-lattice, $\hat{g} = E[g \mid \mathscr{U}]$, and G and $\hat{G}$ are the measures associated with g and $\hat{g}$, respectively, then

$$\begin{cases} G(U) \le \hat{G}(U) \ \forall\ U \in \mathscr{U} \\ G(U) = \hat{G}(U) \ \forall\ U = \hat{g}^{-1}(B) \quad \text{for some Borel set } B. \end{cases} \tag{3.2}$$

(Actually, (3.2) holds if $\mu(U) < \infty$ and B does not contain the origin, but that is the case for all sets in the following argument.)

Let $x_0 \notin h_{\mathbf{e}\mathscr{L}}$. Without further comment, we assume that the event of probability one occurs for which Lemma 3.2 and Theorem 3.6 hold. We treat two cases.

CASE 1: $h_{\mathbf{e}}(\mathbf{x}_0) > 0$. Let $\varepsilon > 0$, $\quad 0 < a < h_{\mathbf{e}}(\mathbf{x}_0) < b$,

$$\delta_a = lim_{\mathbf{e}_n \to \mathbf{e}} \inf_{\{\hat{L} \in \hat{\mathscr{L}}^{-e^n}; \mathbf{x}_0 \notin L\}} \mu(T_a h_{\mathbf{e}} - L) > 0,$$

$$\delta_b = \lim_{\mathbf{e}_n \to \mathbf{e}} \inf_{\{L \in \hat{\mathscr{L}}^{\mathbf{e}-\mathbf{e}_n}; \mathbf{x}_0 \notin L\}} \mu(L - P_b\, h_{\mathbf{e}}) > 0.$$

Then for all n sufficiently large, we have the following inequalities:

$$\begin{aligned}
a &\le \frac{H_{\mathbf{e}}(T_a h_{\mathbf{e}} - (P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)^{\mathbf{e}-\mathbf{e}_n})}{\mu(T_a h_{\mathbf{e}} - (P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)^{\mathbf{e}-\mathbf{e}_n})} \le \frac{F(T_a h_{\mathbf{e}} - (P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)^{\mathbf{e}-\mathbf{e}_n})}{\mu(T_a h_{\mathbf{e}} - (P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)^{\mathbf{e}-\mathbf{e}_n})} \\
&\le \frac{F_n^{\mathbf{e}-\mathbf{e}_n}(T_a h_{\mathbf{e}} - (P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)^{\mathbf{e}-\mathbf{e}_n}) + \varepsilon\, \delta_a/4}{\mu(T_a h_{\mathbf{e}} - (P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)^{\mathbf{e}-\mathbf{e}_n})} = \frac{F_n(T_a h_{\mathbf{e}}^{\mathbf{e}_n-\mathbf{e}} - P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n) + \varepsilon\, \delta_a/4}{\mu((T_a h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}} - P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)} \\
&\le \frac{\hat{F}_n((T_a h_e)^{\mathbf{e}_n-\mathbf{e}} - P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n) + \varepsilon\, \delta_a/4}{\mu((T_a h_e)^{\mathbf{e}_n-\mathbf{e}} - P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)} \le \hat{f}_n(\mathbf{x}_0) + \frac{\varepsilon\, \delta_a/4}{\mu((T_a h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}} P_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)} \le \hat{f}_n(\mathbf{x}_0) + \frac{\varepsilon}{2} \\
&\le \frac{\hat{F}_n(T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n - (P_b h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}})}{\mu(T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n - (P_b h_e)^{\mathbf{e}_n-\mathbf{e}})} + \frac{\varepsilon}{2} \le \frac{F_n(T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n - (P_b h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}})}{\mu(T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n - (P_b h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}})} + \frac{\varepsilon}{2} \\
&= \frac{F_n^{\mathbf{e}-\mathbf{e}_n}((T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)^{\mathbf{e}-\mathbf{e}_n} - P_b h_{\mathbf{e}})}{\mu(T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n - (P_b h_e)^{\mathbf{e}_n-\mathbf{e}})} + \frac{\varepsilon}{2} \le \frac{F((T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)^{\mathbf{e}-\mathbf{e}_n} - P_b h_{\mathbf{e}}) + \varepsilon\, \delta_b/4}{\mu(T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n - (P_b h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}})} + \frac{\varepsilon}{2} \\
&\le \frac{H_{\mathbf{e}}((T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n)^{\mathbf{e}-\mathbf{e}_n} - P_b h_{\mathbf{e}}) + \varepsilon\, \delta_b/4}{\mu(T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n - (P_b h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}})} + \frac{\varepsilon}{2} \le b + \frac{\varepsilon\, \delta_b/4}{\mu(T_{\hat{f}_n(\mathbf{x}_0)} \hat{f}_n - (P_b h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}})} + \frac{\varepsilon}{2} \le b + \varepsilon.
\end{aligned}$$

In this chain of 14 equalities and inequalities, the 1st, 6th, 8th, and 13th need no comment; the 2nd, 5th, 9th, and 12th are obtained from (3.2); the 3rd and 11th from Lemma 3.2; the 4th and 10th from the definition of $F_n^{\mathbf{e}-\mathbf{e}_n}$ and μ-invariance under the translation; the 7th and 14th require more explanation. To cancel the δ_a in inequality 7, we must insure the membership of $\mathbf{x}_0$ in the denominator set preceding 7. To this end, note that

$$|\mu(T_a h_{\mathbf{e}} - P_{\hat{f}_n(\mathbf{x}_0)}\hat{f}_n) - \mu((T_a h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}} - P_{\hat{f}_n(\mathbf{x}_0)}\hat{f}_n)| \leq \mu(T_a h_{\mathbf{e}} \Delta (T_a h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}}).$$

But this symmetric difference goes to zero because Lemma 3.2 implies $|H_{\mathbf{e}}(T_a h_{\mathbf{e}}) - H_e(T_a h_e)^{\mathbf{e}_n-\mathbf{e}}| \to 0$ (which, because $h_{\mathbf{e}}(\mathbf{x}) > a$ iff $\mathbf{x} \in T_a h_{\mathbf{e}}$ and because $\mu(T_a h_{\mathbf{e}}) = \mu((T_a h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}})$, would be inconsistent with the symmetric difference remaining positive). For inequality 14, show analogously that $\mu(T_{\hat{f}_n(\mathbf{x}_0)}\hat{f}_n - (P_b h_{\mathbf{e}})^{\mathbf{e}_n-\mathbf{e}})$ is close to $\mu(T_{\hat{f}_n(\mathbf{x}_0)} f_n - P_b h_{\mathbf{e}})$.

Since $\mathbf{x}_0 \notin h_{\mathbf{e}\mathscr{L}}$, a and b are arbitrary. Thus $\hat{f}_n(\mathbf{x}_0) \to h_{\mathbf{e}}(\mathbf{x}_0)$.

CASE 2: $h_{\mathbf{e}}(\mathbf{x}_0) = 0$. Apply the argument of Case 1 from inequalities 8–14 only, whenever $0 < \hat{f}_n(\mathbf{x}_0)$ so that the denominator of the terms in the inequalities will be finite. □

COROLLARY 3.8. *Let $\mathscr{L}$ be a σ-lattice and an F-uniformity class. If* $\mathbf{e}_n \to \mathbf{0}$, *then* $\hat{f}_n(\mathbf{x}) \to f(\mathbf{x})$ a.s. $\forall\, \mathbf{x} \notin f_{\mathscr{L}}$, *hence* $\forall\, \mathbf{x}$ *in a set of F-measure one.*

PROOF. $h_0 = f$.

It should be noted that F-uniformity has been used in Theorem 3.7 only through Lemma 3.2. In view of Topsøe's (1970) techniques, however, there seems little to be gained by making Lemma 3.2 an assumption to replace that of F-uniformity.

4. Examples. The first three examples have previously been studied in the literature. The results cited are all easily seen to be special cases of the general methods presented in this paper. In each case $\mathscr{L}$ is an F-uniformity class and $f_{\mathscr{L}}$ consists precisely of the points of discontinuity of f.

EXAMPLE 4.1. Estimating a unimodal univariate density with known mode (Robertson, 1967, particularly Theorem 4.1 and Corollary 4.1), $\mathscr{L}$ consists of all intervals containing the mode θ.

EXAMPLE 4.2. Estimating a unimodal univariate density with unknown mode (Wegman, 1969, particularly Theorem 4.1), $\mathscr{L}^{e_n}$ consists of all intervals containing a consistent estimate e_n (e.g., Sager, 1975) of the mode θ.

EXAMPLE 4.3. Estimating an "increasing" density on the unit square (Robertson, 1967). If $f(x, y)$ is defined on $[0, 1) \times [0, 1)$ and $\mathscr{L}$ is the class of sets with "nonincreasing" curves as boundaries, then the consistency result Theorem 4.4 of Robertson (1967) again follows.

EXAMPLE 4.4. Estimating a "unimodal" density on $\mathbb{R}^k$. Let $f(\mathbf{x})$ be defined on $\mathbb{R}^k$ and $\mathscr{L}$ be the σ-lattice of sets L defined by the following: $L \in \mathscr{L} \Leftrightarrow \mathbf{x} \in L$ implies the k-cell determined by $\mathbf{0}$ and $\mathbf{x}$ is contained in L (if $x_i \geq 0$, $i = 1, \cdots, k$, the k-cell determined by $\mathbf{0}$ and $\mathbf{x}$ is $[0, x_1] \times \cdots \times [0, x_k]$; if any $x_i < 0$, then replace $[0, x_i]$ by $[x_i, 0]$.) It follows from Topsøe (1970) that $\mathscr{L}$ is an $F^{\mathbf{e}}$-uniformity class for all $\mathbf{e}$ for absolutely continuous F. The meaning of unimodality is determined by the class $\mathscr{L}$. In this example, f being $\mathscr{L}$-measurable implies the existence of a "modal set" $L_0 \in \mathscr{L}$ containing $\mathbf{0}$ on which f has the constant value $f(L_0) > f(\mathbf{x})$ for all $\mathbf{x} \notin L_0$. Moreover, f is nonincreasing along every ray emanating from $\mathbf{0}$. But even more structure is imposed by the requirement that modal regions $[f \geq c]$, $\forall\, c$, be in $\mathscr{L}$.

It should be emphasized that the computation of $\hat{f}_n$ with respect to this $\mathscr{L}$ or any location/rotation translation of $\mathscr{L}$ is an easy matter with the maximum upper sets algorithm. The reason for this is that, after $\mathbf{e}_n$-location/rotation of the coordinate systems, the $L(\mathbf{x}_i)$ are simply k-cells determined by $\mathbf{x}_i$ and $\mathbf{0}$, the μ measure of which is the product of the magnitude of the coordinates of $\mathbf{x}_i$. Hence, for small n one could compute $\hat{f}_n$ by hand, if necessary. A consistent multivariate modal estimator $\boldsymbol{\theta}_n$ may be used for locating $\mathscr{L}^{\mathbf{e}_n}$ (see Sager, 1978). Estimating the rotational orientation of $\mathscr{L}^{\mathbf{e}_n}$ may prove more difficult. A search through various rotational orientations until one is found which maximizes the likelihood of $\hat{f}_n$ will be computationally expensive without an efficient algorithm.

Figure 2 was produced by assuming the theoretically correct location and orientation were known. In this map, the σ-lattice $\mathscr{L}$ is obviously misspecified (it should be ellipses) and the map has been processed to consolidate some of the isopleths for visual display. One would probably not use these raw estimated contours as final estimates any more than one would use the empiric cdf as a final estimate of a continuous cdf. One might use this very general σ-lattice for a "first look" under conditions of relative ignorance about the shape of the true contours.

EXAMPLE 4.5. Estimating a density with elliptical contours. Let $f(\mathbf{x})$ be unimodal at $\boldsymbol{\theta}$ in $\mathbb{R}^2$ and have elliptical contours centered at $\boldsymbol{\theta}$. Suppose it is further known that the ellipses all have uniform eccentricity ε and the same major and minor axes. Let $\mathscr{L}(\varepsilon)$ be the class of elliptical regions of eccentricity ε centered at $\mathbf{0}$ with the horizontal and vertical axes as major and minor axes. Then $\mathscr{L}(\varepsilon)^{\mathbf{e}}$ is a σ-lattice and an $F^{-\mathbf{e}}$-uniformity class for all ε and $\mathbf{e}$. Again, if one can estimate the location and orientation of $\mathscr{L}(\varepsilon)$, one can compute $\hat{f}_n$ by the method of Section 2 and obtain consistency from Section 3. Additionally, if the eccentricity ε must be estimated and can be done so consistently then adaptation of the methods of Section 3 will yield consistency of $\hat{f}_n$.

Figure 1 is illustrative of this method. The center of the ellipses was estimated by the vector of means and the orientation and eccentricity were estimated from the eigenvectors of the variance-covariance matrix. Thus, if the data were from a normal distribution (and they are), these estimates would be maximum likelihood. But the contours that are obtained are not restricted to those of a normal distribution. Both heavier and lighter-tailed distributions may be obtained.

This particular example was motivated by work the author has done for the U.S. Navy. A multivariate density estimate was desired for a high-dimensional dataset. It was felt that the true density was similar in shape to (and may in fact be) a multivariate normal. But the assumption of normality was felt to be too strong and a nonparametric estimate was desired. However, standard nonparametric density estimators converge very slowly in higher dimensions—too slowly for the problem in question. By assuming elliptical contours and estimating orientation from principal components, the possibility of normality was allowed while preserving the most relevant nonparametric alternatives. Moreover, since the elliptical contours impose a linear order on $\mathbb{R}^k$ in which "higher" means closer to the center of the ellipses, the algorithm for the estimator essentially reduces the problem to one dimension. In fact, the methods of Prakasa Rao (1969) could probably be used to show that $n^{1/3}\{\hat{f}_n(\mathbf{x}) - f(\mathbf{x})\}$ has an asymptotic distribution. So the rate of convergence does not depend on the dimension, nor does computation speed of the estimate (except for principal components).

APPENDIX

PROOF OF LEMMA 3.1. Topsøe (1967, Theorem 3), showed that $\mathscr{A}$ is an F-uniformity class iff $F(\bigcap_{i=1}^{\infty}\partial_{\delta_i}A_i) = 0$ for every sequence $\{\delta_i\} \downarrow 0$ and every sequence $\{A_i\}$ of sets from $\mathscr{A}$ where $\partial_\delta A$ denotes the set of points within distance δ of both A and the complement of A. Let $\mathscr{A}$ be an F-uniformity class, $\{\delta_i\} \downarrow 0$ and $A_i \in \mathscr{A}$. Then $\forall\, c > 0$, $F(\bigcap_{i=1}^{\infty}\partial_{\delta_i} A_i \cap T_c f) = 0$. Hence, by absolute continuity, $\mu(\bigcap_{i=1}^{\infty}\partial_{\delta_i} A_i \cap T_c f) = 0$ and $H_{\mathbf{e}}(\bigcap_{i=1}^{\infty}\partial_{\delta_i} A_i \cap T_c f) = 0$. Now by the properties of isotonic regression (3.2), $H^{\mathbf{e}}(T_c f) = F(T_c f) \uparrow 1$ as $c \downarrow 0$. Hence $H_{\mathbf{e}}(\bigcap_{i=1}^{\infty}\partial_{\delta_i} A_i) = 0$. The converse is similar. □

PROOF OF LEMMA 3.2. Part (a) is immediate from the definitions. Since a quarter-space (a subset of the plane congruent to the first quadrant) is convex, it follows from

Topsøe (1970) and other sources that the class of quarter-spaces is an $F^{-\mathbf{e}}$-uniformity class. It is easy to see from the nature of the $\mathbf{e}_n$-translation and the absolute continuity of F that $F^{-\mathbf{e}_n}(\mathbf{x}) \to F^{-\mathbf{e}}(\mathbf{x})\ \forall\ \mathbf{x}$. Hence

$$|F^{-\mathbf{e}_n}(\mathbf{x}) - F^{-\mathbf{e}}(\mathbf{x})| \le \sup_{Q\in\mathscr{Q}} |F_n(Q) - F(Q)| + |F^{-\mathbf{e}_n}(\mathbf{x}) - F^{-\mathbf{e}}(\mathbf{x})| \to 0 \text{ a.s.}$$

where $\mathscr{Q}$ is the class of all quarter spaces. Hence $F_n^{-\mathbf{e}_n} \Rightarrow F^{-\mathbf{e}}$. So

$$\sup_{L\in\mathscr{L}^{\mathbf{e}}} |F_n^{\mathbf{e}-\mathbf{e}_n}(L) - F(L)| = \sup_{L\in\mathscr{L}} |F_n^{-\mathbf{e}_n}(L) - F^{-\mathbf{e}}(L)| \to 0 \text{ a.s.}$$

Since $\mathscr{L}$ is assumed to be an $F^{-\mathbf{e}}$-uniformity class, then $\mathscr{L}^{\mathbf{e}}$ is an F-uniformity class and by Lemma 3.1 an $H_{\mathbf{e}}$-uniformity class. As in part (b), $H_{\mathbf{e}}^{\mathbf{e}_n-\mathbf{e}}(\mathbf{x}) \to H_{\mathbf{e}}(\mathbf{x})$ so that $H_{\mathbf{e}}^{\mathbf{e}_n-\mathbf{e}} \Rightarrow H_{\mathbf{e}}$. Part (c) follows. □

PROOF OF THEOREM 3.5. Let $g_{\mathscr{L}}^i$, $i = 1, 2$, denote the subsets of $g_{\mathscr{L}}$ for which condition (i) or (ii) holds, respectively. We first consider $g_{\mathscr{L}}^1$. For each $\mathbf{x} \in g_{\mathscr{L}}^1$, let $a_{\mathbf{x}}$ denote the inf of all a for which (i) holds. For each rational r, let $Q_r = \{\mathbf{x} \in g_{\mathscr{L}}^1, a_{\mathbf{x}} < r < g(\mathbf{x})\}$. The following two properties of Q_r are immediate:

(a) $F_n(Q_r) \to F(Q_r)$ a.s.

(b) If $\mathbf{x} \in Q_r$ then $\lim_{\mathbf{d}\to\mathbf{0}} \inf_{\{L\in\mathscr{L};\notin L^{\mathbf{d}}\}} \mu(T_r g - L) = 0$.

Now let $\varepsilon > 0$. By absolute continuity, $\exists\ \delta$ such that $\mu(A) < \delta \Rightarrow F(A) < \varepsilon/3$. For a given realization, let us denote the observations in Q_r by $\mathbf{x}_1, \cdots, \mathbf{x}_p$. By (b), we can find a $\mathbf{d}$ with $|\mathbf{d}|$ small and $L_1, \cdots, L_p$ such that $\mathbf{x}_i \notin L_i \in \mathscr{L}^{\mathbf{d}}$ and $\mu(T_r g - L_i) < \delta p^{-1}$. Letting $L = L_i \cap \cdots \cap L_p \in \mathscr{L}^{\mathbf{d}}$, we have $\mathbf{x}_i \notin L^{\mathbf{d}}$ and $F(T_r g - L) < \varepsilon/3$. It is easy to see from Topsøe's characterization that $\{T_r g \cap L;\ L \in \mathscr{L}^{\mathbf{d}}\}$ is an F-uniformity class. By applying this and Lemma 3.2, we have $F_n(Q_r) = F_n(T_r g) - F_n(L \cap T_r g) \le |F_n(T_r g) - F(T_r g)| + |F(T_r g) - F(L \cap T_r g)| + |F(L \cap T_r g) - F_n(L \cap T_r g)| < \varepsilon$. Since ε is arbitrary, $F(Q_r) = 0$ by (a). Now for each $\mathbf{x} \in g_{\mathscr{L}}^1$, $\mathbf{x} \in Q_r$ for some rational r. Hence $F(g_{\mathscr{L}}^1) = 0$. An analogous argument shows that $F(g_{\mathscr{L}}^2) = 0$.

A pathology. To see that the set $g_{\mathscr{L}}$ of Theorem 3.5 is not necessarily null, consider the following example: Let x_1 and x_2 be i.i.d. $N(0, 1)$. Let $\mathscr{L}_1$ consist of all balls centered at the origin. Let $\mathscr{L}_2$ consist of the sets B_i, $i = 1, 2, \cdots$ where B_i is the ball of radius i centered at the origin but with the "wedge" between $\pm 1/i$ radians removed from the ball. And let $\mathscr{L}$ be the smallest σ-lattice containing $\mathscr{L}_1 \cup \mathscr{L}_2$. Clearly, $\mathscr{L}$ is an F-uniformity class and the density of (x_1, x_2) is $\mathscr{L}$-measurable. Now, Robertson's (1966) representation theorem can be used to show that $\hat{f}_n(\mathbf{x}) \equiv 0$ a.s. for all n and for all $\mathbf{x}$ on the nonnegative half of the horizontal axis! (By adding wedges to B_i, the same outcome could be obtained.)

REFERENCES

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M., and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

BONEVA, L., KENDALL, D. and STEFANOV, I. (1971). Spline transformations: three new diagnostic aids for the statistical data analyst. *J. Roy. Statist. Soc. Ser. B* **33** 1–70.

BRUNK, H. D. (1963). On an extension of the concept of conditional expectation. *Proc. Am. Math. Soc.* **14** 298–304.

DAVIS, J. C. and CULLAGH, M. J. [Editors] (1975). *Display and Analysis of Spatial Data.* Wiley, New York.

DYKSTRA, RICHARD (1981). Personal communication.

MATHÉRON, G. (1973). The intrinsic random functions and their applications. *Adv. in Appl. Probab.* **5** 439–468.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhya (A)* **31** 23–36.

ROBERTSON, T. (1966). A representation for conditional expectations given σ-lattices. *Ann. Math. Statist.* **37** 1279–1283.

ROBERTSON, T. (1967). On estimating a density which is measurable with respect to a σ-lattice. *Ann. Math. Statist.* **38** 482–493.

SAGER, T. (1975). Consistency in nonparametric estimation of the mode. *Ann. Statist.* **3** 698–706.

SAGER, T. (1978). Estimation of a multivariate mode. *Ann. Statist.* **6** 802–812.

SAGER, T. (1979). An iterative procedure for estimating a multivariate mode and isopleth. *J. Am. Statist. Assoc.* **74** 329–339.

SILVERMAN, B. W. (1978). Choosing the window width when estimating a density. *Biometrika* **65** 1–11.

TAPIA, RICHARD A. and THOMPSON, JAMES R. (1978). *Nonparametric Probability Density Estimation.* Johns Hopkins University Press, Baltimore and London.

TARTER, M. E. and KRONMAL, R. A. (1976). An introduction to the implementation and theory of nonparametric density estimation, *Amer. Statist.* **30** 105–112.

TOPSØE, F. (1967). On the connection between *P*-continuity and *P*-uniformity in weak convergence. *Theor. Probab. Appl.* **12** 241–250 [English].

TOPSØE, F. (1970). On the Glivenko-Cantelli theorem. *Z. Wahrsch. verw. Gebiete* **14** 239–250.

WEGMAN, E. (1969). A note on estimating a unimodal density. *Ann. Math. Statist.* **40** 1661–1667.

DEPARTMENT OF GENERAL BUSINESS
BUSINESS-ECONOMICS BUILDING 600
UNIVERSITY OF TEXAS
AUSTIN, TEXAS 78712