Journal of Multivariate Analysis **76**, 90–109 (2001)
doi:10.1006/jmva.2000.1921, available online at http://www.idealibrary.com on IDEAL®

On Local Moments

Hans-Georg Müller and Xin Yan

University of California, Davis

Received November 11, 1998; published online November 7, 2000

We introduce the concept of local moments for a distribution in $\Re^p$, $p \geqslant 1$, at a point $\mathbf{z} \in \Re^p$. Local moments are defined as normalized limits of the ordinary moments of a truncated version of the distribution, ignoring the probability mass falling outside a window centered at $\mathbf{z}$. The limit is obtained as the size of the window converges to 0. Corresponding local sample moments are obtained via properly normalized ordinary sample moments calculated from those data falling into a small window. The most prominent local sample moments are the local sample mean which is simply the standardized mean vector of the data falling into the window, and the local covariance, which is a standardized version of the covariance matrix of the data in the window. We establish consistency with rates of convergence and asymptotic distributions for local sample moments as estimates of the local moments. First and second order local moments are of particular interest and some applications are outlined. These include locally based iterative estimation of modes and contours and the estimation of the strength of local association. © 2000 Academic Press

AMS 1991 subject classifications: 62G07, 62G20.

Key words and phrases: contour, density estimation, dependence function, derivative estimation, local covariance, local geometry, local mean, mode estimation, multivariate data analysis.

1. INTRODUCTION

Local geometric features of a probability density function of p-variate data, such as its mode and its contours, are important elements of nonparametric multivariate data analysis. Specific examples are discussed for example in Scott [14]. As we aim to show in this paper, local moments in $\Re^p$ and their sample counterparts, the local sample moments, provide a unifying concept to explore such features. Local moments have the additional appeal of simplicity as they are obtained by localizing the basic statistical notions of ordinary moments, sample moments and averages.

Of particular interest are first order local moments or local means and second order local moments or local covariance matrices. First order local

0047-259X/00 $35.00

sample moments have been implicitly used previously in a mode finding algorithm called the Mean Update Algorithm (compare Thompson and Tapia [16]). Other applications of local moments include the estimation of local dependence via local covariance matrices and the estimation of density derivatives. Another area of application of local moments is the estimation of contours, tangent and normal directions, including directions of steepest ascent.

For the two-dimensional case, the idea of a (k_1, k_2)-th order local moment $\mu_{k_1, k_2}(\mathbf{z})$, at $\mathbf{z} = (z_1, z_2)$, corresponds to the following prescription: Define a sequence of rectangular neighborhoods or windows $S_n(\mathbf{z}) = [z_1 - \gamma, z_1 + \gamma] \times [z_2 - \gamma, z_2 + \gamma]$, which shrink towards $\mathbf{z}$ as $\gamma = \gamma(n) \to 0$ with increasing sample size $n \to \infty$. Calculate the centered moments of the data, conditional upon their falling into $S_n(\mathbf{z})$, normalize and take the limit as $n \to \infty$. For a fixed $\mathbf{z} = (z_1, z_2) \in \Re^2$, given bivariate random variables (X_1, X_2),

$$\mu_{k_1, k_2}(\mathbf{z}) = \lim_{n \to \infty} \frac{1}{4\gamma^2} E[(X_1 - z_1)^{k_1} (X_2 - z_2)^{k_2} \mid (X_1, X_2) \in S_n(\mathbf{z})].$$

We define and investigate local moments in $\Re^p$ and their relation to the local geometry of the density at $\mathbf{z}$ in the following Section 2. The sample versions of local moments are the local sample moments, which are introduced in Section 3. Local sample moments are obtained as ordinary sample moments computed only from those data which fall into windows S_n. Our main result is that local sample moments converge to the local moments under mild conditions.

The various applications of local moments are briefly discussed in Section 4. Technical results and proofs are compiled in Section 5.

2. LOCAL MOMENTS

For some $p \geqslant 1$, let $\mathbf{X} = (X_1, ..., X_p) \in \Re^p$ be a random vector with twice differentiable density f. Given a fixed point $\mathbf{z} = (z_1, ..., z_p) \in \Re^p$, and a sequence of factors $\gamma = \gamma_n > 0$, $\gamma_n \to 0$ as $n \to \infty$, define the rectangular neighborhood of $\mathbf{z}$,

$$S = S_n = S(\mathbf{z}, \gamma) = \prod_{j=1}^{p} [z_j - \gamma, z_j + \gamma].$$

In the following, we use the multiindex notation

$$\alpha=(\alpha_1, ..., \alpha_p) \text{ with integers } \alpha_i \geqslant i, \qquad |\alpha| = \sum_{i=1}^{p} \alpha_i, \qquad \mathbf{x}^\alpha = \prod_{i=1}^{p} x_i^{\alpha_i}$$

for a vector $\mathbf{x} \in \mathfrak{R}^p$, and

$$D^\alpha f(\mathbf{x}) = \frac{\partial^{\alpha_1 + \cdots + \alpha_p}}{\partial^{\alpha_1} x_1 \cdots \partial^{\alpha_p} x_p} f(\mathbf{x}).$$

For a multiindex α, we define $\lambda\alpha = (\lambda\alpha_1, ..., \lambda\alpha_p)$ for any real number λ, and

$$|\alpha^+| = |\alpha| + \tfrac{1}{2} \sum_{i=1}^{p} [1-(-1)^{\alpha_i}].$$

This means that $|\alpha^+|$ corresponds to $|\alpha|$, increased by one for each occurrence of an odd index among $(\alpha_1, ..., \alpha_p)$. Further, let $\mathbf{e}_i = (0, ..., 1, ..., 0)$ be the multiindex consisting of zeros except for a 1 at the ith position.

Given a multiindex α and a sequence $\gamma = \gamma_n \to 0$ as $n \to \infty$, we define the local moment of order α at $\mathbf{z}$,

$$\mu_\alpha = \lim_{\gamma \to 0} \frac{1}{\gamma^{|\alpha^+|}} E\{(\mathbf{X}-\mathbf{z})^\alpha \mid \mathbf{X} \in S\}. \tag{2.1}$$

The normalization factor $\gamma^{|\alpha^+|}$ in the denominator is motivated by the following result, which shows that the local moments are well defined, and which reveals the relations with the local geometry of the density at $\mathbf{z}$. A basic assumption for the derivation of asymptotic properties of local moments is:

(A1) The point $\mathbf{z} \in \mathfrak{R}^p$ is an interior point of the support of the density f, and f is twice continuously differentiable at $\mathbf{z}$, with $f(\mathbf{z}) > 0$.

THEOREM 2.1. *If for a given multiindex $\alpha = (\alpha_1, ..., \alpha_p)$ all of the α_i are even, then the local moment of order α at a point $\mathbf{z} \in \mathfrak{R}^p$ is given by*

$$\mu_\alpha(\mathbf{z}) = \prod_{i=1}^{p} \frac{1}{\alpha_i + 1}; \tag{2.2}$$

if all indices α_i *are even except for one index which is odd, then*

$$\mu_\alpha(\mathbf{z}) = \prod_{i=1}^{p} \frac{1}{\alpha_i+1} \sum_{j=1}^{p} \frac{\alpha_j+1}{\alpha_j+2} \frac{D^{\mathbf{e}_j} f(\mathbf{z})}{f(\mathbf{z})} 1_{\{\alpha_j \text{ is odd}\}}; \tag{2.3}$$

if all indices α_i *are even except for two indices which are odd, then*

$$\mu_\alpha(\mathbf{z}) = \prod_{i=1}^{p} \frac{1}{\alpha_i+1} \sum_{j,k=1,\, j\neq k}^{p} \frac{(\alpha_j+1)(\alpha_k+1)}{(\alpha_j+2)(\alpha_k+2)} \times \frac{D^{\mathbf{e}_j+\mathbf{e}_k} f(\mathbf{z})}{f(\mathbf{z})} 1_{\{\alpha_j \text{ and } \alpha_k \text{ are odd}\}}. \tag{2.4}$$

The proof and more detailed results can be found in Section 5. From the proof of Lemma 5.1 it becomes clear how results for local moments of orders with more than two odd indices can be derived as well; such results are, however, somewhat unwieldy, and the details are omitted.

Some special cases are of particular interest. If $|\alpha| = 1$, the local moment μ_α is of first order. Then $\alpha = \mathbf{e}_k$ for a k, $1 \leqslant k \leqslant p$, and Theorem 2.1(2.3) yields

$$\mu_{\mathbf{e}_k}(\mathbf{z}) = \frac{1}{3} \frac{D^{\mathbf{e}_k} f(\mathbf{z})}{f(\mathbf{z})}. \tag{2.5}$$

Regarding second order local moments, where $|\alpha| = 2$, we can write each such α as $\alpha = \mathbf{e}_j + \mathbf{e}_k$, for some j, k with $1 \leqslant j,\ k \leqslant p$. Theorem 2.1(2.2) and (2.4) then imply

$$\mu_{\mathbf{e}_j+\mathbf{e}_k}(\mathbf{z}) = \begin{cases} \dfrac{1}{3} & \text{for} \quad j = k, \\ \dfrac{1}{9} \dfrac{D^{\mathbf{e}_j+\mathbf{e}_k} f(\mathbf{z})}{f(\mathbf{z})} & \text{for} \quad j \neq k. \end{cases} \tag{2.6}$$

A more detailed analysis is called for to obtain the local analogue of a covariance matrix. Define $\sum = (\sigma_{jk})_{1 \leqslant j,\, k \leqslant p}$ by

$$\sigma_{jj} = \frac{1}{\gamma^2} Var(X_j \mid \mathbf{X} \in S), \tag{2.7}$$

and

$$\sigma_{jk} = \frac{1}{\gamma^4} Cov((X_j, X_k) \mid \mathbf{X} \in S), \qquad 1 \leqslant j, k \leqslant p, \quad j \neq k. \tag{2.8}$$

THEOREM 2.2. *As $\gamma \to 0$, it holds that*

$$\sigma_{jj} = \frac{1}{3} - \left[\left(\frac{D^{\mathbf{e}_j} f(\mathbf{z})}{3 f(\mathbf{z})} \right)^2 - 2 \left(\frac{D^{2\mathbf{e}_j} f(\mathbf{z})}{45 f(\mathbf{z})} \right) \right] \gamma^2 + o(\gamma^2) \tag{2.9}$$

and

$$\sigma_{jk} = \frac{1}{9} \left[\frac{D^{\mathbf{e}_j + \mathbf{e}_k} f(\mathbf{z})}{f(\mathbf{z})} - \frac{D^{\mathbf{e}_j} f(\mathbf{z})\, D^{\mathbf{e}_k} f(\mathbf{z})}{f(\mathbf{z})^2} \right] + o(1), \qquad j \neq k. \tag{2.10}$$

For the case $p = 2$, using $\mathbf{e}_1 = (1\ 0)'$, $\mathbf{e}_2 = (0\ 1)'$ and the abbreviations $D^{\mathbf{e}_1} f(\mathbf{z}) = f^{(10)}$, $D^{\mathbf{e}_2} f(\mathbf{z}) = f^{(01)}$, $D^{\mathbf{e}_1 + \mathbf{e}_2} f(\mathbf{z}) = f^{(11)}$, $D^{2\mathbf{e}_1} f(\mathbf{z}) = f^{(20)}$, $D^{2\mathbf{e}_2} f(\mathbf{z}) = f^{(02)}$ and $f(\mathbf{z}) = f$, we obtain the following local first order moments and local covariances:

$$\mu_{\mathbf{e}_1}(\mathbf{z}) = \frac{f^{(10)}}{3f}, \qquad \mu_{\mathbf{e}_2}(\mathbf{z}) = \frac{f^{(01)}}{3f},$$

$$\sigma_{11} = \frac{1}{3} + \left\{ \left(\frac{f^{(10)}}{3f} \right)^2 - \frac{2 f^{(20)}}{45 f} \right\} \gamma^2 + o(\gamma^2), \tag{2.11}$$

$$\sigma_{12} = \sigma_{21} = (f f^{(11)} - f^{(10)} f^{(01)}) / 9 f^2 + o(\gamma^2).$$

In the limit, eliminating the constant terms and renormalizing, we obtain for instance

$$\tilde{\sigma}_{11} = \lim_{\gamma \to 0} \frac{1}{\gamma^4} \left\{ var(X_1 \mid \mathbf{X} \in S) - \frac{\gamma^2}{3} \right\} = \left(\frac{f^{(10)}}{3f} \right)^2 - \frac{2 f^{(20)}}{45 f}, \tag{2.12}$$

$$\tilde{\sigma}_{12} = \lim_{\gamma \to 0} \frac{1}{\gamma^4} Cov\{ (X_1, X_2) \mid \mathbf{X} \in S \}$$

$$= (f f^{(11)} - f^{(10)} f^{(01)}) / 9 f^2. \tag{2.13}$$

Similar expressions hold for the p-dimensional case. These results demonstrate the close relationship between the local geometry of f at $\mathbf{z}$ and the local moments. We note that the constants such as 1/3, 1/9 in Theorem 2.2 result from the fact that we use regular unweighted moments in the definition of local moments. An obvious generalization would be weighted moments, in which case these constants depend on the particular choice of weight function. The corresponding calculations are similar to the ones presented here.

3. LOCAL SAMPLE MOMENTS

The local sample moment of order α at $\mathbf{z} \in \Re^p$, based on a sample of data vectors $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n \in \Re^p$, serves as an estimator of $\mu_\alpha(\mathbf{z})$. It is natural to define it by

$$\hat{\mu}_\alpha(\mathbf{z}) = \left[\sum_{i=1}^n (\mathbf{X}_i - \mathbf{z})^\alpha \, 1_S(\mathbf{X}_i) \right] \Big/ \left[\gamma^{|\alpha^+|} \sum_{i=1}^n 1_S(\mathbf{X}_i) \right]. \tag{3.1}$$

Here $|\alpha^+|$ is as defined before (2.1), and $1_S(\mathbf{X}) = 1$ if $\mathbf{X} \in S$, $1_S(\mathbf{X}) = 0$ if $\mathbf{X} \notin S$. These local sample moments are analogous to the regular sample moments except that they are conditional for those data falling into the window S, and are rescaled by the factor $\gamma^{|\alpha^+|}$. For example, first order local sample moments are given by

$$\hat{\mu}_{\mathbf{e}_j}(\mathbf{z}) = \frac{1}{\gamma^2} \sum_{i=1}^n (X_{ij} - z_j) \, 1_S(\mathbf{X}_i) \Big/ \sum_{i=1}^n 1_S(\mathbf{X}_i),$$

and second order local sample moments by

$$\hat{\mu}_{2\mathbf{e}_j}(\mathbf{z}) = \frac{1}{\gamma^2} \sum_{i=1}^n (X_{ij} - z_j)^2 \, 1_S(\mathbf{X}_i) \Big/ \sum_{i=1}^n 1_S(\mathbf{X}_i), \qquad 1 \leqslant j \leqslant p,$$

and

$$\hat{\mu}_{\mathbf{e}_j + \mathbf{e}_k}(\mathbf{z}) = \frac{1}{\gamma^4} \sum_{i=1}^n (X_{ij} - z_j)(X_{ik} - z_k) \, 1_S(\mathbf{X}_i) \Big/ \sum_{i=1}^n 1_S(\mathbf{X}_i),$$

$$1 \leqslant j, k \leqslant p, \quad j \neq k.$$

Some guidelines for choices of the window S can be derived from the following asymptotic result. Certain rates of shrinkage of the window S_n to zero ensure the stochastic convergence of local sample moments to the local moments. Our main result on local sample moments concerns their convergence in distribution to a normal limit. The following shorthand notations will be useful: Given a multiindex α, define

$$\beta_\alpha = \frac{1}{3f(\mathbf{z})} \prod_{i=1}^p (\alpha_i + 1)^{-1} \sum_{i=1}^p \frac{\alpha_i}{\alpha_i + 3} D^{2\mathbf{e}_i} f(\mathbf{z})$$

and

$$v_\alpha = \frac{1}{2^p f(\mathbf{z})} \left\{ \prod_{i=1}^{p} (2\alpha_i + 1)^{-1} - \prod_{i=1}^{p} (\alpha_i + 1)^{-2} \right\}, \qquad \text{if all } \alpha_i \text{ are even,}$$

$$v_\alpha = \frac{1}{2^p f(\mathbf{z})} \prod_{i=1}^{p} (2\alpha_i + 1)^{-1}, \qquad \text{if at least one } \alpha_i \text{ is odd,} \quad 1 \leqslant i \leqslant p.$$

Furthermore, denote by $\xrightarrow{p}$ convergence in probability and by $\xrightarrow{\mathscr{D}}$ convergence in distribution as $n \to \infty$.

THEOREM 3.1. *Let α be a multiindex with at most two odd elements and assume that $(A1)$ holds and that $\mu_{2\alpha}(\mathbf{z})$ exists. As $n \to \infty$, assume that*

$$\gamma \to 0, \qquad n\gamma^{2(|\alpha^+| - |\alpha|) + p} \to \infty \tag{3.2}$$

and

$$n\gamma^{2(|\alpha^+| - |\alpha|) + p + 4} \to \lambda^2 \tag{3.3}$$

for some constant $\lambda \geqslant 0$. Then

$$\{n\gamma^{2(|\alpha^+| - |\alpha|) + p}\}^{1/2} (\hat{\mu}_\alpha(\mathbf{z}) - \mu_\alpha(\mathbf{z})) \xrightarrow{\mathscr{D}} \mathscr{N}(\lambda\beta_\alpha, v_\alpha). \tag{3.4}$$

The proof is in Section 5. If $\lambda = 0$, the asymptotic limiting distribution will be centered at 0. The asymptotic mean squared error (AMSE) is obtained by summing squared asymptotic bias and asymptotic variance as given in the limiting distribution (3.4), i.e.,

$$AMSE = (\gamma^2 \beta_\alpha)^2 + \frac{v_\alpha}{n\gamma^{2(|\alpha^+| - |\alpha|) + p}}. \tag{3.5}$$

It is easy to see that this is minimized for

$$\begin{aligned} \gamma^* &= \left\{ \frac{v_\alpha [2(|\alpha^+| - |\alpha|) + p]}{4\beta_\alpha^2 n} \right\}^{1/[2(|\alpha^+| - |\alpha|) + p + 4]} \\ &\sim n^{-1/[2(|\alpha^+| - |\alpha|) + p + 4]}, \end{aligned} \tag{3.6}$$

which gives the rate of convergence

$$AMSE \sim n^{-4/[2(|\alpha^+| - |\alpha|) + p + 4]}$$

for the asymptotic mean squared error of the local sample moments. The optimal scaling factor γ^* depends on unknown quantities, which could be estimated from the data, leading to a plug-in type bandwidth choice.

As a more specific example for formula (3.6), consider the important two-dimensional case $p=2$ and let $\alpha=(1,0)$. Then we obtain

$$\beta_{(10)}=\frac{f^{(20)}}{24f},\qquad v_{(10)}=\frac{1}{12f},$$

and

$$\gamma^*_{(10)}=\left(\frac{48f}{[f^{(20)}]^2\,n}\right)^{1/8},$$

as well as $AMSE\sim n^{-1/2}$.

Another consequence of (3.4) is the consistency of the local sample moments $\hat{\mu}_\alpha(\mathbf{z})$ for local moments $\mu_\alpha(\mathbf{z})$ as long as the bandwidths γ satisfy (3.2), (3.3).

COROLLARY 3.1. *Under the assumptions of Theorem* 3.1,

$$\hat{\mu}_\alpha(\mathbf{z})\xrightarrow{p}\mu_\alpha(\mathbf{z}). \tag{3.7}$$

As special cases, we obtain from (2.5), (2.6) and (3.7),

$$\hat{\mu}_{\mathbf{e}_k}(\mathbf{z})\xrightarrow{p}\frac{1}{3}\,\frac{D^{\mathbf{e}_k}f(\mathbf{z})}{f(\mathbf{z})}. \tag{3.8}$$

and

$$\hat{\mu}_{\mathbf{e}_j+\mathbf{e}_k}(\mathbf{z})\xrightarrow{p}\frac{1}{9}\,\frac{D^{\mathbf{e}_j+\mathbf{e}_k}f(\mathbf{z})}{f(\mathbf{z})},\qquad \text{for}\quad j\neq k,\ 1\leqslant j,k\leqslant p. \tag{3.9}$$

We turn now to local sample covariance matrices. Forming empirical covariances within local windows, one obtains

$$\begin{aligned}\hat{\sigma}_{jj}&=\frac{1}{\gamma^2}\left\{\frac{\sum_{i=1}^n (X_{ij}-z_j)^2\,1_S(\mathbf{X}_i)}{\sum_{i=1}^n 1_S(\mathbf{X}_i)}-\left[\frac{\sum_{i=1}^n (X_{ij}-z_j)\,1_S(\mathbf{X}_i)}{\sum_{i=1}^n 1_S(\mathbf{X}_i)}\right]^2\right\}\\&=\hat{\mu}_{2\mathbf{e}_j}(\mathbf{z})-\gamma^2\hat{\mu}^2_{\mathbf{e}_j}(\mathbf{z}),\qquad \text{for}\quad 1\leqslant j\leqslant p,\end{aligned}\tag{3.10}$$

and

$$\begin{aligned}\hat{\sigma}_{jk}&=\left\{\frac{\sum_{i=1}^n (X_{ij}-z_j)(X_{ik}-z_k)\,1_S(\mathbf{X}_i)}{\sum_{i=1}^n 1_S(\mathbf{X}_i)}\right.\\&\qquad\left.-\frac{\sum_{i=1}^n (X_{ij}-z_j)\,1_S(\mathbf{X}_i)}{\sum_{i=1}^n 1_S(\mathbf{X}_i)}\;\frac{\sum_{i=1}^n (X_{ik}-z_k)\,1_S(\mathbf{X}_i)}{\sum_{i=1}^n 1_S(\mathbf{X}_i)}\right\}\\&=\hat{\mu}_{\mathbf{e}_j+\mathbf{e}_k}(\mathbf{z})-\hat{\mu}_{\mathbf{e}_j}(\mathbf{z})\,\hat{\mu}_{\mathbf{e}_k}(\mathbf{z}),\qquad \text{for}\quad 1\leqslant j,k\leqslant p,\quad j\neq k.\end{aligned}\tag{3.11}$$

With covariances σ_{jk} (2.8) (see also (2.9), (2.10)), we have

COROLLARY 3.2. *Under the assumptions of Theorem* 3.1, *for* $\alpha = 2\mathbf{e}_j$, *assuming* $n\gamma^{p+4} \to 0$,

$$(n\gamma^p)^{1/2} (\hat{\sigma}_{jj} - \tfrac{1}{3}) \xrightarrow{\mathscr{D}} \mathscr{N}(0, v_{2\mathbf{e}_j}), \qquad 1 \leqslant j \leqslant p, \tag{3.12}$$

and for $\alpha = \mathbf{e}_j + \mathbf{e}_k$, *assuming* $n\gamma^{p+8} \to \lambda^2$, $\lambda \geqslant 0$,

$$(n\gamma^{p+4})^{1/2} (\hat{\sigma}_{jk} - \sigma_{jk}) \xrightarrow{\mathscr{D}} \mathscr{N}(\lambda\beta_{\mathbf{e}_j+\mathbf{e}_k}, v_{\mathbf{e}_j+\mathbf{e}_k}), \qquad 1 \leqslant j, k \leqslant p, \quad j \neq k. \tag{3.13}$$

The proof is in Section 5. It is clear that $\hat{\sigma}_{jj} \xrightarrow{p} \frac{1}{3}$ and $\hat{\sigma}_{jk} \xrightarrow{p} \sigma_{jk}$. Along the same lines, $\frac{1}{\gamma^2}\{\hat{\mu}_{2\mathbf{e}_j}(\mathbf{z}) - \hat{\sigma}_{jj}\} \xrightarrow{p} \mu^2_{\mathbf{e}_j}(\mathbf{z})$. We note in passing that there is an obvious alternative definition for $\hat{\sigma}_{jj}$, which is based on $\bar{\sigma}_{jj} = \hat{\mu}_{2\mathbf{e}_j}(\mathbf{z}) - \hat{\mu}^2_{\mathbf{e}_j}(\mathbf{z})$. Analogously to the proof of (3.12), one can show that $(n\gamma^p)^{1/2} (\bar{\sigma}_{jj} - 1/3) \xrightarrow{\mathscr{D}} \mathscr{N}(\lambda\beta_{2\mathbf{e}_j}, v_{2\mathbf{e}_j})$.

Specializing these results to the two-dimensional case, we infer from (3.12) and (3.13) that

$$\hat{\sigma}_{ii} = \tfrac{1}{3} + O_p((n\gamma^2)^{-1/2}), \qquad i = 1, 2, \tag{3.14}$$

and

$$\hat{\sigma}_{12} = \frac{1}{9} \frac{f^{(11)}f - f^{(10)}f^{(01)}}{f^2} + O_p((n\gamma^6)^{-1/2}). \tag{3.15}$$

4. APPLICATIONS OF LOCAL MOMENTS

We review here some statistical estimation problems which can be phrased in terms of the unifying concept of local moments. These problems concern the definition and estimation of a local dependence function; the estimation of derivatives of a density function; the problem of estimating tangents and normals of the density contours and the uphill search for modes of the density by a mode climbing algorithm.

4.1. *Measures of Local Dependence*

The local covariance (2.10) immediately suggests the following measure of local dependence between components X_j and X_k of a p-dimensional distribution:

$$\rho_{jk}(\mathbf{z}) = \left\{ \frac{D^{\mathbf{e}_j+\mathbf{e}_k} f(\mathbf{z})}{f(\mathbf{z})} - \frac{D^{\mathbf{e}_j} f(\mathbf{z})\, D^{\mathbf{e}_k} f(\mathbf{z})}{f(\mathbf{z})^2} \right\}, \qquad j \neq k. \tag{4.1}$$

For the two-dimensional case, this becomes

$$\rho(\mathbf{z}) = \frac{1}{f^2}\{ff^{(11)} - f^{(10)}f^{(01)}\}. \tag{4.2}$$

This latter measure of local dependence was considered previously as a local dependence function by Holland and Wang [8] and Jones [10], compare also Lancaster [11], Scarsini and Venetoulias [12] and Wang [17] for related discussions. We note that nonparametric measures of local dependence also found renewed interest in the context of regression analysis (Bjerve and Doksum [2]; Doksum *et al.* [5]). We have seen that for $\tilde{\sigma}_{12}$ (2.13), $9\tilde{\sigma}_{12} = \rho(\mathbf{z})$ (4.2), and that for the local sample covariance $\hat{\sigma}_{12}$ (3.11), $\hat{\sigma}_{12} \xrightarrow{p} \tilde{\sigma}_{12}$, according to Corollary 3.2, so that $9\hat{\sigma}_{12} \xrightarrow{p} \rho(\mathbf{z})$, see (3.15). This shows that the intuitive local sample covariance indeed provides a consistent estimate of the local dependence function $\rho(\cdot)$, and this holds for any two components of the p-variable random vector.

4.2. *Density Derivatives*

We consider the two-dimensional case, $p = 2$. Define the two-dimensional kernel density estimator with rectangular kernel

$$\hat{f} = \hat{f}(\mathbf{z}) = \frac{1}{4n\gamma^2} \sum_{i=1}^{n} 1_S(\mathbf{X}_i). \tag{4.3}$$

Abbreviating $f^{(10)} = f^{(10)}(\mathbf{z})$, $\hat{f}^{(10)} = \hat{f}^{(10)}(\mathbf{z})$, $f = f(\mathbf{z})$, (3.10) and (3.11) motivate the following estimators for derivatives, $\hat{f}^{(10)} = 3\hat{\mu}_{(10)}\hat{f}$, $\hat{f}^{(01)} = 3\hat{\mu}_{(01)}\hat{f}$ and $\hat{f}^{(11)} = 9\hat{\mu}_{(11)}\hat{f}$. As Corollary 3.1 yields $\hat{\mu}_\alpha \xrightarrow{p} \mu_\alpha$, consistency of these estimates is ensured by $\hat{f} \xrightarrow{p} f$, which follows from classical results on nonparametric density estimation when $\gamma \to 0$, $n\gamma^2 \to \infty$ (see for instance Cacoullos [4]).

Regarding the asymptotic limit distributions for estimates $\hat{f}^{(10)}$, $\hat{f}^{(01)}$ and $\hat{f}^{(11)}$, we obtain the following Corollary of Theorem 3.1.

COROLLARY 4.1. *If* $\gamma \to 0$, $n\gamma^4 \to \infty$, $n\gamma^8 \to 0$, *then*

$$(n\gamma^4)^{1/2}(\hat{f}^{(10)} - f^{(10)}) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \tfrac{3}{4}f), \tag{4.4}$$

and the same holds when replacing $\hat{f}^{(10)}$, $f^{(10)}$ *by* $\hat{f}^{(01)}$, $f^{(01)}$. *If* $\gamma \to 0$, $n\gamma^6 \to \infty$, $n\gamma^{10} \to 0$, *then*

$$(n\gamma^6)^{1/2}(\hat{f}^{(11)} - f^{(11)}) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \tfrac{9}{4}f). \tag{4.5}$$

These special cases are comparable to results obtained for kernel estimators of partial derivatives, see Singh [15]. Analogous results obviously hold for $\Re^p$, $p > 2$. Direct kernel estimation of derivatives in one dimension can be interpreted as using a basic nonnegative weight function multiplied with a polynomial so that the product satisfies certain moment conditions (Müller [13]). In local polynomial fitting, such kernels are implicitly constructed (Jones [9]). The "equivalent" kernels to the proposed moment based derivative estimates are the uniform kernels $1_{[-1,1]}$.

An interesting problem occurs when one wishes to estimate derivatives $f^{(20)}$, $f^{(02)}$ via local moments. These derivatives cannot be directly estimated, as they do not appear among the leading terms of the local moments. They do, however, appear in higher order terms such as

$$\mu_{(20)} = \frac{1}{3} + \frac{2}{45}\gamma^2 \frac{f^{(20)}}{f} + o(\gamma^2), \tag{4.6}$$

according to Lemma 5.1. This motivates the estimates

$$\hat{f}^{(20)} = \frac{45\hat{f}}{2\gamma^2}\left(\hat{\mu}_{(20)} - \frac{1}{3}\right), \tag{4.7}$$

analogously for $f^{(02)}$.

COROLLARY 4.2. *If* $\gamma \to 0$, $n\gamma^6 \to \infty$, *then*

$$\hat{f}^{(20)} \to f^{(20)}. \tag{4.8}$$

We note that the scaling for (4.7) requires that $n\gamma^6 \to \infty$ which means that $(1/\gamma^2)/(n\gamma^2)^{1/2} \to 0$. Thus, the scaling γ^{-2} in (4.7) is of smaller order than the scaling $(n\gamma^2)^{1/2}$ used for the corresponding asymptotic normality result (3.4) for $\hat{\mu}_{(20)}$, where it was required that $n\gamma^6 \to \lambda^2 \geqslant 0$.

4.3. *Density Contours, Tangents, and Normals*

We develop these concepts in $\Re^2$, keeping in mind that extensions to $\Re^p$ are readily available. The contours or level curves of a density f are defined by

$$C_\gamma = \{\mathbf{z} \in \Re^2 \mid f(\mathbf{z}) = \gamma\} \qquad \text{for} \quad \gamma > 0.$$

We assume that the density is smooth and does not have any plateaus, i.e., there are no sets with positive measure where the density has a constant value, so that the contours C_γ are simple differentiable curves in $\Re^2$.

A vector pointing in the direction of the steepest ascent of the density function at $\mathbf{z} \in \Re^2$ is referred to as normal vector. Its direction is given by the differential $(f^{(10)}(\mathbf{z}), f^{(01)}(\mathbf{z}))^T$; the line (or, in general, plane) orthogonal to this vector passing through $\mathbf{z}$ is the tangent. In $\Re^2$, the tangent consists of all $\mathbf{x}$ satisfying $f^{(10)}(\mathbf{z})(x_1 - z_1) + f^{(01)}(\mathbf{z})(x_2 - z_2) = 0$.

The direction of the normal vector can be easily estimated via local sample moments, observing that, according to Theorem 3.1, see also (3.9),

$$\begin{pmatrix} \hat{\mu}_{(10)} \\ \hat{\mu}_{(01)} \end{pmatrix} = \frac{1}{3f(\mathbf{z})} \begin{pmatrix} f^{(10)}(\mathbf{z}) \\ f^{(01)}(\mathbf{z}) \end{pmatrix} + O_p([n\gamma^4]^{-1/2}). \tag{4.9}$$

Accordingly, any vector (plane) orthogonal to $(\hat{\mu}_{(10)}, \hat{\mu}_{(01)})^T$ which contains $\mathbf{z}$ assumes the role of an estimated tangent vector (plane). The estimator (4.9) essentially is the rescaled sample mean of the data falling into the neighborhood S, centered at $\mathbf{z}$,

$$\bar{\mathbf{X}}(\mathbf{z}) = \sum_{i=1}^{n} (\mathbf{X}_i - \mathbf{z})\, 1_S(\mathbf{X}_i) \Big/ \sum_{i=1}^{n} 1_S(\mathbf{X}_i), \tag{4.10}$$

noting that $\bar{\mathbf{X}}(\mathbf{z}) = \gamma^2(\hat{\mu}_{(10)}, \hat{\mu}_{(01)})^T$.

4.4. *Mode Climbing*

One can employ the simple approach provided by (4.9), (4.10) to estimate normal directions for steepest ascent methods by noting that the normal direction corresponds to the direction of steepest ascent. When starting at an arbitrary point $\mathbf{z}$, iterative local averaging can be used to iteratively climb up the density slope towards the mode. The first version of a corresponding "Mean Update Algorithm" apparently goes back to Fwu, Tapia and Thompson [7]. Versions of such iterative mode finding algorithms were proposed in [7] and further discussed in Boswell [3], Elliott and Thomson [6] and Thompson and Tapia [16].

One version of this algorithm is as follows: (1) Start with a grid of points $\mathbf{z} \in \Re^2$ or with the observed data themselves; (2) For each starting point $\mathbf{z}$, obtain $\bar{\mathbf{X}}(\mathbf{z})$ (4.10); (3) Iterate step (2) by choosing $\bar{\mathbf{X}}(\mathbf{z})$ as new starting point, until there is no or only a small change in the starting point, i.e., $\mathbf{z}$ and $\bar{\mathbf{X}}(\mathbf{z})$ are close according to some criterion.

The starting points and their updates can be connected by an arrow, and these *mode climbing paths* help to visualize features of the density surface. An example of such paths can be seen in Fig. 1, which is based on $n = 500$ data sampled from the bivariate normal distribution with parameters $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$ and $\rho = cov(X_1, X_2) = 0.2$. The true mode is at the point (0, 0). We use $\gamma = 2$ in constructing the paths. The estimated mode is at (0.037, 0.047).

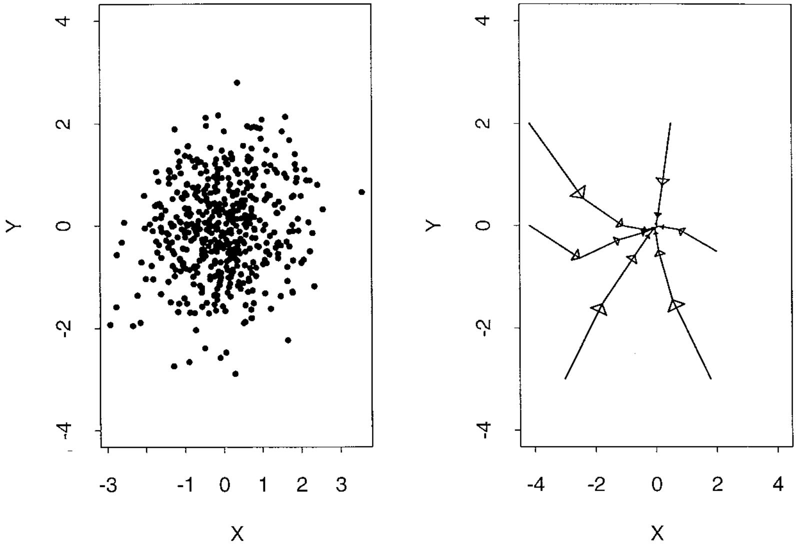

FIG 1. Data sample (left) and mode climbing path (right) for $\gamma = 2$ and 500 data randomly drawn from the bivariate normal density with $\mu_1 = \mu - 2 = 0$, $\sigma_1^2 = \alpha_2^2 = 1$, and $\rho = 0.2$.

For very large and high-dimensional continuous data sets such as one may encounter in "data mining" applications, the mode has thus the attractive property that it can be computed using exclusively "local" data repeatedly. Updating requires only a small fraction of the data. Only a subset of the data needs to be accessible at a time, which reduces memory requirements. This subset consists of data which are contained either in the updated window or in the previous window, when using updating formulas for the means. Using fast updating, the mode may be reached quickly. Furthermore, the length of mode climbing paths can be used as a distance measure from mode to data points, and thus these paths may be useful for cluster analysis and related applications.

Mode climbing paths could also be useful for tracking resources. Assume local information on the occurrence of certain phenomena is available, such as the frequency of a biological species in a local neighborhood. The goal is to find areas of peak occurrence by moving forward sequentially. The available information on local distribution can be used to form local means and to determine the normal direction. After progressing in that direction, one then would obtain a new local sample and normal direction. Iterating this process, one ultimately arrives in high-density areas. Of course, this approach works also to find the direction of steepest descent, for instant if the goal is to move away from danger zones corresponding to high density areas.

5. PROOFS AND AUXILIARY RESULTS

For all of the proofs, we denote the volume of a set A by $|A|$. For p-variate multiindices, we define the indicator sets

$$N_0 = \{\alpha\colon \alpha_i \text{ are even, } 1 \leqslant i \leqslant p\}$$
$$N_1 = \{\alpha\colon \text{exactly one of the } \alpha_i \text{ is odd, } 1 \leqslant i \leqslant p\}$$
$$N_2 = \{\alpha\colon \text{exactly two of the } \alpha_i \text{ are odd, } 1 \leqslant i \leqslant p\}.$$

In addition, the following abbreviations will be convenient. For a given multiindex α, let

$$\sigma_0(\alpha) = \sum_{i=1}^{p} \frac{\alpha_i + 1}{\alpha_i + 3} \frac{D^{2\mathbf{e}_i} f(z)}{f(\mathbf{z})},$$

$$\sigma_1(\alpha) = \sum_{i=1}^{p} \frac{\alpha_i + 1}{\alpha_i + 2} \frac{D^{2\mathbf{e}_i} f(\mathbf{z})}{f(\mathbf{z})} 1_{\{\alpha_i \text{ odd}\}}$$

$$\sigma_2(\alpha) = \sum_{j,\, k=1,\, j \neq k}^{p} \frac{\alpha_j + 1}{\alpha_j + 2} \frac{\alpha_k + 1}{\alpha_k + 2} \frac{D^{\mathbf{e}_j + \mathbf{e}_k} f(\mathbf{z})}{f(\mathbf{z})} 1_{\{\alpha_j \text{ and } \alpha_k \text{ odd}\}},$$

$$\sigma_3(\alpha) = \sum_{i=1}^{p} \frac{\alpha_i}{\alpha_i + 3} D^{2\mathbf{e}_i} f(\mathbf{z}),$$

$$\pi(\alpha) = \prod_{i=1}^{p} (\alpha_i + 1)^{-1},$$

and the conditional moment

$$M(\alpha) = E\{(\mathbf{x} - \mathbf{z})^\alpha \mid \mathbf{x} \in S\}.$$

5.1. *Proof of Theorem* 2.1

Theorem 2.1 is an immediate consequence of the following lemma, which provides a more detailed result for conditional moments from which local moments are obtained by taking the limit as $\gamma \to 0$. Following the details of the proof, more general local moments for the cases where more than two indices α_i are odd can be easily obtained.

LEMMA 5.1. *For a multiindex* α *with* $\alpha \in N_0 \cup N_1 \cup N_2$,

$$M(\alpha) = \gamma^{|\alpha^+|} \pi(\alpha) \left\{ \left[1 + \frac{\gamma^2}{3} \sum_{i=1}^{p} \frac{\alpha_i}{\alpha_i + 3} \frac{D^{2\mathbf{e}_i} f(\mathbf{z})}{f(\mathbf{z})} + o(\gamma^2) \right] 1_{\{\alpha \in N_0\}} \right.$$
$$\left. + [\sigma_1(\alpha) + O(\gamma^2)]\, 1_{\{\alpha \in N_1\}} + [\sigma_2(\alpha) + O(\gamma^2)]\, 1_{\{\alpha \in N_2\}} \right\}. \quad (5.1)$$

Proof. By a Taylor expansion of f around $\mathbf{z}$,

$$f(\mathbf{x}) = f(\mathbf{z}) + \sum_{i=1}^{p} D^{\mathbf{e}_i} f(\mathbf{z})(x_i - z_i)$$
$$+ \tfrac{1}{2} \sum_{i,\, j=1}^{p} D^{\mathbf{e}_i + \mathbf{e}_j} f(\mathbf{z})(x_i - z_i)(x_j - z_j) + o(\gamma^2), \tag{5.2}$$

and thus

$$\int_S f(\mathbf{x})\, d\mathbf{x} = |S|\, f(\mathbf{z}) \left[1 + \frac{\gamma^2}{6 f(\mathbf{z})} \sum_{i=1}^{p} D^{2\mathbf{e}_i} f(\mathbf{z}) + o(\gamma^2) \right],$$

whence

$$\left(\int_S f(\mathbf{x})\, d\mathbf{x} \right)^{-1} = \frac{1}{|S|\, f(\mathbf{z})} \left[1 - \frac{\gamma^2}{6 f(\mathbf{z})} \sum_{i=1}^{p} D^{2\mathbf{e}_i} f(\mathbf{z}) + o(\gamma^2) \right]. \tag{5.3}$$

Observe

$$\int_S (\mathbf{x} - \mathbf{z})^\alpha \, d\mathbf{x} = |S|\, \gamma^{|\alpha|} \pi(\alpha)\, 1_{\{\alpha \in N_0\}},$$

$$\int_S (\mathbf{x} - \mathbf{z})^\alpha \sum_{i=1}^{p} D^{2\mathbf{e}_i} f(\mathbf{z})(x_i - z_i)\, d\mathbf{z} = |S|\, \gamma^{|\alpha|+1} \pi(\alpha)\, f(\mathbf{z})\, \sigma_1(\alpha)\, 1_{\{\alpha \in N_1\}}$$

and

$$\int_S (\mathbf{x} - \mathbf{z})^\alpha \sum_{i,\, j=1}^{p} D^{\mathbf{e}_i + \mathbf{e}_j} f(\mathbf{z})(x_i - z_i)(x_j - z_j)\, d\mathbf{x}$$
$$= |S|\, \gamma^{|\alpha|+2} \pi(\alpha)\, f(\mathbf{z}) [\sigma_0(\alpha)\, 1_{\{\alpha \in N_0\}} + \sigma_2(\alpha)\, 1_{\{\alpha \in N_2\}}].$$

With (5.2), this leads to

$$E[(\mathbf{X} - \mathbf{z})^\alpha\, 1_S(\mathbf{X})] = |S|\, \gamma^{|\alpha^+|} \pi(\alpha)\, f(\mathbf{z}) \left[\left(1 + \frac{\gamma^2}{2} \sigma_0(\alpha) \right) 1_{\{\alpha \in N_0\}} \right.$$
$$\left. + \sigma_1(\alpha)\, 1_{\{\alpha \in N_1\}} + \frac{1}{2} \sigma_1(\alpha)\, 1_{\{\alpha \in N_2\}} + o(\gamma^2) \right]. \tag{5.4}$$

Combining (5.4) with (5.3), observing

$$M(\alpha) = \frac{\int_S (\mathbf{x} - \mathbf{z})^\alpha f(\mathbf{x})\, d\mathbf{x}}{\int_S f(\mathbf{x})\, d\mathbf{x}},$$

and collecting the terms according to whether $\alpha \in N_0$, $\alpha \in N_1$, or $\alpha \in N_2$, then leads to the result.

5.3. *Proof of Theorem* 2.2

Choosing $\alpha = \mathbf{e}_k \in N_1$ in the conditional moment $M(\alpha)$, such that $|\alpha| = 1$, $|\alpha^+| = 2$, we find from (5.1) that

$$M(\mathbf{e}_k) = \frac{D^{\mathbf{e}_k} f(\mathbf{z})}{3f(\mathbf{z})} \gamma^2 + o(\gamma^2). \tag{5.5}$$

Analogously, choosing $\alpha = 2\mathbf{e}_k$, such that $\alpha \in N_0$, $|\alpha| = 2$, $|\alpha^+| = 2$, we find from (5.1) that

$$M(2\mathbf{e}_k) = \frac{\gamma^2}{3} + \frac{2D^{2\mathbf{e}_k} f(\mathbf{z})}{45 f(\mathbf{z})} \gamma^4 + o(\gamma^4). \tag{5.6}$$

Furthermore, choosing $\alpha = \mathbf{e}_j + \mathbf{e}_k$ for $j \neq k$, one has $\alpha \in N_2$, $|\alpha| = 2$, $|\alpha^+| = 4$, and, according to (5.1),

$$M(\mathbf{e}_j + \mathbf{e}_k) = \frac{D^{\mathbf{e}_j + \mathbf{e}_k} f(\mathbf{z})}{9 f(\mathbf{z})} \gamma^4 + o(\gamma^4). \tag{5.7}$$

We conclude

$$Var(X_k \mid \mathbf{X} \in S) = M(2\mathbf{e}_k) - [M(\mathbf{e}_k)]^2,$$

$$Cov(X_j, X_k \mid \mathbf{X} \in S) = M(\mathbf{e}_j + \mathbf{e}_k) - M(\mathbf{e}_j)\, M(\mathbf{e}_k),$$

and (2.9), (2.10) follow by inserting (5.5)–(5.7) into these expressions.

5.3. *Proof of Theorem* 3.1

It follows from $\gamma \to 0$ that

$$E\left[\frac{1}{n\,|S|} \sum_{i=1}^{n} 1_S(\mathbf{X}_i)\right] = \frac{1}{|S|} \int_S f(\mathbf{x})\, d\mathbf{x} \to f(\mathbf{z}),$$

$$E\left[\frac{1}{n\,|S|} \sum_{i=1}^{n} 1_S(\mathbf{X}_i)\right]^2 = \frac{1}{|S|^2} \left(\int_S f(\mathbf{x})\, d\mathbf{x}\right)^2 \to f(\mathbf{z})^2, \qquad \text{as} \quad n \to \infty.$$

This implies that

$$\frac{1}{n\,|S|} \sum_{i=1}^{n} 1_S(\mathbf{X}_i) \xrightarrow{p} f(\mathbf{z}). \tag{5.8}$$

Defining a triangular scheme of random variables

$$U_{i,n} = \frac{1}{n\gamma^{|\alpha^+|}\,|S|}\,[(\mathbf{X}_i - \mathbf{z})^\alpha - \gamma^{|\alpha^+|}\mu_\alpha(\mathbf{z})]\,1_S(\mathbf{X}_i),$$

we can write for local sample moments (3.1)

$$\begin{aligned}\hat{\mu}_\alpha(\mathbf{z}) - \mu_\alpha(\mathbf{z}) &= \sum_{i=1}^{n} \frac{(\mathbf{X}_i - \mathbf{z})^\alpha\, 1_S(\mathbf{X}_i) - \gamma^{|\alpha^+|}\mu_\alpha(\mathbf{z})\, 1_S(\mathbf{X}_i)}{\gamma^{|\alpha^+|}\sum_{i=1}^{n} 1_S(\mathbf{X}_i)}\\ &= \sum_{i=1}^{n} U_{i,n} \Big/ \frac{1}{n\,|S|}\sum_{i=1}^{n} 1_S(\mathbf{X}_i). \qquad (5.9)\end{aligned}$$

By Theorem 2.1 and (5.4), using

$$E\,1_S(\mathbf{X}_i) = |S|\; f(\mathbf{z}) \left[1 + \frac{\gamma^2}{6} \sum_{i=1}^{p} \frac{D^{2\mathbf{e}_i} f(\mathbf{z})}{f(\mathbf{z})} + o(\gamma^2) \right],$$

we find

$$\begin{aligned}&E[(\mathbf{X} - \mathbf{z})^\alpha\, 1_S(\mathbf{X}) - \gamma^{|\alpha^+|}\mu_\alpha(\mathbf{z})\, 1_S(\mathbf{X})]\\ &\quad = |S|\; \gamma^{|\alpha^+|}\pi(\alpha)\; f(\mathbf{z}) \left[\frac{\gamma^2}{3}\,\sigma_3(\alpha)/f(\mathbf{z}) + o(\gamma^2) \right]. \qquad (5.10)\end{aligned}$$

Furthermore, noting that $\gamma^{|\alpha^+|} = o(\gamma^{|\alpha|})$ if $|\alpha^+| > |\alpha|$, and that $|\alpha^+| = |\alpha|$ for $\alpha \in N_0$, we find

$$\begin{aligned}&E[(\mathbf{X} - \mathbf{z})^\alpha\, 1_S(\mathbf{X}) - \gamma^{|\alpha^+|}\mu_\alpha(\mathbf{z})\, 1_S(\mathbf{X})]^2\\ &\quad = |S|\; f(\mathbf{z}) \left[\gamma^{2\,|\alpha|}\pi(2\alpha) \left\{ 1 + \frac{\gamma^2}{2}\,\sigma_0(2\alpha) + o(\gamma^2) \right\} - \gamma^{2\,|\alpha^+|}(\pi(\alpha))^2 \right.\\ &\qquad \left. \times \{ 1_{\{\alpha \in \mathbf{N}_0\}} + (\sigma_1(\alpha))^2\, 1_{\{\alpha \in \mathbf{N}_1\}} + (\sigma_2(\alpha))^2\, 1_{\{\alpha \in \mathbf{N}_2\}} + O(\gamma^2) \} \right]\\ &\quad = |S|\; f(\mathbf{z})\, \gamma^{2\,|\alpha|}[\pi(2\alpha) - (\pi(\alpha))^2\, 1_{\{\alpha \in \mathbf{N}_0\}}] + o(|S|\, \gamma^{2\,|\alpha|}). \qquad (5.11)\end{aligned}$$

Defining now the triangular array of i.i.d. random variables

$$V_{i,n} = [n\gamma^{p + 2(|\alpha^+| - |\alpha|)}]^{1/2}\, U_{i,n}, \qquad 1 \leqslant i \leqslant n, \quad n = 1, 2, ...$$

we obtain from (5.10) and (3.3) that

$$E(U_{i,n}) = \frac{\gamma^2}{3n}\,\pi(\alpha)\,\sigma_3(\alpha) + o\left(\frac{\gamma^2}{n}\right) \qquad (5.12)$$

and

$$E\left(\sum_{i=1}^{n} V_{i,n}\right)=\frac{1}{3}\lambda^{1/2}\pi(\alpha)\sum_{i=1}^{p}\frac{\alpha_i}{\alpha_i+3}D^{2\mathbf{e}_i}f(\mathbf{z})+o(1).$$

Similarly, observing $|S|=(2\gamma)^p$ and (5.11),

$$\begin{aligned} var(U_{i,n}) &= E(U_{i,n}^2)(1+o(1)) \\ &= (n^2\gamma^{2(|\alpha^+|-|\alpha|)}|S|)^{-1} \\ &\quad\times\{f(\mathbf{z})[\pi(2\alpha)-(\pi(\alpha))^2 1_{\{\alpha\in\mathbf{N}_0\}}]+o(1)\} \end{aligned} \tag{5.13}$$

and

$$var\left(\sum_{i=1}^{n} V_{i,n}\right)=2^{-p}f(\mathbf{z})[\pi(2\alpha)-(\pi(\alpha))^2 1_{\alpha\in\mathbf{N}_0}]+o(1).$$

Applying Lindeberg's Central Limit Theorem (see, e.g., Billingsley [1]) then shows

$$\sum_{i=1}^{n} V_{i,n}\xrightarrow{\mathscr{D}}\mathscr{N}\left(\tfrac{1}{3}\lambda^{1/2}\pi(\alpha)\,\sigma_3(\alpha),\,2^{-p}f(\mathbf{z})[\pi(2\alpha)-(\pi(\alpha))^2 1_{\{\alpha\in\mathbf{N}_0\}}]\right).$$

The result follows by applying (5.8), (5.9) and Slutsky's theorem.

5.4. *Proof of Corollary* 3.2

Observing

$$(n\gamma^p)^{1/2}(\hat{\mu}_{2\mathbf{e}_j}(\mathbf{z})-\mu_{2\mathbf{e}_j}(\mathbf{z}))\xrightarrow{\mathscr{D}}\mathscr{N}(0,v_{2\mathbf{e}_j}),$$

$\mu_{2\mathbf{e}_j}(\mathbf{z})=\frac{1}{3}$ and $(n\gamma^p)^{1/2}\gamma^2\hat{\mu}_{\mathbf{e}_j}(\mathbf{z})\xrightarrow{p}0$ then implies (3.12). For (3.13), we use

$$\begin{aligned}(n\gamma^{p+4})^{1/2}(\hat{\sigma}_{jk}-\sigma_{jk}) &= (n\gamma^{p+4})^{1/2}\{(\hat{\mu}_{\mathbf{e}_j+\mathbf{e}_k}(\mathbf{z})-\mu_{\mathbf{e}_j+\mathbf{e}_k}(\mathbf{z})) \\ &\quad -[(\hat{\mu}_{\mathbf{e}_1}(\mathbf{z})-\mu_{\mathbf{e}_1}(\mathbf{z}))\hat{\mu}_{\mathbf{e}_2}(\mathbf{z})-(\hat{\mu}_{\mathbf{e}_2}(\mathbf{z}))\mu_{\mathbf{e}_1}(\mathbf{z})]\}.\end{aligned}$$

Observing $\hat{\mu}_{\mathbf{e}_i}(\mathbf{z})\xrightarrow{p}\mu_{\mathbf{e}_i}(\mathbf{z})$, $\hat{\mu}_{\mathbf{e}_i}(\mathbf{z})-\mu_{\mathbf{e}_i}(\mathbf{z})=O_p((n\gamma^{p+2})^{-1/2})$ and Theorem 3.1 implies the result.

5.5. *Proof of Corollary* 4.1

Note that

$$\hat{f}^{(10)}-f^{(10)}=3\hat{\mu}_1(\hat{f}-f)+3f\left(\hat{\mu}_1-\frac{f^{(10)}}{3f}\right)$$

and

$$(n\gamma^{p+4})^{1/2}\left(\hat{\mu}_1 - \frac{f^{(10)}}{3f}\right) \xrightarrow{\mathscr{D}} \mathscr{N}\left(0, \frac{1}{12f}\right),$$

according to Theorem 3.1.

Basic results on kernel density estimation (see, e.g., Scott [14]) imply that $\hat{f} - f = O_p((n\gamma^2)^{-1/2})$, so that

$$(n\gamma^4)^{1/2}\,(3\hat{\mu}_1)(\hat{f} - f) = o_p(1).$$

This implies the result for $\hat{f}^{(10)}$, and the other cases are proved similarly.

5.6. *Proof of Corollary* 4.2

The result follows from the following more general Lemma. To see this, choose $p = 2$ and $\alpha = (2, 0)$ and observe that $\mu_{(20)} = 1/3$ and $2(|\alpha^+| - |\alpha|) + p + 4 = 6$.

LEMMA 5.2. *Assume that* α *is a multiindex with at most two odd elements*, $(A1)$ *holds and* $\mu_{2\alpha}(\mathbf{z})$ *exists. Furthermore, assume that* $\gamma \to 0$, *and* $n\gamma^{2(|\alpha^+| - |\alpha|) + p + 4} \to \infty$. *Then*

$$\frac{1}{\gamma^2}\,(\hat{\mu}_\alpha - \mu_\alpha) \xrightarrow{p} \frac{\pi(\alpha)}{3}\,\frac{\sigma_3(\alpha)}{f(\mathbf{z})}. \tag{5.14}$$

Proof. With $U_{i,n}$ as in (5.9),

$$\frac{1}{\gamma^2}\,(\hat{\mu}_\alpha - \mu_\alpha) = \frac{1}{n\,|S|}\sum_{i=1}^{n} 1_S(\mathbf{X}_i)\,\frac{1}{\gamma^2}\sum_{i=1}^{n} U_{i,n}.$$

By (5.13),

$$var\left(\frac{1}{\gamma^2}\sum_{i=1}^{n} U_{i,n}\right) \sim [n\gamma^{2(|\alpha^+| - |\alpha|) + p + 4}]^{-1}, \tag{5.15}$$

and by (5.12),

$$E\left(\frac{1}{\gamma^2}\sum_{i=1}^{n} U_{i,n}\right) \to \frac{\pi(\alpha)}{3}\,\sigma_3(\alpha). \tag{5.16}$$

Now (5.14) follows by combining (5.15) and (5.16) with (5.8).

ACKNOWLEDGMENTS

We thank Professor David W. Scott for pointing out the related work on iterative mode finding algorithms and Professor James R. Thompson for sending his papers on this topic. We are particularly grateful to a referee for detailed comments which led to many improvements in the presentation. This research was supported in part by NSF Grants DMS-96-25984 and NSA Grant MDA 904-96-10026.

REFERENCES

1. P. Billingsley, "Probability and Measure," Wiley, New York, 1986.
2. S. Bjerve and K. Doksum, Measures of association as functions of covariate values, *Ann. Statist.* **21** (1993), 890–902.
3. S. B. Boswell, Nonparametric mode estimation for higher dimensional densities, Doctoral dissertation, Rice University, 1983.
4. T. Cacoullos, Estimation of a multivariate density, *Ann. Inst. Statist. Math.* **18** (1966), 179–189.
5. K. Doksum, S. Blyth, E. Bradlow, X. L. Meng, and H. Zhao, Correlation curves as local measures of variance explained by regression, *J. Amer. Statist. Assoc.* **87** (1994), 571–582.
6. M. N. Elliott and J. R. Thompson, An exploratory algorithm for the estimation of mode location and numerosity in multidimensional data, *in* "Proceedings of the 38th Conference on Design of Experiments in Army Research Development and Testing," pp. 229–244, 1993.
7. C. Fwu, R. A. Tapia, and J. R. Thompson, The nonparametric estimation of probability densities in ballistics research, *in* "Proceedings of the 26th Conference of the Design of Experiments in Army Research Development and Testing," pp. 309–326, 1981.
8. P. W. Holland and Y. J. Wang, Dependence function for continuous bivariate densities, *Comm. Statist.-Theory Methods A* **16** (1987), 863–876.
9. M. C. Jones, On kernel density derivative estimation, *Comm. Statist. Theory Methods A* **23** (1994), 2133–2139.
10. M. C. Jones, The local dependence function, *Biometrika* **83** (1996), 899–904.
11. H. O. Lancaster, Correlation and complete dependence of random variables, *Ann. Math. Statist.* **34** (1963), 1315–1321.
12. H. G. Müller, On the boundary kernel method for nonparametric curve estimation near endpoints, *Scand. J. Statist.* **20** (1993), 313–328.
13. A. Scarsini and A. Venetoulias, Bivariate distributions with nonmonotone dependence structure, *J. Amer. Statist. Assoc.* **88** (1993), 338–344.
14. D. W. Scott, "Multivariate Density Estimation: Theory, Practice and Visualization," Wiley, New York, 1992.
15. R. S. Singh, Nonparametric estimation of mixed partial derivates of a multivariate density, *J. Multivariate Anal.* **6** (1976), 111–122.
16. J. R. Thompson and R. A. Tapia, "Nonparametric Function Estimation, Modelling and Simulation," SIAM, Philadelphia, 1990.
17. Y. J. Wang, Construction of continuous bivariate density functions, *Statist. Sinica* **3** (1993), 173–187.