MASTER

NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION
OF SPATIAL PATTERNS

Thomas W. Sager

TECHNICAL REPORT NO. 29
OCTOBER 1979

PREPARED UNDER THE AUSPICES OF
SIAM INSTITUTE FOR MATHEMATICS AND SOCIETY

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Prepared For
THE U.S. DEPARTMENT OF ENERGY
UNDER CONTRACT NO. EY-76-S-02-2874

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION
OF SPATIAL PATTERNS

BY

THOMAS W. SAGER

DISCLAIMER
This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, or completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

TECHNICAL REPORT NO. 29

OCTOBER 1979

STUDY ON STATISTICS AND ENVIRONMENTAL
FACTORS IN HEALTH (SIMS)

PREPARED UNDER SUPPORT TO SIMS FROM

DEPARTMENT OF ENERGY (DOE)

ROCKEFELLER FOUNDATION

SLOAN FOUNDATION

ENVIRONMENTAL PROTECTION AGENCY (EPA)

NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Abstract

Let $\tilde{X}$ be an absolutely continuous random variable in $\mathbb{R}^k$ with distribution function $F(\tilde{x})$ and density $f(\tilde{x})$. Let $\tilde{X}_1, \dots, \tilde{X}_n$ be independent random variables distributed according to F . Mapping the spatial distribution of $\tilde{X}$ 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 $\mathcal{L}$ of sets whose boundaries constitute admissible isopleths and then maximizing the likelihood $\prod_{i=1}^n g(\tilde{x}_i)$ over all g whose isopleths are boundaries of $\mathcal{L}$ -sets. The only restrictions on $\mathcal{L}$ 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 $\mathcal{L}$ may be data-dependent up to locational and/or rotational translations. Strong consistency of the estimator is shown in the most general case.

Abbreviated title: NONPARAMETRIC MLE OF SPATIAL PATTERNS

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.

Nonparametric Maximum Likelihood Estimation of Spatial Patterns

1. Introduction

Let $F(\underline{x})$ be the distribution function of an absolutely continuous random variable $\underline{X}$ on $\mathbb{R}^k$, $k = 1, 2, 3, \dots$, with density $f(\underline{x})$. Let $X_1, \dots, X_n$ be independent random variables distributed according to F . It is often of interest, particularly when $k = 2$, to map the distribution of $\underline{X}$ 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 $\underline{X}$ 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 $f_1, \dots, f_n$ of the density taken with error at corresponding fixed sampling points $\underline{x}_1, \dots, \underline{x}_n$. 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. On the other hand, the data may consist of a large number of locations $\underline{x}_1, \dots, \underline{x}_n$ at each of which an occurrence of the phenomenon $\underline{X}$ is observed. The locations 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.

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 use ad hoc fitting methods with completely unknown statistical properties. There are a few notable exceptions to these remarks, particularly packages based on Kriging or splines.

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 $\mathcal{L}$ denote this class of sets. Since an isopleth is a level curve $\{x; f(x) = c\}$ and normally $f \geq c$ on the interior of an isopleth, we call sets of the form $\{x; f(x) \geq c\}$ modal regions of f . Restricting the statistician's choice of modal regions [isopleths] to $\mathcal{L}$ [boundaries of $\mathcal{L}$ -sets] means that any estimate g of f should satisfy $\{x; g(x) \geq c\} \in \mathcal{L}$ for all c . We say that g is $\mathcal{L}$ -measurable if and only if this condition is satisfied. The fitting criterion we propose is maximum likelihood: Given $\mathcal{L}$ and observations $X_1 = x_1, \dots, X_n = x_n$, we say $\hat{f}_n$ is an $\mathcal{L}$ -maximum

likelihood estimate ($\mathcal{L}$ -mle) if and only if $\prod_{i=1}^n \hat{f}_n(x_i) = \max\{\prod_{i=1}^n g(x_i); g \text{ is } \mathcal{L}\text{-measurable}\} < \infty$. $\mathcal{L}$ plays the same role in this conceptualization of maximum likelihood estimation as the more usual specification of a (parameterized) family of densities $\mathcal{F}$ over which $\prod_{i=1}^n g(x_i)$, $g \in \mathcal{F}$, 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 $\mathcal{L}$. The first is that $\mathcal{L}$ be a complete σ -lattice of Borel or Lebesgue sets and the second is that $\mathcal{L}$ 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. $\mathcal{L}$ is a complete σ -lattice if and only if $\mathcal{L}$ contains the empty set and the whole space and is closed under arbitrary union and intersection. (We also require measurability for the members of $\mathcal{L}$.)

Definition 2. A class $\mathcal{G}$ of measurable sets is a P -uniformity class if and only if $\sup_{P \Rightarrow P} \limsup_{n \rightarrow \infty} \sup_{A \in \mathcal{G}} |P_n(A) - P(A)| = 0$, 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. 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 $\mathcal{L}$). 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 $\mathcal{L}$ be a σ -lattice than by the F-uniformity structure. 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 $(\Omega, \mathcal{B}, \mu)$ be a totally finite measure space and let $\mathcal{L}$ be a complete σ -lattice of subsets of Ω . Let $\omega_1, \dots, \omega_n$ be a fixed set of points in Ω . Let $\sigma(\mathcal{S})$ be the smallest σ -lattice containing $\mathcal{S}$, for any class of sets $\mathcal{S}$.

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

Solution. Let $L(\omega_i) = \bigcap \{L; \omega_i \in L \in \mathcal{L}\}$. First consider the above problem with the additional restriction that g be a simple function on $\mathcal{S} = \{\Lambda_1, \dots, \Lambda_m\}$, the maximal disjoint base of nonempty sets for $\sigma\{L(\omega_1), \dots, L(\omega_n)\}$ of which every such disjoint base is a refinement.

Now a simple function on $\mathcal{G}$ is $\mathcal{L}$ -measurable if and only if it is isotonic with respect to the quasi order $\leq$ on $\mathcal{G}$ defined by

$A_i \leq A_j \Leftrightarrow \forall L \in \mathcal{L} \cap \sigma(\mathcal{G})$ such that $A_i \subset L$, we have $A_j \subset L$. For

if the simple g is $\mathcal{L}$ -measurable and $A_i \leq A_j$, then

$A_j \subset \{\omega; g(\omega) \geq g(A_i)\} \in \mathcal{L} \cap \sigma(\mathcal{G}) \Rightarrow g(A_i) \leq g(A_j)$. And if the simple

g is isotonic, then observing that $g \geq g(A_i)$ on $\cap\{L \in \mathcal{L} \cap \sigma(\mathcal{G}); A_i \subset L\}$, we see that $\{\omega; g(\omega) \geq c\} = \bigcup_{\{A_i; g(A_i) \geq c\}} \cap\{L \in \mathcal{L} \cap \sigma(\mathcal{G}); A_i \subset L\} \in \mathcal{L}$.

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

minimize $\sum_{i=1}^n \frac{F_n(A_i)}{\mu(A_i)} \cdot \mu(A_i) \log g_i$ subject to $g_i \geq 0$,

$\sum_{i=1}^m g_i \cdot \mu(A_i) = 1$ and $g_i \leq g_j$ iff $A_i \leq A_j$. (Here, F_n denotes

empirical measure.) By example 1.10, p. 45-46 of Barlow, et al (1972),

the solution to this modified problem is given by the isotonic regression

of $\left\{ \frac{F_n(A_i)}{n \mu(A_i)} ; i = 1, \dots, n \right\}$ with weights $\mu(A_i)$, $i = 1, \dots, n$ with

respect to the quasi order $\leq$. To show that this also solves the

general problem, let h be any $\mathcal{L}$ -measurable function and let

$\xi_1, \dots, \xi_n$ be the rearranged values of $\omega_1, \dots, \omega_n$ such that

$h(\xi_1) \leq h(\xi_2) \leq \dots \leq h(\xi_n)$. If we define

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

where c is a normalization constant, then h^* is $\mathcal{L}$ -measurable, simple

on $\mathcal{G}$ and $\prod_{i=1}^n h^*(\omega_i) = c^n \prod_{i=1}^n h(\omega_i) \geq \prod_{i=1}^n h(\omega_i)$ since

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

□

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 $(\Omega, \mathcal{G}, \mu)$ is not totally finite (e.g., $\mathbb{R}^k$), restrict the space to $\bigcup_{i=1}^n L(\omega_i)$, 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), p. 484. In fact, $\mathcal{L}$ is freed from any connection with any quasi order on Ω (although such an order may always be defined after being given $\mathcal{L}$). 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. In our use of the general framework, Ω will become $\mathbb{R}^k$, μ will denote Lebesgue measure, and $\omega_1, \dots, \omega_n$ will be realizations $x_1, \dots, x_n$ of $X_1, \dots, X_n$.

Calculating the solution $\hat{f}_n$.

$\hat{f}_n$ may be readily calculated provided the sets $L(\omega_1), \dots, L(\omega_n)$ can be identified. An enormous computational simplification may then be achieved by essentially replacing the class $\mathcal{L}$ by $\sigma\{L(\omega_1), \dots, L(\omega_n)\}$ and iteratively reducing and modifying the latter. The technique is the maximum upper sets algorithm (or equivalently, minimum lower sets algorithm; see Barlow, et al (1972), p. 76 ff.):

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

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 $\mathcal{L}$ 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 $\mathcal{L}$ according to

those estimates. Any such location/rotation translation of members of $\mathcal{L}$ is rigid and preserves their Lebesgue measure. In the sequel, we shall suppose without loss of generality that the location and orientation of $\mathcal{L}$ are correct ones so that f is $\mathcal{L}$ -measurable. The error made by the statistician in estimating the vector of location/rotation parameters is $\underline{e}_n$. Thus $\hat{f}_n$ is calculated with respect to the $\underline{e}_n$ -translation of $\mathcal{L}$. Other notation used in this section:

μ Lebesgue measure

$\mathcal{L}^{\underline{e}_n}$ the $\underline{e}_n$ -location/rotation translation of $\mathcal{L}$

$A^{\underline{e}_n}$ the $\underline{e}_n$ -translation of the set A

$F^{\underline{e}_n}$ the measure $F^{\underline{e}_n}(A) = F(A^{\underline{e}_n})$

$f^{\underline{e}_n}$ the density of $F^{\underline{e}_n}$

$F_n^{\underline{e}_n}$ the empiric measure $F_n^{\underline{e}_n}(A) = F_n(A^{\underline{e}_n})$

$\hat{F}_n$ the measure $\hat{F}_n(A) = \int_A \hat{f}_n d\mu$

$T_a g$ the set $\{x; g(x) \geq a\}$ for any function g

$P_a g$ the set $\{x; g(x) > a\}$ 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 $\underline{e}_n \rightarrow 0$. 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 $\underline{e}_n \rightarrow \underline{e}$. Let $h_{\underline{e}} = E[f|\mathcal{L}^{\underline{e}}]$, the conditional expectation of f with

respect to the σ -lattice $\mathcal{L}^e$ (see Barlow, et al (1972), chapter 7).

Also, let $H_e(A) = \int_A h_e d\mu$. Note that H_e and F^e are not the same, in general, unless $e = 0$. Throughout this section we suppose that $\int f^2 d\mu < \infty$.

We begin with some simple lemmas.

Lemma 3.1. Let $\mathcal{G}$ be a class of measurable sets in $\mathbb{R}^k$. $\mathcal{G}$ is an F-uniformity class if and only if $\mathcal{G}$ is an H_e -uniformity class.

Proof. Topsøe (1967, Theorem 3), showed that $\mathcal{G}$ is an F-uniformity class iff $F(\cap_{i=1}^{\infty} \partial_{\delta_i} A_i) = 0$ for every sequence $\{\delta_i\} \downarrow 0$ and every sequence $\{A_i\}$ of sets from $\mathcal{G}$ where $\partial_{\delta} A$ denotes the set of points within distance δ of both A and the complement of A . Let $\mathcal{G}$ be an F-uniformity class, $\{\delta_i\} \downarrow 0$ and $A_i \in \mathcal{G}$. Then $\forall c > 0$, $F(\cap_{i=1}^{\infty} \partial_{\delta_i} A_i \cap T_c f) = 0$. Hence, by absolute continuity, $\mu(\cap_{i=1}^{\infty} \partial_{\delta_i} A_i \cap T_c f) = 0$ and $H_e(\cap_{i=1}^{\infty} \partial_{\delta_i} A_i \cap T_c f) = 0$. Now by the properties of isotonic regression (3.2) below, $H_e(P_0 f) = F(P_0 f) = 1$. Hence $H_e(\cap_{i=1}^{\infty} \partial_{\delta_i} A_i) = 0$. The converse is similar.

□

Lemma 3.2. Let $\mathcal{L}$ be an $F^{\tilde{e}}$ -uniformity class and $e_n \rightarrow e$. Then

$$a) \sup_{L \in \mathcal{L}^e} |F_n(L) - F(L)| \rightarrow 0 \text{ a.s.}$$

$$b) \sup_{L \in \mathcal{L}^e} |F_n^{\tilde{e}-e_n}(L) - F(L)| \rightarrow 0 \text{ a.s.}$$

$$c) \sup_{L \in \mathcal{L}^e} |H_n^{\tilde{e}-e}(L) - H_e(L)| \rightarrow 0.$$

Proof: Part a) is immediate from the definitions. Since a quarter-space is convex, it follows from Topsøe (1970) and other sources that the class of quarter-spaces is an F^{-e} -uniformity class. It is easy to see from the nature of the e_n -translation and the absolute continuity of F that $F_n^{-e}(x) \rightarrow F^{-e}(x) \forall x$. Hence

$$|F_n^{-e}(x) - F^{-e}(x)| \leq \sup_{Q \in \mathcal{Q}} |F_n(Q) - F(Q)| + |F_n^{-e}(x) - F^{-e}(x)| \rightarrow 0 \text{ a.s.}$$

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

$$\sup_{L \in \mathcal{L}^e} |F_n^{-e}(L) - F^{-e}(L)| = \sup_{L \in \mathcal{L}} |F_n^{-e}(L) - F^{-e}(L)| \rightarrow 0 \text{ a.s.}$$

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

It is rather remarkable that F -uniformity of the σ -lattice $\mathcal{L}$ 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 $\frac{F_n(L \cap U)}{\mu(L \cap U)}$ used in defining $\hat{f}_n(x)$ will be close to $\frac{F(L \cap U)}{\mu(L \cap U)}$ which will be close to $f(x)$. However, there are interesting exceptions to this behavior. Fortunately, these exceptions may occur only on a set of F -measure zero. The following example is illustrative.

Example 3.3. Let $f(x_1, x_2) = (2\pi)^{-1} \exp\{-\frac{1}{2}(x_1^2 + x_2^2)\}$. Let $\mathcal{L}_1$ consist of $\emptyset, \mathbb{R}^2$ and all balls centered at the origin. Let $\mathcal{L}_2$ consist of the sets B_i , $i = 1, 2, \dots$, where B_i is the ball centered at 0 of radius i with the segment between $\pm i^{-1}$ radians removed from the ball. Let $\mathcal{L}$ be the smallest σ -lattice containing $\mathcal{L}_1 \cup \mathcal{L}_2$. Clearly, f is $\mathcal{L}$ -measurable. Moreover, it is not hard to verify that

$\mathcal{L}$ is an F-uniformity class. Set $e_n = 0$. The conditions of Robertson's (1966) representation theorem are satisfied, so

$$\hat{f}_n(x) = \sup_{L; x \in L} \inf_{L'; \mu(L-L') > 0} \frac{F_n(L-L')}{\mu(L-L')} .$$

Now let x_0 be any point on the nonnegative half of the horizontal axis. If $x_0 \in L \in \mathcal{L}$, then L is a ball of radius at least $\|x_0\|$. For any such L and any realization $x_1, \dots, x_n$ having no points on the horizontal axis, one can always find a B_i for i sufficiently large such that $\mu(L - B_i \cap L) > 0$ but $F_n(L - B_i \cap L) = 0$. Hence $\hat{f}_n(x) \equiv 0$ a.s. for all x on the nonnegative half of the horizontal axis and for all n !

It is clear that the example will generalize to more than 2 dimensions. Moreover, changing $\mathcal{L}_2$ to the class of balls B_i centered at 0 of radius α with removal of the segment between $\pm i^{-1}$ radians, $i = 1, 2, \dots$, allows one to make $\hat{f}_n(x_0)$ asymptotically any positive value less than $f(x_0)$ by varying α . For a one-dimensional example, let f be the standard normal density, $\mathcal{L}_1$ be the symmetric intervals about 0, and let $\mathcal{L}_2$ be the sets $[-i, x_0 - i^{-1}] \cup [x_0 + i^{-1}, i]$, $i = 1, 2, \dots$. In these examples, the difficulty stems from being able to "pack" the modal regions of f to arbitrary fullness with $B_i \cap L \in \mathcal{L}$, each of which misses part of the deeper interior of the modal region. Note that the examples could be reworked to the same outcome by adding wedges onto the exteriors of balls in $\mathcal{L}_2$ rather than by removing wedges. These considerations lead us to distinguish the set of points for which exceptional behavior of this sort may occur.

Definition 3.4. Let $\mathcal{L}$ be a class of sets and g be $\mathcal{L}$ -measurable.

Let $g_{\mathcal{L}}$ denote the set of all $\tilde{x}$ for which at least one of the following holds:

$$i) \exists a < g(\tilde{x}) \text{ with } \inf_{\{L \in \mathcal{L}; \tilde{x} \notin L\}} \mu(T_a g - L) = 0$$

$$ii) \exists b > g(\tilde{x}) \text{ with } \inf_{\{L \in \mathcal{L}; \tilde{x} \in L\}} \mu(L - P_b g) = 0.$$

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

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

$$a) F_n(Q_r) \rightarrow F(Q_r) \text{ a.s.}$$

$$b) \text{ If } \tilde{x} \in Q_r \text{ then } \inf_{\{L \in \mathcal{L}; \tilde{x} \notin L\}} \mu(T_r g - L) = 0.$$

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 $\tilde{x}_1, \dots, \tilde{x}_p$. By b), we can find $L_1, \dots, L_p$ such that $\tilde{x}_i \notin L_i \in \mathcal{L}$ and $\mu(T_r g - L_i) < \delta p^{-1}$. Letting $L = T_r g \cap L_1 \cap \dots \cap L_p \in \mathcal{L}$, we have $\tilde{x}_i \notin L$ and $F(T_r g - L) < \varepsilon/3$. If n has been chosen large enough, by Lemma 3.2, we have

$$\begin{aligned} F_n(Q_r) &= F_n(T_r g) - F_n(L) \leq |F_n(T_r g) - F(T_r g)| + |F(T_r g) - F(L)| \\ &+ |F(L) - F_n(L)| < \varepsilon. \end{aligned}$$

Since ε is arbitrary, $F(Q_r) = 0$ by a).

Now for each $\underline{x} \in g_{\underline{f}}^1$, $\underline{x} \in Q_r$ for some rational r . Hence $F(g_{\underline{f}}^1) = 0$.

An analogous argument shows that $F(g_{\underline{f}}^2) = 0$.

□

Existence of the mle.

For the sake of notational convenience, we shall prove the existence of the mle only for the case $e_n = 0$. There is little loss of generality in this, as noted below, provided we take care that the conditioning event, $0 < \mu(L(X_i)) < \infty$, in Theorem 3.6, always be satisfied for e_n different from 0. For example, $\underline{g}$ may be a nested class of sets decreasing to a point. If $\underline{g}^{e_n}$ is centered on an estimate of that point and the estimate is an observation $\underline{x}_i$, then $\mu(L(\underline{x}_i)) = 0$. Whenever $\mu(L(\underline{x}_i)) = 0$ [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. $P[\text{mle exists} | 0 < \mu(L(X_i)) < \infty, i = 1, \dots, n] = 1$.

Proof: By induction on n . The case $n = 1$ is trivial, so suppose the theorem true for n . Within the event $[\hat{f}_n \text{ exists}] \cap [0 < \mu(L(X_i)) < \infty, i = 1, \dots, n+1]$, condition on $X_1 = \underline{x}_1, \dots, X_n = \underline{x}_n$. It suffices to show

$$(3.1) \quad P\{\underline{x}_{n+1}; L(\underline{x}_i) \Delta L(\underline{x}_{n+1}) \neq \emptyset \text{ and } \mu(L(\underline{x}_i) \Delta L(\underline{x}_{n+1})) = 0\} = 0$$

for $i = 1, \dots, n$ where Δ is the symmetric difference operation.

For if $L(\underline{x}_i) = L(\underline{x}_{n+1})$ then $\mathcal{g} = \{A_1, \dots, A_m\}$, the maximal disjoint base for $\sigma\{L(\underline{x}_1), \dots, L(\underline{x}_{n+1})\}$ (see Section 2), is the same as the

base for $\sigma\{L(\underline{x}_1), \dots, L(\underline{x}_{n+1})\}$ and the induction hypothesis applies. Or if $\mu(L(\underline{x}_i) \Delta L(\underline{x}_{n+1})) > 0$, $i = 1, \dots, n$, 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 $L(\underline{x}_i)$ 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 $\underline{e}_n \equiv \underline{e}$ (free of n). It does not strictly apply to general $\underline{e}_n$ because the addition of $\underline{x}_{n+1}$, by changing the σ -lattice from $\underline{\mathcal{L}}^{\underline{e}_n}$ to $\underline{\mathcal{L}}^{\underline{e}_{n+1}}$, may change the sets $L(\underline{x}_1), \dots, L(\underline{x}_n)$. However, this difficulty may be overcome by insisting that $\underline{e}_{n+1}$ be chosen so that, for each n , the condition $0 < \mu(L(\underline{x}_i)) < \infty$, $i = 1, \dots, n$ would have been met if $\underline{e}_{n+1}$ had been used at stage n in lieu of the translation $\underline{e}_n$. This is a relatively modest restriction which allows the induction argument to work in the case of general $\underline{e}_n$.

□

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

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

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

Let $x_0 \notin h_{\tilde{e}}^{\sim}$. 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: $h_{\tilde{e}}^{\sim}(x_0) > 0$.

Let $\varepsilon > 0$, $0 < a < h_{\tilde{e}}^{\sim}(x_0) < b$, $\delta_a = \inf_{\{L \in \mathcal{L}; x_0 \notin L\}} \mu(T_a h_{\tilde{e}}^{\sim} - L) > 0$, $\delta_b = \inf_{\{L \in \mathcal{L}; x_0 \in L\}} \mu(L - P_b h_{\tilde{e}}^{\sim}) > 0$. Then for all n sufficiently large, we have the following inequalities:

$$\begin{aligned} a &\leq \frac{H_{\tilde{e}}(T_a h_{\tilde{e}}^{\sim} - P_{\hat{f}_n(x_0)} \hat{f}_n^{e-e})}{\mu(T_a h_{\tilde{e}}^{\sim} - P_{\hat{f}_n(x_0)} \hat{f}_n^{e-e})} \leq \frac{F(T_a h_{\tilde{e}}^{\sim} - P_{\hat{f}_n(x_0)} \hat{f}_n^{e-e})}{\mu(T_a h_{\tilde{e}}^{\sim} - P_{\hat{f}_n(x_0)} \hat{f}_n^{e-e})} \\ &\leq \frac{F_n^{e-e}(T_a h_{\tilde{e}}^{\sim} - P_{\hat{f}_n(x_0)} \hat{f}_n^{e-e}) + \varepsilon \delta_a / 4}{\mu(T_a h_{\tilde{e}}^{\sim} - P_{\hat{f}_n(x_0)} \hat{f}_n^{e-e})} \\ &= \frac{F_n(T_a h_{\tilde{e}}^{e-e} - P_{\hat{f}_n(x_0)} \hat{f}_n) + \varepsilon \delta_a / 4}{\mu(T_a h_{\tilde{e}}^{e-e} - P_{\hat{f}_n(x_0)} \hat{f}_n)} \\ &\leq \frac{\hat{F}_n(T_a h_{\tilde{e}}^{e-e} - P_{\hat{f}_n(x_0)} \hat{f}_n) + \varepsilon \delta_a / 4}{\mu(T_a h_{\tilde{e}}^{e-e} - P_{\hat{f}_n(x_0)} \hat{f}_n)} \end{aligned}$$

$$\leq \hat{f}_n(x_0) + \frac{\varepsilon \delta_a/4}{\mu\left(T_a h_e^{\tilde{n}-e} - P_{\hat{f}_n(x_0)} \hat{f}_n\right)} \leq \hat{f}_n(x_0) + \frac{\varepsilon}{2}$$

$$\leq \frac{\hat{F}_n\left(T_{\hat{f}_n(x_0)} \hat{f}_n - P_b h_e^{\tilde{n}-e}\right)}{\mu\left(T_{\hat{f}_n(x_0)} \hat{f}_n - P_b h_e^{\tilde{n}-e}\right)} + \frac{\varepsilon}{2}$$

$$\leq \frac{F_n\left(T_{\hat{f}_n(x_0)} \hat{f}_n - P_b h_e^{\tilde{n}-e}\right)}{\mu\left(T_{\hat{f}_n(x_0)} \hat{f}_n - P_b h_e^{\tilde{n}-e}\right)} + \frac{\varepsilon}{2}$$

$$= \frac{\tilde{F}_n^{\tilde{n}-e}\left(T_{\hat{f}_n(x_0)} \hat{f}_n^{\tilde{n}-e} - P_b h_e^{\tilde{n}-e}\right)}{\mu\left(T_{\hat{f}_n(x_0)} \hat{f}_n - P_b h_e^{\tilde{n}-e}\right)} + \frac{\varepsilon}{2}$$

$$\leq \frac{F\left(T_{\hat{f}_n(x_0)} \hat{f}_n^{\tilde{n}-e} - P_b h_e^{\tilde{n}-e}\right) + \varepsilon \delta_b/4}{\mu\left(T_{\hat{f}_n(x_0)} \hat{f}_n - P_b h_e^{\tilde{n}-e}\right)} + \frac{\varepsilon}{2}$$

$$\leq \frac{H_e\left(T_{\hat{f}_n(x_0)} \hat{f}_n^{\tilde{n}-e} - P_b h_e^{\tilde{n}-e}\right) + \varepsilon \delta_b/4}{\mu\left(T_{\hat{f}_n(x_0)} \hat{f}_n - P_b h_e^{\tilde{n}-e}\right)} + \frac{\varepsilon}{2}$$

$$\leq b + \frac{\varepsilon \delta_b/4}{\mu\left(T_{\hat{f}_n(x_0)} \hat{f}_n - P_b h_e^{\tilde{n}-e}\right)} + \frac{\varepsilon}{2} \leq b + \varepsilon.$$

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^{e-e} and μ -invariance under the translation; the 7th and 14th require more explanation. To cancel the δ_a in number 7, we must insure the membership of x_0 in the denominator set preceding 7. To this end, note that

$$|\mu(T_a h_e - P_{\hat{f}_n(x_0)} \hat{f}_n) - \mu(T_a h_e^{e-n} - P_{\hat{f}_n(x_0)} \hat{f}_n)| \leq \mu(T_a h_e \Delta T_a h_e^{e-n}).$$

But this symmetric difference goes to zero because Lemma 3.2 implies

$$|H_e(T_a h_e) - H_e(T_a h_e^{e-n})| \rightarrow 0 \text{ (which, because } h_e(x) \geq a \text{ iff } x \in T_a h_e \text{ and because } \mu(T_a h_e) = \mu(T_a h_e^{e-n}), \text{ would be inconsistent with the symmetric difference remaining positive). For 14, show analogously that } \mu(T_{\hat{f}_n(x_0)} \hat{f}_n - P_b h_e^{e-n}) \text{ is close to } \mu(T_{\hat{f}_n(x_0)} \hat{f}_n - P_b h_e).$$

Since $x_0 \notin h_{e\mathcal{L}}$, a and b are arbitrary. Thus $\hat{f}_n(x_0) \rightarrow h_e(x_0)$.

Case 2: $h_e(x_0) = 0$.

Apply the argument of case 1 from inequalities 8-14 only, whenever $0 < \hat{f}_n(x_0)$ so that the denominator of the terms in the inequalities will be finite.

□

Corollary 3.8. Let $\mathcal{L}$ be a σ -lattice and an F-uniformity class. If $e_n \rightarrow 0$, then $\hat{f}_n(x) \rightarrow f(x)$ a.s. $\forall x \notin f_{\mathcal{L}}$, hence $\forall x$ 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

Several examples follow in which the generality and power of the methods developed in sections 2 and 3 are illustrated.

Example 4.1. Estimating a unimodal univariate density with known mode.

Let $f(x)$ be unimodal at θ and $\mathcal{J}$ be the σ -lattice of all intervals containing θ . Let $e_n = 0$. $\mathcal{J}$ is easily shown to be an F-uniformity class. $f_{\mathcal{J}}$ consists precisely of the points of discontinuity of f . Hence, $\hat{f}_n(x) \rightarrow f(x)$ a.s. if f is continuous at x . From this one obtains uniform convergence a.s. of $\hat{f}_n$ to f on closed intervals (Corollary 4.1 of Robertson (1967)). Moreover, by choosing $u_m \uparrow x$ and $v_m \downarrow x$, $u_m, v_m \notin f_{\mathcal{J}}$ for $x < \theta$ and applying the monotonicity of $\hat{f}_n$, one obtains for each m

$$f(u_m) = \liminf_n \hat{f}_n(u_m) \leq \liminf_n \hat{f}_n(x) \leq \limsup_n \hat{f}_n(x) \leq \limsup_n \hat{f}_n(v_m) = f(v_m) \text{ a.s.}$$

Hence $f(x^-) \leq \liminf_n \hat{f}_n(x) \leq \limsup_n \hat{f}_n(x) \leq f(x^+) \text{ a.s.}$ and similarly with reversed inequalities when $x > \theta$, even when $x \in f_{\mathcal{J}}$. (Theorem 4.1 of Robertson (1967).)

Example 4.2. Estimating a unimodal univariate density with unknown mode.

Let $f(x)$ be unimodal at θ (unknown) and $\mathcal{J}$ be the σ -lattice of

all intervals containing θ . Suppose $\theta_n \rightarrow \theta$ so that $e_n = \theta_n - \theta \rightarrow 0$ (e.g., suppose θ_n is a consistent estimator of the mode, e.g., Sager (1975)). Thus $\mathcal{L}^{e_n}$ is the σ -lattice of all intervals containing θ_n . As in example 4.1, one obtains $\hat{f}_n(x) \rightarrow f(x)$ a.s. if f is continuous at x and $f(x^-) \leq \liminf_n \hat{f}_n(x) \leq \limsup_n \hat{f}_n(x) \leq f(x^+)$ a.s. when $x < \theta$ even when $x \in f_{\mathcal{L}}$. And if $\theta_n - \theta \rightarrow e$, one obtains the above with $h_e(x^\pm)$ replacing $f(x^\pm)$ (Theorem 4.1 of Wegman (1969)).

Example 4.3. Estimating an "increasing" density on the unit square.

Let $f(\underline{x})$ be defined on $[0,1) \times [0,1)$ and $\mathcal{L}$ be the σ -lattice of sets L defined by $L \in \mathcal{L} \Leftrightarrow (x_1, y_1) \in L$ and $x_2 \geq x_1, y_2 \geq y_1$ imply $(x_2, y_2) \in L$. Topsøe (1970) shows that $\mathcal{L}$ is an F^e -uniformity class for all e for absolutely continuous F . Once again $f_{\mathcal{L}}$ is precisely the set of discontinuities of f , and similarly to examples 4.1 and 4.2 we obtain $f(\underline{x}^-) \leq \liminf_n \hat{f}_n(\underline{x}) \leq \limsup_n \hat{f}_n(\underline{x}) \leq f(\underline{x}^+)$ a.s. even when $\underline{x} \in f_{\mathcal{L}}$ (Theorem 4.4 of Robertson (1967)).

Example 4.4. Estimating a "unimodal" density on $\mathbb{R}^k$.

Let $f(\underline{x})$ be defined on $\mathbb{R}^k$ and $\mathcal{L}$ be the σ -lattice of sets L defined by the following: $L \in \mathcal{L} \Leftrightarrow \underline{x} \in L$ implies the k -cell determined by $\underline{0}$ and $\underline{x}$ is contained in L (if $x_i \geq 0, i = 1, \dots, k$, the k -cell determined by $\underline{0}$ and $\underline{x}$ is $[0, x_1] \times \dots \times [0, x_k]$; if any $x_i < 0$, then replace $[0, x_i]$ by $[x_i, 0]$.) It follows from Topsøe (1970) that $\mathcal{L}$ is an F^e -uniformity class for all e for absolutely continuous F . The meaning of unimodality is determined by

the class $\mathcal{L}$. In this example, f being $\mathcal{L}$ -measurable implies the existence of an $L_0 \in \mathcal{L}$ containing $\underline{0}$ such that $f(L_0) > f(\underline{x})$ for all $\underline{x} \notin L_0$. Moreover, f is nonincreasing along every ray emanating from $\underline{0}$. But even more structure is imposed by the requirement that modal regions $[f \geq c], \forall c$, be in $\mathcal{L}$.

It should be emphasized that the computation of $\hat{f}_n$ with respect to $\mathcal{L}$ or any location/rotation translation of $\mathcal{L}$ is an easy matter with the maximum upper sets algorithm. The reason for this is that, after e_n -location/rotation of the coordinate systems, the $L(\underline{x}_i)$ are simply k -cells determined by $\underline{x}_i$ and $\underline{0}$, the μ measure of which is the product of the magnitude of the coordinates of $\underline{x}_i$. Hence, for moderate n one could readily compute $\hat{f}_n$ by hand, if necessary. A consistent multivariate modal estimator $\hat{\theta}_n$ may be used for locating $\mathcal{L}_n^{e_n}$ (see Sager (1978)). Estimating the rotational orientation of $\mathcal{L}_n^{e_n}$ may prove more difficult. We suggest one scheme without investigating its consistency properties: after having estimated the location of $\mathcal{L}_n$, search through various rotational orientations until one is found which maximizes the likelihood of $\hat{f}_n$. If $e_n \rightarrow \underline{0}$, then $\hat{f}_n(\underline{x}) \rightarrow f(\underline{x})$ for each $\underline{x} \in f_{\mathcal{L}}$.

Example 4.5. Estimating a density with elliptical contours.

Let $f(\underline{x})$ be unimodal at $\underline{\theta}$ in $\mathbb{R}^2$ and have elliptical contours centered at $\underline{\theta}$. Suppose it is further known that the ellipses all have uniform eccentricity ϵ and the same major and minor axes. Let $\mathcal{L}(\epsilon)$ be the class of elliptical regions of eccentricity ϵ centered at $\underline{0}$ with the horizontal and vertical axes as major and minor axes.

Then $\mathcal{L}(\varepsilon)^e$ is a σ -lattice and an F^e -uniformity class for all ε and e . Again, if one can estimate the location and orientation of $\mathcal{L}(\varepsilon)$, 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 $\hat{f}_n$.

REFERENCES

- Barlow, R. E., Bartholomew, D. J., Bremner, J. M., Brunk, H. D. (1972). Statistical Inference under Order Restrictions, Wiley, New York.
- Brunk, H. D. (1963). "On an extension of the concept of conditional expectation". Proc. Am. Math. Soc., 14, 298-304.
- 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". Annals of Statistics, 3, 698-706.
- Sager, T. (1978). "Estimation of a multivariate mode". Annals of Statistics, 6, 802-812.
- Sager, T. (1979). "An iterative procedure for estimating a multivariate mode and isopleth". Jour. Am. Stat. Assoc., 74, 329-339.
- Topsøe, F. (1967). "On the connection between P-continuity and P-uniformity in weak convergence". Theor. Prob. Appl., 12, 241-250 [English].
- Topsøe, F. (1970). "On the Glivenko-Cantelli theorem". Z. Wahr. verw. Geb., 14, 239-250.
- Wegman, E. (1969). "A note on estimating a unimodal density". Ann. Math. Statist., 40, 1661-1667.