Biometrika (1991), **78**, 1, *pp.* 197–205
Printed in Great Britain

Feasibility of multivariate density estimates

BY DAVID W. SCOTT
Department of Statistics, Rice University, Houston, Texas, 77251, U.S.A.
AND M. P. WAND
Department of Statistics, Texas A&M University, College Station, Texas, 77843, U.S.A.

SUMMARY

The 'curse of dimensionality' has been interpreted as suggesting that kernel methods have limited applicability in more than several dimensions. In this note, qualitative and quantitative performance measures for multivariate density estimates are examined. Optimal pointwise and global window widths for mean absolute and mean squared errors are compared for multivariate data. One result is that the optimal pointwise absolute and squared error window widths are nearly equal for all dimensions. We also show that sample size requirements predicted by absolute rather than squared error criterion are substantially less. Further reductions are realized by using a coefficient of variation criterion. Finally, an example of a 10-dimensional kernel density estimate is given. It is suggested that the true nature of the curse of dimensionality is as much the lack of full rank as sparseness of the data.

Some key words: Coefficient of variation; Curse of dimensionality; Kernel density estimation; Mean absolute error; Mean squared error; Projection pursuit.

1. INTRODUCTION

It is widely accepted that the 'curse of dimensionality' (Huber, 1985) prohibits application of nonadaptive kernel smoothing methodology with multivariate data, unless an unusually large data set is available. After all, kernel smoothers use local averaging and multivariate neighbourhoods tend to be empty. Friedman & Stuetzle (1981) provided a vivid illustration of this fact for points uniformly distributed in a unit cube in R^{10}. They observed that a cubical histogram bin with edges of length 0·1 would contain only 10^{-10} of the sample; on the other hand, if the bin is to cover 10% of the sample, the bin edge length should be $(0{\cdot}1)^{1/10} \doteq 0{\cdot}79$. Thus the bins are either empty or not local. Friedman, Stuetzle & Schroeder (1984) performed a simulation in R^{10} with $n = 225$ using a kth nearest density neighbour estimator (Loftsgaarden & Quesenberry, 1965) and demonstrated the poor quality of even adaptive kernel estimates with small samples; see § 4 below.

Tables computed by Epanechnikov (1969) and Silverman (1986) based upon squared error provide some theoretical basis for pessimism. Both considered the case where the density f is $N(0, I_d)$. Squared error is not dimensionless, making comparisons of errors among different dimensions difficult. Epanechnikov attempted to create a meaningful dimensionless quantity based on mean integrated squared error, MISE, by dividing by

$$\int_{R^d} f(x)^2 \, dx.$$

It is not immediately clear how useful this normalization really is. However, Silverman (1986) avoided this normalization question by simply computing the optimal exact mean squared error, MSE, at the mode $x=0$ relative to $\{f(0)\}^2$. An equivalent and also dimensionless quantity is the relative root mean squared error

$$\text{RRMSE}\,(x) \equiv \frac{[\text{MSE}\,\{\hat{f}(x;\,h)\}]^{\frac{1}{2}}}{f(x)}. \tag{1·1}$$

For a relative root mean squared error of $10^{\frac{1}{2}}\% = 31{\cdot}6\%$ at the origin, Silverman computed that sample sizes of 4, 223, 10 730 and 841 565 are required in 1, 4, 7 and 10 dimensions, respectively; $n = 3{\cdot}04$ is the exact one-dimensional requirement. Silverman remarks that to form a comparable table using the global error measure of Epanechnikov, the counts above should be increased by approximately 70%. This observation lends validity to Epanechnikov's choice of criterion. The progression is even more pessimistic when a smaller value of the relative root mean squared error is chosen. For example, the respective samples sizes for RRMSE $= 14{\cdot}5\%$ are 50, $7{\cdot}15\times10^3$, $1{\cdot}25\times10^6$ and $3{\cdot}69\times10^8$; see Table 3 below.

For many applications such as clustering, identifying peaks is of some importance. It is well-known that peaks are flattened and spread out by kernel methods. In this situation the relative root mean squared error criterion can be replaced by a more relevant criterion, the sample root coefficient of variation, RCV, defined by

$$\text{RCV}\,(x) \equiv \frac{[\text{var}\,\{\hat{f}(x;\,h)\}]^{\frac{1}{2}}}{E\{\hat{f}(x;\,h)\}}, \tag{1·2}$$

which is the inverse of the signal-to-noise criterion. A small value of this criterion in high dimensions would suggest that widely separated peaks will be identifiable even if the estimates are biased downward with a large window width. This criterion will be examined in § 3.

However interesting pointwise results are, a realistic picture of global estimation error would be more useful, since the tails of a density dominate in higher dimensions. Hall & Wand (1988) derived some useful approximations for mean integrated absolute error in the univariate setting. Previously, only upper bounds to the mean integrated absolute error were available (Devroye & Györfi, 1985). Hall & Wand found that the optimal smoothing parameters for absolute and squared errors are similar for many univariate densities. Since absolute error is a dimensionless quantity and has an appealing misclassification interpretation for any dimension (Devroye & Györfi, 1985), similar multivariate results would be useful. This subject is discussed below in § 2.

Finally, we reconsider the 10-dimensional simulation example of Friedman et al. (1984), but using a nonadaptive fixed kernel estimator. As we show in § 4, the estimates reveal the 10-dimensional structure quite well. We conclude with some thoughts on the contradictory evidence and speculate on other notions and consequences of the curse of dimensionality.

2. MULTIVARIATE THEORY FOR MEAN ABSOLUTE ERROR

2·1. *Optimal mean absolute error window widths*

Let $X_1, \ldots, X_n$ be a set of independent R^d-valued random vectors each having probability density f. In this section we assume that f is bounded with all second-order partial derivatives bounded and continuous. For a vector $z = (z_1, \ldots, z_d)$ we define

$\nabla^2 f(z) = \Sigma_j (\partial^2/\partial z_j^2) f(z)$. We consider the class of kernel estimators of $f(x)$ at a point $x \in R^d$ given by

$$\hat{f}(x; h) = (nh^d)^{-1} \sum_{i=1}^{n} K\{(x - X_i)/h\},$$

where the kernel K is a d-variate probability density satisfying

$$\int_{R^d} zK(z)dz = 0$$

and $\kappa \equiv \int z_j^2 K(z)dz$, where the integral is over the range R^d, is nonzero and independent of $j = 1, \ldots, d$. The window width $h = h_n$ is a positive, deterministic sequence for which $h \to 0$ and $nh^d \to \infty$ as $n \to \infty$.

The mean absolute error, MAE, of $f(x)$ is

$$\text{MAE}\{\hat{f}(x; h)\} = E|\hat{f}(x; h) - f(x)|.$$

As stated in (2·9) of Hall & Wand (1988) we have

$$\hat{f}(x; h) - f(x) = h^2 b_d(x) + (nh^d)^{-\frac{1}{2}} \sigma_d(x) Z_1 + o(h^2),$$

where

$$b_d(x) \equiv \tfrac{1}{2}\kappa \nabla^2 f(x), \quad \sigma_d^2(x) \equiv f(x) \int_{R^d} K(z)^2 dz, \quad Z_1 = Z_1(x)$$

is asymptotically standard univariate normal. Define the function ψ to be $\psi(u) = E|Z - u|$ for a $N(0, 1)$ random variable Z. It follows that

$$\text{MAE}\{\hat{f}(x; h)\} = \text{AMAE}\{\hat{f}(x; h)\} + o\{h^2 + (nh^d)^{-\frac{1}{2}}\},$$

where the asymptotic mean absolute error is

$$\text{AMAE}\{\hat{f}(x; h)\} = (nh^d)^{-\frac{1}{2}} \sigma_d(x) \psi\left\{\frac{b_d(x)(nh^{4+d})^{\frac{1}{2}}}{\sigma_d(x)}\right\}. \tag{2·1}$$

It can be shown that $\psi(u) = 2u\Phi(u) + 2\phi(u) - u$, where Φ and ϕ are, respectively, the distribution function and density of the standard normal distribution. The expression in (2·1) is minimized when

$$h_1^*(x) \equiv [\alpha_d^2 \sigma_d^2(x)/\{b_d^2(x)n\}]^{1/(4+d)},$$

where α_d is the unique positive solution to $4\alpha_d\{\Phi(\alpha_d) - \frac{1}{2}\} - \phi(\alpha_d)d = 0$. Examples of α_d include $\alpha_1 = 0{\cdot}4809$, $\alpha_{10} = 1{\cdot}2196$ and $\alpha_{100} = 2{\cdot}1308$.

Standard squared error theory (Cacoullos, 1966) reveals that the window width that minimizes the asymptotic mean squared error of $\hat{f}(x; h)$ is

$$h_2^*(x) = [d\sigma_d^2(x)/\{4b_d^2(x)n\}]^{1/(4+d)}.$$

The ratio of these two asymptotically optimal window widths is

$$\tau(d) \equiv h_1^*(x)/h_2^*(x) = (4\alpha_d^2/d)^{1/(4+d)},$$

which, remarkably, depends only on d and not on x, f or K. This ratio is within a few percent of unity for all values of d, the difference being greatest when $d = 9$; see Table 1. Also $\tau(d) \to 1$ as $d \to \infty$.

2·2. *Optimal mean integrated absolute error window widths*

For global estimation an appropriate measure of absolute error is the mean integrated absolute error, MIAE. This is simply the expected L_1 distance between $\hat{f}(\,.\,; h)$ and f and

Table 1. *Constants for the asymptotic mean absolute error, mean integrated absolute error, and mean integrated squared error criteria for multivariate independent normal data and kernel*

d	α_d	Ratio $\tau(d)$	c_{1d}^*	C_{1d}^*	c_{2d}^*
1	0·4809	0·9846	1·030	1·042	1·059
2	0·6573	0·9760	1·043	1·319	1·000
3	0·7808	0·9708	1·078	1·607	0·969
4	0·8769	0·9677	1·117	1·884	0·951
5	0·9558	0·9658	1·151	2·146	0·940
6	1·0228	0·9646	1·179	2·393	0·933
7	1·0809	0·9639	1·201	2·628	0·929
8	1·1322	0·9636	1·220	2·853	0·926
9	1·1781	0·9635	1·235	3·070	0·925
10	1·2196	0·9636	1·249	3·281	0·925
11	1·2575	0·9638	1·260	3·485	0·924
12	1·2922	0·9640	1·271	3·685	0·925
13	1·3244	0·9644	1·280	3·879	0·925
14	1·3542	0·9647	1·290	4·069	0·926
15	1·3821	0·9651	1·298	4·254	0·927
100	2·1308	0·9837	—	—	0·969
1000	2·9091	0·9966	—	—	0·995

is given by

$$\text{MIAE}\{\hat{f}(\,.\,;h)\} \equiv E\int_{R^d} |\hat{f}(x;h)-f(x)|dx.$$

If f satisfies a weak moment condition then it follows from (2·1) that

$$\text{MIAE}\{\hat{f}(\,.\,;h)\} = \text{AMIAE}\{\hat{f}(\,.\,;h)\} + o\{h^2+(nh^d)^{-\frac{1}{2}}\},$$

where the asymptotic mean integrated absolute error is (Hall & Wand, 1988)

$$\text{AMIAE}\{\hat{f}(\,.\,;h)\} = (nh^d)^{-\frac{1}{2}}\int_{R^d} \sigma_d(x)\psi\{b_d(x)(nh^{4+d})^{\frac{1}{2}}/\sigma_d(x)\}dx. \tag{2·2}$$

The window width which minimizes (2·2) is therefore

$$h_1^* = (v^*)^{2/(4+d)}n^{-1/(4+d)}, \tag{2·3}$$

where v^* is the unique positive solution to

$$\int_{R^d} \sigma_d(x)(4vr_d(x)[\Phi\{vr_d(x)\}-\tfrac{1}{2}]-d\phi\{vr_d(x)\})dx = 0 \tag{2·4}$$

with $r_d(x) \equiv b_d(x)/\sigma_d(x)$.

As has previously been established (Epanechnikov, 1969), the $\text{MISE}\{\hat{f}(\,.\,;h)\}$ has its asymptotic version minimized at $h_2^* = c_{2d}^* n^{-1/(4+d)}$, where

$$c_{2d}^* = \left[d\int_{R^d} K^2(z)dz \Big/ \left\{4\int_{R^d} b_d^2(x)dx\right\}\right]^{1/(4+d)}. \tag{2·5}$$

3. MULTIVARIATE NORMAL EXAMPLE

Following Epanechnikov (1969) we consider the case where both f and K are $N(0, I_d)$ densities. In this case h_1^* is found from (2·3), where v^* is the solution to (2·4) with

$$\sigma_d(x) = 2^{-3d/4}\pi^{-\frac{1}{2}d}\exp(-\tfrac{1}{4}x^{\mathrm{T}}x), \quad r_d(x) = 2^{-(5+d)/4}(x^{\mathrm{T}}x-d)\exp(-\tfrac{1}{4}x^{\mathrm{T}}x).$$

The computed coefficients for

$$h_1^* = c_{1d}^* n^{-1/(4+d)}, \quad \text{AMIAE}\{\hat{f}(\,.\,; h_1^*)\} = C_{1d}^* n^{-2/(4+d)} \quad (1 \leqslant d \leqslant 15)$$

are given in Table 1. It can also be shown that $c_{2d}^* = \{4/(2+d)\}^{1/(4+d)}$ in eqn (2·5); see Table 1. The absolute error bandwidths are somewhat larger than the squared error bandwidths except in one dimension.

For normal densities, it is possible to avoid relying on the asymptotic expressions for mean squared and mean integrated squared error. The optimal window width for the latter is obtained by numerically minimizing the exact formula given by Worton (1989)

$$(4\pi)^{\frac{1}{2}d} \text{ MISE} = n^{-1}h^{-d}\{1-(1+h^{-2})^{-\frac{1}{2}d}\} + \{1-2(1+\tfrac{1}{2}h^2)^{-\frac{1}{2}d} + (1+h^2)^{-\frac{1}{2}d}\}. \quad (3\cdot1)$$

The exact Epanechnikov criterion discussed in § 1 now follows. The exact formulae for criteria (1·1) and (1·2) of $\hat{f}(0; h)$ may also be minimized numerically since

$$\begin{aligned} \text{var}\{\hat{f}(0; h)\} &= n^{-1}(2\pi h)^{-d}\{(2+h^2)^{-\frac{1}{2}d} - (h+h^{-1})^{-d}\}, \\ E\{\hat{f}(0; h)\} &= (2\pi)^{-\frac{1}{2}d}(1+h^2)^{-\frac{1}{2}d}. \end{aligned} \quad (3\cdot2)$$

In Table 2, we now extend and slightly modify Silverman's Table 4.2 (1986, p. 94). When $n = 4$ exactly and $d = 1$, the optimal RRMSE (0) = 0·296. The equivalent sample sizes for $2 \leqslant d \leqslant 10$ are computed for this relative root mean squared error value. Similarly, the criteria in the other columns are calibrated so that $n = 4$ is optimal in one dimension. The rapid increase in sample size is evident. The point $x = 0$ is the most difficult point at which to estimate $f(x)$, which partially explains the larger sample sizes in the RRMSE (0) column. The RCV (0) sample sizes are larger than those for the absolute error criterion.

Table 2. *Equivalent sample sizes across dimension d according to several criteria*: RRMSE (0) = 0·296, Epan. = 0·393, AMIAE = 0·5987, RCV (0) = 0·2625. *The first and last criteria are local at* $x = 0$ *while the Epanechnikov and* AMIAE *criteria are global*

	Sample size n			
d	RRMSE (0)	Epan.	AMIAE	RCV (0)
1	4	4	4	4
2	24	17	11	25
3	88	52	32	84
4	303	155	98	249
5	1082	480	312	725
6	4080	1563	1020	2153
7	16324	5382	3415	6596
8	69124	19558	11719	20893
9	308263	74746	41203	68337
10	1441050	299149	148366	230303

Admittedly, the value of these sample sizes is somewhat doubtful since good density estimation with 4 sample points is not possible. In Table 3, the computations are repeated but for the more practical choice of errors corresponding to $n = 50$ in one dimension. We believe this is a fairer depiction of comparable sample sizes for increasing dimension. The increase is even more rapid in this situation. The absolute error figures indicate a 21–36% reduction in sample size compared to the squared error criterion. The most interesting observation is that the figures based upon the RCV (0) criterion in equation (1·2)

Table 3. *Equivalent sample sizes across dimension d according to several criteria*: RRMSE (0) = 0·145, AMIAE = 0·2180, RCV (0) = 0·1265

	Sample size n		
d	RRMSE (0)	AMIAE	RCV (0)
1	50	50	50
2	307	222	284
3	1468	1087	1131
4	7149	5582	4317
5	37242	29462	16872
6	208441	159440	68559
7	1245809	884474	289832
8	7894492	5029367	1270584

are substantially less than the absolute error sample sizes beyond 4 dimensions. Thus broadly defined features should be estimable by the kernel method with these sample sizes. This is illustrated below in the next section.

4. TEN-DIMENSIONAL EXAMPLE

We reconsider a 10-dimensional example introduced by Friedman et al. (1984). In the (x_1, x_2)-space is an equal mixture of bivariate independent unit-variance normal densities with means $(c, 0)$, $(-c, 3)$, $(-c, -3)$ with $c = 3^{3/2}/2$. It is easy to check that the marginal variances are both 7. Eight pure noise dimensions are added that are independent normal random variables with zero mean and variance 7. The task is to estimate the following slice of the 10-dimensional density

$$f(x_1, x_2, 0, 0, 0, 0, 0, 0, 0, 0). \tag{4·1}$$

In their paper, the authors estimated this slice from a sample with $n = 225$ using a kth nearest neighbour and their projection pursuit estimator. We generated a sample and computed a product kernel estimate using the triweight kernel $K(t) = \frac{35}{32}(1-t^2)^3$, $|t| \leqslant 1$:

$$\hat{f}(x; h) = \left(n \prod_{j=1}^{10} h_j\right)^{-1} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{10} K\left(\frac{x_j - X_{ij}}{h_j}\right)\right\}, \tag{4·2}$$

where X_{ij} is the jth component of the ith observation and $h = (h_1, \ldots, h_{10})$ is a window width vector. The optimal window widths for the mixture of 225 points is approximately the same as for 75 points from any one of the three 10-dimensional components since there is virtually no overlap. The asymptotically optimal window widths for $n = 75$ with the component densities, if all are constrained to be equal, can be shown to be $h_j = 4·0$; without the constraint, the best choices are $h_j = 2·0$ and 5·25 for dimensions $j = 1$, 2 and $j = 3, \ldots, 10$, respectively. For multivariate normal data, the asymptotic mean integrated squared error approximations are nearly exact. Observe how large these window widths are. A rough calculation indicates that the sample root coefficient of variation, criterion (1·2), at the modes is less than a third, which gives some hope for resolving the peaks. Several k-nearest neighbour estimates were also computed for k between 1 and 100.

In the four panels of Fig. 1 we display contour plots of the slice indicated in equation (4·1) of (a) the true density, (b) a kernel estimate with $h_j = 2·0$ or 5·25, (c) a 15th nearest neighbour estimate, and (d) a kernel estimate with $h_j = 4·0$. The trimodal structure is

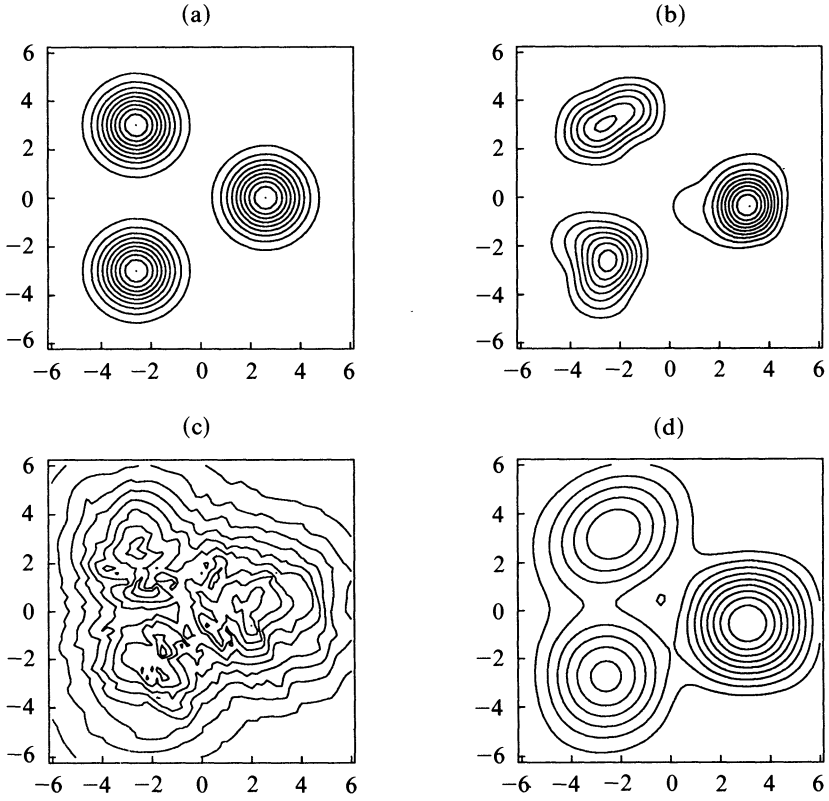

Fig. 1. Contours of slice $(x, y, 0, 0, 0, 0, 0, 0, 0, 0)$ of four density functions; (a) true density function, (b) kernel estimate with $h_1 = h_2 = 2{\cdot}0$ and $h_3 = \ldots = h_{10} = 5{\cdot}25$; (c) k-nearest neighbour estimate with $k = 15$, (d) kernel estimate with $h_1 = \ldots = h_{10} = 4{\cdot}0$.

clearly depicted in the kernel estimate shown in Fig. 1(b). Of course, since all the marginal variances are equal, it is unlikely that the optimal set of unequal smoothing parameters would be used. However, even if all $h_j = 4{\cdot}0$, the trimodal structure is still apparent in Fig. 1(d). The k-nearest neighbour estimate in Fig. 1(c) is also revealing, although the estimate is very rough. We remark that when $k \leqslant 3$, the k-nearest neighbour estimate looks very smooth like a kernel estimate, but this is an artifact of the fact that the 1-6 bumps displayed represent individual samples; see the k-nearest neighbour examples of Friedman et al. (1984). By way of contrast, at the sample mode in Fig. 1(b), 40 of the 225 had nonzero contributions in the kernel sum (4·2). This represents about half the 75 points from that component density.

Our first sample turned out to be the most favourable in ten repetitions, although the same qualitative behaviour was noted in all. The heights of the three peaks sometimes differed by as much as 3 to 1, 1·5 to 1 in Fig. 1(b). Several of the component bumps had an extra secondary bump. It is helpful to imagine what happens if a 226th point is added to the data set at the 10-dimensional origin. The figure with the modified data is exactly like Fig. 1(b) with an additional fourth peak at the origin that is about the same height as the other peaks but slightly narrower. Of course, the original three estimated peaks are the sum of between 20 and 40 weighted and shifted kernels.

The kernel estimate is not as good as the Fig. 1(b) suggests. The sample mode of the kernel estimate is only 34% of the true mode. Thus both the global absolute and squared

error values are certainly quite large. However, we wish to emphasize that the nonadaptive kernel method could still resolve the coarse structure.

The most difficult aspect of the analysis in this example is knowing where to slice the 10-dimensional estimate. For this purpose, the projection pursuit method is clearly superior.

5. DISCUSSION

Table 3 suggests that, for a smooth multivariate density, reasonably accurate estimates of the entire density surface are feasible for at least 5–6 dimensions. Use of the absolute error criterion permits an accurate extrapolation across dimension of equivalent sample sizes with the same global error. Since it is not possible to scrutinize the 6-dimensional surface of a 5-dimensional density, it may not be unreasonable to demand less accuracy than in the bivariate case, for example.

We have also shown that in higher dimensions, accurate estimation of the entire density function does require large samples. However, small samples can still be used to estimate coarse structure in the unknown density. This may be adequate for identifying interesting subspaces of dimension less than 7. Clearly much of the probability mass is 'leaking' to the tails causing substantial bias, but not necessarily resulting in spurious estimates.

Finally, the 'curse of dimensionality' as it applies to kernel methods needs to be examined more closely. Kernel estimates may usefully be thought of as 'local averaging' methods, but the estimate need not be too 'local' to work in higher dimensions. Avoiding the uniform cubical kernel in favour of a continuous kernel that is nearly spherically symmetric is good practical advice. The more significant aspect of data analysis in higher dimensions is the tendency of data to fall into manifolds of lower dimension. Kernel methods are inconsistent for dimensionally deficient data. We believe that alternatives to direct kernel methods, such as projection pursuit and additive models, will also prove to be adversely affected by dimensionally deficient data. This is because any method that is consistent for a wide class of functions must ultimately be 'local'. It is our feeling that more emphasis should be put towards developing linear and nonlinear projection methodology to full rank solutions. Kernel methods should prove quite useful with such projected data in dimensions well beyond the self-imposed 2-dimensional limitation in place today. A difficult challenge is distinguishing rank deficiency and full rank but sparse data. The absolute error theory, while significantly less pessimistic than the squared error theory, still seems to be overly pessimistic relative to some limited experience with multivariate kernel procedures; see Scott (1991) for a review and discussion. The 10-dimensional example in § 4 with only 225 points may not be too realistic, but it lends credence to the finding of 3 sample modes in a kernel estimate of the 4-dimensional Anderson–Fisher iris data where n is only 150; see Scott (1986).

ACKNOWLEDGEMENTS

This research was partially supported by the Office of Naval Research and the Army Research Office.

REFERENCES

CACOULLOS, T. (1966). Estimation of a multivariate density. *Ann. Inst. Statist. Math.* **18**, 179–89.

DEVROYE, L. & GYÖRFI, L. (1985). *Nonparametric Density Estimation: The L_1 View.* New York: Wiley.

EPANECHNIKOV, V. A. (1969). Nonparametric estimation of a multidimensional probability density. *Theory Prob. Applic.* **14**, 153–8.

FRIEDMAN, J. H. & STUETZLE, W. (1981). Projection pursuit regression. *J. Am. Statist. Assoc.* **76**, 817–23.

FRIEDMAN, J. H., STUETZLE, W. & SCHROEDER, A. (1984). Projection pursuit density estimation. *J. Am. Statist. Assoc.* **79**, 599–608.

HALL, P. & WAND, M. P. (1988). Minimizing L_1 distance in nonparametric density estimation. *J. Mult. Anal.* **26**, 59–88.

HUBER, P. J. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13**, 435–75.

LOFTSGAARDEN, D. O. & QUESENBERRY, C. P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.* **36**, 1049–51.

SCOTT, D. W. (1986). Data analysis in 3 and 4 dimensions with nonparametric density estimation. In *Statistical Image Processing and Graphics*, Ed. E. J. Wegman and D. DePriest, pp. 291–305. New York: Marcel Dekker.

SCOTT, D. W. (1991). Some reflections on estimation and visualization of multivariate density functions. In *IMA Volumes on Mathematics and its Applications*, Ed. P. Tukey and A. Buja. To appear. New York: Springer-Verlag.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

WORTON, B. J. (1989). Optimal smoothing parameters for multivariate fixed and adaptive kernel methods. *J. Statist. Comp. Simul.* **32**, 45–57.

[*Received March* 1990. *Revised June* 1990]