AD-A099 369 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
ON A TEST FOR MULTIMODALITY BASED ON KERNEL DENSITY ESTIMATES.(U)
FEB 81 B W SILVERMAN DAAG29-80-C-0041
UNCLASSIFIED MRC-TSR-2181 NL

END
DATE FILMED
81
DTIC

MRC Technical Summary Report # 2181

# ON A TEST FOR MULTIMODALITY BASED ON KERNEL DENSITY ESTIMATES

B. W. Silverman

**Mathematics Research Center**
**University of Wisconsin—Madison**
**610 Walnut Street**
**Madison, Wisconsin 53706**

February 1981

(Received January 14, 1981)





**Approved for public release**
**Distribution unlimited**

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

# ON A TEST FOR MULTIMODALITY BASED ON KERNEL DENSITY ESTIMATES

B. W. Silverman*

Technical Summary Report #2181
February 1981

ABSTRACT

Kernel probability density estimates can be used to construct a test of the hypothesis that the density underlying a given univariate data set has at most $k$ modes, for any given $k > 1$. The test is based on the critical value of the smoothing parameter for $k$ modes to occur in the estimate. The theoretical properties of this test are investigated; the asymptotic properties of the test statistic show that the test is consistent. Furthermore the rate of convergence of the test statistic to zero gives some theoretical insight into a bootstrap technique previously suggested by the author, and also into observed properties of kernel density estimates.

AMS (MOS) Subject Classifications: 60F99, 62G10, 41A25

Key Words: kernel density estimate, cluster analysis, mode, critical smoothing, bump hunting.

Work Unit Number 4 - Statistics and Probability

---

* Permanent address: School of Mathematics, University of Bath, BATH BA2 7AY, United Kingdom.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

## SIGNIFICANCE AND EXPLANATION

An important question in cluster analysis is the determination of the number of clusters into which a given population should be divided. This problem arises in almost every area where data are collected, for example in physics, geology, medicine and psychology.

Frequently, particularly when certain specific clustering methods are being used, the number of clusters is taken to be equal to the number of modes, or local maxima, in the probability density function underlying the given data set. The author has previously suggested a technique for investigating the number of modes underlying a given population. In this paper, the mathematical properties of this procedure are investigated. The results obtained confirm various intuitive remarks made in the original presentation of the method, and also suggest that the technique may cast light on another important problem, that of determining how much to smooth a sample in order to estimate its underlying probability density.

---

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

ON A TEST FOR MULTIMODALITY BASED ON

KERNEL DENSITY ESTIMATES

B. W. Silverman*

## 1. Introduction

Silverman (1981) suggested and illustrated a way that kernel probability density estimates can be used to investigate the number of modes in the density underlying a given independent identically distributed real sample. Given an independent sample $X_1,\ldots,X_n$ from a univariate probability density $f$, define the kernel density estimate $f_n$ with Gaussian kernel by

$$f_n(t,h) = \sum_{i=1}^{n} h^{-1}\phi\{(t-X_i)/h\} \quad ,$$

where the parameter $h$ is the smoothing parameter or window width and $\phi$ is the standard normal density function. Kernel density estimates were introduced by Rosenblatt (1956) and Parzen (1962); the restriction to Gaussian kernels in this work is made for reasons given in Silverman (1981). Often the explicit dependence of $f_n$ on $h$ will be suppressed.

Consider the problem of testing the null hypothesis that $f$ has $k$ or fewer modes against the alternative that $f$ has more than $k$ modes. The statistic suggested for constructing such a test was the k-critical window width $h_{crit}(k)$, defined by

$$h_{crit}(k) = \inf\{h : f_n(\cdot,h) \text{ has at most } k \text{ modes}\} \quad .$$

*Permanent address: School of Mathematics, University of Bath, BATH BA2 7AY, United Kingdom.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

In Silverman (1981) it was stated heuristically that large values of $h_{crit}$ will tend to reject the null hypothesis. The results of this paper show that this procedure does indeed lead to a consistent test.

Subject to certain regularity conditions, it is shown that, under the null hypothesis, $h_{crit}$ converges stochastically to zero, while this is not the case under the alternative hypothesis. The exact rate of convergence of $h_{crit}$ to zero under the null hypothesis is found. It is perhaps interesting that this rate of convergence has precisely the same order as the rate of convergence for the optimum choice of window width for the uniform estimation of the density given, for example, by Silverman (1978b).

In Silverman (1981) a smoothed bootstrap procedure for assessing the significance of an observed value of $h_{crit}$ was suggested and illustrated by an application. The representative of the null hypothesis constructed from the data is obtained from the density estimate with window width $h_{crit}$; the estimate is rescaled, as suggested by Efron (1979), to have variance equal to the sample variance of the data. The remarks above show that $f_n(\cdot, h_{crit})$ is, in a certain sense, optimally uniformly consistent as an estimate of the true density $f$. It follows that, on the null hypothesis, the bootstrap procedure is likely, at least for large samples, to provide an estimate of the true underlying density which is accurate in the uniform norm. A possible drawback for small samples is the fact that the implied constant in the rate of convergence does not necessarily take its optimum value.

An interesting open question raised by this discussion is the possibility of using $h_{crit}(k)$ for some value of $k$ in developing an automatic method for choosing the smoothing parameter in density estimation. Boneva, Kendall and Stefanov (1971) suggested choosing the window width where 'rabbits' or rapid fluctuations just started to appear. Such a window width would perhaps correspond to $h_{crit}(k)$ for some $k > j$; since $h_{crit}(k)$ converges to zero at the optimum rate for all $k > j$, a suitable formalization of the Boneva-Kendall-Stefanov procedure would give estimates which converged at the optimal rate, though not necessarily with the optimal constant multiplier. The fact that $h_{crit}(k)$ has the same rate of convergence for all $k > j$ provides some explanation for the observation made by Boneva, Kendall and Stefanov that the estimate seems suddenly to become noisy as the window width is reduced.

The use of kernel density estimates in mode estimation was originated by Parzen (1962). The 'gradient method' of cluster analysis is based on clustering towards modes in the estimated density; see, for example, Andrews (1972), Fukunaga and Hostetler (1975), and Bock (1977). Papers related to tests of multimodality are Cox (1966) and Good and Gaskins (1980).

## 2. The main result

In this section, the main result of this paper is stated and proved. It is convenient to use the convention throughout that all limits and implied limits are taken as $n$ tends to infinity. Varying conventions will apply to unqualified suprema and infima in Propositions 1 and 2 below, and these will be introduced where necessary. The notations p lim inf and p lim sup will be used to signify the corresponding limits in probability as $n$ tends to infinity, and $O_p$ and $o_p$ will denote probability orders of magnitude. Define, for $h > 0$,

$$\alpha(h) = h^{-5} \log(h^{-1}) \quad . \tag{1}$$

The main results are all contained in the following theorem.

<u>Theorem</u>

<u>Suppose</u> $f$ <u>is a bounded density with bounded support</u> $[a,b]$, <u>and suppose that the following conditions are satisfied:</u>

(i) $f$ <u>is twice continuously differentiable on</u> $[a,b]$

(ii) $f$ <u>has exactly</u> $j$ <u>local maxima on</u> $(a,b)$

(iii) $f'(a+) > 0,\ f'(b-) < 0$

(iv) $\displaystyle\min_{\{z:f'(z)=0\}} \frac{f''(z)^2}{f(z)} = c_0 > 0.$

<u>Let</u> $h_{crit}(k)$ <u>be the</u> $k$-<u>critical window width constructed from an i.i.d. sample of size</u> $n$ <u>from</u> $f$. <u>Then, if</u> $k \geq j$, <u>defining</u> $\alpha$ <u>as in</u> (1) <u>above,</u>

$$\text{p lim inf } n^{-1}\alpha\{h_{crit}(k)\} \geq \frac{2}{3}\pi\sqrt{2}\, c_0 \tag{2}$$

<u>and</u>

$$\text{p lim sup } n^{-1}\alpha\{h_{crit}(k)\} < \infty \tag{3}$$

<u>while if</u> $k < j$ <u>then there exists a constant</u> $h_0(f,k)$ <u>such that</u>

$$P\{h_{crit}(k) > h_0\} \to 1 \quad . \tag{4}$$

Note that condition (iv) is equivalent, in the presence of the other conditions, to the condition that $f$ is strictly positive on $[a,b]$ and $f'$ has no multiple zeroes on $[a,b]$.

It is convenient to prove the various assertions of the theorem separately. Except where otherwise stated, the conditions of the theorem on $f$ will be assumed to be true throughout. The first proposition facilitates the proof of (2).

Proposition 1. Given any $c_1$ with

$$0 < c_1 < \frac{2}{3}\pi\sqrt{2}\, c_0 \quad ,$$

suppose the sequence of window widths $h_n$ satisfies

$$n^{-1}\alpha(h_n) \to c_1 \quad . \tag{5}$$

Then the number of maxima of $f_n$ tends in probability to $j$.

It follows from Proposition 1 and Silverman (1981) that, for all $k > j$, provided (5) holds,

$$P\{h_{crit}(k) < h_n\} \to 1$$

and hence that (2) is satisfied.

The proof of Proposition 1 makes use of several lemmas, the first of which shows that, under certain conditions, maxima and minima of $f_n$ can, eventually, only occur arbitrarily close to those of $f$.

Lemma 1. Let $I$ be any closed interval contained in $[a,b]$, such that $I$ contains none of the zeroes of $f'$. Then, provided $h_n \to 0$ and $n^{-1}h_n^2\alpha(h_n) \to 0$, it will follow that

$$P(f_n \text{ monotonic on } I \text{ in the same sense as } f) \to 1 \quad .$$

Proof. By slight adaptation of the results of Silverman (1978a), it can be seen that, provided $f$ is bounded, we will have, if $h_n$ satisfies the assumptions of Proposition 1,

$$\begin{aligned} \sup_{(-\infty,\infty)} |f_n' - Ef_n'| &= \underline{O}_p\{n^{-\frac{1}{2}}h_n^{-1}\alpha(h_n)^{\frac{1}{2}}\} \\ &= \underline{o}_p(1) \quad . \end{aligned} \tag{6}$$

In Silverman (1978a) the uniform continuity of $f$ was additionally assumed, but careful examination of the proofs of that paper shows that the derivation of the rate of stochastic convergence, though not of the exact constant implied in the $\underline{O}_p$, goes through under the assumption of bounded $f$.

Supposing without loss of generality that $f$ is increasing on $I$, it follows from the continuity of $f'$ on $[a,b]$ that $f'$ is bounded away from zero on $I$ and is non-negative on a neighborhood of $I$, and hence by elementary analysis that

$$\liminf \inf_I Ef_n' > 0 \quad . \tag{7}$$

Combining (6) and (7) completes the proof of Lemma 1.

The next lemma shows that, under suitable conditions, $f_n$ will eventually have exactly one maximum and no minima near each maximum of $f$, and exactly one minimum and no maxima near each minimum of $f$.

Lemma 2. *Suppose* $f'(z) = 0$ *and* $f$ *has a local maximum (respectively minimum) at* $z$. *Suppose* $h_n \to 0$ *and*

$$n^{-1}\alpha(h_n) \to c_2 \in (0, \tfrac{2}{3}\pi\sqrt{2}\, f''(z)^2/f(z)) \quad . \tag{8}$$

*Then, for all sufficiently small* $\varepsilon > 0$, *the probability that* $f_n'$ *has exactly one zero in* $(z-\varepsilon, z+\varepsilon)$, *and that this zero is a maximum (respectively minimum) of* $f_n$, *tends to one as* $n$ *tends to infinity.*

Proof. Only the case of a local maximum will be considered. The proof for a minimum proceeds very similarly and is omitted. Throughout this proof unqualified infima and suprema will be taken to be over $x$ in $[z-\varepsilon, z+\varepsilon]$. By the continuity of $f$ and $f''$, choose $\varepsilon$ sufficiently small that

$$\frac{\inf f''(x)^2}{\sup f(x)} > \frac{3c_2}{2\pi\sqrt{2}} \tag{9}$$

and also $[z-\varepsilon, z+\varepsilon] \subseteq (a,b)$. It is then immediate that $f'(z-\varepsilon) > 0$ and $f'(z+\varepsilon) < 0$ since, by (9), $f''$ cannot cross zero in $(z-\varepsilon, z+\varepsilon)$. Since $f'$ is continuous at $z \pm \varepsilon$, by standard results on the consistency of $f_n'$ (a combination of Parzen (1962) and Bhattacharya (1967))

$$P\{f_n'(z-\varepsilon) > 0 \text{ and } f_n'(z+\varepsilon) < 0\} \to 1 \quad . \tag{10}$$

Very slightly adapting the proofs of Silverman (1976 and 1978a) to cope with the fact that $f''$ is only uniformly continuous on a neighborhood of $[z-\varepsilon, z+\varepsilon]$ gives

$$n^{-\frac{1}{2}}\alpha(h)^{\frac{1}{2}} \sup|f_n''(x) - Ef_n''(x)| \xrightarrow{P} K_1$$

where

$$K_1^2 = 2 \sup f \int \phi''^2$$

$$= 3(2\pi\sqrt{2})^{-1} \sup f \quad .$$

Since, by elementary analysis, $\sup|Ef_n''(x) - f''(x)|$ converges to zero, it follows from (8) that $p \lim_n \sup \sup|f_n''(x) - f(x)| < K_1 c_2^{\frac{1}{2}}$

$$< \inf|f''(x)|$$

by (9). It is immediate that

$$P\{f_n''(x) < 0 \text{ for all } x \text{ in } [z-\varepsilon, z+\varepsilon]\} \to 1 \quad . \tag{11}$$

Combining (10) and (11) completes the proof of Lemma 2.

To complete the proof of Proposition 1, note first that no maxima of $f_n$ can occur outside the interval $(a,b)$. Let $z_1,\ldots,z_{2j-1}$ be the zeroes of $f'$ in $(a,b)$ and choose $\varepsilon$ sufficiently small to satisfy the conclusion of

Lemma 2 for all $z_i$ and to ensure that

$$a < z_1-\varepsilon < z_1+\varepsilon < z_2-\varepsilon < \cdots < z_{2j-1}+\varepsilon < b \quad . \tag{12}$$

Applying either Lemma 1 or Lemma 2 as appropriate to each of the intervals in the partition (12) of the interval $(a,b)$ completes the proof of Proposition 1.

The next proposition leads to the proof of assertion (3), in a similar way to the derivation of (2) from Proposition 1.

Proposition 2

Defining $\alpha$ as in (1) above, suppose that

$$n^{-1}\alpha(h_n) \to \infty \quad \underline{\text{and}} \quad n^{-1}h_n^{-5} \to 0 \quad . \tag{13}$$

Then the number of maxima in $f_n$ tends in probability to infinity.

Given any $k$, it follows from this result and the corollary of Silverman (1981) that, provided (13) holds,

$$P\{h_{crit}(k) > h_n\} \to 1 \quad ;$$

assertion (2) follows at once.

To prove Proposition 2, suppose without loss of generality that $f$ has a maximum at $0$ in $(a,b)$. Choose a sequence $\ell_n$ which satisfies

$$\begin{aligned} &\ell_n \to 0, \quad h_n^{-1}\ell_n = \underline{o}\{n^{-1}\alpha(h_n)\} \quad , \\ &h_n^{-1}\ell_n \to \infty \quad \text{and} \quad |\log \ell_n|\,|\log h_n|^{-1} \to 1 \quad . \end{aligned} \tag{14}$$

The explicit dependence of $h$ and $\ell$ on $n$ will often be suppressed. Let $I_{j,n}$ be the interval $[(j-1)\ell, j\ell]$ for integer $j > 0$.

Following Silverman (1978a) apply Theorem 3 of Komlos, Major and Tusnady (1975) to obtain

$$f_n'(x) = Ef_n'(x) + h^{-1}n^{-\frac{1}{2}}\rho_1(x) + \varepsilon_n'(x)$$

where $\rho_1$ is a Gaussian process with the same covariance structure as $n^{\frac{1}{2}}h(f_n' - Ef_n')$ and $\varepsilon_n'$ is a secondary random error. The process $\rho_1$ is obtained by putting $\delta(u)$ equal to $\phi'(u)$ in Proposition 1 of Silverman (1978a). By elementary analysis and the arguments of Silverman (1978a) we have, in a neighborhood of 0,

$$|Ef_n'(x) - f'(x)| = \underline{O}(h) \quad ;$$

$$|\varepsilon_n'(x)| = \underline{O}(n^{-1}h^{-2}\log n) \qquad \text{a.s.}$$

$$= \underline{o}(h^2) \quad \text{from (13) above} \quad ;$$

$$\text{and} \quad |f'(x)| = \underline{O}(x) \quad ,$$

since $f'(0) = 0$ and $f''$ exists. It follows that, a.s.,

$$\begin{aligned} \sup|Ef_n'(x) + \varepsilon_n'(x)| &= \underline{O}(j\ell) + \underline{O}(h) \\ &= \underline{o}\{n^{-1}h^{-5}\log(\ell/h)\}^{\frac{1}{2}} \end{aligned} \tag{15}$$

by (13) and (14) above, where we adopt the convention, here and subsequently in this proof, that unqualified suprema are taken to be over the interval $I_{j,n}$, and that a fixed $j$ is being considered.

We slightly adapt the argument of Silverman (1976) pp. 138-140 to investigate $\sup \rho_1$. Define

$$\sigma^2(x) = \text{var}\, \rho_1(x) = h^{-1}f(x) \int \phi'^2(1 + \underline{o}(1))$$

$$= h^{-1}f(0) \int \phi'^2(1 + \underline{o}(1)) \quad \text{for } x \text{ in } I_{j,n} \quad ,$$

since the end points of $I_{j,n}$ both converge to zero. Analogously to (12) of Silverman (1976), given any $\lambda$ in $(0,2)$,

$$P[\sup \sigma^{-1}\rho_1 < (1 - \tfrac{1}{2}\lambda)\{2 \log (h^{-1}\ell)\}^{\frac{1}{2}}]$$

$$< \underline{O}(\ell^{-2})\log(h^{-1}\ell) \tag{16}$$

$$\times \iint_{I_{j,n}} |\chi| \exp\{2 \log(h^{-1}\ell)(1 - \tfrac{1}{2}\lambda)^2 |\chi|/(1 + |\chi|)\}$$

where $\chi(x,y) = \mathrm{corr}\{\rho(x),\rho(y)\}$. Using a similar argument to that following (12) of Silverman (1976), but allowing the interval $I$ to vary, shows that the expression in (16) is dominated by

$$\underline{O}(\ell^{-2}) \log(h^{-1}\ell) \{\sigma^2(0) + \underline{o}(1)\}^{-1} \{h^{-1}\ell\}^{(1 - \frac{1}{2}\lambda)^2} \underline{O}(\ell)$$

$$= (h^{-1}\ell)^{-\lambda + \frac{1}{4}\lambda^2} \log(h^{-1}\ell) \to 0$$

by (14) above.

It follows that, setting $K = \{2f(0) \int \phi'^2\}^{\frac{1}{2}}$,

$$p \lim \inf \sup\{h^{-1} \log(h^{-1}\ell)\}^{\frac{1}{2}}\rho_1 > K \tag{17}$$

and that the same result holds if $\rho_1$ is replaced by $-\rho_1$, giving a corresponding result for $\inf \rho_1$. It follows from (15), (17) and the corresponding result for $\inf \rho_1$ that

$$P\{\rho_1 \text{ crosses } -n^{\frac{1}{2}}h(Ef_n' + \varepsilon_n') \text{ in } I_{j,n}\} \to 1 ,$$

and hence that

$$P\{f_n' \text{ crosses zero in } I_{j,n}\} \to 1 . \tag{18}$$

Since (18) holds for all $j$, the number of maxima in $f_n$ tends in probability to infinity, completing the proof of Proposition 2.

The final proposition of this section deals with the case where the alternative hypothesis is true, and shows that $h_{crit}$ will remain bounded away from zero.

Proposition 3

*If* $k < j$ *then there exists a constant* $h_o > 0$, *depending on* $f$ *and* $k$, *such that*

$$P\{h_{crit}(k) > h_o\} \to 1 \quad .$$

Proof

By arguments analogous to those of the proof of the theorem of Silverman (1981), making use of the variation diminishing properties of the Gaussian kernel and the continuity properties of $Ef_n$, the number of maxima in $Ef_n(\cdot,h)$ is a right continuous decreasing function of $h$, for $h > 0$. By choosing $h_0$ sufficiently small, we can ensure that $Ef_n(\cdot,h_0)$ has, independently of $n$, exactly $j$ maxima. Because of the conditions imposed on $f$ in the statement of the Theorem above, we can also ensure that $Ef_n''(\cdot,h_0)$ is non-zero at all stationary points of $Ef_n(\cdot,h_0)$.

The argument of Lemma 2.2 of Schuster (1969), which does not in fact require the convergence to zero of the sequence of window widths, then implies that, with probability one,

$$f_n'(x,h_0) - Ef_n'(x,h_0) \quad \text{and} \quad f_n''(x,h_0) - Ef_n''(x,h_0)$$

both converge to zero uniformly over $x$. By an argument similar to that used in Proposition 1 above, it follows that the number of maxima of $f_n(\cdot,h_0)$ on $[a,b]$ tends almost surely to $j$, the number of maxima of $Ef_n(\cdot,h_0)$. Applying the corollary of Silverman (1981) completes the proof of Proposition 3.

## Discussion

It is natural to enquire to what extent the conditions of the theorem above can be relaxed without affecting the conclusions. In particular it seems intuitively clear that the condition of bounded support for the density $f$ should be able to be replaced by some condition on the tails of $f$, though the present method of proof cannot deal with this case. Condition (iv) appears to be more fundamental to the result; if, for example, $f'(0) = f''(0) = 0 \neq f'''(0)$, then an examination of $f_n$ and $Ef_n$ near zero seems to indicate that, under suitable regularity conditions, there will be no maximum of $f_n$ near zero provided $|f_n''' - Ef_n'''|$ remains small. A heuristic argument suggests that a result corresponding to the theorem of Section 2 can be proved, but with $\alpha(h)$ replaced by $h^{-7}\log(h^{-1})$, so that $h_{crit}$ converges to zero more slowly. Even slower convergence will occur for higher order zeroes in $f'$.

The interest in this discussion lies in the fact that the bootstrap density constructed using the critical window width will not only have infinite tails of similar weight to those of the corresponding normal kernels but will also have a stationary point which is a point of inflexion. The slower convergence to zero of $h_{crit}$ provides support for the remark of Silverman (1981) that the bootstrap test may be conservative; it also bears out the intuition of P. Huber (private communication) that the bootstrap procedure may be excessively conservative, though the difference between $n^{-\frac{1}{5}}$ and $n^{-\frac{1}{7}}$ convergence is very slight in practice.

The methods of this paper can also be used to study the asymptotic properties of a corresponding test for the number of points of inflexion in the density. Both Cox (1966) and Good and Gaskins (1980) prefer to use points of inflexion as an indication that the density is a mixture. The critical window width will now be the smallest window width for which the density has $k$ maxima. Under suitable conditions a result corresponding to the theorem of Section 2 can be proved, but again, among other changes, $\alpha(h)$ will be replaced by $h^{-7}\log(1/h)$ since $f_n''$ will be replaced by $f_n'''$ in much of the argument of the proofs of Propositions 1 and 2.

## Acknowledgements

The author is most grateful to H. H. Bock, P. Deheuvels and P. J. Huber for helpful discussions and suggestions.

## REFERENCES


Andrews, H. C. (1972), Introduction to mathematical techniques in pattern recognition. Wiley, New York.

Bhattacharya, P. K. (1967), Estimation of a probability density function and its derivatives. Sankhyā, Series A, 29, 373-382.

Bock, H. H. (1977), On tests concerning the existence of a classification. Proc. First International Symposium on Data Analysis and Informatics, Versailles, 1977, Institut de Recherche d'Informatique et d'Automatique, Domaine de Voulceau, Le Chesnay, France, 449-464.

Boneva, L. I., Kendall, D. G. and Stefanov, I. (1971), Spline Transformations. J. Roy. Statist. Soc. B, 33, 1-70.

Cox, D. R. (1966), Notes on the analysis of mixed frequency distributions. Brit. J. Math. Statist. Psych., 19, 39-47.

Efron, B (1979), Bootstrap methods - another look at the jack-knife. Ann. Statist., 7, 1-26.

Fukunaga, K. and Hostetler, L. D. (1975), The estimation of the gradient of a density function with applications in pattern recognition. IEEE Trans. Inform. Theory, IT-21, 32-40.

Good, I. J. and Gaskins, R. A. (1980), Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. J. Amer. Stat. Assoc., 75, 42-56.

Komlos, J., Major, P. and Tusnady, G. (1975), An approximation of partial sums of independent random variables and the sample distribution function. Z. Wahrscheinlichkeitsth. und Verw. Gebiete. 32, 111-131.

Parzen, E. (1962), On estimation of a probability density function and mode. Ann. Math. Statist., 33, 1065-1076.

Rosenblatt, M. (1956), Remarks on some non-parametric estimates of a density function. Ann. Math. Statist. 27, 832-837.

Schuster, E. F. (1969), Estimation of a probability density and its derivatives. Ann. Math. Statist., 40, 1187-1196.

Silverman, B. W. (1976), On a Gaussian process related to multivariate probability density estimation. Math. Proc. Cambridge Philos. Soc., 80, 135-144.

Silverman, B. W. (1978a), Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Ann. Statist., 6, 177-184.

Silverman, B. W. (1978b), Choosing the window width when estimating a density. Biometrika, 65, 1-11.

Silverman, B. W. (1981), Using kernel density estimates to investigate multimodality. J. Roy. Statist. Soc. B, 43.

BWS/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

**1. REPORT NUMBER**
#2181

**2. GOVT ACCESSION NO.**
AD-A099369

**3.** ~~RECIPIENT'S CATALOG NUMBER~~
(9) Technical

**4. TITLE (and Subtitle)**
(6) On a Test for Multimodality Based on Kernel Density Estimates.

**5.** ~~TYPE OF REPORT & PERIOD~~ COVERED
Summary ~~Report~~, no specific reporting period

**6. PERFORMING ORG. REPORT NUMBER**

**7. AUTHOR(s)**
(10) B. W. Silverman

**8. CONTRACT OR GRANT NUMBER(s)**
(15) DAAG29-80-C-0041

**9. PERFORMING ORGANIZATION NAME AND ADDRESS**
Mathematics Research Center, University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53706

**10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS**
Work Unit Number 4 - Statistics and Probability

**11. CONTROLLING OFFICE NAME AND ADDRESS**
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, North Carolina 27709

**12.** ~~REPORT DATE~~
(11) Feb[redacted]81

**13. NUMBER OF PAGES**
15

**14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)**
(12) 19

**15. SECURITY CLASS. (of this report)**
UNCLASSIFIED

**15a. DECLASSIFICATION/DOWNGRADING SCHEDULE**

**16. DISTRIBUTION STATEMENT (of this Report)**
Approved for public release; distribution unlimited.

(14) MRC-TSR-2181

**17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)**

**18. SUPPLEMENTARY NOTES**

**19. KEY WORDS (Continue on reverse side if necessary and identify by block number)**
kernel density estimate, cluster analysis, mode, critical smoothing, bump hunting

**20. ABSTRACT (Continue on reverse side if necessary and identify by block number)**

Kernel probability density estimates can be used to construct a test of the hypothesis that the density underlying a given univariate data set has at most $k$ modes, for any given $k \geq 1$. The test is based on the critical value of the smoothing parameter for $k$ modes to occur in the estimate. The theoretical properties of this test are investigated; the asymptotic properties of the test statistic show that the test is consistent. Furthermore the rate of convergence of the test statistic to zero gives some theoretical insight into a bootstrap technique previously suggested by the author, and also into observed properties of kernel density estimates.


DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

221200 -W

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DATE
ILMED
8