Choosing the Window Width when Estimating a Density

Author(s): B. W. Silverman

Source: *Biometrika*, Apr., 1978, Vol. 65, No. 1 (Apr., 1978), pp. 1-11
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2335270

Biometrika (1978), **65**, 1, *pp.* 1–11
With 6 *text-figures*
Printed in Great Britain

Choosing the window width when estimating a density

By B. W. SILVERMAN
*Statistical Laboratory, University of Cambridge**

Summary

A practical method is discussed for determining the amount of smoothing when using the kernel method to estimate a probability density from independent identically distributed observations. Both the univariate and the multivariate cases are considered. The method is illustrated by the analysis of several sets of data; the theoretical motivation and justification are also provided.

Some key words: Data analysis; Density estimation; Graphical methods; Kernel; Smoothing; Test graph.

1. Introduction

The problem of estimating the probability density function underlying independent identically distributed observations has received considerable attention. Rosenblatt (1956) introduced the kernel estimator, defined for all real x by

$$f_n(x) = n^{-1} \sum_{i=1}^{n} h(n)^{-1} \delta\{(x - X_i)/h(n)\},$$

where $X_1, \ldots, X_n$ are independent identically distributed real observations, δ is a kernel function and $h(n)$ is a sequence of window widths, assumed to tend to zero as n tends to infinity. The kernel estimator has been widely discussed; for a survey see Rosenblatt (1971).

When applying the method in practice it is of course necessary to choose a kernel and a window width. The choice of kernel was considered by Epachenikov (1969) who showed that there is in some sense an optimal kernel, which is part of a parabola, but that any reasonable kernel gives almost optimal results. Therefore the choice of kernel is not as important a problem in practice as might be supposed. It is quite satisfactory to choose a kernel for computational convenience, as below, or for any other attractive reason, such as, for example, the argument leading to the quadratic spline kernel used by Boneva, Kendall & Stefanov (1971) in their 'spline transform' technique.

While the choice of kernel does not seem to lead to much difficulty, at least for reasonably large sample sizes, the choice of window width is quite a different matter. The results of Silverman (1978) show that the kernel estimate is uniformly consistent under quite mild conditions on the rate of convergence of the window width to zero, but that the rate of consistency can be very slow. The very interesting practical work of Boneva *et al.* (1971) shows that the estimates can change dramatically under quite small variations in window width. Thus there seems to be considerable need for objective methods of determining the window width appropriate to a given sample. The method developed below, while being subjective to some extent, goes a long way towards resolving this difficulty in certain cases. First the method is described and some applications to sets of data are considered. The application of the method to multivariate data is then discussed. Finally, the theoretical justification of the method is obtained.

* Present address: Mathematical Institute, Oxford.

2. DESCRIPTION OF THE METHOD: THE UNIVARIATE CASE

The practical method developed depends on the following result, which will be proved in § 5 below. The notations sup and inf, when unqualified, are taken to be over the whole real line.

THEOREM 1. *Suppose that the true real density f has uniformly continuous second derivative and that the kernel δ satisfies conditions (a)–(h) of § 5 below. Choose the sequence $h(n)$ to give the best possible rate of uniform consistency of f_n as n tends to infinity. Then*

$$\sup\{f_n'' - E(f_n'')\} \to k \sup|f''|, \quad \inf\{f_n'' - E(f_n'')\} \to -k \sup|f''|$$

in probability as n tends to infinity. The constant k depends on the kernel and is defined explicitly in § 5 below.

Before exploring the consequences of this theorem it is convenient to concentrate attention on a particular kernel. The kernel used in this paper is the piecewise quartic

$$\delta(x) = \begin{cases} \frac{1}{4}|x|^4 - \frac{1}{2}|x|^3 + \frac{1}{2} & \text{for} \quad |x| \leqslant 1, \\ \frac{1}{4}|x|(2-|x|)^3 & \text{for} \quad 1 \leqslant |x| \leqslant 2, \\ 0 & \text{for} \quad |x| \geqslant 2. \end{cases} \tag{1}$$

This kernel is chosen because it is close to optimal in the sense of Epachenikov (1969), it satisfies conditions (a)–(h) of § 5, and, being a piecewise polynomial of narrow support, it leads to an estimate which can be computed quickly. The constant k of the theorem is almost exactly 0·4 for this kernel.

The theorem can be rephrased as follows. If the window width is chosen to give the best estimate of the density then the random fluctuations in the second derivative of the estimate will be asymptotically of maximum size $\pm k \sup|f''|$. This follows from the fact that $E(f_n'')$ is a smoothed version of f'' and so any fluctuations in f_n'' are due to the random error $f_n'' - E(f_n'')$, while any systematic variation is due to the variation of $E(f_n'')$ and hence of f''.

The method suggested is to draw graphs of the second derivative of the estimate for various window widths. These graphs will be called test graphs. Choose the window width which gives fluctuations of the right size in the test graph, and use this window width to construct the estimate of the original density. In the actual calculation the test graph is expressed as

$$f_n''(x) = n^{-1} h^{-3} \sum_{i=1}^{n} \delta''\{h^{-1}(x - X_i)\}$$

and so is easily calculated. The subjectivity of the method is, of course, in the assessment of the right size of fluctuations in the test graph. The test graph f_n'' is an estimate of the second derivative f'' of the density. Therefore the systematic variation of the test graph gives an estimate of f'' and hence of the constant $\sup|f''|$. The fluctuations about $E(f'')$ should, for the ideal window width, be of height $\pm 0{\cdot}4 \sup|f''|$ for kernel (1) and so the ideal test graph should have fluctuations which are quite marked but do not obscure the systematic variation completely. This will be called the test graph principle.

The first example illuminates the principle. One hundred observations were simulated from the standard normal distribution. The test graphs for various window widths are shown in Fig. 1. The test graph for window width 0·5 has fluctuations which dominate the systematic variation; this graph is an example of a test graph which is too rough. On the other hand, the test graph for window width 0·9 has very small fluctuations and so is too smooth. The other graphs are examples with marked fluctuations which are nevertheless smaller than the systematic variation. It was discovered empirically that 0·675 is the window width which gives the density estimate uniformly closest to the true density, though very acceptable estimates were obtained with window widths between 0·625 and 0·75, particularly considering the relatively small sample size. The estimates for window widths 0·5 and 0·9 were much further from the true density.

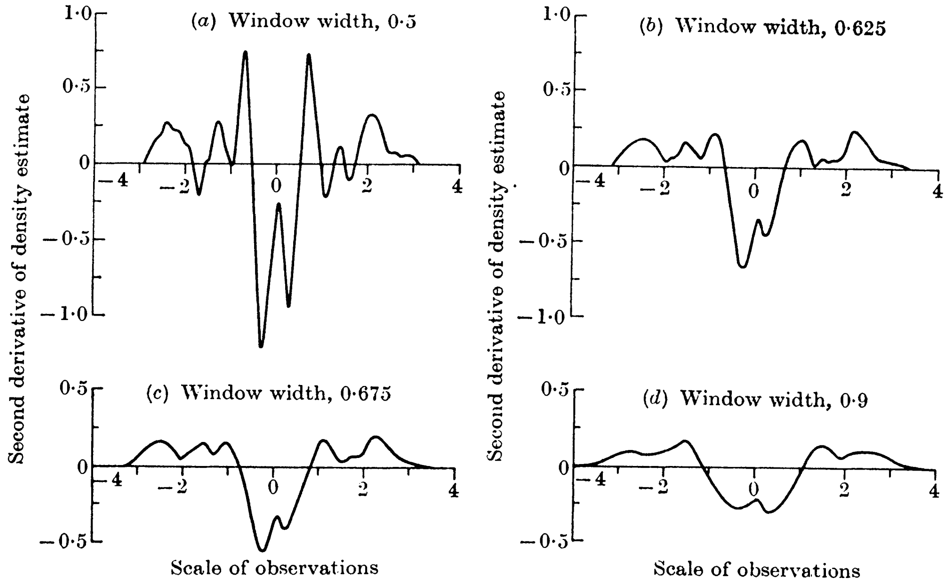

Fig. 1. Test graphs for 100 normal observations for window widths, 0·5, 0·625, 0·675 and 0·9.

Notice that the fluctuations in the test graph grow very quickly as the window width is decreased and so it is easy to get a close idea of the appropriate window width. When dealing with samples of around 100, it may be useful not only to apply the test graph principle stated above, but also to compare the test graphs with those in Fig. 1 to see what sort of behaviour an appropriate test graph should exhibit. For larger samples the number of fluctuations will be large; therefore it will be very easy to distinguish between random fluctuation and systematic variation in order to apply the test graph principle.

When the sample is so large that it is very expensive of computer time to draw several test graphs, it is worth considering a refinement to the procedure. It is known from the results of Silverman (1976, 1978) that the height of the random fluctuations in f_n'' is, for fixed n, asymptotically proportional to $\{h^{-5}\log(1/h)\}^{\frac{1}{2}}$. Therefore the effect on the test graph of altering the window width can be predicted approximately: to multiply the height of the fluctuations by a factor of λ, the window width should be multiplied by about $\lambda^{-\frac{2}{5}}$. Notice also that, for large sample sizes, a test graph drawn with a slightly larger window width will give a good estimate of f'', since the amount of bias introduced will not be too large. Therefore the explicit measurement of the amounts of systematic variation and random fluctuation will be much easier.

3. Some applications

To see how the procedure works with real data, some circular data of M. A. Stephens, concerning the orientation of turtles, were analysed. These data have been considered by Boneva *et al.* (1971) and by Mardia (1975), among others. The sample size is 76 in this case.

To plot the graphs the data are replicated twice, on [−360°, 0°] and [360°, 720°], and then the graphs are drawn for the portion [0°, 360°]. Further replication is not necessary since the width of the support of the kernel is less than 720° for all the window widths considered. Test graphs for various window widths were drawn; see Fig. 2 for some of these. Those for

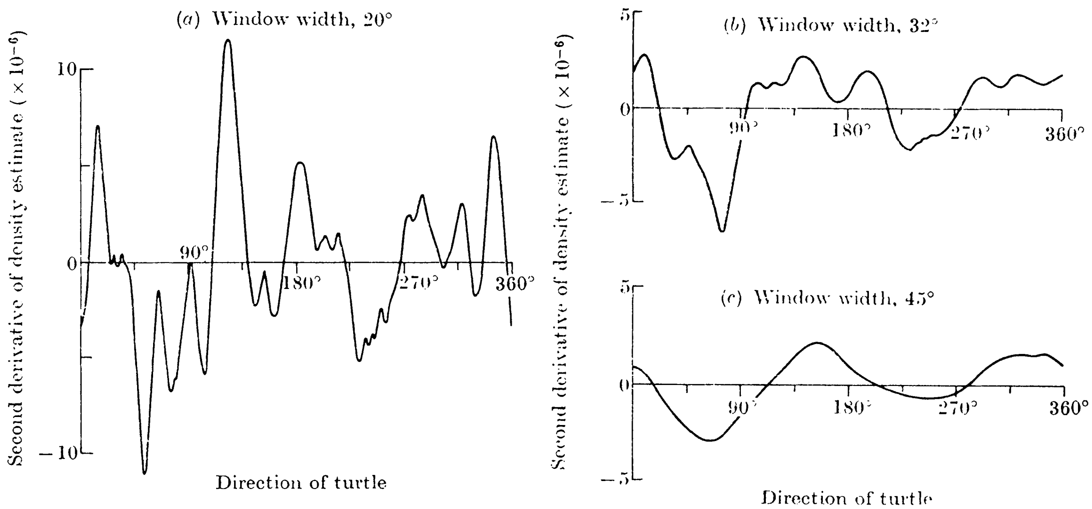

Fig. 2. Test graphs for turtle data for window widths 20°, 32° and 45°.

widths 20° and 45° are clearly unacceptable; comparison with Fig. 1 and application of the test graph principle leads to the choice of 32° as the ideal window width. The density estimate for this window width is shown in Fig. 3, together with the spline transform estimate of Boneva *et al.* for the same data. The window width used for the spline transform is the same as the one used by Boneva *et al* and is 60°; it is impossible to compare this with the 32° used for our kernel because the kernels are different. It is perhaps interesting that the second mode in our analysis is weaker than in the analyses of Mardia and of Boneva *et al*; this provides evidence that the spline kernel of Boneva *et al.* can overemphasize the strength of modes because of the negativity in the cells adjoining the central one.

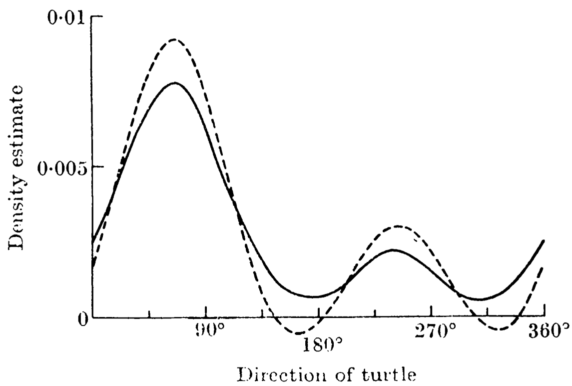

Fig. 3. Solid line: density estimate for turtle data, kernel (1), window width 32°. Dashed line: spline transform of turtle data, window width 60°.

As another example, some data obtained and kindly provided by John Kent in the investigation of the distribution of the maximum of the cosine quantogram (Kendall, 1974) have been considered. The set of data is that used for the construction of curve 1 in Kent's

(1976) discussion of the paper by P. R. Freeman. The test graphs for various window widths are shown in Fig. 4(a), (b) and (c) and the final estimate in Fig. 4(d). The skewness of the data is clearly visible.

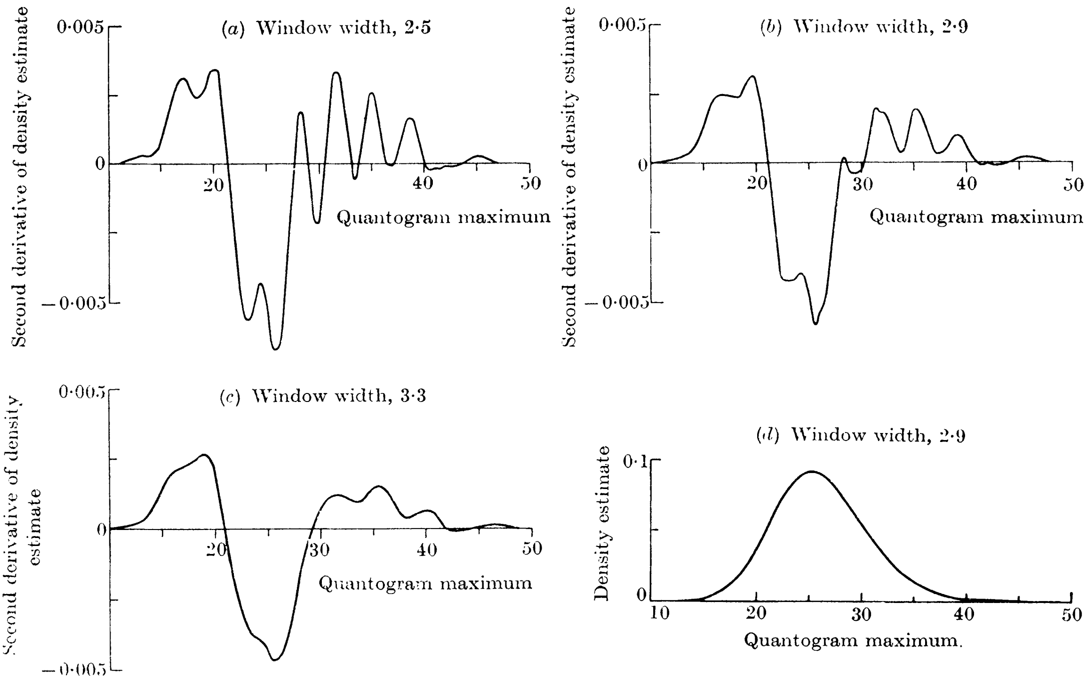

Fig. 4. (a), (b) and (c) Test graphs for cosine quantogram data with window widths 2·5, 2·9 and 3·3. (d) Density estimate for cosine quantogram data, window width 2·9.

4. The multivariate case

The test graph method generalizes easily to the multivariate case. Suppose that $X_1, \ldots, X_n$ are independent identically distributed observations from a d-dimensional density f. Suppose δ is a d-dimensional kernel. Define the estimate f_n by

$$f_n(x) = \sum_{j=1}^{n} n^{-1} h^{-d} \delta\{h^{-1}(x - X_j)\}.$$

The test graph in this case is

$$\nabla^2 f_n(x) = \sum_{j=1}^{n} n^{-1} h^{-d-2} \nabla^2 \delta\{h^{-1}(x - X_j)\}.$$

Because of some technical difficulties the proof of the basic theorem is partly heuristic, but the result remains the same; the best window width for the estimation of the density is the one which gives fluctuations of length $k \sup|\nabla^2 f|$ in the test graph, which is of course in fact a d-dimensional hypersurface. As before, the constant k can be calculated from the kernel and the value of $\sup|\nabla^2 f|$ can be estimated from the test graph, so that the statement of the test graph principle is the same.

Notice that the same window width is used in each coordinate direction. A full treatment would, of course, involve an arbitrary positive-definite symmetric window width matrix and is avoided for reasons of theoretical and practical simplicity.

It is only fair to point out some drawbacks of the method. The first is the difficulty of assessing the test graphs; the best that can conveniently be done in the two-dimensional case is to look at a contour plot or some similar representation of the test graph. In higher dimensions this difficulty becomes much more serious. The second drawback is the fact that the ordinate of the test graph has to be evaluated at a considerable number of points to get even a rough idea of its shape. Because also a moderately large number of observations are needed before we can have any hope of uniformly estimating a multivariate density at all accurately, the method is likely to be quite expensive on the computer. For example, in order to construct each of the test contour diagrams in Fig. 5 it was necessary to evaluate $\nabla^2 f_n$ at 2500 points, taking about 18 seconds of IBM 370/165 time.

Despite these drawbacks, the method applied to a simulated sample seemed to provide quite good results. Two hundred samples were simulated from a mixture of two bivariate normal distributions, with means $(1, 1)$ and $(-1, -1)$ and unit variance–covariance matrices. The remarks above about the choice of kernel in the univariate case still hold; therefore it was decided to choose as kernel the function obtained by rotating the piecewise quartic used

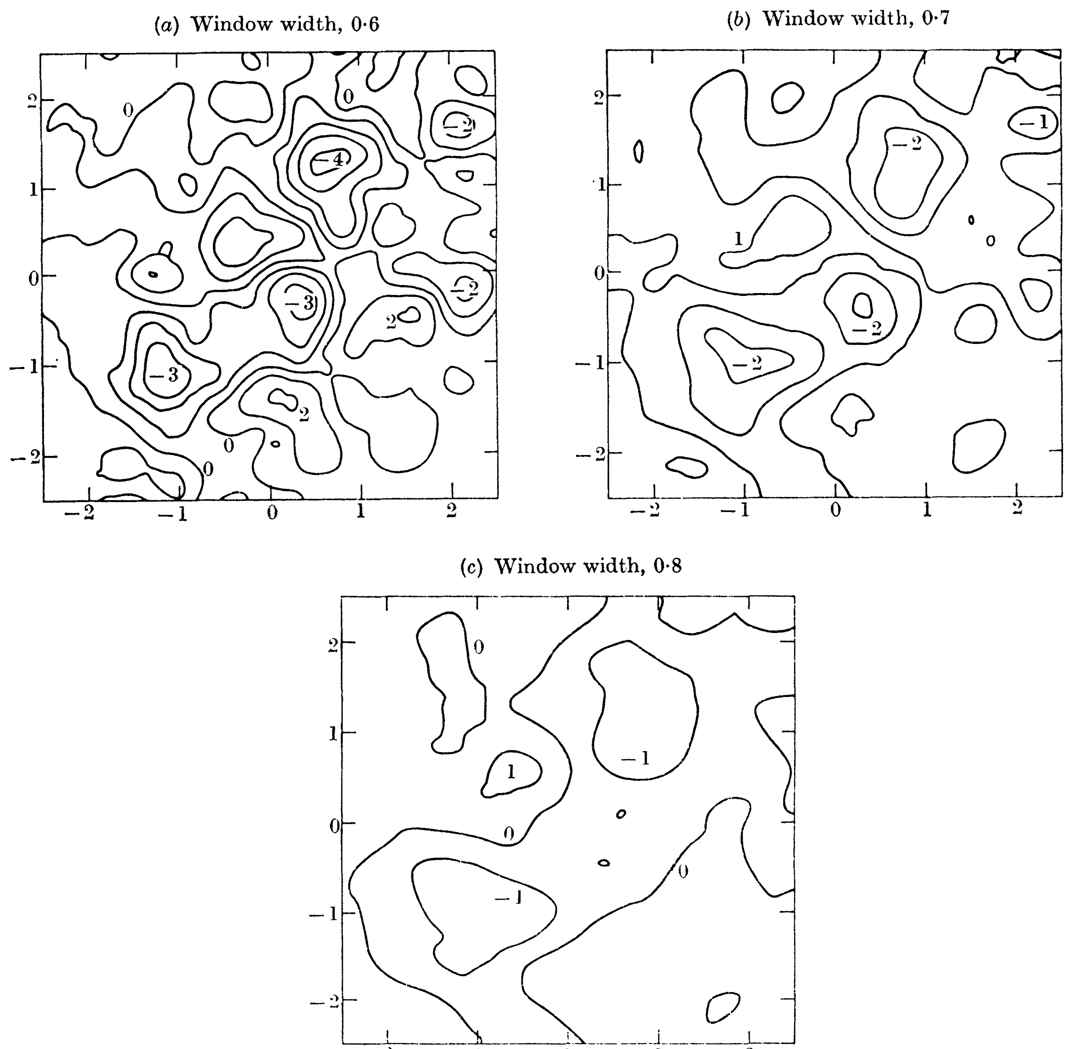

Fig. 5. Test contour diagrams for bivariate normal mixture with window widths, 0·6, 0·7 and 0·8.

in the univariate case and normalizing so that the integral of δ is 1. The resulting kernel is quite close to optimal in the Epachenikov sense. The value of k turns out to be 0·61 in this case, so that the ideal test graph will have random fluctuations which are slightly larger relative to the systematic variation than in the univariate case.

Figure 5 gives test graphs for several window widths. All the graphs are drawn on the square with corners at ($\pm$ 2·5, $\pm$ 2·5). The contour levels are the same on all the test graphs; the contours are numbered on a linear vertical scale. The vertical interval between contours is 0·08.

The graph with window width 0·8 is clearly too smooth; that with window width 0·6 is too rough, particularly considering the fluctuation reaching the -4 level. For the value 0·61 of the constant k, the graph for window width 0·7 is about right, especially bearing in mind the fluctuation near the origin. Experimentation with various window widths for the density estimate shows that window widths around 0·7 give the best estimates. The estimate for window width 0·7 is given in Fig. 6. It picks out both the shape of the density and the position of the modes quite well. In this case the contours are drawn at equal vertical intervals of 0·01 and are in the obvious ascending order. For comparison a corresponding contour diagram for the true density is also shown.

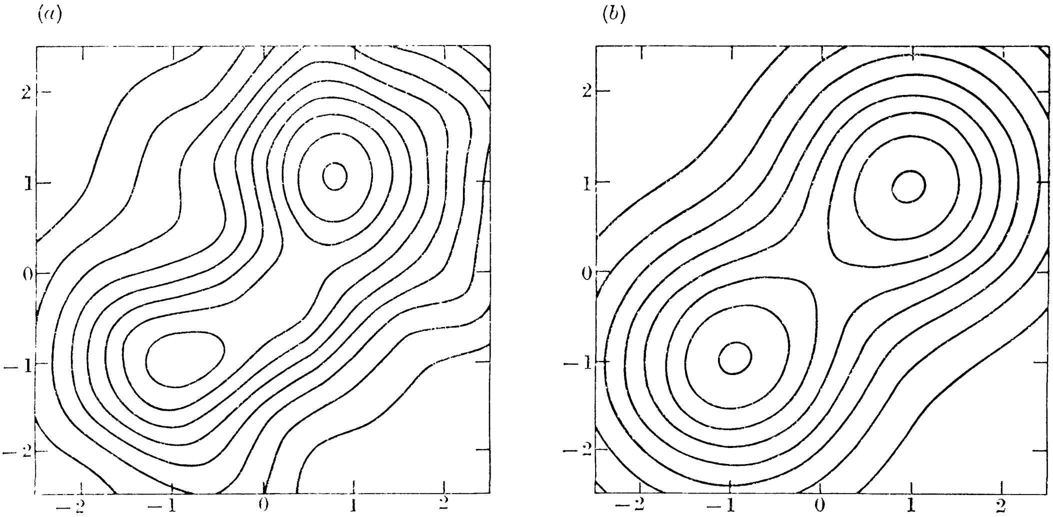

Fig. 6. (a) Density estimate for bivariate normal mixture, window width 0·7. (b) Theoretical bivariate normal mixture density.

In conclusion, it may not be a very good idea to use the test graph alone for choosing the window width for multivariate data; however, the method may be very useful for checking that a chosen window width is sensible.

5. The theoretical background

In order to justify the test graph method, it is necessary to amplify the results of Silverman (1978) on the rate of uniform consistency of the kernel estimate. It is possible to put together various results to find the best possible rate of convergence of f_n to f as n tends to infinity. The technique, used by Rosenblatt (1956) for the mean integrated squared error, is to balance the systematic and random errors; this is done because the bias increases and the random error decreases as h increases. The optimal rates of convergence depend on the amount of regularity assumed for the density. As Bartlett (1963) pointed out, the more derivatives f is assumed to have, the faster the rate of convergence of the bias to zero that can be obtained by choosing a suitable kernel. However, Bartlett also noted that very large sample sizes are necessary before any advantage can be taken of these faster rates. In this paper therefore, f is assumed to have only two uniformly continuous derivatives; the conditions on δ are then satisfied by most smooth symmetric kernels.

Following Silverman (1978), decompose the density estimate as

$$f_n(x) = f(x) + b(x) + n^{-\frac{1}{2}}\rho(x) + \varepsilon(x),$$

where f is the true density, b is the bias $\{E(f_n) - f\}$, ρ is a Gaussian process with the same variance–covariance structure as $n^{\frac{1}{2}}\{f_n - E(f_n)\}$, and ε is a secondary random error term which is uniformly $O(n^{-1}h^{-1}\log n)$.

Taking unspecified integrals to be over $(-\infty, \infty)$ throughout this section, we place the following conditions on the kernel δ:

(a) δ has uniformly continuous second derivative of bounded variation;
(b) $\delta^{(j)}(x) \to 0$ as $x \to \infty$ for $j = 0, 1, 2$;
(c) $\int |\delta^{(j)}|\, dx < \infty$ for $j = 0, 1, 2$; $\int \delta\, dx = 1$;
(d) $\int_0^1 \{\log(1/u)\}^{\frac{1}{2}}\, d\gamma_j(u) < \infty$ for $j = 0$ and 2, where γ_j is the positive square root of the modulus of continuity of $\delta^{(j)}$;
(e) $\int |x \log |x||^{\frac{1}{2}} |\delta'(x)|\, dx < \infty$;
(f) the Fourier transform of δ is not identically zero in any neighbourhood of zero;
(g) $\int x\delta(x)\, dx = 0$;
(h) $\int x^2 \delta(x)\, dx \neq 0$.

Although these conditions appear restrictive, most of them are very mild. They are satisfied by a large variety of kernels, including the one used in this paper. The spline kernel of Boneva *et al.* (1971) unfortunately fails to satisfy conditions (a), (d) and (h).

The following lemma gives the asymptotic behaviour of the window width for the fastest rate of uniform consistency of the kernel density estimate.

LEMMA. *Suppose that f has uniformly continuous and bounded second derivative and that the kernel δ satisfies conditions (a)–(h). For the best rate of uniform convergence in probability of f_n to f,*

$$n^{-1}h^{-5}\log(1/h) \to C(\sup|f''|)^2/(\sup f)$$

as n tends to infinity, where

$$C = \frac{1}{8}\left\{\int x^2\,\delta(x)\,dx\right\}^2\left(\int \delta^2\,dx\right)^{-1},$$

a constant depending only on the kernel δ.

Proof. It is necessary for uniform consistency of f_n to have h tending to zero as n tends to infinity; for a proof see Silverman (1978). Following Bartlett (1963), by elementary manipulation applying Taylor's theorem, we have that as $h \to 0$

$$h^{-2}\sup|b| \to \tfrac{1}{2}\sup|f''| \int x^2\,\delta(x)\,dx. \tag{2}$$

It can be assumed that $(nh)^{-1}(\log n)^2\{\log(1/h)\}^{-1}$ tends to zero and hence (Silverman, 1978) that $\sup|\varepsilon|$ is negligible compared with $n^{-\frac{1}{2}}\sup|\rho|$. For otherwise, letting $\tilde{f}_n$ be the estimate based on all the observations except X_1,

$$f_n(x) - (n-1)\,n^{-1}\tilde{f}_n(x) = n^{-1}h^{-1}\,\delta\{h^{-1}(x - X_1)\},$$

so that it is impossible for $nh \sup|f_n - f|$ to tend to zero in probability. This would imply that the rate of uniform consistency of f_n was at best $(\log n)^{-2}$, a worse rate than that obtained below.

With this restriction on h, the best rate of uniform convergence of f_n to f will be attained when h tends to zero in such a way that in probability as $n \to \infty$

$$n^{-\frac{1}{2}} \sup|\rho| / \sup|b| \to 1. \tag{3}$$

Any other rate for h will cause one of b and $n^{-\frac{1}{2}}\rho$ to converge to zero more slowly than the optimal rate for f_n and the other to converge more quickly, so that f_n will converge to f more slowly. It was shown by Silverman (1976) that

$$\{h^{-1}\log(1/h)\}^{-\frac{1}{2}} \sup|\rho| \to \left(2 \sup f \int \delta^2\right)^{\frac{1}{2}} \tag{4}$$

in probability as n tends to infinity. Put (2), (3) and (4) together to complete the proof of the lemma.

The generalization of Theorem B of Silverman (1978) to the estimation of f'' states that, in probability as n tends to infinity,

$$\{n^{-1}h^{-5}\log(1/h)\}^{\frac{1}{2}} \sup\{f_n'' - E(f_n'')\} \to C_1,$$

$$\{n^{-1}h^{-5}\log(1/h)\}^{\frac{1}{2}} \inf\{f_n'' - E(f_n'')\} \to -C_1,$$

where

$$C_1 = \left\{2 \sup f \int (\delta'')^2 \, dx\right\}^{\frac{1}{2}}.$$

Substituting the result of the lemma into these results completes the proof of Theorem 1, the constant k being defined by

$$k = \frac{1}{2}\left|\int x^2 \delta(x)\, dx\right| \left\{\int (\delta'')^2 \, dx \Big/ \int \delta^2 \, dx\right\}^{\frac{1}{2}}.$$

6. The multivariate theory

There are some theoretical difficulties in the multivariate case arising from the lack of a suitable generalization of the result of Komlos, Major & Tusnady (1975) on the strong embedding of the empirical distribution function; however it is not unreasonable to assume that the secondary error ε is negligible compared to $n^{-\frac{1}{2}}\rho$ when the shrinking coefficient $h(n)$ is chosen to give the optimum rate of uniform consistency. In addition, attention will be restricted to the part of the density falling within a fixed bounded multidimensional interval, so that Theorem A of Silverman (1976) can be used.

Theorem 2. *Consider the estimate of a d-variate density f given by*

$$f_n(x) = \sum_{j=1}^{n} n^{-1} h^{-d} \delta\{h^{-1}(x - X_j)\},$$

where f is twice uniformly continuously differentiable and the kernel δ *and its second derivative* $\nabla^2\delta$ *satisfy the conditions of Silverman* (1976). *With all unspecified integrals being over* R^d, *suppose that*

$$\int x_i \delta(x)\, dx = 0, \quad \int x_i^2 \delta(x)\, dx = d^{-1}\int \|x\|^2 \delta(x)\, dx \neq 0$$

for $i = 1, \ldots, d$. *Let* $n^{-\frac{1}{2}}\rho$ *be the Gaussian process with mean zero and variance–covariance structure the same as* f_n. *Let* $b(x) = E\{f_n(x)\} - f(x)$. *Suppose I is a compact d-dimensional interval in the* *interior of which* $\sup f$ *and* $\sup|\nabla^2 f|$ *are attained. Suppose that the sequence of shrinking coefficients* $h(n)$ *is chosen so that* $\sup_I|n^{-\frac{1}{2}}\rho|/\sup_I|b|\to 1$ *in probability as* $n\to\infty$. *Then*

$$\sup_I n^{-\frac{1}{2}}\nabla^2\rho\to k\sup|\nabla^2 f|,\quad \inf_I n^{-\frac{1}{2}}\nabla^2\rho\to -k\sup|\nabla^2 f|$$

in probability as n tends to infinity, the constant k being defined by

$$k=(2d)^{-1}\left|\int\|x\|^2\,\delta(x)\,dx\right|\left\{\int(\nabla^2\delta)^2\,dx\bigg/\int\delta^2\,dx\right\}^{\frac{1}{2}}.$$

Proof. Consider first the bias term b. Use o to imply a limit which holds uniformly over x, and f^i and f^{ij} to denote $\partial f/\partial x_i$ and $\partial^2 f/\partial x_i\,\partial x_j$ respectively. By the uniform continuity of the second derivatives of f, and by Taylor's theorem

$$b(x)=\int\{h\Sigma_i t_i f^i(x)+\tfrac{1}{2}h^2\Sigma_i\Sigma_j t_i t_j f^{ij}(x)+o(h^2)\}\,\delta(t)\,dt.$$

Apply the conditions assumed for δ to obtain

$$b(x)=h^2(2d)^{-1}\nabla^2 f(x)\int\|t\|^2\,\delta(t)\,dt+o(h^2),$$

and hence, as h tends to zero,

$$h^{-2}\sup_I|b|\to(2d)^{-1}\sup|\nabla^2 f|\int\|x\|^2\,\delta(x)\,dx.$$

By Theorem A of Silverman (1976),

$$\{n^{-1}h^{-d}\log(1/h)\}^{-\frac{1}{2}}\sup_I|n^{-\frac{1}{2}}\rho|\to\left(2d\sup f\int\delta^2\,dx\right)^{\frac{1}{2}}$$

in probability as n tends to infinity and h tends to zero. Putting the last two relations together, we have that in order that h satisfy the hypotheses of Theorem 2 it is necessary that

$$\lim_{n\to\infty}\{n^{-1}h^{-d-4}\log(1/h)\}^{\frac{1}{2}}=(2d)^{-\frac{3}{2}}\sup|\nabla^2 f|\left|\int\|x\|^2\,\delta(x)\,dx\right|\left(\sup f\int\delta^2\,dx\right)^{-\frac{1}{2}}.$$

In order to work out the corresponding limiting behaviour of $\nabla^2\rho$, apply Theorem A of Silverman (1976) to the process $h^2\nabla^2\rho$ to obtain, in probability as h tends to zero,

$$\{h^{-d-4}\log(1/h)\}^{-\frac{1}{2}}\sup_I\nabla^2\rho\to(2d)^{-\frac{1}{2}}\left\{\sup f\int(\nabla^2\delta)^2\right\}^{\frac{1}{2}}.$$

Putting the last two relations together, and using the analogous result for $\inf\nabla^2\rho$, we complete the proof of Theorem 2.

I acknowledge with great pleasure the encouragement and inspiration of Professor David Kendall and the very helpful remarks of the editor and referees.

REFERENCES

BARTLETT, M. S. (1963). Statistical estimation of density functions. *Sankhyā* A **25**, 245–54.

BONEVA, L. I., KENDALL, D. G. & STEFANOV, I. (1971). Spline transformations: Three new diagnostic aids for the statistical data-analyst. *J. R. Statist. Soc.* B **33**, 1–70.

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

KENDALL, D. G. (1974). Hunting quanta. *Phil. Trans. R. Soc. Lond.* A **276**, 231–66.

KENT, J. (1976). Contribution to discussion of paper by P. R. Freeman. *J. R. Statist. Soc.* A **139**, 39–40.

KOMLOS, J., MAJOR, P. & TUSNADY, G. (1975). An approximation of partial sums of independent random variables, and the sample distribution function. *Z. Wahr. verw. Geb.* **32**, 111–31.

MARDIA, K. V. (1975). Statistics of directional data. *J. R. Statist. Soc.* B **37**, 349–93.

ROSENBLATT, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.* **27**, 832–7.

ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42**, 1815–42.

SILVERMAN, B. W. (1976). On a Gaussian process related to multivariate probability density estimation. *Math. Proc. Camb. Phil. Soc.* **80**, 135–44.

SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6**. To appear.

[*Received April* 1976. *Revised September* 1977]