Statistica Sinica **11**(2001), 515-536

ON THE CALIBRATION OF SILVERMAN'S TEST FOR MULTIMODALITY

Peter Hall and Matthew York

Centre for Mathematics and Its Applications, Australian National University

Abstract: It is known that Silverman's bootstrap test for multimodality tends towards conservatism, even in large samples, in the sense that the actual level tends to be less than the nominal one. In this paper, and in the context of testing for a single mode, we propose a means of calibrating Silverman's test so as to improve its level accuracy. The calibration takes two forms — first, an asymptotic approach, which involves identifying the limiting distribution of the test statistic and adjusting for its departure from a hypothetical Uniform distribution; and second, a Monte Carlo technique, which enables a degree of correction for second-order effects to be incorporated. As an aid to applying Silverman's test to contexts rather different from those he envisaged, for example to the case of testing for the number of modes of a density in a compact interval, we show that the modes of the density estimator remain separated as the bandwidth is decreased, at least until a point is reached where the first three derivatives of the estimator vanish simultaneously. Theoretical and numerical properties of alternative forms of Silverman's test are addressed.

Key words and phrases: Bandwidth, bootstrap, density estimation, kernel methods, mode, Monte Carlo methods, smoothing parameter.

1. Introduction

In seminal work on bootstrap methods and the analysis of modality, Silverman (1981, 1983) proposed a method for testing the null hypothesis that a distribution has j modes, versus the alternative that it has $j+1$ or more modes. It is known that Silverman's test is not asymptotically accurate, in the sense that even for infinite sample sizes its exact significance level is different from the nominal one, although the extent of level errors is not known with any precision. In the present paper, in the important case $j = 1$, we describe the level inaccuracy of Silverman's test in both theoretical and numerical terms, and suggest a way of calibrating the test so as to improve its accuracy.

Calibration produces a test with asymptotically correct level accuracy, and may be conducted in at least two ways. First, a straightforward adjustment, depending only on the significance level and applicable for all sample sizes, may be used. The size of the adjustment is given explicitly by a table, in the case of a formal test based on a preset level, or by an interpolation formula, if p-values are to be determined from data. Second, Monte Carlo methods based on simulation from a unimodal density may be employed to calibrate the test in a manner that adjusts for second-order effects, such as those due to sample size.

It should be noted that Silverman's test has a number of characteristics that make it rather exceptional in the class of bootstrap procedures. In particular, the bootstrap part of the algorithm does not consistently estimate the distribution of the test statistic under the null hypothesis, even up to scale or location changes. Therefore, the operation of calibrating the test amounts to substantially more than adjusting its level, as would be the case in more classical problems. The nature of the test has to be altered, so as to render it asymptotically accurate. Furthermore, the so-called critical bandwidth for the test has a proper limiting distribution under the null hypothesis, unlike more conventional settings where it is asymptotically constant. Therefore, the test really involves a critical distribution, not just a single value. Calibrating the distribution is important to producing a test with accurate level.

We also show that when the kernel function is a normal density, the modes and troughs of the density estimator remain isolated as bandwidth is decreased, unless or until a point is reached where the first *three* derivatives of the density vanish simultaneously. (This is a very rare event, particularly for bandwidths of the critical size used in bootstrap testing for modality.) The fact that turning points remain separate provides motivation for using the normal kernel in the testing problem, additional to that offered by Silverman (1981) in his demonstration of the monotonicity of the total number of modes. Silverman's result, although very important, applies only to the case where the number of modes on the whole line is addressed, and there the testing problem is often made more difficult by spurious modes arising from outlying data values. The ability to track modes as the bandwidth decreases enables the mode testing problem to be addressed within a compact interval, even if the density has unbounded support. We study this modified version of Silverman's problem, as well as its more traditional form.

Related work on the problem of testing for modality, or bump hunting as it is often called, includes the penalised likelihood approach suggested by Good and Gaskins (1980), Hartigan and Hartigan's (1985) DIP test (see also Hartigan (1985)), and Müller and Sawitzki's (1991) excess mass method. The DIP and excess mass approaches are numerically equivalent, and a method for calibrating them has been suggested by Cheng and Hall (1998). Work of Polonik (1995a,b), developing the ideas behind the excess mass approach, should also be mentioned in this context. Minotte and Scott (1993) and Minotte (1997) have introduced and developed the concept of a mode tree, as both an exploratory tool for bump hunting and an aid to formally testing hypotheses about modality.

Mammen, Marron and Fisher (1992) conducted an extensive theoretical study of properties of Silverman's method; see also Mammen (1991a, b). The same authors (Fisher, Mammen and Marron (1994)) addressed numerical properties of the technique. Silverman (1986, Section 6.3) reviewed methods for analysing modality, and Izenman (1991) discussed testing for multimodality in his account of more recent developments in density estimation. Cuevas and Gonzales–Manteiga (1991) described methods for bandwidth choice when it is desired to match the number of modes of the estimated and true densities.

Assessment of modality is often an important part of the analysis of mixture models, for example in the work of Roeder (1990, 1994) and Escobar and West (1995). Izenman and Sommer (1988) provided an extensive description of the application of Silverman's and other methods to a problem on identifying the number of modes in a mixture. They included an account of adaptations of Silverman's approach for dealing with spurious modes caused by outlying data clusters. Our work in testing for the number of modes in a compact interval is relevant to these adaptations.

Section 2 introduces Silverman's test, in both its original and its modified forms for testing on a compact interval. In that section we introduce our calibrated form of the test, and describe its general properties; we discuss and make explicit the feature that turning points remain separated as bandwidth is decreased; and we briefly address the general case where the null hypothesis is the existence of j modes. Section 3 develops theory describing the test, showing explicitly that the limiting bootstrap distribution of the test statistic does not depend on unknowns. This confirms our claim that the calibrated form of the test has asymptotically correct level. Section 4 presents numerical work that corroborates and complements these theoretical conclusions. There we also tabulate the degree of adjustment necessary for the non–Monte Carlo approach to calibration. Section 5 outlines technical details behind the results in Section 3.

2. Bootstrap Test and Its Properties

2.1. Testing the hypothesis of a single mode

Given a dataset $\mathcal{X} = \{X_1, \ldots, X_n\}$ from a distribution with unknown density f, we wish to test the null hypothesis H_0 that f has a single mode in the interior of a given closed interval $\mathcal{I}$, and no local minimum in $\mathcal{I}$, against the alternative hypothesis H_1 that f has more than one mode in $\mathcal{I}$. To this end, construct the kernel density estimator

$$\hat{f}_h(x) = (nh)^{-1} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{h}\Big), \tag{2.1}$$

where h is a bandwidth and K a kernel function. Following Silverman (1981) we take K to be the standard normal density function, for which the number of modes of $\hat{f}_h$ on the whole line is always a nonincreasing function of h. Furthermore, $\hat{f}_h$ is unimodal for all sufficiently large h, and so the quantity

$$\hat{h}_{\text{crit}} = \inf \{h : \hat{f}_h \text{ has precisely one mode in } \mathcal{I}\} \tag{2.2}$$

is well-defined, at least when $\mathcal{I}$ is the whole line. The definition of $\hat{h}_{\text{crit}}$ for more general $\mathcal{I}$ will be discussed in Sections 2.3 and 3.4.

If H_0 is true and $\mathcal{I}$ is bounded, then under appropriate conditions on f (see e.g. Corollary 2.1 of Mammen, Marron and Fisher (1992)), $\hat{h}_{\text{crit}}$ is of size $n^{-1/5}$, in the sense that $\lim_{C_1 \downarrow 0, C_2 \uparrow \infty} \liminf_{n\to\infty} P(C_1 n^{-1/5} < \hat{h}_{\text{crit}} < C_2 n^{-1/5}) = 1$. On the other hand, if H_1 is true then $\hat{h}_{\text{crit}}$ may not even converge to zero with n (see Section 3.3). Therefore, unduly large values of $\hat{h}_{\text{crit}}$ provide evidence of the invalidity of H_0. Rigorous assessment of the size of $\hat{h}_{\text{crit}}$ may be conducted via the bootstrap, as follows.

Let $\hat{f}_{\text{crit}}$ denote the version of $\hat{f}_h$ that we obtain by putting $h = \hat{h}_{\text{crit}}$. Conditional on $\mathcal{X}$, let $X_1^*, \ldots, X_n^*$ be a resample drawn from the distribution with density $\hat{f}_{\text{crit}}$, and put

$$\hat{f}_h^*(x) = (nh)^{-1} \sum_{i=1}^{n} K\Big(\frac{x - X_i^*}{h}\Big).$$

Let $\hat{h}_{\text{crit}}^*$ denote the version of $\hat{h}_{\text{crit}}$ in this setting, i.e. the infimum of all bandwidths h such that $\hat{f}_h^*$ has precisely one mode. The test statistic is the bootstrap distribution of $\hat{h}_{\text{crit}}^*/\hat{h}_{\text{crit}}$, and an α-level test of H_0 against H_1 is to reject H_0 if, for an appropriate quantity λ_α, $P(\hat{h}_{\text{crit}}^*/\hat{h}_{\text{crit}} \le \lambda_\alpha | \mathcal{X}) \ge 1 - \alpha$.

It is common to set up the bootstrap test a little differently from this, producing a test which is equivalent to that above when $\lambda_\alpha \equiv 1$. This is based on a notion that the distribution of $U_n = P(\hat{h}_{\text{crit}}^* \le \hat{h}_{\text{crit}} | \mathcal{X})$ is not far from being uniform on the interval $(0, 1)$, at least for large values of n. This is not strictly correct, as we note in Section 3. Our reformulation of the test allows it to be calibrated for level accuracy arising from nonuniformity of the distribution of U_n.

Specifically, given a value of α we show in Section 4 how to choose λ_α in a completely deterministic way (in fact, using a rational-polynomial approximation to express λ_α in terms of α) such that the test has asymptotic level α. Note particularly that we are not calibrating by adjusting the nominal level, α; rather, we are calibrating by selecting an appropriate λ_α for any given value of α.

2.2. Properties of the test

We shall note in Section 3 that, under H_0, the bootstrap distribution function $\hat{G}_n$, defined by $\hat{G}_n(\lambda) = P(\hat{h}_{\text{crit}}^*/\hat{h}_{\text{crit}} \le \lambda | \mathcal{X})$, converges weakly to a stochastic process $\hat{G}$ whose distribution does not depend on unknowns. (Each realization of $\hat{G}$ is a distribution function.) The finite-dimensional distributions of $\hat{G}$ are absolutely continuous, and the realizations of $\hat{G}$ are continuous functions with probability 1. Hence, there exists a unique absolute constant λ_α such that $P\{\hat{G}(\lambda_\alpha) \geq 1-\alpha\} = \alpha$. Using this definition of λ_α, tabulated in Section 4, the bootstrap test is asymptotically correct, in that

$$P\{P(\hat{h}^*_{\text{crit}}/\hat{h}_{\text{crit}} \leq \lambda_\alpha|\mathcal{X}) \geq 1-\alpha\} \rightarrow P\{\hat{G}(\lambda_\alpha) \geq 1-\alpha\} = \alpha. \tag{2.3}$$

This approach will be referred to as "Method 1" in Section 4.

Alternatively, since $\hat{G}$ does not depend on unknowns then we may estimate λ_α by Monte Carlo methods. This involves drawing samples of size n from a distribution whose density is unimodal with its mode interior to $\mathcal{I}$, and applying the bootstrap test to each sample. That approach is potentially better able to correct for second-order effects, for example by taking into account the influence of a specific sample size. It will be referred to as "Method 2" in Section 4.

We show in Section 3.3 that, under H_1, the bootstrap distribution converges to a degenerate mass at the origin, in the sense that $P\{\hat{G}_n(\lambda) \leq x\} \rightarrow 0$ for all $\lambda > 0$ and all $x < 1$. Therefore, the bootstrap test is consistent.

2.3. Spurious modes, and the definition of $\hat{h}_{\text{crit}}$, for general intervals $\mathcal{I}$

If both the support of f and the interval $\mathcal{I}$ are unbounded then properties of $\hat{h}_{\text{crit}}$ are generally determined by extreme values in the sample, not by the modes of f. For example, if f is a unimodal density for which the upper tail decreases like a constant multiple of $x^{-\beta-1}$ for some $\beta > 0$ (e.g. if the sampling density is Student's t), then the spacings between consecutive pairs of extremes in the sample $\mathcal{X}$ are of size $n^{1/\beta}$, from which it may be proved that if $\mathcal{I}$ is unbounded on the right then $\hat{h}_{\text{crit}}$ *diverges* at least as fast as a constant multiple of $n^{1/\beta}$. For similar reasons, if f is a normal density and $\mathcal{I}$ is unbounded on either the left or the right then $\hat{h}_{\text{crit}}$ cannot decrease to zero any faster than $(\log n)^{-1/2}$. Likewise, if the support of f is compact and lies within $\mathcal{I}$, and if f decreases to zero sufficiently quickly at the extremities of its support, then the size of $\hat{h}_{\text{crit}}$ is driven by the rate of decrease, since that determines the spacings of the extreme order statistics of f.

To avoid these problems one would usually, in practice, take $\mathcal{I}$ to be a compact interval within which f does not vanish. Even for such a choice of $\mathcal{I}$, and even if K is the normal kernel, the number $N(h)$ of modes of $\hat{f}_h$ within $\mathcal{I}$ is not always a monotone function of h. The position of a mode can migrate slightly as h is altered; if it was just inside or outside $\mathcal{I}$ for some h, it can switch to the other side as we alter h in one direction or another.

This causes few practical problems, however, since the positions of modes are readily monitored as h is varied. Indeed, a turning point remains isolated — that is, it does not merge with other turning points — as h is decreased, at least until a bandwidth is reached where the first three derivatives of $\hat{f}_h$ vanish simultaneously, as the result below shows.

Theorem 2.1. *Assume that K is the standard normal density. Given $h_1 > 0$, let x_{h_1} denote a point such that $\hat{f}'_{h_1}(x_{h_1}) = 0$ and $\hat{f}''_{h_1}(x_{h_1}) \neq 0$. As h is decreased through positive values less than h_1, x_h varies continuously through points x satisfying $\hat{f}'_h(x) = 0$, $\hat{f}''_h(x) \neq 0$ and $\mathrm{sgn}\{\hat{f}''_h(x)\} = \mathrm{sgn}\{\hat{f}''_{h_1}(x_{h_1})\}$, at least until a value $h_2 < h_1$ is encountered with the property that $\hat{f}'_{h_2}(x_{h_2}) = \hat{f}''_{h_2}(x_{h_2}) = \hat{f}'''_{h_2}(x_{h_2}) = 0$.*

Moreover, provided f has no turning point on the boundary of $\mathcal{I}$, the probability that $N(h)$ is monotone in h, within a wide range of h's, converges to 1 under relatively general conditions. This result and others enable us to establish the asymptotic validity of the bootstrap test under quite general conditions. See Section 3.4.

2.4. Testing the hypothesis of j modes

The problem of testing the null hypothesis H_{0j} that f has precisely j modes in $\mathcal{I}$, against the alternative that it has $j+1$ or more modes there, is in principle similar for all values of j; see Silverman (1981). Indeed, defining $\hat{h}_{\mathrm{crit},j} = \inf\{h : \hat{f}_h$ has precisely j modes in $\mathcal{I}\}$, $n^{1/5}\,\hat{h}_{\mathrm{crit},j}$ converges in distribution under H_{0j} to $\max_{0\le i\le 2j-1}(c_i R_i)$, where $c_i = f(t_i)^{1/5}/|f''(t_i)|^{2/5}$, $t_1, \ldots, t_{2j-1}$ are the turning points of f in $\mathcal{I}$ (assumed to all satisfy $f''(t_i) \neq 0$), and $R_1, \ldots, R_{2j-1}$ are independent and identically distributed random variables with a distribution that we shall define at (3.2). Intuitively, the independence here follows from the fact that the turning points are well separated, at least in asymptotic terms, and the density estimation is local. The bootstrap distribution function, $\hat{G}_n(\lambda|j) = P(\hat{h}^*_{\mathrm{crit},j}/\hat{h}_{\mathrm{crit},j} \le \lambda|\mathcal{X})$, again converges weakly to a stochastic process, $\hat{G}(\cdot|j)$, but unless $j = 1$ the distribution of the latter process depends on the $2j-2$ unknowns $\{c_i/c_1 : 2 \le i \le 2j-1\}$. Therefore, if $j \ge 2$ then the bootstrap test cannot be calibrated by simply forming the ratio $\hat{h}^*_{\mathrm{crit},j}/\hat{h}_{\mathrm{crit},j}$.

3. Theoretical Results

3.1. Distribution of $\hat{h}_{\mathrm{crit}}$ under H_0

We define a *level point* of f to be a real number t such that $f'(t) = 0$, and a turning point to be a level point such that $\mathrm{sgn}\{f'(t+)\} = -\mathrm{sgn}\{f'(t-)\}$. Under

H_0, f has a unique turning point t_0 in $\mathcal{I}$. Put $c = f(t_0)^{1/5}/|f''(t_0)|^{2/5}$, and let

$$Z(r,s) = r^{-3} \int K''(s+u)\, W(ru)\, du, \tag{3.1}$$

where K is the standard normal kernel, W is a standard Wiener process, $r > 0$ and $-\infty < s < \infty$. Define

$$R = \inf\{r > 0 : \text{the function } Z(r,s) + s \text{ changes sign exactly once in the range } -\infty < s < \infty\}, \tag{3.2}$$

and let S be the unique point at which $Y(s) = Z(R,s) + s$ changes sign. (There exists another point S_1 such that $Y(S_1) = 0$, but $\mathrm{sgn}\{Y(S_1+)\} = \mathrm{sgn}\{Y(S_1-)\}$). The variation-diminishing property of the integral operator with kernel K'' ensures that with probability 1 the number of sign changes of $Z(r,s) + s$ is a right-continuous, nonincreasing function of r; see Schoenberg (1950) and Silverman (1981). Similarly, if $\mathcal{I} = (-\infty, \infty)$, then the property ensures that $\hat{h}_{\mathrm{crit}}$ is well-defined by (2.2).

Our first result describes the limiting distribution of $\hat{h}_{\mathrm{crit}}$. Following Mammen, Marron and Fisher (1992) we impose regularity conditions that require f to be compactly supported, and allow $\mathcal{I}$ to be the whole real line. Theorem 3.4 will address alternative settings, where $\mathcal{I}$ is compact and f may be infinitely supported.

Theorem 3.1. *Assume that f is supported on a compact interval $\mathcal{S} = [a,b]$ and has two continuous derivatives there; that it has a mode t_0, giving a local maximum, in the interior of $\mathcal{S}$, with $f''(t_0)\, f(t_0) \neq 0$; that f has no other level points in $\mathcal{S}$; and that $f'(a+) > 0$ and $f'(b-) < 0$. Take $\mathcal{I} = (-\infty, \infty)$. Then, $n^{1/5}\, \hat{h}_{\mathrm{crit}}$ converges in distribution to cR as $n \to \infty$.*

3.2. Bootstrap distribution of $\hat{h}^*_{\mathrm{crit}}/\hat{h}_{\mathrm{crit}}$ under H_0

For each n it is possible to construct the Wiener process W above, depending in part on the data, such that, with R defined at (3.2), we have $n^{1/5}\hat{h}_{\mathrm{crit}} = cR + o_p(1)$. The reader is referred to the "Hungarian embedding" introduced during the proof of Theorem 3.1 in Section 5.2, and in particular to the last paragraph of that section, for details. Let W^* denote a second standard Wiener process, independent of W, and let Z^* have the definition at (3.1) except that W there should now be replaced by W^*. Put

$$\zeta^*(r,s) = Z^*(r,s) + r^{-1}\, R\, Z(R, S + R^{-1}rs) + r^{-1}RS + s,$$
$$R^* = \inf\{r > 0 : \text{the function } \zeta^*(r,s) \text{ changes sign exactly once in the range } -\infty < s < \infty\}.$$

Our next result describes the limiting bootstrap distribution of $\hat{h}^*_{\rm crit}$.

Theorem 3.2. *Assuming the conditions of Theorem* 3.1*,*

$$\sup_{-\infty<x<\infty} |P(n^{1/5}\, \hat{h}^*_{\rm crit} \le cx|\mathcal{X}) - P(R^* \le x|W)| \to 0$$

in probability as $n \to \infty$*.*

Taking $x = n^{1/5}\, c^{-1}\, \lambda\, \hat{h}_{\rm crit} = \lambda R + o_p(1)$ in Theorem 3.2, we deduce that the stochastic process $\hat{G}_n$, defined by $\hat{G}_n(\lambda) = P(\hat{h}^*_{\rm crit}/\hat{h}_{\rm crit} \le \lambda|\mathcal{X})$, converges weakly to $\hat{G}$, where $\hat{G}(\lambda) = P(R^*/R \le \lambda|W)$. By definition of the distributions of R and R^*, the distribution of the stochastic process $\hat{G}$ does not depend on unknowns. This confirms the asymptotic accuracy of the calibrated test proposed in Section 2.2, in particular the validity of formula (2.3) under the null hypothesis.

3.3. Distribution of $\hat{h}_{\rm crit}$ and $\hat{h}^*_{\rm crit}/\hat{h}_{\rm crit}$ under H_1

We show that $\hat{h}_{\rm crit}$ tends to be much larger, and $\hat{h}^*_{\rm crit}/\hat{h}_{\rm crit}$ much smaller, under H_1 than under H_0. This implies that the asymptotic power of the bandwidth test is 1 under fixed alternatives in H_1. Similar arguments may be used to address power under local alternatives, in particular under bimodal distributions where the second mode is $n^{-1/5}$ distant from the first. These are the so-called "difficult cases" of Cheng and Hall (1999).

Theorem 3.3. *Assume the conditions of Theorem* 3.1*, except that where before* f *had just one local maximum in the interior of* $\mathcal{S}$*, we now ask that it have* m *turning points* $t_1, \ldots, t_m$ *there, that* $2 \le m < \infty$ *and that each* $f(t_i)\, f''(t_i) \ne 0$*. Then* (a) *there exists a constant* $C = C(f) > 0$ *such that* $P(\hat{h}_{\rm crit} > C) \to 1$*, and* (b) *for each* $\lambda > 0$*,* $\hat{G}_n(\lambda) \to 1$ *in probability.*

3.4. Alternative regularity conditions

Here we let $\mathcal{I}$ be a proper subset of the support of f, and show that in such cases the earlier results continue to be valid in an asymptotic sense. By way of notation, let ϵ_n be a sequence of positive constants converging to zero, let $\delta > 0$ be fixed, put $\mathcal{H}_n = [\epsilon_n n^{-1/5}, \delta]$, let $\mathcal{H}'_n$ denote the set of all $h \in \mathcal{H}_n$ such that $\hat{f}_h$ has precisely one mode in $\mathcal{I}$, and redefine $\hat{h}_{\rm crit} = \inf \mathcal{H}'_n$ if $\mathcal{H}'_n$ is nonempty, and $\hat{h}_{\rm crit} = \delta$ otherwise. Let $N(h)$ denote the number of modes of $\hat{f}_h$ within $\mathcal{I}$.

Theorem 3.4. *Assume that* $\mathcal{I}$ *is compact, that* f'' *is bounded and continuous in an open interval containing* $\mathcal{I}$*, that* f *has only a finite number of level points* t *in* $\mathcal{I}$ *at each of which* $f(t)\, f''(t) \ne 0$*, and that neither endpoint of* $\mathcal{I}$ *is a level point of* f*. Then, if* $\epsilon_n \to 0$ *sufficiently slowly and* $\delta > 0$ *is sufficiently small,* (a) *the probability that* $N(\cdot)$ *is nonincreasing on* $\mathcal{H}_n$ *converges to* 1*,* (b) *under*

H_0, the probability that $\mathcal{H}'_n$ is not empty converges to 1*, and* (c) *the conclusions of Theorems* 3.1–3.3 *apply for the above definition of* $\hat{h}_{\text{crit}}$.

The analogues of Theorems 3.1 and 3.2 require that we assume additionally that f have one mode interior to $\mathcal{I}$, and no other turning points there; and the analogue of Theorem 3.3 requires f to have at least two modes interior to $\mathcal{I}$.

It follows that if we restrict attention to bandwidths that are not too small then, with probability tending to 1 as n increases, the calibration methods suggested in Section 2 apply in the case of testing for the number of modes on a compact interval.

4. Numerical Results

In this section we quantify the asymptotic conservatism of the critical bandwidth test, and use two different methods to compute the constant λ_α, required for calibrating the bandwidth test so as to give it correct level. We simulated the calibrated forms of the test on unimodal distributions for sample sizes $n = 50, 100$ and 200, and compared performance with the version of the bandwidth test proposed by Silverman (1981). We studied the power of these different forms of the test by considering a family of bimodal normal mixtures.

In these simulations we determined the critical bandwidth $\hat{h}_{\text{crit}}$ by computing kernel density estimates on an equally spaced grid of 256 points, and searching for the smallest bandwidth that yielded a unimodal density estimate. For unimodal distributions the actual level of the test was estimated by the proportion of times that null hypothesis was rejected, and the same approach was used to estimate the power of the test for bimodal distributions.

To quantify the asymptotic conservatism of the bandwidth test we simulated the test on samples of size $n = 10{,}000$ drawn from a standard normal distribution. Other simulations, not reported here, indicated that samples of this size were sufficiently large to capture the asymptotic behaviour of the test. We drew 5000 samples from the standard normal distribution and, for each sample, we drew 5000 resamples. To avoid problems associated with the detection of spurious modes in the tails of the distribution, discussed in Section 2, we conducted the test over the interval $\mathcal{I} = [-1.5, 1.5]$. The actual test levels, obtained for a range of nominal levels, are given in Table 1. These results indicate that the critical bandwidth test is particularly conservative, even in the limit. When the test is performed at the nominal levels of 0.01, 0.05, 0.1 and 0.2, the actual levels of the test are 0.000, 0.010, 0.032 and 0.102, respectively.

To implement the calibrated forms of the test proposed in this paper, we need to specify the constant λ_α. In Section 2 we discussed two possible approaches to calculating λ_α. The first (Method 1) was an asymptotic correction based on the limiting distribution of the test statistic and the second (Method 2) was based on Monte Carlo techniques, which enabled a degree of correction for second-order effects. To compute the asymptotic version of λ_α we used the results of the simulation which computed the asymptotic level accuracy. The constant λ_α was chosen to produce a test with correct level accuracy for $\alpha = 0.001, 0.002, \ldots, 0.999$. These values of λ_α, determined by Monte Carlo simulation, are plotted on Figure 1. The somewhat erratic nature of the plotted points there is the result of stochastic error. If we were able to do an infinite number of simulations for each value of α, the plot would be a one-to-one function of α and would be indistinguishable from the smooth curve in Figure 1, on the scale of that figure. A function of the form

$$\lambda_\alpha = \frac{a_1\alpha^3 + a_2\alpha^2 + a_3\alpha + a_4}{\alpha^3 + a_5\alpha^2 + a_6\alpha + a_7} \tag{4.1}$$

was fitted to the output to provide a means of approximating λ_α for arbitrary α. The coefficients are listed in Table 2, and the approximating curve is included in Figure 1.

Table 1. Asymptotic level of Silverman's critical bandwidth test.

Nominal level	Actual level	Nominal level	Actual level
0.005	0.000	0.130	0.050
0.010	0.000	0.140	0.057
0.020	0.002	0.150	0.062
0.030	0.004	0.160	0.070
0.040	0.006	0.170	0.079
0.050	0.010	0.180	0.088
0.060	0.012	0.190	0.094
0.070	0.016	0.200	0.102
0.080	0.021	0.250	0.149
0.090	0.025	0.300	0.202
0.100	0.032	0.350	0.252
0.110	0.038	0.400	0.308
0.120	0.043	0.500	0.423

Table 2. Estimated coefficients for the approximating function for λ_α.

$a1$	$a2$	$a3$	$a4$	$a5$	$a6$	$a7$
0.94029	−1.59914	0.17695	0.48971	−1.77793	0.36162	0.42423

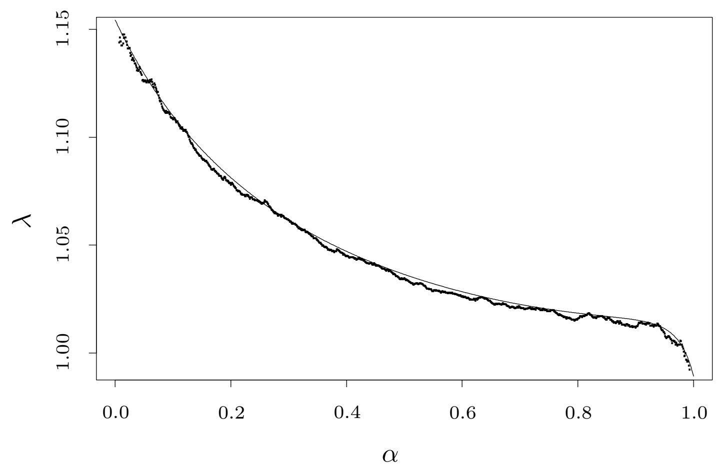

Figure 1. Asymptotic value of λ_α. The points are the empirically computed values and the line is the approximating function.

This asymptotic value of λ_α, needed for Method 1, could have been calculated by simulating the limiting distributions of $\hat{h}_{\text{crit}}$ and $\hat{h}^*_{\text{crit}}$ rather than simulating the test on very large samples. However, while these limiting distributions are known in theory they are extremely complicated, particularly that of $\hat{h}^*_{\text{crit}}$. We found it more straightforward to approximate the asymptotic distributions of $\hat{h}_{\text{crit}}$ and $\hat{h}^*_{\text{crit}}$ by repeatedly applying the test to large samples. For this purpose we took $n = 10{,}000$ and sampled from the standard normal distribution.

In the case of Method 2 one performs Monte Carlo simulation for a sample size equal to that of the dataset, rather than referring to the (effectively asymptotic) case $n = 10{,}000$. Specifically, in order to determine the appropriate value of $\lambda_\alpha(n)$ one would simulate from a unimodal distribution resembling the sampled one, in particular with scale chosen empirically. Simulation would be from that part of the "model" distribution over the interval where the test was being conducted. We found that Monte Carlo simulation from the normal distribution produced a test with good level accuracy over a wide class of sampled distributions, although in some cases the test was anticonservative, as we shall shortly see.

It should be stressed that for none of these methods are we calibrating by adjusting the nominal level, α, of the test. Instead, we are enhancing coverage accuracy by choosing an appropriate λ_α for a specified value of α, rather than simply taking $\lambda_\alpha = 1$ as in Silverman's (1981) standard test.

In our simulation study we considered three versions of the critical bandwidth test. The first was the form of the test proposed by Silverman (1981), equivalent to taking $\lambda_\alpha \equiv 1$. In the description of the bandwidth test in Section 2 we stated that the resampled data are drawn from the distribution with density $\hat{f}_{\rm crit}$. In practice, $\hat{f}_{\rm crit}$ is normally rescaled so that the resampling distribution has the same mean and variance as the data. This adjustment improves level accuracy of the test for small to medium samples sizes, so we used the adjustment in our study. Silverman's test is much more conservative when this rescaling is not used. Our second version of the test, Method 1, also used the rescaled version of $\hat{f}_{\rm crit}$. However for the third testing procedure, Method 2, we did not rescale $\hat{f}_{\rm crit}$, since the value of λ_α used here implicitly incorporates a correction for the variance inflation effect of the kernel density estimator.

To compare the level accuracy of Method 1, Method 2 and Silverman's version of the test we illustrate performance for samples of sizes $n = 50$ and 200, drawn from unimodal distributions. We drew 1000 samples and, for each sample, we drew 1000 resamples. The results for the standard normal, Beta (3,4) and Gamma (3) distributions are shown in the first, second and third rows, respectively, of Figure 2. For the standard normal distribution we conducted the test over the interval $\mathcal{I} = [-1.5, 1.5]$, for the Beta (3,4) we used $\mathcal{I} = [0, 1]$, and for the Gamma (3) we took $\mathcal{I}$ to be [0.5,5]. The normal and gamma distributions had infinite support so we chose $\mathcal{I}$ to be sufficiently wide to contain the majority of the data while still being narrow enough to avoid problems caused by detecting spurious modes in the tails. While the beta distribution had bounded support its right-hand tail tended to zero sufficiently quickly for there to be, potentially, tail problems. We experimented with a range of choices for $\mathcal{I}$, and found that there was little difference in the performance of the test and that the level accuracy of the test actually deteriorated as $\mathcal{I}$ was shortened from [0,1].

The first panel in each row of Figure 2 shows the respective sampling density. The next two panels depict level accuracy for $n = 50$ and 200. The plots reveal that Silverman's method was conservative for all three distributions, for the three sample sizes considered. Methods 1 and 2 produced tests with good level accuracy. Method 2 tended to perform slightly better than Method 1, particularly for the Beta (3,4) distribution. For beta and gamma data both methods are slightly anticonservative when $n = 50$ or 100 (not shown here), although at least for Method 2 this has virtually vanished by $n = 200$.

To examine the increase in apparent power achieved by calibrating the bandwidth test we studied mixtures of normal distributions with densities of the form $f(x) = 0.5\,\phi(x-\mu) + 0.5\,\phi(x+\mu)$, where ϕ is the standard normal density, for $\mu = 1.1, 1.2, \ldots, 2.0$. When $\mu = 1.1$ the two modes are only barely discernible and, as μ increases to 2, the separation and size of the modes increase. The

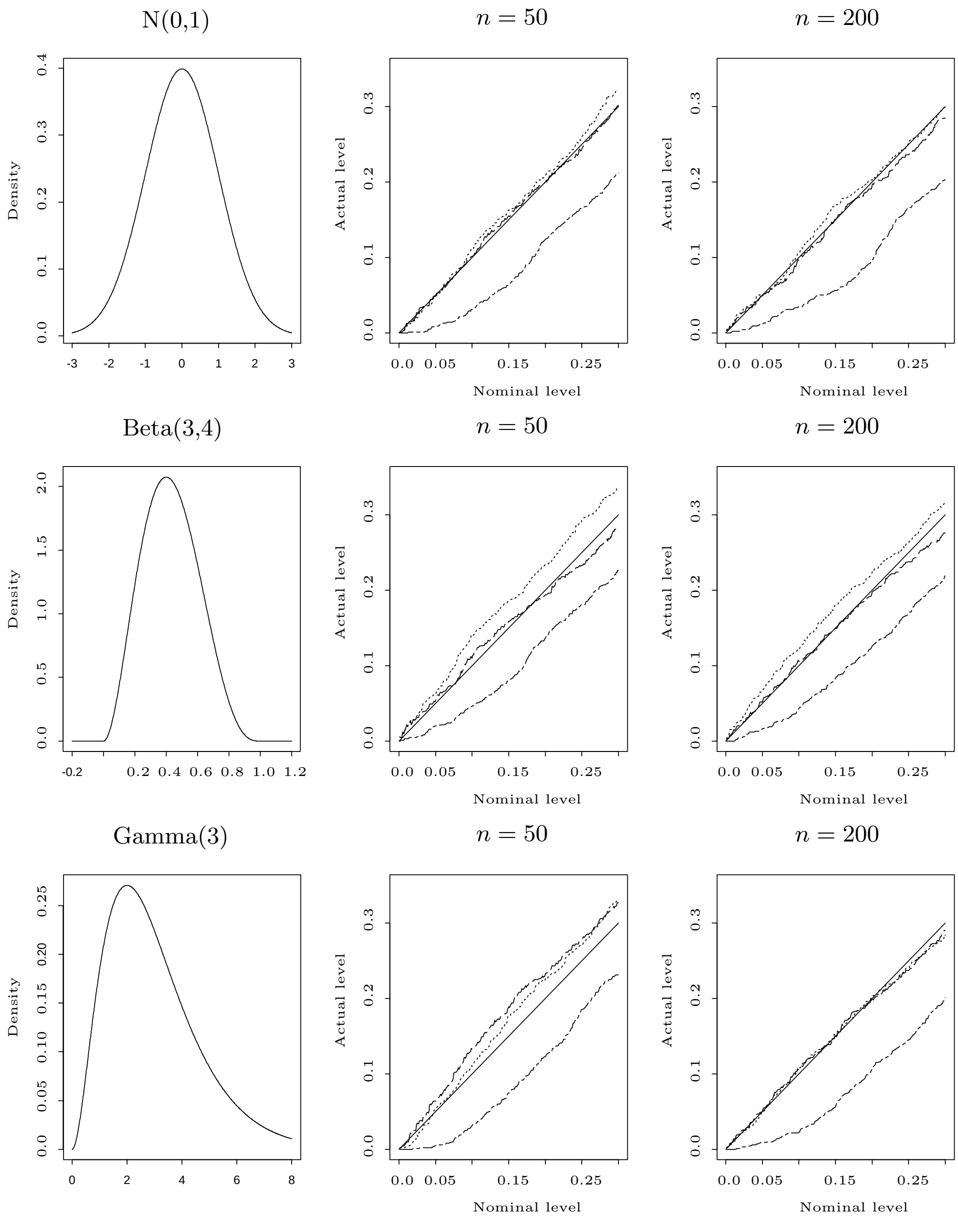

Figure 2. Level accuracy. The three rows show results for the normal, Beta (3,4) and Gamma (3) distributions, respectively. The three panels in each row show the density of the sampling distribution, and results for $n = 50$ and $n = 200$, respectively. In each of the last two panels in each row, level accuracy for Silverman's critical bandwidth test, and for the Method 1 and Method 2 tests, are depicted by a dot-dashed line, a dotted line and a dashed line, respectively.

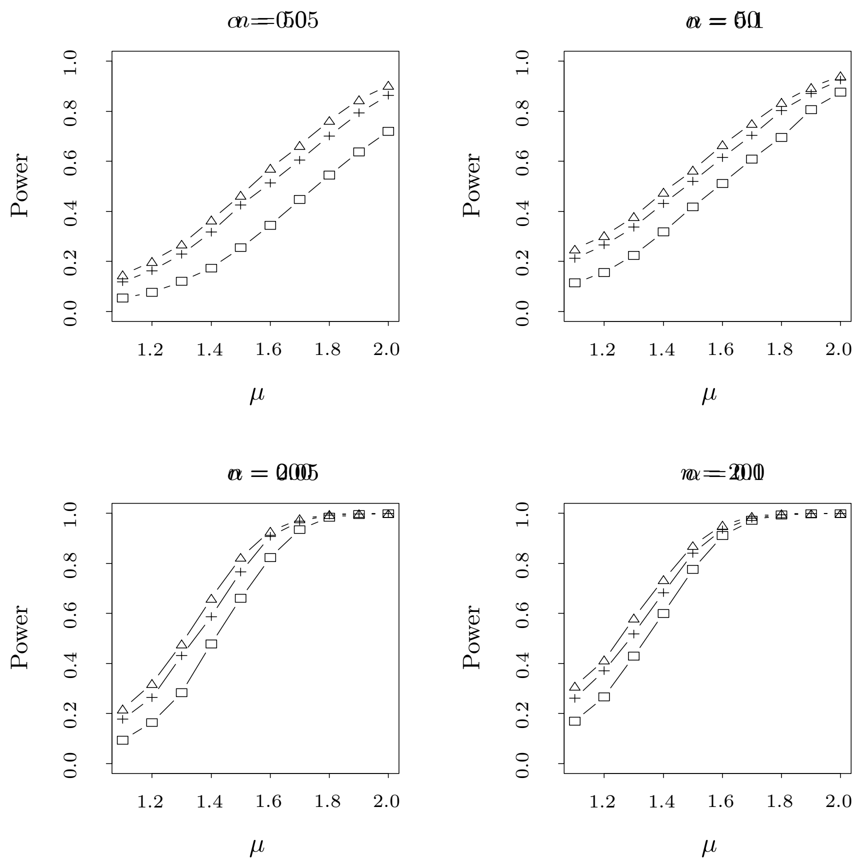

Figure 3. Power for bimodal mixtures of normal distributions. The power of Method 1 (triangle), Method 2 (cross) and Silverman's critical bandwidth test (square) for bimodal mixtures with densities of the form $f(x) = 0.5\,\phi(x-\mu) + 0.5\,\phi(x+\mu)$. The value of μ is plotted on the x-axis. The four panels show the power for samples of size $n = 50$ and $n = 200$ for the levels $\alpha = 0.05$ and $\alpha = 0.1$.

apparent power of the three forms of the test is plotted in Figure 3 for levels $\alpha = 0.05$ and 0.1, and for sample sizes $n = 50$ and 200. The apparent power of all three methods increases as μ increases. Silverman's test is the least powerful, Method 1 and Method 2 generally have about 10% greater apparent power. For the difficult cases of $\mu = 1.1, 1.2$ and 1.3, where the modes are not well-separated, the calibrated versions of the test have over twice the apparent power of the uncalibrated one. Method 1 has more apparent power than Method 2 under the alternative hypothesis, but in some cases it tends to be anti-conservative relative to Method 2; see Figure 2.

5. Technical Arguments

5.1. Outline proof of Theorem 2.1

By properties of derivatives of the normal density, $(\partial/\partial h)\hat{f}_h^{(i)}(x)=h\hat{f}_h^{(i+2)}(x)$. Let x_h denote a turning point of $\hat{f}_h$, so that $\hat{f}_h'(x_h)\equiv 0$, and put $x_h'=(\partial/\partial h)\,x_h$. In this notation,

$$0=\frac{\partial}{\partial h}\,\hat{f}_h'(x_h)=\frac{\partial}{\partial h}\,\hat{f}_h'(x)\Big|_{x=x_h}+\hat{f}_h''(x_h)\,x_h'=h\,\hat{f}_h^{(3)}(x_h)+\hat{f}_h''(x_h)\,x_h'.$$

Therefore, $x_h'=-h\,\hat{f}_h^{(3)}(x_h)/\hat{f}_h''(x_h)$, whence

$$\frac{\partial}{\partial h}\,\hat{f}_h^{(i)}(x_h)=h\,\hat{f}_h^{(i+2)}(x_h)-h\,\hat{f}_h^{(i+1)}(x_h)\,\hat{f}_h^{(3)}(x_h)\,\{\hat{f}_h''(x_h)\}^{-1}. \tag{5.1}$$

As we alter h, x_h varies continuously through values of x such that $\hat{f}_h'(x)=0$. Suppose that $\hat{f}_{h_1}''(x_{h_1})<0$ for some h_1. If it is not true that $\hat{f}_h''(x_h)<0$ for all $0<h<h_1$, then the quantity $h_2=\sup\{h<h_1:\hat{f}_h''(x_h)=0\}$ is well-defined, and $0<h_2<h_1$. Suppose $\hat{f}_{h_2}^{(3)}(x_{h_2})\neq 0$. Taking $i=2$ in (5.1), we have that $(\partial/\partial h)\,\hat{f}_h''(x_h)\to+\infty$ as $h\downarrow h_2$. This means that $\hat{f}_h''(x_h)$ decreases through negative values as $h\downarrow h_2$, contradicting the hypothesis that $\hat{f}_{h_2}''(x_{h_2})=0$.

Similarly, if $\hat{f}_{h_1}''(x_{h_1})>0$ for some h_1 then $\hat{f}_h'(x_h)>0$ for all $h<h_1$, unless or until a point is reached at which the first three derivatives of $\hat{f}_h$ vanish simultaneously; call this event $\mathcal{E}$. The result proved in the previous paragraph means that we may track a mode as h decreases, and it keeps the mode property. It will be uniquely defined unless it merges with another mode. In this event, since there must be at least one local minimum between the two modes, the local minimum must merge with the two modes at the same time as the modes merge, which is precluded by the result in the first sentence of this paragraph unless or until h has become so small that $\mathcal{E}$ occurs. Hence, each mode (and similarly, each local minimum) remains distinct as we decrease h, unless or until $\mathcal{E}$ occurs.

5.2. Outline proof of Theorem 3.1

Put $h_0=n^{-1/5}$. It is known from Corollary 2.1 of Mammen, Marron and Fisher (1992) that, under the conditions of Theorem 3.1,

$$\lim_{C_1\downarrow 0,C_2\uparrow\infty}\ \liminf_{n\to\infty}\ P(C_1h_0\le\hat{h}_{\rm crit}\le C_2h_0)=1 \tag{5.2}$$

$$\lim_{C_1\downarrow 0,C_2\uparrow\infty}\ \limsup_{n\to\infty}\ P\{\hat{f}_h'(x)=0\text{ for some }h\in[C_1h_0,\,C_2h_0]\text{ and }x\notin[t_0-C_2h_0,\,t_0+C_2h_0]\}=0. \tag{5.3}$$

By applying the approximation of Komlós, Major and Tusnády (1975) we may prove that in the neighbourhood of any turning point t_0 of f one can approximate $\hat{f}_h'(t) - E\{\hat{f}_h'(t)\}$ by

$$Z_1(h,t) = -n^{-1/2}h^{-3} \int K''\Big(\frac{t-x}{h}\Big)\, W^0\{F(x)\}\, dx, \tag{5.4}$$

where W^0 is a standard Brownian bridge and F is the distribution function associated with f. During the proof of Theorem 4 of Mammen, Marron and Fisher (1992) it is shown that if $h = h_0 R^\dagger$ denotes the infimum of values of h such that $Z_1(h,t) + f''(t_0)\,(t-t_0)$ (as a function of t) has precisely one zero, then $\hat{h}_{\rm crit}/h_0 = R^\dagger + o_p(1)$ as $n \to \infty$.

In the integrand at (5.4) we may replace $W^0\{F(x)\}$ by $D(x) = W^0\{F(x)\} - W^0\{F(t_0)\}$ without affecting the value of the integral. Using properties of the modulus of continuity of a Gaussian process (see e.g. Garsia (1970)) one may prove that $D(t_0 + hs) = W^0\{F(t_0) + hs\, f(t_0)\} - W^0\{F(t_0)\} + O_p\{h_0\,(\log n)^{1/2}\}$ uniformly in $|hs| = O(h_0)$. Therefore, defining

$$W_1(u) = \{h_0 f(t_0)\}^{-1/2}\, [W^0\{F(t_0) - h_0 u\, f(t_0)\} - W^0\{F(t_0)\}]$$

(a standard Wiener process) and

$$Z_2(\rho, s) = -f(t_0)^{1/2}\, \rho^{-3} \int K''(s+u)\, W_1(\rho u)\, du\,,$$

we have

$$Z_1(h, t_0 + hs) = h\, Z_2(h/h_0, s) + O_p\Big\{(h_0^3 \log n)^{1/2}\Big\}\,,$$

uniformly in $h \in [C_1 h_0,\, C_2 h_0]$ and $s \in [-C_2 h_0/h,\, C_2 h_0/h]$ for any $0 < C_1 < C_2 < \infty$. Noting (5.2) and (5.3) we see that this limitation on h and s is no impediment, and thence that $R^\dagger = R^{\dagger\dagger} + o_p(1)$, where $R^{\dagger\dagger}$ denotes the infimum of values of ρ such that $Z_2(\rho, s) + f''(t_0)\, s$ (as a function of s) has precisely one zero.

With c as defined in Section 3, put $W(t) = {\rm sgn}\{f''(t_0)\}\, c^{-1/2} W_1(ct)$ and, for this W, define $Z(r,s)$ as in Section 3. Then W is a standard Wiener process, and $Z_2(\rho,s)/f''(t_0) = Z(\rho/c, s)$. Therefore, $R^{\dagger\dagger} = cR$, and so $\hat{h}_{\rm crit}/h_0 = R^\dagger + o_p(1) = R^{\dagger\dagger} + o_p(1) = cR + o_p(1)$, which completes the proof of Theorem 3.1.

5.3. Outline proof of Theorem 3.2

Work through the above argument a second time, with $\hat{f}_h$ and f replaced by $\hat{f}_h^*$ and $\hat{f}_{\rm crit}$, respectively, and with all probabilities computed in the bootstrap distribution, conditional on $\mathcal{X}$. The theory is similar, with several changes. First, W_1 should be replaced by W_1^*, which (conditional on $\mathcal{X}$) is a Wiener process.

Second, the quantity $Z_2(\rho, s)$ in the formula $Z_2(\rho, s) + sf''(t_0)$ should be replaced by

$$Z_2^*(\rho, s) = -f(t_0)^{1/2}\, \rho^{-3} \int K''(s+u)\, W_1^*(\rho u)\, du\,.$$

Third, the quantity $sf''(t_0)$ in $Z_2(\rho, s) + sf''(t_0)$, formerly an approximation to $f'(t_0+hs)/h$, should be replaced by $\hat{f}'_{\rm crit}(\hat{t}_0+hs)/h$, or a suitable approximation, where $\hat{t}_0$ is the turning point of $\hat{f}_{\rm crit}$ nearest to t_0. Now, for bandwidths h of size $h_0 = n^{-1/5}$,

$$\hat{f}'_{\rm crit}(\hat{t}_0 + hs) = \hat{h}_{\rm crit}\, Z_2\{\hat{h}_{\rm crit}/h_0, (h/\hat{h}_{\rm crit})\,(\hat{s}_h + s)\} + h\,(\hat{s}_h + s)\, f''(t_0) + o_p(h_0),$$

where $\hat{s}_h = (\hat{t}_0 - t_0)/h$. Furthermore, $\hat{h}_{\rm crit} = cRh_0 + o_p(h_0)$ and $(\hat{t}_0 - t_0)/\hat{h}_{\rm crit} = S + o_p(1)$, and so, writing $h = crh_0$,

$$\begin{aligned}\frac{\hat{f}'_{\rm crit}(\hat{t}_0 + hs)}{h\, f''(t_0)} &= \frac{\hat{h}_{\rm crit}}{h}\, Z\left\{\frac{\hat{h}_{\rm crit}}{c\, h_0}, \frac{h}{\hat{h}_{\rm crit}}\,(\hat{s}_h + s)\right\} + (\hat{s}_h + s) + o_p(1)\\ &= (R/r)\, Z(R, S + R^{-1}rs) + r^{-1}RS + s + o_p(1).\end{aligned}$$

We may write $Z_2^*(\rho, s)/f''(t_0) = Z^*(\rho/c, s)$, and so the analogue of $Z_2(cr, s) \times f''(t_0)^{-1} + s = Z(r, s) + s$ in the present setting, which is $Z_2^*(\rho, s)\, f''(t_0)^{-1} + \hat{f}'_{\rm crit}(\hat{t}_0 - hs)\,\{h\, f''(t_0)\}^{-1}$, may be written as $Y^*(s) = Z^*(r, s) + (R/r)\, Z(R, S + R^{-1}rs) + r^{-1}RS + s$, plus terms which converge to zero.

5.4. Outline proof of Theorem 3.3

Under the conditions of the theorem, there exist constants $0 < h_1 \le h_2 < \infty$ such that (a) for all $h < h_1$, $f_h(x) = \int K(y)\, f(x - hy)\, dy$ has at least two turning points interior to $\mathcal{I}$, and (b) for all $h > h_2$, $f_h(x)$ has at most one turning point interior to $\mathcal{I}$. If $C < h_1$ then $P(\hat{h}_{\rm crit} > C) \to 1$.

Let $\hat{h}^*(h)$ be the version of $\hat{h}^*_{\rm crit}$ computed if the resample $X_1^*, \ldots, X_n^*$ is drawn from the distribution with density $\hat{f}_h$, instead of from $\hat{f}_{\rm crit}$. It may be shown that for each $\epsilon, \delta > 0$ with $\epsilon < h_1$, $P(h_1 - \epsilon < \hat{h}_{\rm crit} < h_2 + \epsilon) \to 1$ as $n \to \infty$ and

$$\sup_{h_1 - \epsilon < h < h_2 + \epsilon} P\{\hat{h}^*(h) \le \delta | \mathcal{X}\} \to 1$$

in probability. The theorem follows from these results.

5.5. Outline proof of Theorem 3.4

Choose $\delta > 0$ so small that the equation $f'_h(x) = f''_h(x) = 0$ has no solutions with $h \in (0, \delta]$ and $x \in \mathcal{I}$. In view of Theorem 2.1, if part (a) of Theorem 3.4 fails then there exists a constant $C_1 > 0$, and sequences of random variables $h(n) \in [C_1 n^{-1/5}, \delta]$ and $x_{h(n)} \in \mathcal{I}$, such that, with $\mathcal{E}(n) = \{\hat{f}'_{h(n)}(x_{h(n)}) =$

$\hat{f}''_{h(n)}(x_{h(n)}) = 0\}$, the probability $P\{\mathcal{E}(n)\}$ is bounded away from 0 along an infinite sequence $\{n_k\}$ of n's. (We choose to neglect the third derivative.) We show that this leads to a contradiction. Two cases need separate treatment — where $h(n)$ does not converge to zero, and where it does.

Case 1. For some $\epsilon > 0$, $P\{h(n_k) > \epsilon\}$ does not converge to 0. Suppose $P\{h(n_k) > \epsilon\}$ is bounded away from zero. Along a subsequence of values of k, the random vector $(x(n_k), h(n_k))$ converges in distribution. Let (x_1, h_1) be a point of support of its limiting distribution. Necessarily, $h_1 \in [\epsilon, \delta]$. Since $(\hat{f}'_h(x), \hat{f}''_h(x))$ converges to $(f'_h(x), f''_h(x))$, with probability one, uniformly in values of (x, h) in any sufficiently small neighbourhood of (x_1, h_1), then $f'_{h_1}(x_1) = f''_{h_1}(x_1) = 0$, which contradicts our choice of δ. Therefore, Case 1 cannot arise.

Case 2. $h(n_k) \to 0$ in probability. Since $P\{\hat{f}'_{h(n)}(x_{h(n)}) = 0\}$ is bounded away from 0 along a subsequence, and

$$\sup_{C_1 n^{-1/5} \le h \le \delta,\, x \in \mathcal{I}} |\hat{f}'_h(x) - f'_h(x)| \to 0$$

in probability, and because $h(n) \to 0$ in probability, then for some $t_0 \in \mathcal{I}$ with $f'(t_0) = 0$, some sequence of constants $\delta_n \to 0$, and all $\epsilon > 0$,

$$\limsup_{n\to\infty} P\{\mathcal{E}(n)\,,\, h(n) \in [C_1 h_0, \delta_n]\,,\, |x_{h(n)} - t_0| \le \epsilon\} > 0. \tag{5.5}$$

Next we prove the following lemma.

Lemma 5.1. *Assume the conditions of Theorem* 3.4*, let $t_1, \ldots, t_m$ denote the turning points of f interior to $\mathcal{I}$, put $h_0 = n^{-1/5}$, and let $\delta_n \to 0$ so slowly that $h_0/\delta_n \to 0$. Then*

$$\lim_{C_2\to\infty} \limsup_{n\to\infty} P\Big\{\hat{f}'_h(x) = \hat{f}''_h(x) = 0 \text{ for some } h \in [C_2 h_0, \delta_n]$$
$$\text{and some } x \in \mathcal{I} \text{ satisfying } \inf_{1\le k\le m} |x - t_k| > C_2 h_0\Big\} = 0.$$

Proof. Let $t_0 \in \mathcal{I}$ be a turning point of f, and let $j = 1$ or 2. We begin by proving that for all $C_1 > 0$ and $0 < \epsilon < \frac{1}{2}$,

$$E\Big\{\sup_{u: t_0 + \rho h_0 u \in \mathcal{I}} \sup_{C_1 \le \rho \le \delta_n/h_0} (1+|u|)^{-(1/2)-\epsilon} \rho^j\, n^{(2-j)/5}$$
$$\times |\hat{f}^{(j)}_{\rho h_0}(t_0 + \rho h_0 u) - f^{(j)}_{\rho h_0}(t_0 + \rho h_0 u)|\Big\} = O(1). \tag{5.6}$$

The approximation of Komlós, Major and Tusnády (1975) gives

$$\hat{f}^{(j)}_h(x) - f^{(j)}_h(x) = (n^{1/2} h^{j+1})^{-1} \int K^{(j+1)}(y)\, W^0\{F(x - hy)\}\, dy$$
$$+ R_{1jn}(x, h)\, (n h^{j+1})^{-1} \log n$$

for $n \geq 2$, where

$$E\Big\{\sup_{x\in\mathcal{I}}\ \sup_{C_1 h_0\leq h\leq\delta_n} |R_{1jn}(x,h)|\Big\} = O(1).$$

Properties of the modulus of continuity of W^0 (e.g. Garsia (1970)) yield, for $n \geq 2$,

$$\int |K^{(j+1)}(y)|\,\Big|W^0[F\{t_0+h(u-y)\}] - W^0\{F(t_0)+h(u-y)\,f(t_0)\}\Big|\,dy$$
$$= R_{2jn}(u,h)\,h\,(\log n)\,(1+|u|),$$

where

$$E\Big\{\sup_{u:t_0+hu\in\mathcal{I}}\ \sup_{C_1 h_0\leq h\leq\delta_n} |R_{2jn}(u,h)|\Big\} = O(1).$$

Hence, with $\rho = h/h_0$ and $W_2(v) = \{h_0\,f(t_0)\}^{-1/2}[W^0\{F(t_0) - h_0 v\,f(t_0)\} - W^0\{F(t_0)\}]$ we have

$$\hat f^{(j)}_{\rho h_0}(t_0+\rho h_0 u) - f^{(j)}_{\rho h_0}(t_0+\rho h_0 u)$$
$$= (n^{1/2}\,\rho^{j+1} h_0^{j+(1/2)})^{-1} f(t_0)^{1/2}\int K^{(j+1)}(y)\,W_2\{\rho(y-u)\}dy$$
$$+R_{3jn}(u,\rho)(n^{1/2}h_0^j\rho^j)^{-1}\log n,$$

where $R_{3jn}(u,h)$ satisfies

$$E\Big\{\sup_{u:t_0+\rho h_0 u\in\mathcal{I}}\ \sup_{C_1\leq\rho\leq\delta_n/h_0} |R_{kjn}(u,h)|\Big\} = O(1). \tag{5.7}$$

Let $0<\epsilon<\frac{1}{2}$, and define

$$V_1 = \sup_{|t|\leq 1}|W_2(t)|,\quad V_2 = \sup_{|t|\geq 1}|t|^{-(1/2)-\epsilon}|W_2(t)| \quad\text{and}\quad V=\max(V_1,V_2).$$

Then $|W_2(t)| \leq V\,(1+|t|)^{(1/2)+\epsilon}$ for all t. Hence, for $j=1,2$,

$$\int |K^{(j+1)}(y)\,W_2\{\rho(y-u)\}|\,dy \leq V\int |K^{(j+1)}(y)|\,(1+\rho|y-u|)^{(1/2)+\epsilon}\,dy$$
$$\leq C\,V\,(1+\rho)^{(1/2)+\epsilon}\,(1+|u|)^{(1/2)+\epsilon},$$

where the constant C depends only on ϵ. Therefore,

$$|\hat f^{(j)}_{\rho h_0}(t_0+\rho h_0 u) - f^{(j)}_{\rho h_0}(t_0+\rho h_0 u)| = R_{4jn}(u,\rho)\,n^{-(2-j)/5}\,\rho^{-j}\,(1+|u|)^{(1/2)+\epsilon},$$

where R_{4jn} satisfies (5.7). This implies (5.6).

Returning to the proof of Lemma 5.1, we first apply (5.6) for $j=1$. Given $x\in\mathcal{I}$, let $t_0 = t_0(x)\in\{t_1,\ldots,t_m\}$ denote the turning point to which x is nearest. Since $f''(t_0) \neq 0$, there exists $C > 0$ such that $|f_h'(x)| > C\,|x - t_0|$ uniformly in values $x \in \mathcal{I}$ that are nearer to t_0 than to any other turning point in $\mathcal{I}$. Hence, with $u = u(x, h) = (x - t_0)/h$, $|\hat{f}_h'(x)| \geq |f_h'(x)| - |\hat{f}_h'(x) - f_h'(x)| \geq C\,h\,|u| - R_{4,1,n}(u, \rho)\,h_0^2\,h^{-1}\,(1 + |u|)^{(1/2)+\epsilon}$, and so by (5.6), for all $C_1 > 0$,

$$\lim_{C_3\to\infty} \limsup_{n\to\infty} P\Big\{\hat{f}_h'(x) = 0 \text{ for some } h \in [C_1 h_0, \delta_n] \text{ and some } x \in \mathcal{I} \quad \text{satisfying } \inf_{1\leq k\leq m} |x - t_k| > C_3 h\Big\} = 0. \tag{5.8}$$

Next we apply (5.6) when $j = 2$. There exists a constant $C > 0$ such that $|f_h''(x)| \geq C$ uniformly in $|x - t_0| \leq C_3 h$ and $h \in [C_1 h_0, \delta_n]$, and for such values of (x, h), $|\hat{f}_h''(x)| \geq |f_h''(x)| - |\hat{f}_h''(x) - f_h''(x)| \geq C - R_{4,2,n}(u, \rho)\,\rho^{-2}\,(1 + C_3)^{(1/2)+\delta}$, whence by (5.6) we have for all $C_3 > 0$,

$$\lim_{C_2\to\infty} \limsup_{n\to\infty} P\Big\{\hat{f}_h''(x) = 0 \text{ for some } h \in [C_2 h_0, \delta_n] \text{ and some } x \in \mathcal{I} \quad \text{satisfying } \inf_{1\leq k\leq m} |x - t_k| \leq C_3 h\Big\} = 0. \tag{5.9}$$

Lemma 5.1 follows from (5.8) and (5.9).

In view of Lemma 5.1, result (5.5) will fail if, for all $0 < C_1 < C_2 < \infty$,

$$P\Big\{\hat{f}_h'(x) = \hat{f}_h''(x) = 0 \text{ for some } h \in [C_1 h_0, C_2 h_0] \text{ and some } x \in \mathcal{I} \quad \text{satisfying } |x - t_0| \leq C_2 h_0\Big\} \to 0.$$

The Wiener−process approximation arguments used during the proof of the lemma may be employed to show that the probability on the left-hand side converges to

$$P\Big[\int K^{(2)}(y + u)\,W_2(\rho y)\,dy = -\rho^3\,p\,u\,,\ \int K^{(3)}(y + u)\,W_2(\rho y)\,dy = -\rho^3\,p \quad \text{for some } \rho \in [C_1, C_2] \text{ and some } u \text{ satisfying } |u| \leq C_2\Big],$$

where $p = f''(t_0)/f(t_0)^{1/2}$. This probability is zero. Therefore, (5.5) cannot be correct, and so part (a) of Theorem 3.4 must be valid.

In conclusion, we outline derivation of parts (b) and (c) of Theorem 3.4. Suppose f has just m turning points $t_1, \ldots, t_m$ in $\mathcal{I}$. Choose $\epsilon > 0$ so small that $\mathcal{J}_j = [t_j - \epsilon, t_j + \epsilon] \subseteq \mathcal{I}$ and $\mathcal{J}_j$ does not include any t_k's with $k \neq j$, let $h(u) = n^{-1/5}u$, let $M_j(u)$ equal the number of turning points of $\hat{f}_{h(u)}$ in $\mathcal{J}_j$, and put $c_j = f(t_j)^{1/5}/|f''(t_j)|^{2/5}$. The methods employed to derive Theorem 3.1 may be used to show that for arbitrary fixed $0 < C_1 < C_2 < \infty$ the vector $(M_1, \ldots, M_m)$ of processes $M_j(u)$, $u \in [C_1, C_2]$, has a joint weak limit $(L_1, \ldots, L_m)$, where (i) the processes L_j are stochastically independent, and (ii) each L_j has the distribution of the number of zeros of $Y_j(s) = Z(u/c_j, s) + s$. (The processes L_j and M_j are right-continuous step functions with left-hand limits.) The probability that a mode "migrates" away from a turning point converges to 0 as $n \to \infty$. Indeed, if $\delta > 0$ is sufficiently small and $\mathcal{J} \subseteq \mathcal{I}$ denotes a set on which $|f_h'|$ is bounded away from zero uniformly in $0 < h \leq \delta$, then, since $\hat{f}_h' - f_h'$ converges to 0 uniformly in $h \in [\epsilon_n n^{-\frac{1}{5}}, \delta]$, provided $\epsilon_n \to 0$ sufficiently slowly, we may show that

$$P\{\text{there exists no } x \in \mathcal{J} \text{ or } h \in [\epsilon_n n^{-1/5}, \delta] \text{ such that } \hat{f}_h'(x) = 0\} \to 1.$$

In view of these results, parts (b) and (c) of Theorem 3.4 may be derived by making minor changes to the proofs of Theorems 3.1—3.3.

Acknowledgement

We are grateful to a referee and an associate editor for helpful comments.

References

Cheng, M.-Y. and Hall, P. (1998). Calibrating the excess mass and dip tests of modality. *J. Roy. Statist. Soc. Ser. B* **60**, 579-590.

Cheng, M.-Y. and Hall, P. (1999). Mode testing in difficult cases. *Ann. Statist.* **27**, 1294-1315.

Cuevas, A. and Gonzales-Manteiga, W. (1991). Data-driven smoothing based on convexity properties. In *Nonparametric Functional Estimation and Related Topics* (Edited by G. Roussas), 225-240. Kluwer, Dordrecht.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90**, 577-588.

Fisher, N. I., Mammen, E. and Marron, J. S. (1994). Testing for multimodality. *Comput. Statist. Data Anal.* **18**, 499-512.

Garsia, A. M. (1970). Continuity properties of a Gaussian process with multidimensional time parameter. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2**, 369-374. University of California Press.

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. Statist. Assoc.* **75**, 42-73.

Hartigan, J. A. and Hartigan, P. M. (1985). The DIP test of unimodality. *Ann. Statist.* **13**, 70-84.

Hartigan, P. M. (1985). Algorithm AS217: computation of the Dip statistic to test for unimodality. *Appl. Statist.* **34**, 320-325.

Izenman, A. J. (1991). Recent developments in nonparametric density estimation. *J. Amer. Statist. Assoc.* **86**, 205-224.

Izenman, A.J. and Sommer, C. (1988). Philatelic mixtures and multimodal densities. *J. Amer. Statist. Assoc.* **83**, 941-953.

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's, and the sample DF. I. *Z. Wahr. verw. Gebiete* **32**, 111-131.

Mammen, E. (1991a). Estimating a smooth monotone regression function. *Ann. Statist.* **19**, 724-740.

Mammen, E. (1991b). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19**, 741-759.

Mammen, E., Marron, J. S. and Fisher, N. I. (1992). Some asymptotics for multimodality tests based on kernel density estimates. *Probab. Theory Related Fields* **91**, 115-132.

Minotte, M. C. (1997). Nonparametric testing of the existence of modes. *Ann. Statist.* **25**, 1646-1660.

Minotte, M. C. and Scott, D. W. (1993). The mode tree: a tool for visualization of nonparametric density features. *J. Comput. Graph. Statist.* **2**, 51-68.

Müller, D. W. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86**, 738-746.

Polonik, W. (1995a). Measuring mass concentrations and estimating density contour clusters — an excess mass approach. *Ann. Statist.* **23**, 855-881.

Polonik, W. (1995b). Density estimation under qualitative assumptions in higher dimensions. *J. Multivariate Anal.* **55**, 61-81.

Roeder, K. (1990). Density estimation with confidence sets. *J. Amer. Statist. Assoc.* **85**, 617-624.

Roeder, K. (1994). A graphical technique for determining the number of components in a mixture of normals. *J. Amer. Statist. Assoc.* **89**, 487-495.

Schoenberg, I. J. (1950). On Pólya frequency functions, II: variation diminishing integral operators of the convolution type. *Acta Scientarum Mathematicarum Szeged* **12B**, 97-106.

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B* **43**, 97-99.

Silverman, B. W. (1983). Some properties of a test for multimodality based on kernel density estimates. In *Probability, Statistics and Analysis* (Edited by J. F. C. Kingman and G. E. H. Reuter), 248-259. Cambridge University Press, Cambridge, UK.

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

Centre for Mathematics and Its Applications, Australian National University, Canberra, A.C.T. 0200, Australia.

E-mail: halpstat@pretty.anu.edu.au

(Received August 1999; accepted September 2000)