Annals of Statistics, **27** (1999), 1294–1315.

MODE TESTING IN DIFFICULT CASES

Ming-Yen Cheng, Peter Hall

Centre for Mathematics and its Applications, Australian National University

ABSTRACT. Usually, when testing the null hypothesis that a distribution has one mode, against the alternative that it has two, the null hypothesis is interpreted as entailing that the density of the sampling distribution has a unique point of zero slope, which is a local maximum. In this paper we argue that a more appropriate null hypothesis is that the density has two points of zero slope, of which one is a local maximum and the other is a shoulder. We show that when a test for a mode-with-shoulder is properly calibrated, so that it has asymptotically correct level, it is generally conservative when applied to the case of a mode without a shoulder. We suggest methods for calibrating both the bandwidth and dip/excess mass tests in the setting of a mode with a shoulder. We also provide evidence in support of the converse: a test calibrated for a single mode without a shoulder tends to be anticonservative when applied to a mode with a shoulder. The calibration method involves resampling from a 'template' density with exactly one mode and one shoulder. It exploits the following asymptotic factorisation property for both the sample and resample forms of the test statistic: all dependence of these quantities on the sampling distribution cancels asymptotically from their ratio. In contrast to other approaches, the method has very good adaptivity properties.

KEYWORDS. Bandwidth, bootstrap, calibration, curve estimation, level accuracy, local maximum, shoulder, smoothing, turning point.

SHORT TITLE. Mode testing

AMS SUBJECT CLASSIFICATION. Primary 62G07, Secondary 62G09.

1. INTRODUCTION

Testing for modality is one way of finding evidence of sub-populations in the population from which data are drawn. Early tests were often based on parametric mixture models (e.g. Cox 1966), but during the last two decades several nonparametric methods have been developed. They are generally conservative, however, and increasing interest is being shown in ways of calibrating them so that their levels are closer to those prescribed. Heuristically, it is to be expected that improving the level accuracy of a conservative test would lead to increased power.

It is usually necessary to have at least an approximate model for densities f representing the "null hypothesis" that is being tested, since we need to calibrate the test under the null. For example, in the case of testing for unimodality against the alternative of multimodality, the null hypothesis is generally that f has one local maximum, no local minima, and no places of zero gradient that do not correspond to turning points. We shall call this the "classic null hypothesis", $H_{0,\mathrm{class}}$; it is tested against the alternative, H_1, that f has two or more modes.

Such alternative hypotheses are generally relatively easy to distinguish from the null, however. We argue that a test of modality will have better performance if it works well against distributions that are 'marginal', or 'most difficult' to tell apart from the null — this is the sense in which we use the term 'difficult' in our paper. The difficult cases are densities that represent the *boundary* between one and two modes — that is, those where f has one local maximum, no local minima, and exactly one point x for which $f'(x) = 0$ but x is a shoulder point (defined by $f''(x) = 0$ and $f'''(x) \neq 0$) rather than a local maximum or local minimum. We term this the 'boundary null hypothesis', $H_{0,\mathrm{bound}}$. The issue of which null hypothesis is employed determines the type of theory which best describes properties of tests for modality, and affects the tests' level accuracy and power.

Figure 1.1 illustrates some of these issues. Panels (a) and (c) depict densities that are unimodal and bimodal, satisfying $H_{0,\mathrm{class}}$ and H_1 respectively; and panel (b) shows a "shoulder" density which in a sense is midway between the other two, and satisfies $H_{0,\mathrm{bound}}$. Intuitively, when an empirical test finds it hard to distinguish between panels (a) and (c), the problem really arises because the test can't solve the more difficult problem of deciding between panels (b) and (c). To optimise performance in these difficult cases the test should be constructed so that it addresses the harder problem, not the easier one.

[*Put Figure 1.1 about here, please*]

It is helpful to consider the related, parametric problem of testing composite, one-sided hypotheses, of the form $\theta \leq \theta_0$ versus $\theta > \theta_0$, where θ denotes a scalar parameter. There it is common to construct first a test of the simple null hypothesis, $\theta = \theta_0$, against the alternative hypothesis $\theta > \theta_0$, and then use the same test in the case of the composite one-sided null hypothesis. When the likelihood ratio is monotone, this approach is optimal and gives uniformly most powerful tests; see Kendall and Stuart (1979, Chapter 23). The null hypothesis $\theta = \theta_0$ is more difficult than $\theta < \theta_0$ to distinguish from $\theta > \theta_0$, and the optimal approach is to construct the test in the more difficult case.

In the context of the mode testing problem, $H_{0,\text{bound}}$ represents the simple null hypothesis $\theta = \theta_0$ at the boundary, and $H_{0,\text{class}}$ plays the role of the null hypothesis $\theta < \theta_0$. Following the line suggested in the previous paragraph, we argue that the test should be developed for the more difficult null hypothesis, $H_{0,\text{bound}}$. Section 2.4 establishes that, analogously to the conclusions reached in the previous paragraph for the parametric case, our test is also appropriate for $H_{0,\text{class}}$; Figure 3.3 indicates the conservatism of a test of $H_{0,\text{bound}}$ when applied to $H_{0,\text{class}}$; and Figure 3.4 illustrates the anticonservatism of a test for $H_{0,\text{class}}$ when applied to $H_{0,\text{bound}}$.

In this paper we suggest methods, and develop theory, pertaining to this view of testing for modality. We employ two particular tests as examples, the bandwidth test of Silverman (1981) and the dip/excess mass test of Hartigan and Hartigan (1985) and Müller and Sawitzki (1991). Both involve rejecting the null hypothesis if the test statistic exceeds a certain critical point. For either test we discuss a bootstrap calibration method that produces the asymptotically correct level under $H_{0,\text{bound}}$, and is slightly conservative under $H_{0,\text{class}}$. Related methods, inspired by work of Hartigan (1997), will also be noted. Importantly, the level of the test under $H_{0,\text{class}}$ does not converge to zero as sample size increases, and so the bootstrap procedure is relatively adaptive to both null hypotheses. In comparison, alternative methods for calibrating tests of $H_{0,\text{bound}}$ have a level which converges to zero under $H_{0,\text{class}}$.

Our theoretical description of mode testing under the boundary null hypothesis is in contradistinction to existing accounts in the literature, which seem always to assume the classic null hypothesis. Examples include Silverman (1983), Mammen, Marron and Fisher (1992) and Cheng and Hall (1998). The results in the two cases are quite different, with respect to order of magnitude as well as asymptotic distribution. For example, under $H_{0,\text{class}}$ the critical value for the bandwidth test is of size $n^{-1/5}$, where n is the number of data values (Mammen, Marron and Fisher 1992); but under $H_{0,\text{bound}}$ it is of size $n^{-1/7}$. The analogues for critical points in the case of the dip/excess mass tests are $n^{-3/5}$ and $n^{-4/7}$, respectively. The limiting distributions in the four cases are all different and non-Normal. These facts alone demonstrate that calibration methods developed specifically for $H_{0,\text{class}}$ can be inappropriate for $H_{0,\text{bound}}$, and so can suffer problems when $H_{0,\text{class}}$ is only "just true", unless they have the adaptivity property noted in the previous paragraph.

Specifically, suppose $H_{0,\text{class}}$ is true, but only just true (that is, $H_{0,\text{bound}}$ is "almost" true); and the test is constructed so as to reject the null hypothesis when the test statistic exceeds a critical point whose asymptotic size is appropriate to $H_{0,\text{class}}$. (Therefore, the critical point is of size $n^{-1/5}$ if the bandwidth test is used, and of size $n^{-3/5}$ for the excess mass test.) Then the test will tend to incorrectly reject the null hypothesis, for the simple reason that $n^{-1/5} < n^{-1/7}$ and $n^{-3/5} < n^{-4/7}$. Our adaptive tests based on bootstrap calibration does not suffer from this problem.

Because of the light which these theoretical results shed on the importance of distinguishing between the two types of null hypothesis, we shall discuss our theoretical work first, in Section 2. Section 3 will summarise the results of a simulation study that assesses the performance of our adaptive tests. Section 2.1 will describe alternative, non-adaptive approaches. Technical arguments for Section 2 will be placed into Section 4. For simplicity we shall consider only the case of testing for unimodality. There is no technical difficulty in stating and deriving analogues of our theory for testing the hypothesis of m modes against that of $m+1$ modes, where $m \geq 1$, although notation becomes rather complex in that case. The versions of our adaptive tests in that general setting seem prohibitively complex, however. In this multimodal setting, recent work of Hartigan (1997) is particularly deserving of mention. There, a novel sequential (in m) approach to using the excess mass test is suggested.

2. THEORETICAL PROPERTIES OF TEST STATISTICS

2.1. Summary and conclusions. The bandwidth test, which will be introduced and discussed in Section 2.2, involves rejecting the null hypothesis if a critical bandwidth, $\hat{h}_{\text{crit}}$, is too large; and the dip/excess mass test, to be described in Section 2.3, rejects the null hypothesis if a test statistic Δ is too large. When the sampling density f satisfies the null hypothesis $H_{0,\text{class}}$, and appropriate regularity conditions hold, $n^{1/5}\hat{h}_{\text{crit}}$ has a proper limiting distribution that may be written as that of a random variable C_1R_1, where the nonzero constant C_1 depends only on f, and the distribution of the random variable R_1 does not depend on f. See Mammen, Marron and Fisher (1992). By way of contrast, we shall point out in Section 2.2 that under $H_{0,\text{bound}}$ and appropriate conditions on f, $n^{1/7}\,\hat{h}_{\text{crit}} \to C_2R_2$ in distribution, where (here and below) C_j and R_j have the properties ascribed to C_1 and R_1 above.

Analogous results hold for the dip/excess mass test, where, under $H_{0,\text{class}}$ and regularity conditions on f, $n^{3/5}\Delta \to C_3R_3$ in distribution (see Cheng and Hall 1998); and, under $H_{0,\text{bound}}$ and regularity conditions, $n^{4/7}\Delta \to C_4R_4$ in distribution (see Section 2.3).

The formulae for $C_1,\ldots,C_4$ are very different from one another, as too are the distributions of $R_1,\ldots,R_4$. However, in each case the principle is the same: the distribution of the test statistic factorises, asymptotically, into a constant that depends only on f and a random variable whose distribution is continuous and is in principle known. Note particularly that even the order of magnitude of the critical points, let alone the constants C_j and the random variables R_j, depends not only on the type of test but also on the particular form of null hypothesis that is chosen.

For both the bandwidth and dip/excess mass tests, the factorisation property may be exploited to construct a test that adapts itself well to either $H_{0,\text{class}}$ or $H_{0,\text{bound}}$. It amounts to computing the ratio of the test statistic (either $\hat{h}_{\text{crit}}$ or Δ) and its bootstrap form; and rejecting the null hypothesis if the bootstrap distribution of the ratio assumes values that are too large. On account of the factorisation, the unknown constants C_j cancel from the ratio in all four cases, and so the bootstrap distribution function of the ratio (a stochastic process) does not depend asymptotically on any unknowns. Unlike the case of more standard statistical problems (such as percentile-t statistics) where scale parameters cancel, the bootstrap versions of the distributions of variables R_j are not particularly close to those of the respective R_j's, and so the stochastic process noted just above is not degenerate. Nevertheless, its properties may be determined by Monte Carlo methods, and after suitable calibration it has asymptotically correct level under both $H_{0,\mathrm{bound}}$ and $H_{0,\mathrm{class}}$. Adaptive tests will be introduced in Sections 2.2 (for the bandwidth method) and 2.3 (dip/excess mass method), and Section 2.4 will discuss their properties.

An alternative way to proceed would be to directly estimate that one of the unknown constants $C_1, \ldots, C_4$ which is appropriate to the context (e.g. C_1 if we were using the excess mass test under $H_{0,\mathrm{class}}$), use Monte Carlo methods to calculate the distribution of the respective variable R_j, and thereby approximate the asymptotic distribution of the test statistic under the null hypothesis. If the bootstrap method described in the previous paragraph is likened to Studentizing so to cancel the effects of scale, then this approach is similar to using standard asymptotic approximations after "plugging in" an estimate of scale. However, by its very construction the latter approach is highly sensitive to choice of null hypothesis, be it $H_{0,\mathrm{class}}$ or $H_{0,\mathrm{bound}}$, and in particular it does not enjoy the adaptivity of the bootstrap approach. If it is constructed so that it gives an asymptotically correct test under $H_{0,\mathrm{class}}$ [respectively, $H_{0,\mathrm{bound}}$], then the level of the test under $H_{0,\mathrm{bound}}$ [or $H_{0,\mathrm{class}}$] will be 0 [or 1].

Moreover, even if these problems are overcome, it is likely that the bootstrap approach captures at least some of the first-order features of the distribution of the test statistic that a purely asymptotic method misses. In the context of bootstrap versus asymptotic approximations to critical points for Silverman's (1981) bandwidth test, York (1998) has demonstrated this numerically. The bootstrap approach, through taking the resample size equal to the sample size, n, offers a significantly better approximation than does taking $n = \infty$, even if the template density is not the true density.

2.2. Bandwidth test. To introduce the test, let $\mathcal{X} = \{X_1, \ldots, X_n\}$ denote a random sample drawn from a distribution with unknown density f, and construct the kernel 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. As in Silverman (1981) we take K to be the standard Normal density, for which the number of modes of $\hat{f}_h$ on the whole line is a nonincreasing function of h. Furthermore, $\hat{f}_h$ is unimodal for all sufficiently large h. Let $\hat{h}_{\text{crit}}$ denote the infimum of bandwidths such that $\hat{f}_h$ has only one mode. A test of the null hypothesis of unimodality consists of rejecting unimodality if $\hat{h}_{\text{crit}}$ is too large.

Mammen, Marron and Fisher (1992) proved that under $H_{0,\text{class}}$, and assuming appropriate regularity conditions on f, $\hat{h}_{\text{crit}}$ is of size $n^{-1/5}$. We show next that it is of size $n^{-1/7}$ under $H_{0,\text{bound}}$. First we state an analogue of Mammen, Marron and Fisher's (1992) regularity conditions (corresponding also to the conditions of Silverman (1983)) in the case of $H_{0,\text{bound}}$:

> f is supported on a compact interval $[a,b]$, and has two derivatives there; $f'=0$ at distinct points $x_0, x_1 \in (a,b)$, and $f' \neq 0$ at all other points in (a,b); f has respectively two and three Hölder-continuous derivatives in neighbourhoods of x_0 and x_1;
> $$f''(x_0) < 0\,,\; f''(x_1) = 0\,,\; f'''(x_1) \neq 0\,,\; f'(a+) > 0\,,\; f'(b-) < 0\,. \tag{2.2}$$

For $0 < r < \infty$ and $-\infty < s < \infty$, define

$$Z(r,s) = r^{-4} \int K''(s+u)\, W(ru)\, du + \tfrac{1}{2}(1+s^2)\,,$$

where W is a standard Wiener process. Put $C_2 = \{f(x_1)/|f'''(x_1)|^2\}^{1/7}$, where x_1 is the shoulder point noted in (2.2), and let R_2 denote the infimum of all values of r such that the function $Z(r,\cdot)$ does not change sign on $(-\infty,\infty)$. (In view of total positivity properties of K'' (see Schoenberg, 1950), if $Z(r,\cdot)$ does not change sign on $(-\infty,\infty)$ then, with probability 1, neither does $Z(r',\cdot)$ for any $r' > r$.)

Theorem 2.1. *Assume condition* (2.2). *Then* $n^{1/7}\hat{h}_{\text{crit}} \to C_2 R_2$ *in distribution as* $n \to \infty$.

We should comment on the nature of condition (2.2), which asks that f decrease linearly to zero at the ends of its support. This ensures that the likelihood of spurious bumps in the tails of the density estimator $\hat{f}_h$ is very small. Therefore, the size of $\hat{h}_{\text{crit}}$ is determined by properties of f at points of zero slope interior to (a,b). More generally, when f might not satisfy (2.2), one would either confine attention to testing for unimodality away from the tails, or use larger bandwidths in the tails so as to suppress bumps that arise from data sparseness.

Next we define the bootstrap version of $\hat{h}_{\rm crit}$, and show that it satisfies a limit law similar to that in Theorem 2.1. Conditional on $\mathcal{X}$, let $\mathcal{X}^* = \{X_1^*, \ldots, X_n^*\}$ denote a resample drawn randomly, with replacement, from the distribution with density $\hat{f}_{\rm crit} = \hat{f}_{\hat{h}_{\rm crit}}$, and define $\hat{f}_h^*$ by (2.1) except that X_i there is replaced by X_i^*. Write $\hat{h}_{\rm crit}^*$ for the infimum of bandwidths such that $\hat{f}_h^*$ is unimodal.

Our proof of Theorem 2.1 in Section 4 will involve constructing W (depending on n) such that

$$n^{1/7}\, \hat{h}_{\rm crit} \to C_2 R_2 \quad \text{in probability}\,. \tag{2.3}$$

For this W, let W^* be a standard Wiener process independent of W, and let S be the unique point at which $Z(R_2, \cdot)$ vanishes. Define

$$Z^*(r,s) = (rR_2)^{-2} \int K''(s+u)\, W^*(ru)\, du + \int Z(R_2, S - R_2^{-1}\, ru)\, K(u)\, du\,,$$

and let R_2^* denote the infimum of all values of r such that the function $Z^*(r,\cdot)$ does not change sign on $(-\infty,\infty)$. It is straightforward to prove that R_2^* is strictly positive with probability 1.

Theorem 2.2. *Assume condition* (2.2), *and that* W *is constructed so that (2.3) holds. Then,*

$$\sup_{0\le x<\infty} \left|P\big(n^{1/7}\, \hat{h}_{\rm crit}^* \le C_2 x \big| \mathcal{X}\big) - P(R_2^* \le x | W)\right| \to 0$$

in probability as $n \to \infty$.

Theorem 2.2 and (2.3) together imply that, under $H_{0,\rm bound}$,

$$\sup_{0\le x<\infty} \left|P\big(\hat{h}_{\rm crit}^*/\hat{h}_{\rm crit} \le x \big| \mathcal{X}\big) - P(R_2^*/R_2 \le x | W)\right| \to 0 \tag{2.4}$$

in probability. It follows that the distribution of the stochastic process $\widehat{G}(x) = P(R_2^*/R_2 \le x|W)$ does not depend on f, which makes it possible to develop an asymptotically correct test of $H_{0,\rm bound}$. This could be based on tabulation of the distribution of $\widehat{G}$, and applying an asymptotic test, but alternatively it may be accomplished by Monte Carlo methods, as follows. Put $\widehat{G}_n(x) = P(\hat{h}_{\rm crit}^*/\hat{h}_{\rm crit} \le x|\mathcal{X})$, let f_0 denote a "template" density with a shoulder, and let $\widehat{G}_{0n}$ denote the version of $\widehat{G}_n$ that results from an n–sample drawn randomly from f_0. Using Monte Carlo methods we may compute to arbitrary accuracy the value of a constant $t_\alpha =$

$t_\alpha(n)$ such that $P\{\widehat{G}_{0n}(t_\alpha) \geq 1-\alpha\} = \alpha$, where α is the desired significance level of the test. Then, the test with the form: reject $H_{0,\mathrm{bound}}$ in favour of H_1 if $\widehat{G}_n(t_\alpha) \geq 1-\alpha$, has asymptotically correct level under $H_{0,\mathrm{bound}}$.

One would expect the template approach to capture second-order effects better than a purely asymptotic argument. This may be confirmed by simulation. To capture second-order effects even more accurately one could use a skewed template (for example) if there was evidence that the sampling distribution was skewed, although it is difficult to ensure both the right degree of skewness and the right value of C_2.

2.3. Dip/excess mass test. It suffices to consider the excess mass test statistic, Δ, which equals twice the dip test statistic. Let $\widehat{F}$ be the empirical distribution function of the n–sample $\mathcal{X}$ introduced in Section 2.2, and for $m \geq 1$ and $\lambda > 0$ define

$$E_{nm}(\lambda) = \sup_{C_1,\ldots,C_m} \sum_{j=1}^{m} \left\{\widehat{F}(C_j) - \lambda\,\|C_j\|\right\},$$

where the supremum is over disjoint intervals $C_1, \ldots, C_m$, $\widehat{F}(C)$ is the $\widehat{F}$–measure of C, and $\|C\|$ equals the length of C. Put $D_{nm}(\lambda) = E_{nm}(\lambda) - E_{n,m-1}(\lambda)$ and $\Delta = \sup_\lambda D_{n2}(\lambda)$. We reject the null hypothesis of unimodality if Δ is too large.

Cheng and Hall (1998) established that under $H_{0,\mathrm{class}}$, Δ is of size $n^{-3/5}$. We show next that under $H_{0,\mathrm{bound}}$ it is of size $n^{-4/7}$, for which purpose we augment (2.2) by the condition:

$$\begin{gathered} f' \text{ is Hölder-continuous within a neighbourhood of} \\ \text{the unique point } x_2 \neq x_1 \text{ satisfying } f(x_2) = f(x_1). \end{gathered} \tag{2.5}$$

Let W be as in Section 2.2, and define $C_4 = \{f(x_1)^4/|f'''(x_1)|\}^{1/7}$, $\Delta(t_1, t_2, u) = \{W(t_1) - W(t_2)\} - (t_2^4 - t_1^4) - u(t_2 - t_1)$ and

$$\begin{aligned} R_4 = 24^{1/7} \sup_{-\infty<u<\infty} \Bigg[& \sup_{-\infty<t_1<t_2<t_3<\infty} \{\Delta(0,t_1,u) + \Delta(t_2,t_3,u)\} \\ & - \sup_{-\infty<t_1<\infty} \Delta(0,t_1,u)\Bigg]. \end{aligned} \tag{2.6}$$

It may be proved that R_4 is finite and positive with probability one, and that its distribution has no atoms.

Theorem 2.3. *Assume conditions* (2.2) *and* (2.5). *Then* $n^{4/7}\Delta \to C_4 R_4$ *in distribution as* $n \to \infty$.

The bootstrap setting for Theorem 2.3 is similar to that for Theorem 2.1. Let Δ^* be the bootstrap version of Δ, computed using the resample $\mathcal{X}^*$ drawn by sampling from the distribution with density $\hat{f}_{\rm crit}$. For a suitable construction of W, Theorem 2.3 may be stated in the stronger sense that $n^{4/7}\Delta \to C_4 R_4$ in probability. We assume this construction below. Let W^* be another Wiener process, independent of W; define

$$U(r,s) = r^{-4} \int K''(s+u)\, W(ru)\, du\,;$$

let R denote the infimum of all $r>0$ such that $U(r,s)+\frac{1}{2}(1+s^2)$, as a function of s, does not change sign on the real line; and let S be the unique point at which $U(R,s)+\frac{1}{2}\,(1+s^2)$ vanishes. Put

$$\begin{aligned}\Psi(y_1,y_2,u) = W^*(y_1) &- W^*(y_2)\\ &- R^2 \int_0^1 t\,\Big[y_2^2\,U\{R, S+R^{-1}(1-t)\,y_2\}\\ &\qquad\qquad - y_1^2\,U\{R, S+R^{-1}(1-t)\,y_1\}\Big]\,dt\\ &- \tfrac{1}{2}\,(1+S^2)\,\big(y_2^2-y_1^2\big) - \tfrac{1}{6}\,RS\,\big(y_2^3-y_1^3\big)\\ &- \tfrac{1}{24}\,\big(y_2^4-y_1^4\big) - u\,(y_2-y_1)\,,\end{aligned}$$

and, with $\Psi/24^{1/7}$ replacing Δ, define R_4^* by (2.6). With probability one, R_4^* is finite and positive, and its distribution has no atoms.

Theorem 2.4. *Assume conditions* (2.2) *and* (2.5), *and that* W *is constructed so that* $n^{4/7}\Delta \to C_4 R_4$ *in probability. Then,*

$$\sup_{0\le x<\infty} \big|P\big(n^{4/7}\Delta^* \le C_4 x \big| \mathcal{X}\big) - P(R_4^* \le x|W)\big| \to 0$$

in probability as $n\to\infty$.

Theorem 2.4 is directly analogous to Theorem 2.2, and implies the obvious analogue of (2.4):

$$\sup_{0\le x<\infty} \big|P\big(\Delta^*/\Delta \le x \big| \mathcal{X}\big) - P(R_4^*/R_4 \le x|W)\big| \to 0 \qquad (2.7)$$

Therefore, bootstrap calibration applied to the ratio Δ^*/Δ produces tests of $H_{0,\rm class}$ with asymptotically correct level. Specifically, if f_0 is the template density introduced in Section 2.2, if $\widehat{H}_n(x) = P(\Delta^*/\Delta \le x|\mathcal{X})$, if $\widehat{H}_{0n}$ is the version of $\widehat{H}_n$ when the n-sample is drawn from f_0 rather than f, and if the constant u_α is defined by $P\{\widehat{H}_{0n}(u_\alpha) \geq 1-\alpha\} = \alpha$, then the test which rejects $H_{0,\text{bound}}$ if $\widehat{H}_n(u_\alpha) \geq 1-\alpha$ has asymptotically correct level under $H_{0,\text{bound}}$.

Hartigan (1997) has suggested an asymptotic test based on the results in Theorem 2.4, normalising the test statistic using the square root of the number of data values interior to the shoulder segment. If one calibrates via the asymptotic distribution then this ingenious approach avoids using the template density. In order to better capture second-order effects, however, one could compute the template density and then, simulating from that distribution (taking the Monte Carlo sample size equal to the actual sample size), compute an approximation to the distribution of the test statistic under the null hypothesis.

2.4. Adaptivity of bootstrap calibration methods. The factorisation which forms the basis for our bootstrap calibration method is also valid under $H_{0,\text{class}}$, where instead of (2.4) and (2.7) it produces results of the form:

$$\sup_{0\leq x<\infty} \left|P\big(\hat{h}^*_{\text{crit}}/\hat{h}_{\text{crit}} \leq x\big|\mathcal{X}\big) - P(R_1^*/R_1 \leq x|W)\right| \to 0\,, \tag{2.8}$$

$$\sup_{0\leq x<\infty} \left|P\big(\Delta^*/\Delta \leq x\big|\mathcal{X}\big) - P(R_3^*/R_3 \leq x|W)\right| \to 0\,. \tag{2.9}$$

A suitable regularity condition for each of these results is the following version of (2.2), where the shoulder point x_1 is no longer permitted, thereby ensuring that $H_{0,\text{class}}$ (rather than $H_{0,\text{bound}}$) obtains:

> f is supported on a compact interval $[a,b]$, and has two derivatives there; $f'=0$ at $x_0 \in (a,b)$, and $f' \neq 0$ at all other points in (a,b); f has two Hölder-continuous derivatives in a neighbourhood of x_0; $f''(x_0)<0$, $f'(a+)>0$, $f'(b-)<0$.

Result (2.8) is discussed in an ANU PhD thesis by M. York (1998), and (2.9) appears in Cheng and Hall (1996). As in the case of R_2 and R_4, the variables R_1 and R_3 are functionals of a standard Weiner process W; R_1^* and R_3^* are functionals of W and an independent Wiener process W^*; and all variables R_j and R_j^* have continuous distributions. It follows from (2.8) and (2.9) that if $H_{0,\text{class}}$ holds instead of $H_{0,\text{bound}}$, yet we apply the bootstrap test suggested when $H_{0,\text{bound}}$ is valid, the asymptotic level of the test lies strictly between 0 and 1. In this sense, the tests suggested in Sections 2.2 and 2.3 are adaptive; other approaches to calibration, such as that discussed towards the end of Section 2.1, do not enjoy this property. Moreover, bootstrap calibration under $H_{0,\text{bound}}$ turns out to be conservative when $H_{0,\text{class}}$ is true, as we shall show in the next section.

3. NUMERICAL STUDY

The bandwidth and dip/excess mass tests for $H_{0,\text{bound}}$ were applied to three Normal mixture densities: the two unimodal-with-shoulder densities given by

$$\left\{8e^{9/8}\left(1+8e^{9/8}\right)^{-1}\right\} * N(0,1) + \left(1+8e^{9/8}\right)^{-1} * N(-9\sqrt{3}/8, 0.0625)\,, \qquad (3.1)$$

$$(100/109) * N(0,1) + (9/109) * N(1.3, 0.09) \qquad (3.2)$$

and illustrated in panels (a) and (b), respectively, of Figure 3.1; and the unimodal-without-shoulder standard Normal density, depicted in panel (d) of that figure. In all cases the bandwidth and dip/excess mass tests for $H_{0,\text{bound}}$ were calibrated using the methods suggested in Sections 2.2 and 2.3. The template density f_0 employed for calibration was taken as

$$(16/17) * N(0,1) + (1/17) * N(-1.25, 0.0625)\,, \qquad (3.3)$$

and is unimodal with a shoulder. It is illustrated in panel (c) of Figure 3.1.

[*Put Figure 3.1 about here, please*]

The sample sizes used were 50 and 100. In each setting, 500 samples were simulated; and conditional on each of these, 500 resamples were drawn. Then, all the required conditional and unconditional probabilities were approximated by their corresponding empirical values. To obtain values of $\hat{h}_{\text{crit}}$ and $\hat{h}^*_{\text{crit}}$, kernel density estimates were computed over an equally-spaced grid of 512 points. To avoid problems arising from data sparseness in the tails, only modes that occurred within 1.5 standard deviations of the mean were counted. The same rule was followed when evaluating the dip/excess mass statistics.

Figure 3.2 illustrates the actual versus nominal levels when the two tests for $H_{0,\text{bound}}$ (calibrated using the density at (3.3) as the template) were applied to data generated from the two shoulder-densities given by (3.1) and (3.2), respectively. Note that the actual versus nominal curves are close to the diagonal line, especially in the cases illustrated by panels (b), (c) and (d). This indicates that both tests have accurate levels. The figure also suggests that, overall, the excess mass test has better level accuracy than the bandwidth test.

[*Put Figure 3.2 about here, please*]

Figure 3.3 depicts, for both the bandwidth and dip/excess mass tests, the actual versus nominal levels when the true density is standard Normal and the shoulder density f_0 is used to provide calibration. Note particularly that all the curves always lie below the diagonal line, illustrating the conservatism of a method calibrated for $H_{0,\text{bound}}$ when it is applied to test $H_{0,\text{class}}$.

[*Put Figure 3.3 about here, please*]

Figure 3.4 is essentially the obverse of Figure 3.3: in the latter, the sampling density was standard Normal, and we calibrated using f_0, but in Figure 3.4 the sampling density is f_0 and we calibrate using the standard Normal. The fact that the dashed and dotted lines in both panels of Figure 3.4 lie above the diagonal line demonstrates that, as expected, calibrating a test of $H_{0,\text{bound}}$ using a template for $H_{0,\text{class}}$ results in an anticonservative procedure.

[*Put Figure 3.4 about here, please*]

4. TECHNICAL ARGUMENTS

4.1. Proof of Theorem 2.1. Let $\eta = n^{-1/7}$ and write C, R for C_2, R_2, respectively. We shall prove that

> there exist $\epsilon_1, \epsilon_2 > 0$ such that, if $\hat{h}_{\text{crit}} = \hat{h}_{\text{crit}}(\epsilon_1, \epsilon_2)$ is re-defined to be the supremum of the set $\mathcal{H}$ of values $h \le n^{-(1/7)+\epsilon_1}$ such that $\hat{f}(\cdot|h)$ has at least one turning point in $\mathcal{I}(\epsilon_2) = (x_1 - \eta n^{\epsilon_2}, x_1 + \eta n^{\epsilon_2})$, then with probability tending to one, $\mathcal{H}$ is nonempty; and $n^{1/7}\hat{h}_{\text{crit}}$ has the claimed limit distribution. (4.1)

Arguments similar to those of Mammen, Marron and Fisher (1992) may be employed to prove that (a) for each $\epsilon_1 \in (0, 1/7)$, the probability that for some $h \ge n^{-(1/7)+\epsilon_1}$ the function $\hat{f}(\cdot|h_1)$ has more than one turning point in $\mathbb{R}$ converges to 0, (b) for each $c > 0$ and $\epsilon_2 > 0$, the probability that for some $h > cn^{-1/7}$ the function $\hat{f}(\cdot|h)$ has more than one turning point in $\mathbb{R}\backslash\mathcal{I}(\epsilon_2)$ converges to 0, and (c) with probability 1, $\hat{f}(\cdot|h)$ has at least one turning point in $\mathcal{I}(\epsilon_2)$ for each $h < \hat{h}_{\text{crit}}$. The theorem follows from (4.1) and (a)–(c).

The embedding of Komlós, Major and Tusnády (1975) ensures the existence of a standard Wiener process W_1 such that, with $W^0(t) = W_1(t) - tW_1(1)$, the empirical distribution function $\widehat{F}$ of $\mathcal{X}$ may be written as $\widehat{F}(x) = F(x) + n^{-1/2}\, W^0\{F(x)\} + O_p(n^{-1} \log n)$ uniformly in x. It follows that

$$\hat{f}'(x|h) - E\hat{f}'(x|h) = -\left(n^{1/2}\, h^2\right)^{-1} \int \left[W_1\{F(x-hz)\} - W_1\{F(x_1)\}\right] K''(z)\, dz + O_p\left\{(nh^2)^{-1} \log n\right\}$$

uniformly in $-\infty < x < \infty$ and $h > 0$. Writing $x = x_1 + \eta y$ and $h = \eta r_1$, and using standard results on the modulus of continuity of a Wiener process, we deduce that if $\epsilon_1, \epsilon_2 > 0$ are sufficiently small then for some $\epsilon_3 > 0$,

$$\begin{aligned} &\hat{f}'(x_1 + \eta y|\eta r_1) - E\hat{f}'(x_1 + \eta y|\eta r_1) \\ &\quad = -\left(n^{1/2}\eta^2 r_1^2\right)^{-1} \int \left[W_1\{F(x_1) + \eta(y - r_1 z)\, f(x_1)\}\right. \\ &\quad\quad \left. - W_1\{F(x_1)\}\right] K''(z)\, dz + O_p\left(\eta^2 n^{-\epsilon_3} r_1^{-2}\right) \end{aligned}$$

uniformly in $0 < r_1 \le \text{const.}n^{\epsilon_1}$ and $|y| \le \text{const.}n^{\epsilon_1}$, for all values of the constants. Therefore, defining

$$W_2(t) = -\{\eta f(x_1)\}^{-1/2}\left[W_1\{F(x_1) + \eta\, f(x_1)\, t\} - W_1\{F(x_1)\}\right],$$

we find that, uniformly in the same values of r_1 and y,

$$\begin{aligned} &\eta^{-2}\, r_1^2\left\{\hat{f}'(x_1 + \eta y|\eta r_1) - E\hat{f}'(x_1 + \eta y|\eta r_1)\right\} \\ &\quad = f(x_1)^{1/2} \int W_2(y - r_1 z)\, K''(z)\, dz + O_p\left(n^{-\epsilon_3}\right). \end{aligned} \tag{4.2}$$

Using the fact that f'' is Hölder continuous in a neighbourhood of x_1 we see that, for $\epsilon_1, \epsilon_2, \epsilon_3 > 0$ chosen sufficiently small,

$$\begin{aligned} E\hat{f}'(x_1 + \eta y|\eta r_1) &= \int f'\{x_1 + \eta(y - r_1 z)\}\, K(z)\, dz \\ &= \tfrac{1}{2}\, \eta^2\left(y^2 + r_1^2\right) f'''(x_1) + O\left\{\eta^2\left(y^2 + r_1^2\right) n^{-\epsilon_3}\right\} \end{aligned} \tag{4.3}$$

uniformly in $0 < r_1 \le \text{const.}n^{\epsilon_1}$ and $|y| \le \text{const.}n^{\epsilon_2}$. Combining (4.2) and (4.3) we deduce that

$$\begin{aligned} \hat{f}'(x_1 + \eta y|\eta r_1) = \eta^2 &\left[r_1^{-2}\, f(x_1)^{1/2} \int W_2(y - r_1 z)\, K''(z)\, dz\right. \\ &\left. + \tfrac{1}{2}\left(y^2 + r_1^2\right) f'''(x_1) + O_p\left\{\left(r_1^{-2} + y^2 + r_1^2\right) n^{-\epsilon_3}\right\}\right] \end{aligned} \tag{4.4}$$

uniformly in $0 < r_1 \leq \text{const}.n^{\epsilon_1}$ and $|y| \leq \text{const}.n^{\epsilon_2}$.

Let $T = \text{sgn}\{f'''(x_1)\}$, $C = \{f(x_1)/|f'''(x_1)|^2\}^{1/7}$, $C' = \{f(x_1)^2\,|f'''(x_1)|^3\}^{1/7}$, $y = Crs$, $r_1 = Cr$ and $W_2(Ct) = C^{1/2}\,TW(-t)$. Then W is a standard Wiener process, and (4.4) implies that for different values of $\epsilon_1, \epsilon_2, \epsilon_3 > 0$, chosen sufficiently small,

$$\begin{aligned}\hat{f}'(x_1 + \eta Crs|\eta Cr) &= \eta^2\, C'T\Big(r^{-2}\int W\{r(z-s)\}\, K''(z)\, dz \\ &\qquad + \tfrac{1}{2}\, r^2(1+s^2) + O_p[\{r^{-2} + r^2(1+s^2)\}\, n^{-\epsilon_3}]\Big) \\ &= \eta^2\, C'Tr^2\,\big[Z(r,s) + O_p\big\{\big(r^{-4} + 1 + s^2\big)\, n^{-\epsilon_3}\big\}\big]\,, \qquad (4.5)\end{aligned}$$

uniformly in $0 < r \leq \text{const}.n^{\epsilon_1}$ and $|y| \leq \text{const}.n^{\epsilon_2}$. Result (4.1) follows from this formula.

4.2. Proof of Theorem 2.2. We give the proof only in outline, noting the analogues of steps in the proof of Theorem 2.1 and not pausing to give detailed bounds for remainder terms. In the derivation of Theorem 2.1 we should replace $(\hat{f}(\cdot|h), f)$ by $(\hat{f}^*(\cdot|h), \hat{f}_{\text{crit}})$. Let $\hat{x}_1$ denote the shoulder of $\hat{f}_{\text{crit}}$. (Thus, $\hat{f}'_{\text{crit}}(\hat{x}_1) = \hat{f}''_{\text{crit}}(\hat{x}_1) = 0$.) In place of (4.2) we have, conditional on $\mathcal{X}$ and for a standard Wiener process W_2^* independent of W,

$$\begin{aligned}&\eta^{-2} r_1^2\,[\hat{f}^{*\prime}(\hat{x}_1 + \eta y|\eta r_1) - E\{\hat{f}^{*\prime}(\hat{x}_1 + \eta y|\eta r_1)|\mathcal{X}\}] \\ &\qquad = f(x_1)^{1/2}\int W_2^*(y - r_1 z)\, K''(z)\, dz + o_p(1)\,. \qquad (4.6)\end{aligned}$$

By (4.5) and since $\hat{h}_{\text{crit}} - \eta CR = o_p(\eta)$ we have, in notation from the proof of Theorem 2.1,

$$\hat{f}'_{\text{crit}}(x_1 + \eta CRs) = \hat{f}'_{\text{crit}}(x_1 + \eta CRs|\hat{h}_{\text{crit}}) = \eta^2 C'TR^2 Z(R,s) + o_p\big(\eta^2\big)\,.$$

Furthermore, $\hat{x}_1 - (x_1 + \eta CRS) = o_p(\eta)$, and so

$$\begin{aligned}&E\{\hat{f}^{*\prime}(\hat{x}_1 + \eta y|\eta r_1)|\mathcal{X}\} \\ &\quad = \int \hat{f}'_{\text{crit}}\{\hat{x}_1 + \eta(y - r_1 z)\}\, K(z)\, dz \\ &\quad = \eta^2 C'TR^2 \int Z\big\{R, (\eta CR)^{-1}(\hat{x}_1 - x_1) + (CR)^{-1}(\eta - r_1 z)\big\}\, K(z)\, dz + o_p\big(\eta^2\big) \\ &\quad = \eta^2 C'TR^2 \int Z\big\{R, S + (CR)^{-1}(y - r_1 z)\big\}\, K(z)\, dz + o_p\big(\eta^2\big)\,. \qquad (4.7)\end{aligned}$$

Combining (4.6) and (4.7) we deduce that

$$\hat{f}^{*\prime}(\hat{x}_1+\eta y|\eta r_1) = \eta^2\Big[r_1^{-2}\, f(x_1)^{1/2}\int W_2^*(y-r_1 z)\, K''(z)\, dz + C'TR^2\int Z\{R, S+(CR)^{-1}(y-r_1 z)\}\, K(z)\, dz\Big] + o_p\big(\eta^2\big)\,. \tag{4.8}$$

Making the changes of variable $y = Crs$, $r_1 = Cr$ and $W_2^*(Ct) = C^{1/2}\, W^*(-t)$, the right-hand side of (4.8) becomes

$$C'TR^2\eta^2\, Z^*(r,s) + o_p\big(\eta^2\big)\,.$$

The theorem follows from this approximation.

4.3. Proof of Theorem 2.3. Let $a = f(x_1)$ and $b = \frac{1}{24}\,|f'''(x_1)|$. Given $\epsilon_0, \epsilon_1 \in (0, \min(a, 1/7))$, define $\mathcal{J}_1 = (0, a-\epsilon_0]$, $\mathcal{J}_2 = (a-\epsilon_0, a-n^{-(3/7)+3\epsilon_1}]$ and $\mathcal{J}_3 = (a-n^{-(3/7)+3\epsilon_1}, \infty)$. Arguing as in the proof of Theorem 2 of Müller and Sawitzki (1991) we may show that

$$\sup_{\lambda\in\mathcal{J}_1} D_{n2}(\lambda) = O_p\{(n^{-1}\,\log n)^{2/3}\}\,, \quad \sup_{\lambda\in\mathcal{J}_2} D_{n2}(\lambda) = O_p\big(n^{-(4/7)-(\epsilon_1/5)}\big)\,.$$

Therefore,

$$\sup_{\lambda\in\mathcal{J}_1\cup\mathcal{J}_2} D_{n2}(\lambda) = o_p\big(n^{-4/7}\big)\,. \tag{4.9}$$

We prove the theorem in the case $f'''(x_1) > 0$. The case $f'''(x_1) < 0$ may be treated similarly. Since $f'''(x_1) > 0$ and condition (4.2) holds, $x_1 < x_0$ and there exists a point x_2 such that $x_0 < x_2, f(x_2) = f(x_1)$ and $f'(x_2) < 0$. Let $\eta = n^{-1/7}$, $\xi = n^{-\epsilon_3}$ with $\epsilon_3 \geq 1/7$, $\mathcal{I}_0 = (x_1 - \eta n^{\epsilon_1}, x_1 + \eta n^{\epsilon_1})$, $\mathcal{I}_1 = (x_2 - \xi n^{\epsilon_1}, x_2 + \xi n^{\epsilon_1})$ and $\mathcal{I}_2 = (-n^{\epsilon_1}, n^{\epsilon_1})$. Given $t_1, \ldots, t_3 \in \mathcal{I}_0$, put $y_j = (t_j - x_1)/\eta \in \mathcal{I}_2, j = 1, \ldots, 3$. Let $\sup^{(1)}, \ldots, \sup^{(7)}$ denote suprema over, respectively, (1) $-\infty < t_1 < t_2 < \infty$, (2) $t_1 \in \mathcal{I}_0, t_2 \in \mathcal{I}_1$ such that $t_1 < t_2$, (3) $y_1 \in \mathcal{I}_2$, (4) $-\infty < t_1 < \ldots < t_4 < \infty$, (5) $t_1, \ldots, t_3 \in \mathcal{I}_0, t_4 \in \mathcal{I}_1$ such that $t_1 < \ldots < t_4$, (6) $t_1 \in \mathcal{I}_0, t_2 \ldots, t_4 \in \mathcal{I}_1$ such that $t_1 < \ldots < t_4$, and (7) $y_1, \ldots, y_3 \in \mathcal{I}_2$ such that $y_1 < \ldots < y_3$. Write $\lambda = a - b\zeta\eta^3$, where $-\infty < \zeta < \infty$. Given a standard Wiener process W_1, define

$$W(y) = (a\eta)^{-1/2}\,\big[W_1\{F(x_1) + a\eta y\} - W_1\{F(x_1)\}\big]\,,$$

also a standard Wiener process. Using the embedding of Komlós, Major and Tusnády (1975) we may choose W_1, a standard Wiener process depending on n, such that

$$\begin{aligned}\widehat{F}(t_2) - \widehat{F}(t_1) = F(t_2) - F(t_1) + n^{-1/2}\big[W_1\{F(t_2)\} - W_1\{F(t_1)\} \\ - \{F(t_2) - F(t_1)\}\, W_1(1)\big] + O_p\big(n^{-1}\log n\big)\end{aligned}$$

uniformly in all t_1, t_2. Therefore, defining $D(x_1, x_2, \lambda) = \widehat{F}(x_2) - \widehat{F}(x_1) - \lambda(x_2 - x_1)$ and $D_0(y_1, y_2, \zeta) = a^{1/2}\{W(y_1) - W(y_2)\} - b(y_2^4 - y_1^4) - b\zeta(y_2 - y_1)$, and noting the Hölder continuity of f''' in a neighbourhood of x_1, we deduce that if $\epsilon_1 > 0$ is sufficiently small,

$$\begin{aligned}\sup{}^{(2)} D(t_1, t_2, \lambda) &= \sup{}^{(2)}\big\{\widehat{F}(t_2) - \widehat{F}(x_2) + \widehat{F}(x_1) - \widehat{F}(t_1) \\ &\quad + \widehat{F}(x_2) - \widehat{F}(x_1) - \lambda(t_2 - t_1)\big\} \\ &= \sup{}^{(3)}\Big(a\,(\xi y_2 - \eta y_1) + \tfrac{1}{2}\, f'(x_2)\,\xi^2 y_2^2 - b\,\eta^4 y_1^4 \\ &\quad + n^{-1/2}\big[W_1\{F(x_2) + a\xi y_2\} - W_1\{F(x_2)\}\big]\,\{1 + o_p(1)\} \\ &\quad - n^{-1/2}\big[W_1\{F(x_1) + a\eta y_1\} - W_1\{F(x_1)\}\big]\,\{1 + o_p(1)\} \\ &\quad - \big(a - b\zeta\eta^3\big)\,(x_2 - x_1) - \big(a - b\zeta\eta^3\big)\,(\xi y_2 - \eta y_1)\Big) \\ &\quad + \big\{\widehat{F}(x_2) - \widehat{F}(x_1)\big\} + o_p\big\{\eta^4 + \xi^2\big\}\,.\end{aligned}$$

Since $f'(x_2) < 0$ and $f'''(x_1) > 0$ then for any $-\infty < \zeta < \infty$, the above quantity is maximised when $\xi = n^{-3/7}$. Hence,

$$\begin{aligned}\sup{}^{(2)} D(t_1, t_2, \lambda) = \widehat{F}(x_2) - \widehat{F}(x_1) - \big(a - b\zeta\eta^3\big)\,(x_2 - x_1) \\ + \eta^4 \sup{}^{(3)} D_0(0, y_1, \zeta) + o_p(\eta^4)\,.\end{aligned} \tag{4.10}$$

Similarly,

$$\begin{aligned}&\sup{}^{(5)}\big\{D(t_1, t_2, \lambda) + D(t_3, t_4, \lambda)\big\} \\ &\quad = \widehat{F}(x_2) - \widehat{F}(x_1) - \big(a - b\zeta\eta^3\big)\,(x_2 - x_1) \\ &\qquad + \eta^4 \sup{}^{(7)}\big\{D_0(0, y_1, \zeta) + D_0(y_2, y_3, \zeta)\big\} + o_p\big(\eta^4\big)\,,\end{aligned} \tag{4.11}$$

$$\begin{aligned}&\sup{}^{(6)}\{D(t_1, t_2, \lambda) + D(t_3, t_4, \lambda)\} \\ &\quad = \widehat{F}(x_2) - \widehat{F}(x_1) - \big(a - b\zeta\eta^3\big)\,(x_2 - x_1) \\ &\qquad + \eta^4 \sup{}^{(3)} D_0(0, y_1, \zeta) + o_p\big(\eta^4\big)\,,\end{aligned} \tag{4.12}$$

Define $\mathcal{I}_3 = (a - b\,\eta^{3(1+\epsilon_1)}, +\infty)$, $\mathcal{I}_4 = (-\infty, n^{3\epsilon_1})$ (i.e. such that $\mathcal{I}_3 = \{\lambda(\zeta) : \zeta \in \mathcal{I}_4\}$), $S_n = \sup_{\lambda\in\mathcal{I}_3} D_{n2}(\lambda)$,

$$S_n' = \sup_{\lambda\in\mathcal{I}_3} \Big[\max\big(\sup{}^{(5)}\,\{D(t_1,t_2,\lambda) + D(t_3,t_4,\lambda)\},\\ \sup{}^{(6)}\,\{D(t_1,t_2,\lambda) + D(t_3,t_4,\lambda)\}\big) - \sup{}^{(2)}\, D(t_1,t_2,\lambda)\Big]\,.$$

We may show that for sufficiently small ϵ_1, $P(S_n = S_n') \to 1$. From this result, (4.10), (4.11) and (4.12) we deduce that

$$\eta^{-4} S_n = \sup_{\zeta\in\mathcal{I}_4} \Big[\sup{}^{(7)}\{D_0(0,y_1,\zeta) + D_0(y_2,y_3,\zeta)\}\\ - \sup{}^{(3)} D_0(0,y_1,\zeta)\Big] + o_p(1)\,. \quad (4.13)$$

Define

$$Z' = \sup_{-\infty<\zeta<\infty} \bigg(\sup_{-\infty<y_1<\ldots<y_3<\infty} \big[a^{1/2}\{-W(y_3) + W(y_2) - W(y_1) + W(0)\}\\ - b\,\big(y_3^4 - y_2^4 + y_1^4\big) - \zeta\,(y_3 - y_2 + y_1)\big]\\ - \sup_{-\infty<y_1<\infty} \big[a^{1/2}\{W(0) - W(y_1)\} - b\,y_1^4 - \zeta\,y_1\big]\bigg)\,.$$

It can be shown that the difference between Z' and the right-hand side of (4.13) converges in probability to zero. Changing variable from y_i to $t_i = (b^2/a)^{1/7} y_i$, and noting that $W_2(t) = (b^2/a)^{1/14}\ W\{(a/b^2)^{1/7} t\}$ also defines a Wiener process, we deduce that Z' has the same distribution as $(a^4/b)^{1/7} 24^{-1/7}\, Z$. Theorem 2.3 follows from this result and (4.9).

4.4. Proof of Theorem 2.4. Write $\widehat{F}_{\text{crit}}$ for the distribution function corresponding to density $\hat{f}_{\text{crit}}$. Let $\hat{x}_1$ denote the shoulder of $\hat{f}_{\text{crit}}$ (thus, $\hat{f}_{\text{crit}}'(\hat{x}_1) = \hat{f}_{\text{crit}}''(\hat{x}_1) = 0$). Let C, C' be as in Section 4.1. Using the embedding of Komlós, Major and Tusnády (1975) we may prove that for an appropriate choice of W, and with the random function U defined as in Section 2.3,

$$\hat{f}_{\text{crit}}'(\hat{x}_1 + \eta C y) = \eta^2 C' T R^2 \left[U\big(R, S + R^{-1} y\big) + \tfrac{1}{2}\left\{S + (R^{-1} y)^2\right\} + o_p(1)\right].$$

(In this simplified argument it is assumed, here and below, that $|y|, |y_1|, |y_2|$ are all bounded.) Therefore, using the exact form for the remainder in Taylor's Theorem,

$$\widehat{F}_{\text{crit}}(\hat{x}_1 + \eta C y) - \widehat{F}_{\text{crit}}(\hat{x}_1)$$

$$
\begin{aligned}
&= \eta C y\, \hat f_{\text{crit}}(\hat x_1) + (\eta C y)^2 \int_0^1 t\, \hat f'_{\text{crit}}\{\hat x_1 + \eta C y(1-t)\}\, dt \\
&= \eta C y\, \hat f_{\text{crit}}(\hat x_1) + (\eta C y)^2\, \eta^2 C' T R^2 \\
&\qquad \times \int_0^1 t\, \big[U\big\{R, S + R^{-1}(1-t) y\big\} \\
&\qquad + \tfrac{1}{2}\big\{S + R^{-1}(1-t)\, y\big\}^2\big]\, dt + o_p\big(\eta^4\big)\,. \qquad (4.14)
\end{aligned}
$$

Hence, writing $t_i = \hat x_1 + \eta C y_i$ for $i = 1$ and 2, and defining $A(y_1, y_2) = \widehat F_{\text{crit}}(\hat x_1 + \eta C y_2) - \widehat F_{\text{crit}}(\hat x_1 + \eta C y_1)$ and $\lambda = \hat f_{\text{crit}}(\hat x_1) - u(C\eta)^3 C C' T$, we have

$$
\begin{aligned}
&\big\{A(y_1,y_2) - \lambda\,(t_2 - t_1)\big\}\big/\big(\eta^4\, C^2 C' T\big) \\
&\quad = R^2 \int_0^1 t\, \Big[y_2^2\, U\big\{R, S + R^{-1}(1-t)\, y_2\big\} - y_1^2\, U\big\{R, S + R^{-1}(1-t)\, y_1\big\}\Big]\, dt \\
&\qquad + \tfrac{1}{2}\, R^2(1+S^2)\, \big(y_2^2 - y_1^2\big) + \tfrac{1}{6}\, RS\, \big(y_2^3 - y_1^3\big) \\
&\qquad + \tfrac{1}{24}\, \big(y_2^4 - y_1^4\big) + u\,(y_2 - y_1) + o_p(1)\,. \qquad (4.15)
\end{aligned}
$$

Define $\hat x_2 \neq \hat x_1$ by $\hat f_{\text{crit}}(\hat x_2) = \hat f_{\text{crit}}(\hat x_1)$. Since $\hat h_{\text{crit}} = \eta C R + o_p(\eta^2)$ and $\hat x_1 = x_1 + \eta C R S + o_p(\eta)$, we may show that $\hat x_2 = x_2 + O_p(\eta^2)$. This result and the Hölder continuity of f' near x_2 yield that, for any sequence of numbers $\xi = o(1)$ and real number $|y| < \infty$,

$$
\widehat F_{\text{crit}}(\hat x_2 + \xi y) = \widehat F_{\text{crit}}\big(\hat x_2\big) + \xi y \hat f_{\text{crit}}\big(\hat x_2\big) + \tfrac{1}{2}\, f'(x_2) \xi^2 y^2 + o_p(\xi^2)\,. \qquad (4.16)
$$

Using the Komlós–Major–Tusnády embedding again, this time conditional on $\mathcal{X}$ and for the empirical distribution function $\widehat F^*$ of the resample $\mathcal{X}^*$; and noting that

$$
n^{-1/2}\, \big\{\eta C\, \hat f_{\text{crit}}(\hat x_1)\big\}^{1/2} \big(\eta^4\, C^2 C'\big)^{-1} \to 1
$$

in probability as $n \to \infty$; we may establish the existence of standard Wiener processes W^* and W^{**} (conditional on $\mathcal{X}$) such that

$$
\begin{aligned}
\big\{\widehat F^*(\hat x_1 + \eta C y_2) - \widehat F^*(\hat x_1 + \eta C y_1) - A(y_1, y_2)\big\}\big/\big(-\eta^4 C^2 C' T\big)& \\
= W^*(y_2) - W^*(y_1) + o_p(1)& \qquad (4.17)
\end{aligned}
$$

and

$$
\begin{aligned}
&\widehat F^*(\hat x_2 + \xi y_2) - \widehat F^*(\hat x_2 + \xi y_1) \\
&\qquad = \Big[\widehat F_{\text{crit}}(\hat x_2 + \xi y_2) - \widehat F_{\text{crit}}(\hat x_2 + \xi y_1) \\
&\qquad\quad + n^{-1/2}\big\{\xi \hat f_{\text{crit}}(\hat x_2)\big\}^{1/2}\big\{W^{**}(y_2) - W^{**}(y_1)\big\}\Big]\, \{1 + o_p(1)\}
\end{aligned}
$$

The last equality and (4.16) yield that

$$\begin{aligned}
&\widehat{F}_{\text{crit}}\big(\hat{x}_2+\xi y_2\big)-\widehat{F}_{\text{crit}}\big(\hat{x}_2+\xi y_1\big)-\lambda\big(\hat{x}_2+\xi y_2-\hat{x}_2-\xi y_1\big)\\
&\qquad=\tfrac{1}{2}\,f'(x_2)\xi^2\big(y_2^2-y_1^2\big)+\eta^3CC'\xi u(y_2-y_1)\\
&\qquad\quad+n^{-1/2}\big\{\xi\hat{f}_{\text{crit}}(\hat{x}_2)\big\}^{1/2}\big\{W^{**}(y_2)-W^{**}(y_1)\big\}\\
&\qquad\quad+o_p\big(\xi^2+\eta^3\xi+n^{-1/2}\xi^{1/2}\big)\,.
\end{aligned}\tag{4.18}$$

Combining (4.15) and (4.17), and observing that $C^2C'=C_4$, we see that

$$\widehat{F}^*(t_2)-\widehat{F}^*(t_1)-\lambda\,(t_2-t_1)=-\eta^4C_4T\,\big\{\Psi(y_1,y_2,u)+o_p(1)\big\}\,.\tag{4.19}$$

Then (4.18) and (4.19) imply that

$$\begin{aligned}
&\widehat{F}^*(\hat{x}_2+\xi y_2)-\widehat{F}^*(\hat{x}_1+\eta C y_1)-\lambda\,(x_2+\xi y_2-x_1-\eta C y_1)\\
&\qquad=\tfrac{1}{2}\,f'(x_2)\,\xi^2y_2^2+u\eta^3\xi CC'y_2+n^{-1/2}\big\{\xi\hat{f}_{\text{crit}}(\hat{x}_2)\big\}^{1/2}\\
&\qquad\quad\times\big\{W^{**}(y_2)-W^{**}(0)\big\}-\eta^4C_4T\Psi(y_1,0,u)\\
&\qquad\quad+\widehat{F}^*(\hat{x}_2)-\widehat{F}^*(\hat{x}_1)-\lambda(\hat{x}_2-\hat{x}_1)\\
&\qquad\quad+o_p\big(\xi^2+\eta^4+n^{-1/2}\xi^{1/2}\big)\,.
\end{aligned}\tag{4.20}$$

Arguing as in the proof of Theorem 2.3 and observing (4.18) – (4.20),

$$\begin{aligned}
\Delta^*=\eta^4C_4\sup_u\Bigg[&\sup_{y_1<\ldots<y_3}\big\{\Psi(0,y_1,u)+\Psi(y_2,y_3,u)\big\}\\
&\qquad\qquad-\sup_y\,\Psi(0,y,u)\Bigg]+o_p\big(\eta^4\big)\,.
\end{aligned}$$

This result implies Theorem 2.4.

Acknowledgements. We are grateful to the Editor and two anonymous referees for helpful comments.

REFERENCES

CHENG, M.-Y. AND HALL, P. (1996). Calibrating the excess mass test of modality. *Research Report SRR–002–97*. Centre for Mathematics and its Applications, Australian National University, Canberra.

CHENG, M.-Y. AND HALL, P. (1998). Calibrating the excess mass and dip tests of modality. *J. Roy. Statist. Soc. Ser. B* **60**, 579–589.

COX, D.R. (1966). Notes on the analysis of mixed frequency distributions. *Brit. J. Math. Statist. Psychol.* **19**, 39–47.

HARTIGAN, J.A. (1997). Establishing antimodes. Manuscript.

HARTIGAN, J.A. AND HARTIGAN, P.M. (1985). The dip test of unimodality. *Ann. Statist.* **13**, 70–84.

KENDALL, M. AND STUART, A. (1979). *The Advanced Theory of Statistics*, Vol. 2, Fourth Edn. Griffin, London.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's, and the sample DF. I. *Z. Wahrscheinlichkeits-theorie verw. Gebiete* **32**, 111–131.

MAMMEN, E., MARRON, J.S. AND FISHER, N.I. (1992). Some asymptotics for multimodality tests based on kernel density estimates. *Probab. Theory Rel. Fields* **91**, 115–132.

MÜLLER, D.W AND SAWITZKI, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86**, 738–746.

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*, Eds J.F.C. Kingman and G.E.H. Reuter, pp. 248–259. Cambridge University Press, Cambridge, UK.

YORK, M. (1998). *Tests for Modality*. ANU PhD thesis, in preparation.

CAPTIONS FOR FIGURES

Caption for Figure 1.1. Panels (a), (b) and (c) respectively depict the standard Normal density, densities represented by the Normal mixture formulae (3.2) and $0.8*N(-0.3,1)+0.2*N(1.6,0.16)$, giving a unimodal-without-shoulder, a unimodal-with-shoulder and a bimodal density.

Caption for Figure 3.1. Panels (a), (b) and (c) depict the unimodal-with-shoulder densities represented by the Normal mixture formulae (3.1)–(3.3), respectively. Panel (d) illustrates the standard Normal density, which of course is unimodal without a shoulder.

Caption for Figure 3.2. Actual versus nominal levels for the bandwidth (dashed lines) and dip/excess mass (dotted lines) tests, calibrated for $H_{0,\mathrm{bound}}$ using the template density at (3.3), when data are generated from the density at (3.1) (panel (a) for $n = 50$ and panel (b) for $n = 100$) or from the density at (3.2) (panel (c) for $n = 50$ and panel (d) for $n = 100$).

Caption for Figure 3.3. Actual versus nominal levels for the bandwidth (dashed lines) and dip/excess mass (dotted lines) tests, calibrated for $H_{0,\mathrm{bound}}$ using the template density at (3.3), when data are generated from the standard Normal density (panel (a) for $n = 50$ and panel (b) for $n = 100$).

Caption for Figure 3.4. Actual versus nominal levels for the bandwidth (dashed lines) and dip/excess mass (dotted lines) tests, calibrated for $H_{0,\mathrm{class}}$ using the standard Normal density, when data are generated from the density at (3.3) (panel (a) for $n = 50$ and panel (b) for $n = 100$).

Figure 1.1

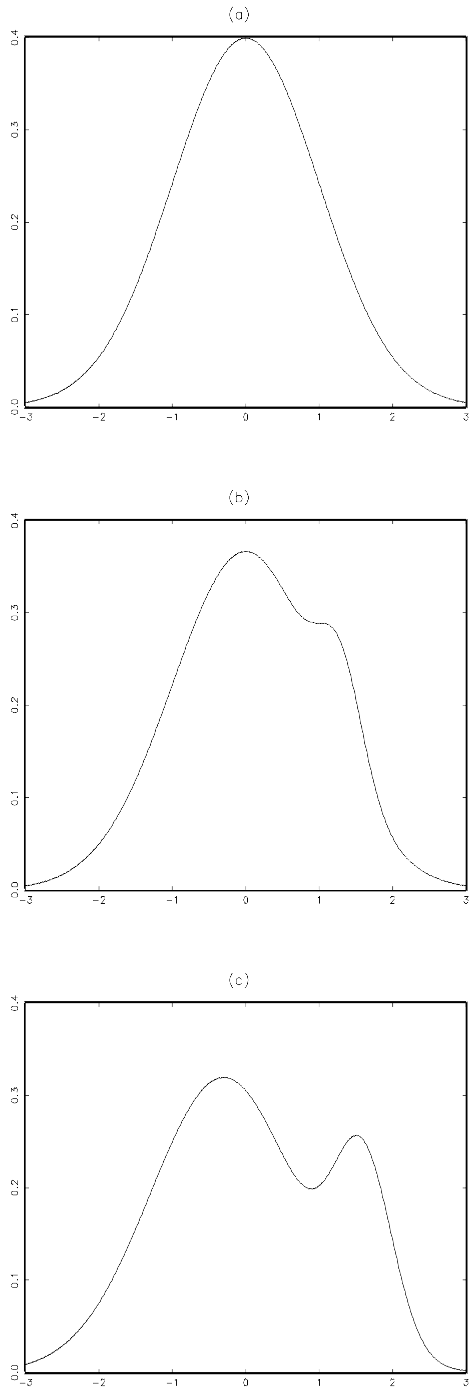

Figure 3.1

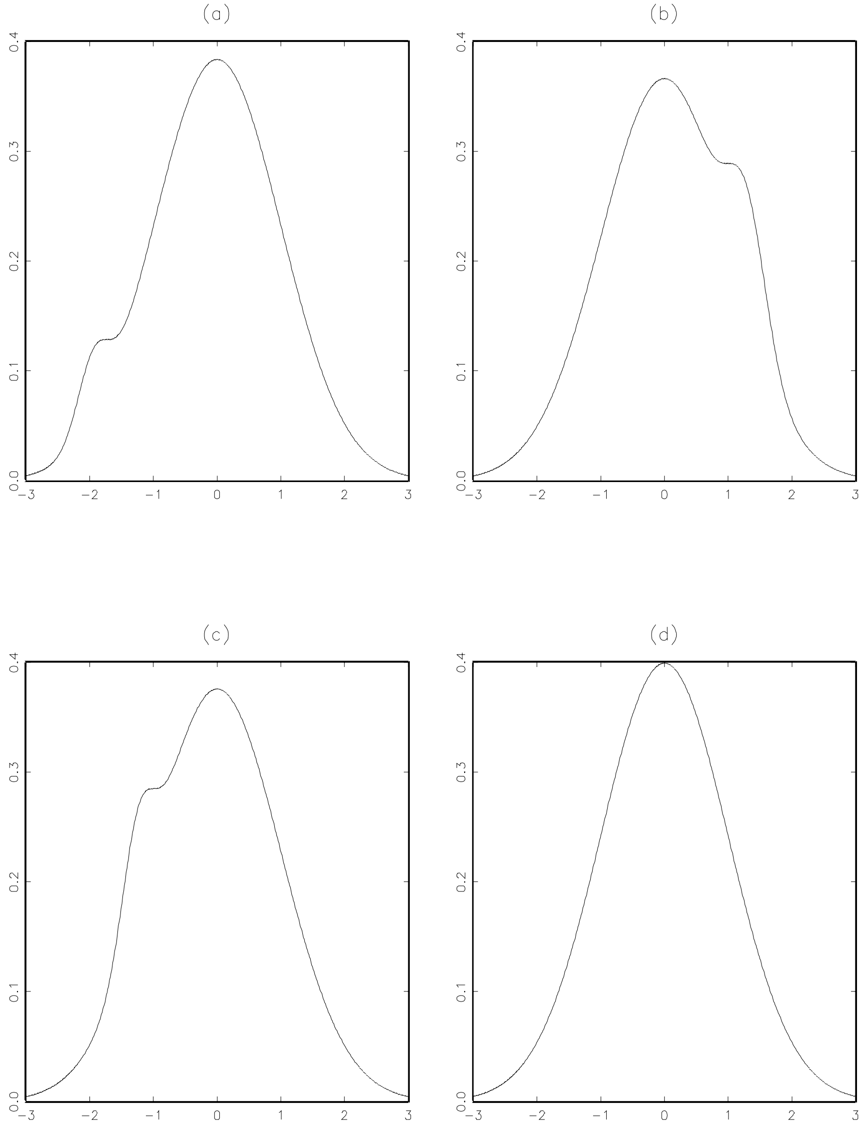

Figure 3.2

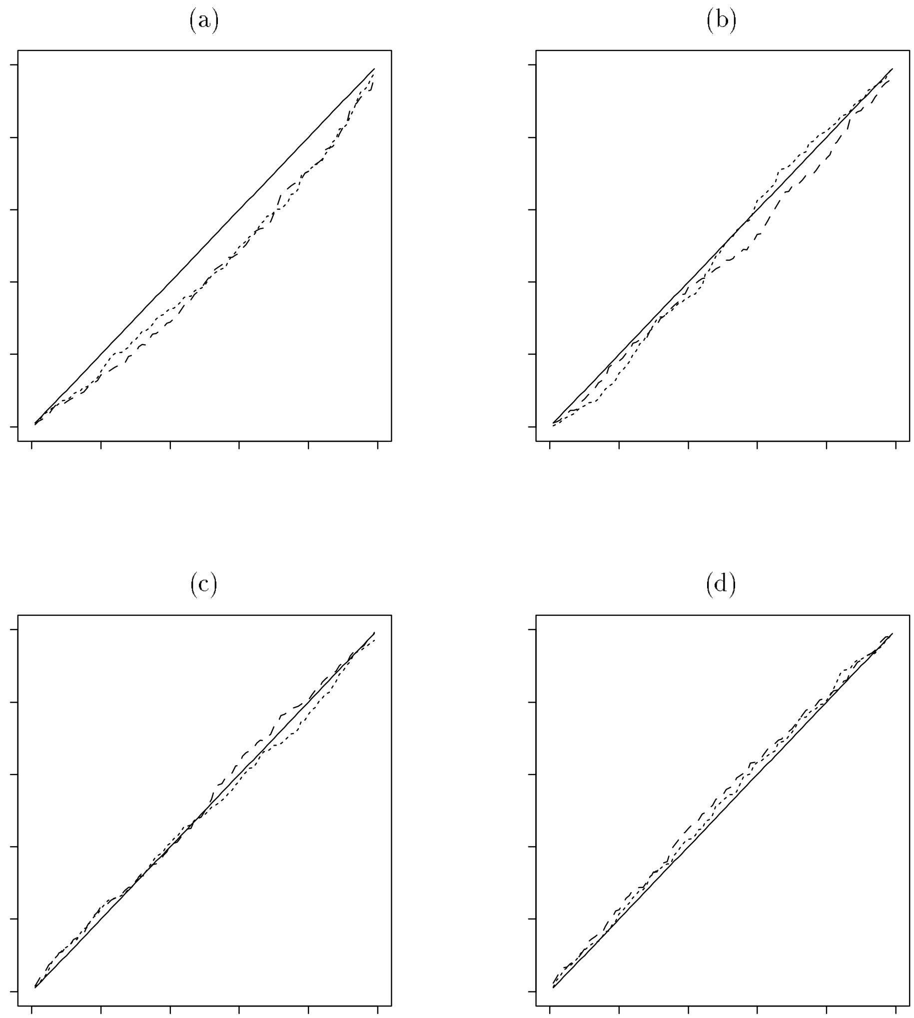

Figure 3.3

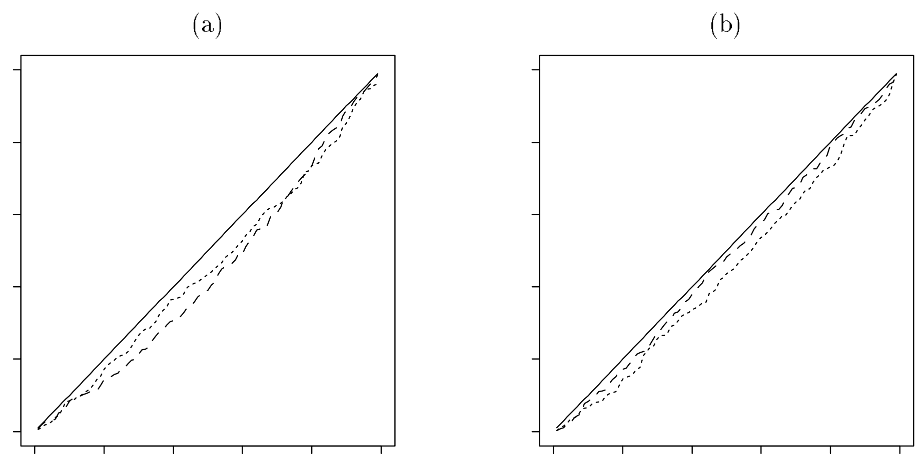

Figure 3.4