Excess Mass Estimates and Tests for Multimodality
Author(s): D. W. Muller and G. Sawitzki
Source: *Journal of the American Statistical Association*, Vol. 86, No. 415 (Sep., 1991), pp. 738-746
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2290406
Accessed: 09-03-2020 14:57 UTC

Excess Mass Estimates and Tests for Multimodality

D. W. MÜLLER and G. SAWITZKI*

We propose and study a method for analyzing the modality of a distribution. The method is based on the excess mass functional. This functional measures excessive empirical mass in comparison with multiples of uniform distribution. By the excess mass approach we separate the investigation about the number of modes from questions concerning their location. For distributions on the line, the excess mass functional can be estimated at a square-root rate. The asymptotic behavior of estimators is studied, and tests for multimodality based on the excess mass are derived.

KEY WORDS: Bump hunting; Cluster analysis.

1. INTRODUCTION

We propose and study a method for analyzing the modality of a distribution. The key is a new way of thinking about modality, which is statistically more relevant than the classical analytical interpretation of peaks in the density. We separate questions about the number of modes from problems concerning their location. Benefits of our approach include guaranteed confidence and the interpretation of the statistical analysis in terms of the empirical probability distribution.

The idea we adopt is this: A mode is present where an excess of probability mass is concentrated. Among the peaks of a distribution there is a difference in strength, or distinctiveness. This will be measured by means of the excess mass functional $\lambda \to \mathbf{E}(\lambda)$ defined in Section 2. This functional can be estimated from the data.

From an analytical point of view, a usual definition relates a mode to a local maximum of the density. With this idea, however, a distribution can have a mode at x_0 while giving arbitrarily small probability to some neighborhood containing x_0 (Figure 1). From a statistical point of view, one would like to associate a mode with a location carrying high probability over a neighborhood. Thus the analytical definition does not capture the statistical idea adequately. This explains why several related definitions have been introduced, like block clusters (Hartigan 1975), modes of a given width (Hartigan 1977), bumps (Good and Gaskins 1980), or density contour clusters (Hartigan 1975; Wong 1985).

Estimating the location and estimating the number of modes are separate problems. It is tempting to solve both problems simultaneously by first estimating the density and then deriving estimates for location and number of modes. These procedures, however, will in general depend on the amount of smoothing chosen in the initial density estimation step. The modality of an estimated density is affected by this choice; see Silverman (1981) for a test of unimodality exploiting this dependence on bandwidth selection. Using excess mass estimates, we take a more direct approach. This approach allows a distinction between "empirical" peaks strong enough to provide sufficient confidence for a peak in the underlying distribution and those that might be attributed to mere sampling fluctuation. Thus the excess mass approach can avoid overestimating the number of modes of a distribution.

It is well known that there is no way of obtaining nonparametric confidence statements about upper bounds for the number of modes (Donoho 1988). This is due to the fact that by an infinitesimal perturbation in total variation norm any density can be changed into one having infinitely many modes. Despite the structural instability that forbids upper bounds for the number of modes, confidence statements about lower bounds can be obtained in various ways, since at least those modes that are distinct enough can be inferred from the empirical distribution. The "distinctiveness" of modes is reflected by the excess mass functional described in the next section.

In Section 2 we define the excess mass functional and discuss elementary properties. Section 3 describes related graphical displays. Tests based on the excess mass functional are discussed in Section 4, asymptotic results in Section 5 with proofs in Appendix A. Appendix B gives an algorithm.

2. EXCESS MASS AND EXCESS MASS ESTIMATES

In this article we consider distributions F on $\mathbf{R}^k$ that have a bounded continuous density f with respect to Lebesgue measure. Later on, we mainly concentrate on $k = 1$. Our tool for investigating the modality of a distribution F is to compare it with a reference measure, typically a multiple of Lebesgue measure. For various values of a constant λ, we consider the mass of F exceeding the λ-multiple of Lebesgue measure (Figure 2). In formal terms, we study the excess mass functional

$$\lambda \to \mathbf{E}(\lambda) = \int (f(x) - \lambda)^+ \, dx.$$

This was considered already in Hartigan (1987) and in Müller and Sawitzki (1987).

$\mathbf{E}(\lambda)$ is a sum of contributions $\mathbf{E}_C(\lambda) \equiv \int_C (f(x) - \lambda)\, dx$ coming from connected sets C $[C \subset \{x : f(x) \geq \lambda\}]$. The connected components of $\{x : f(x) \geq \lambda\}$ will be called density contour clusters at level λ, for short λ-clusters (Figure 3) (cf. Hartigan 1975). As λ increases, the λ-clusters concentrate on modes (local maxima of f). In particular, for any λ, a distribution with m modes has at most m λ-clus-

* D. W. Müller is Professor and G. Sawitzki is Akademischer Rat, Institut für Angewandte Mathematik, Universität Heidelberg, W-6900 Heidelberg, Germany. The authors are indebted to S. Marron for his suggestions leading to an improved presentation; special thanks are due to a referee and W. Polonik for helpful comments.

Journal of the American Statistical Association
September 1991, Vol. 86, No. 415, Theory and Methods

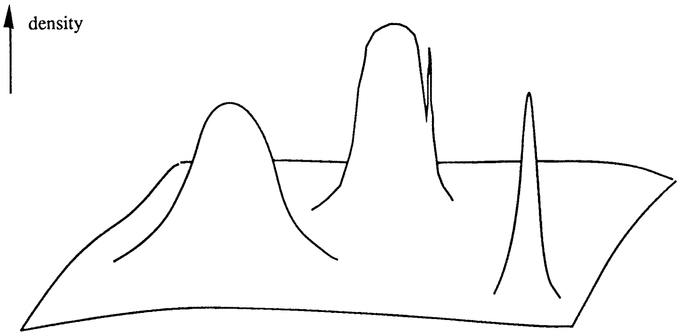


Figure 1. Modes of Different Statistical Quality. Two peaks of this density are strong and interesting features, while the others may not be reflected at all in a sample of moderate size.

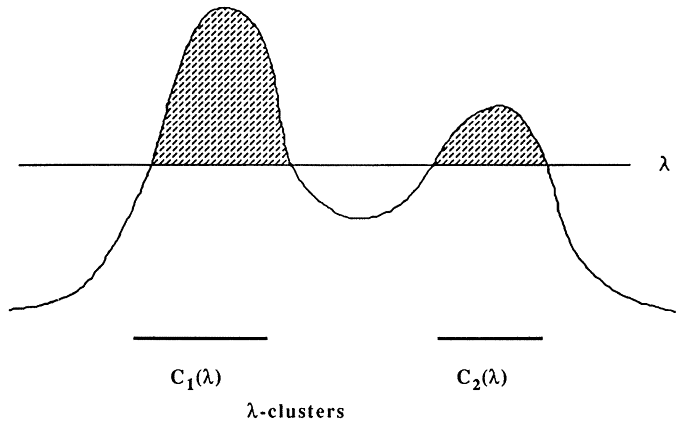


Figure 3. The Excess Mass is the Sum of Contributions Coming From the λ-Clusters, That is, the Connected Components of $\{f(x) \geq \lambda\}$.

ters. If a density has exactly m λ-clusters, the excess $\mathbf{E}(\lambda)$ can be expressed as

$$\mathbf{E}(\lambda) = \int (f(x) - \lambda)^{+}\, dx = \sup \sum \int_{C_j(\lambda)} (f(x) - \lambda)\, dx,$$

where the supremum is taken over all families $\{C_j : j = 1, \ldots, m\}$ of pairwise disjoint connected sets. In general we define

$$\mathbf{E}_m(\lambda) = \sup \sum_{j=1}^{m} \int_{C_j(\lambda)} (f(x) - \lambda)\, dx, \tag{1}$$

where the supremum is as already given. Defining $H_\lambda = F - \lambda \cdot \mathrm{Leb}$ (Leb stands for Lebesgue measure) we can write (1) as

$$\mathbf{E}_m(\lambda) = \sup \sum_{j=1}^{m} H_\lambda(C_j).$$

Given independent observations $x_1, \ldots, x_n$ sampled from F, we substitute the empirical measure F_n to define $H_{n,\lambda} = F_n - \lambda \cdot \mathrm{Leb}$. This gives estimators

$$\mathbf{E}_{n,M}(\lambda) = \sup \sum_{j=1\ldots M}^{M} H_{n,\lambda}(C_j), \tag{2}$$

where the parameter M is an assumed maximum number of modes.

For testing, in general the interesting alternative to unimodality would be multimodality with a sizeable excess

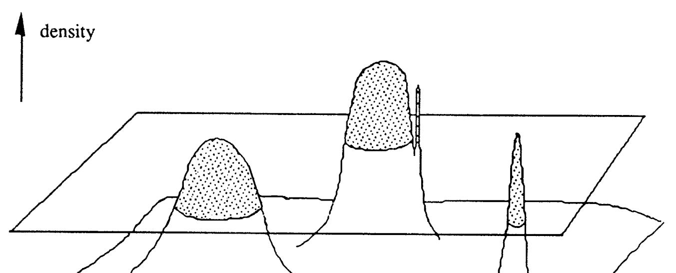


Figure 2. The Excess Mass. The probability mass exceeding a given level measures the strength of a mode.

probability ascribed to the modes. A large difference $\mathbf{D}_{n,M}(\lambda) = \mathbf{E}_{n,M}(\lambda) - \mathbf{E}_{n,1}(\lambda)$ for some $\lambda > 0$ indicates a violation of the hypothesis of unimodality in this direction and suggests $\Delta_{n,M} = \max_\lambda \mathbf{D}_{n,M}(\lambda)$ as a test statistic. This is the test statistic of the *excess difference test* for multimodality.

Components C_j maximizing the sum in (2) will be called "empirical λ-clusters" and denoted by $C_{n,j}(\lambda)$. They may be used as estimators of the λ-clusters of f. In the one-dimensional situation the empirical λ-clusters are closed intervals with endpoints at data points, or empty. In the absence of flat parts of f they consistently estimate the real λ-clusters (cf. Prop. 2, Sec. 5).

The excess mass $\mathbf{E}(\lambda)$ measures the amount of probability mass exceeding density level λ, thus reflecting the distinctiveness of the modes. It defines a *concentration curve* containing information about certain aspects of the shape of f. For the uniform distribution on an interval of the line, it is linear, quadratic for the triangular, and $2\Phi(\theta) - 2\lambda\theta - 1$, where $\theta = (-2 \log(\sqrt{2\pi}\lambda))^{1/2}$, for the standard normal distribution. The excess mass function $\lambda \rightarrow \mathbf{E}(\lambda)$ is continuous, monotone decreasing, $\mathbf{E}(0) = 1$, and $\mathbf{E}(\lambda) = 0$ for $\lambda \geq \sup f(x)$. Of course $\mathbf{E}_i(\lambda) \leq \mathbf{E}_j(\lambda)$ for $i \leq j$. For a density with at most m λ-clusters one has $\mathbf{E}_j(\lambda) = \mathbf{E}_m(\lambda) = \mathbf{E}(\lambda)$ for $j \geq m$. The estimate $\lambda \rightarrow \mathbf{E}_{n,M}(\lambda)$ is continuous, piecewise linear, and monotone decreasing. $\mathbf{E}_{n,M}(0) = 1$, and, in the absence of ties, $\mathbf{E}_{n,M}(\lambda) = M/n$ for large λ.

In what follows we will only consider distributions on the line. It will be convenient to use the symbols $F(x)$, $F_n(x)$, $H_\lambda(x)$, and $H_{n,\lambda}(x)$ for the (right-continuous) distribution functions evaluated at x of the measures F, F_n, H_λ, and $H_{n,\lambda}$, respectively. At any level λ, $\mathbf{E}_{n,1}(\lambda) = \sup_{x' \leq x''} H_{n,\lambda}(x'') - H_{n,\lambda}(x')$. Computing differences $\mathbf{E}_{n,M+1}(\lambda) - \mathbf{E}_{n,M}(\lambda)$ is facilitated by a "splitting property" (for simplicity, $M = 1$ only):

$\mathbf{D}_{n,2}(\lambda) = \mathbf{E}_{n,2}(\lambda) - \mathbf{E}_{n,1}(\lambda)$ is the larger of the two quantities

$$\max\{H_{n,\lambda}(C) : C \subset C_{n,1}(\lambda)^c\},\ \max\{-H_{n,\lambda}(C) : C \subset C_{n,1}(\lambda)\}. \tag{3}$$

Here the maximum is extended over all intervals C, but it suffices to take the first maximum over all closed intervals, the second over all open intervals, with endpoints at data points. Statement (3) becomes clear from considering the function $x \to H_{n,\lambda}(x)$ noting that $H_{n,\lambda}([x, y]) = H_{n,\lambda}(y) - H_{n,\lambda}(x\text{-})$ for $x \leq y$. Figure 4 shows typical graphs of $H_{n,\lambda}$ revealing the two alternatives.

Example. Let U_a be the uniform distribution on $[-a/2 - 1/2, -a/2] \cup [a/2, a/2 + 1/2]$. Then the excess mass $\mathbf{E}(\lambda)$ is composed of contributions coming from the intervals, $H_\lambda([-a/2 - 1/2, -a/2]) = H_\lambda([a/2, a/2 + 1/2]) = 1/2 - \lambda/2$ and a gap contribution $-H_\lambda((-a/2, a/2)) = -\lambda a$. For any λ and small gap size a we have $\mathbf{E}_1(\lambda) = H_\lambda([-a/2 - 1/2, -a/2 + 1/2])$ and $\mathbf{E}_2(\lambda) - \mathbf{E}_1(\lambda) = -H_\lambda((-a/2, a/2))$. If we increase the gap size a so that $-\lambda a < 1/2 - \lambda/2$, then $\mathbf{E}_1(\lambda)$ is the contribution of one interval and $\mathbf{E}_2(\lambda)$ picks up the other. The degree of separation of the modes (determined by their distance a) is reflected in $\max_\lambda(\mathbf{E}_2(\lambda) - \mathbf{E}_1(\lambda)) = a/(2a + 1)$.

For every bimodal distribution F, the quantity $\max_\lambda(\mathbf{E}_2(\lambda) - \mathbf{E}_1(\lambda))$ is half the total variation distance between F and the closest unimodal distribution, $\max_\lambda(\mathbf{E}_2(\lambda) - \mathbf{E}_1(\lambda)) = (1/2)\inf_{G:G\text{ unimodal}}\|F - G\|$. It measures the minimal amount of probability mass that has to be moved in order to convert F into a unimodal distribution (Figure 5). Thus it can be regarded as a measure of strength attributed to the second largest peak of F. An analogous interpretation of $\max_\lambda(\mathbf{E}_3(\lambda) - \mathbf{E}_2(\lambda))$, however, only yields that $\max_\lambda(\mathbf{E}_3(\lambda) - \mathbf{E}_2(\lambda)) \leq (1/2)\inf_{G:G\text{ bimodal}}\|F - G\|$.

On any fixed set the empirical fluctuation of $H_{n,\lambda}$ is that of the empirical process, since $H_{n,\lambda} = H_\lambda + (F_n - F)$. But since we do not know the λ-clusters a priori, we must take into account the random component from the empirical λ-clusters as well. In Section 5 we show that asymptotic normality holds for $\mathbf{E}_{n,M}(\lambda)$, provided that the density f has at most M modes and no flat parts [i.e., $F[f(x) = \lambda] = 0$ $(\lambda \geq 0)$]. The convergence to a rescaled Brownian bridge B (cf. Th. 1, Sec. 5) leads to asymptotic confidence bands for $\lambda \to \mathbf{E}(\lambda)$: for $\Lambda \geq \sup f(x)$

$$\Pr[\sqrt{n}\max_{0\leq\lambda\leq\Lambda}|\mathbf{E}_{n,M}(\lambda) - \mathbf{E}(\lambda)| \leq \kappa] \to \Pr[\max_{0\leq t\leq 1}|B(t)| \leq \kappa].$$

The condition excluding flat parts cannot be omitted without destroying asymptotic normality. As an example let F

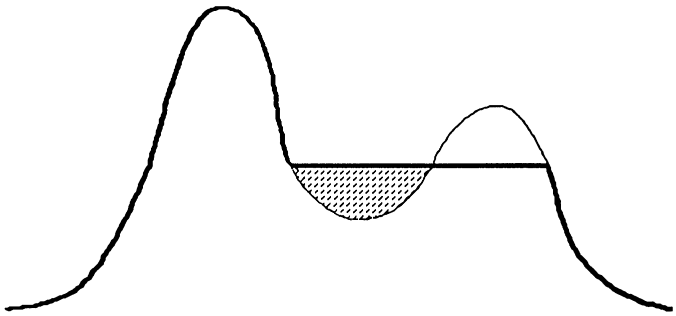


Figure 5. Maximal Excess Mass Difference. The maximal excess mass difference measures the minimal amount of mass to be moved to convert a bimodal distribution into a unimodal one.

be the uniform distribution on [0, 1]. Here $\sqrt{n}(\mathbf{E}_{n,M}(1) - \mathbf{E}(1))$ converges in law to a limit, which is not normal [concentrated on the positive half-line, since $\mathbf{E}(1) = 0$]. It is the distribution of $\max_{s<t}(B(t) - B(s))$ for a standard Brownian bridge B. The exceptional role played by flat parts of a distribution has repeatedly been observed in the contexts of cluster analysis and density estimation (see, for example, Groeneboom 1985).

3. DATA ANALYSIS USING EXCESS MASS ESTIMATES IN ONE DIMENSION

The empirical λ-clusters $C_{n,1}(\lambda), \ldots, C_{n,M}(\lambda)$ show the location of mass concentration and thus can be used for data analysis. The plot of $C_{n,1}(\lambda), \ldots, C_{n,M}(\lambda)$ against λ, more precisely the subset $\mathcal{S}_n = \{(x, \lambda) : x \in \cup_{j=1}^M C_{n,j}(\lambda)\}$ of the plane, will be called the *silhouette*. An example is given in Figure 6.

If the density f has at most M modes, the symmetric differences of $\mathcal{S}_n$ and $\mathcal{S} = \{(x, \lambda) : f(x) \geq \lambda\}$, restricted to $0 \leq \lambda \leq \Lambda$ for finite Λ, tends to 0 in $F \otimes$ Leb measure almost surely as $n \to \infty$. This is a simple consequence of Proposition 2 of Section 5. Another consequence is the approximate monotonicity of the λ-sections of $\mathcal{S}_n$: The F measure of the difference set $\cup_{j=1}^M C_{n,j}(\lambda')\setminus\cup_{j=1}^M C_{n,j}(\lambda)$ tends to 0 almost surely as $n \to \infty$ $(\lambda < \lambda')$. Moreover, the following sharper result holds (proved in App. A).

Let f be a unimodal density. For each $\delta, \epsilon > 0$ there exists n_0 such that with probability at least $1 - \delta$, the fol-

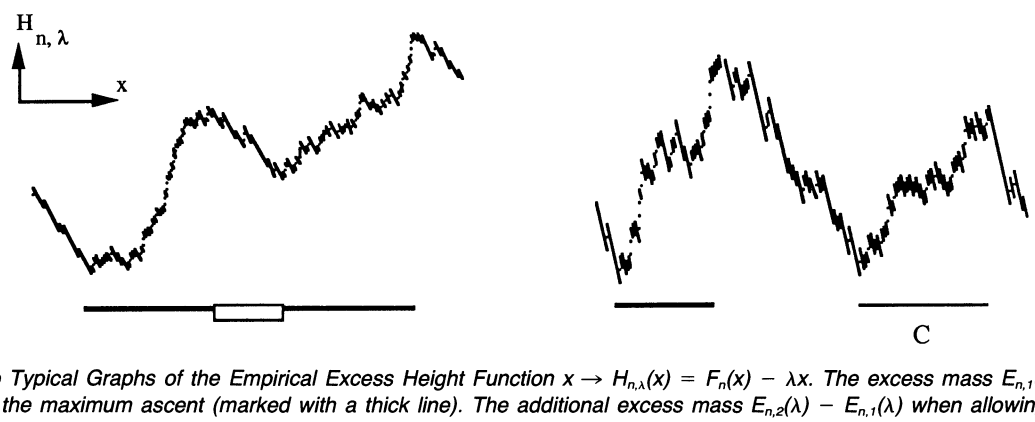


Figure 4. Two Typical Graphs of the Empirical Excess Height Function $x \to H_{n,\lambda}(x) = F_n(x) - \lambda x$. The excess mass $E_{n,1}(\lambda)$ is picked up on the interval with the maximum ascent (marked with a thick line). The additional excess mass $E_{n,2}(\lambda) - E_{n,1}(\lambda)$ when allowing for two modes is gained by either discarding a subinterval (marked by a box, left) or by adding a disjoint interval C (right).

lowing inclusion holds: $C_{n,1}(\lambda') \subset C_{n,1}(\lambda)$ for all $0 \leq \lambda < \lambda' \leq \max f(x) - \epsilon$ and $n \geq n_0$.

A similar statement applies to unimodal parts of a distribution. While monotonicity dominates the picture for values of λ bounded away from $\sup f(x)$, for larger λ the silhouette picture eventually gets blurred by those $C_{n,j}(\lambda)$ carrying only minute excess mass. The picture can be cleared by suppressing the presentation of intervals if their inclusion would result in a negligible change $\mathbf{E}_{n,M+1}(\lambda) - \mathbf{E}_{n,M}(\lambda)$, that is, if this is smaller than some constant μ. One can vary μ by trial and error, watching the picture with changing sensitivity of the procedure. Increasing the cutoff parameter μ yields a picture that concentrates on the main part of the modes (Figure 7).

The silhouette should be complemented by a plot of the excess mass functions $\lambda \to \mathbf{E}_{n,M}(\lambda)$ $(M = 1, 2, \ldots)$ on the same λ-scale. For a bimodal distribution the maximal difference $\max_\lambda(\mathbf{E}_{n,2}(\lambda) - \mathbf{E}_{n,1}(\lambda))$ can be interpreted as the estimated probability mass that distinguishes F from a unimodal distribution, or as the relative number of cases not fitting the hypothesis of unimodality. For a trimodal distribution note, however, that $\max_\lambda(\mathbf{E}_{n,3}(\lambda) - \mathbf{E}_{n,2}(\lambda))$ in general underestimates the probability mass not fitting the hypothesis of bimodality (cf. the remark in Sec. 2).

We illustrate the use of excess mass estimates by the example of the chondrite data (Ahrens 1965) that were introduced as a guinea pig for bump hunters by Good and Gaskins (1980). The data give the percentages of silica in $n = 22$ chondrite meteors. The silhouette is shown in Figure 6, the excess mass estimates in Figure 8. The parameters used were set for $M = 1, 2, 3$. Hence for each λ at most 3 intervals are displayed. The maximal excess mass differences have the following interpretation: Assuming bimodality instead of unimodality recovers an additional mass estimated at 16%. Assuming trimodality instead of bimodality allows explanation of (at least) an additional 14%.

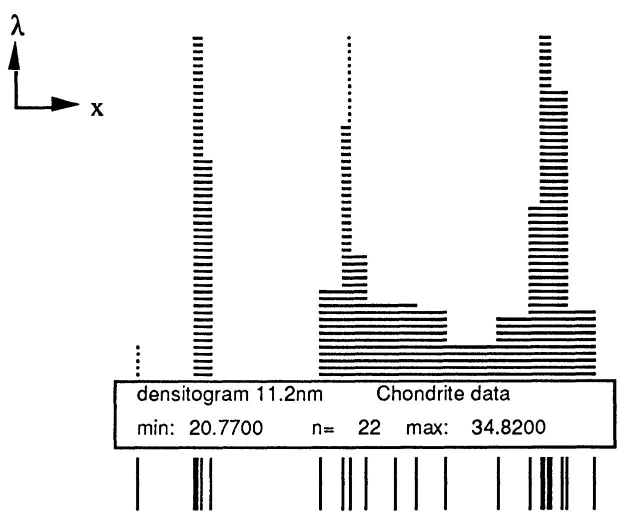


Figure 6. Unfiltered Silhouette of the Chondrite Data. The bottom panel of the graph contains bars marking the observed data points. The silhouette is given in the upper panel: The intervals $C_{n,j}(\lambda)$, the empirical λ-clusters, are presented as horizontal lines $\{(x, \lambda): x \in C_{n,j}(\lambda)\}$ at level λ for a grid of λ-values.

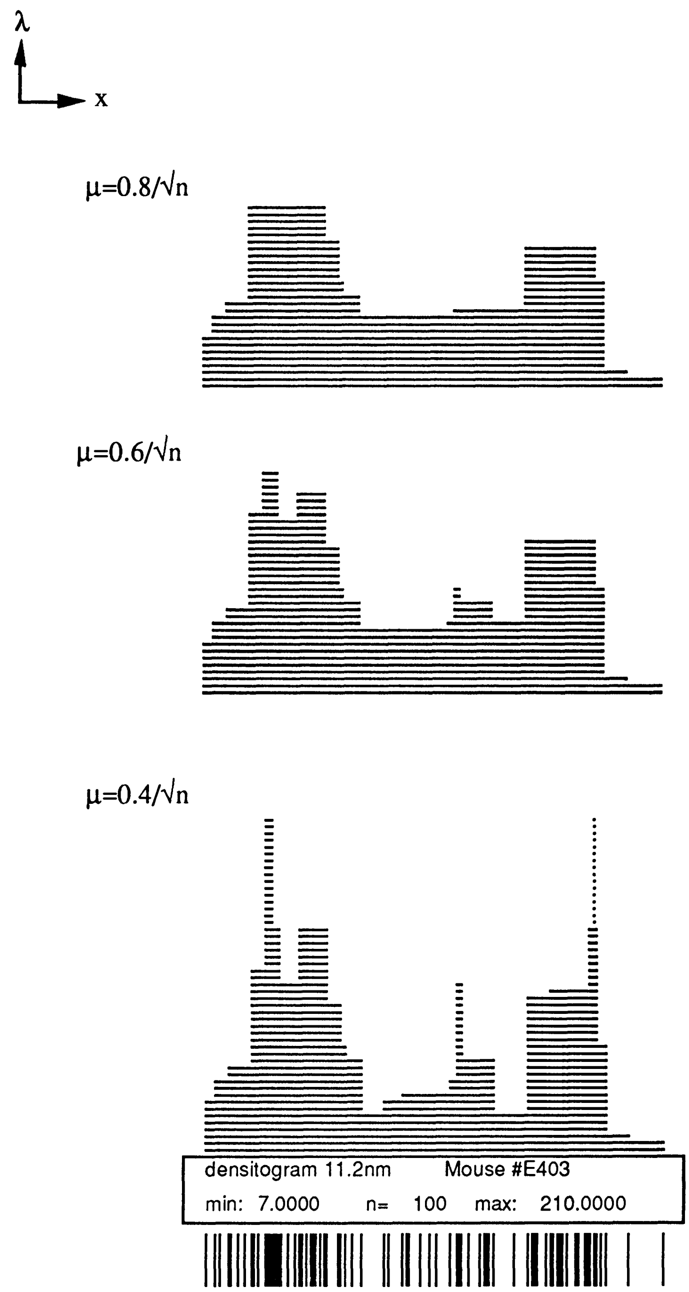


Figure 7. Watching the Silhouette of a Set of Random Input Data With Varying Sensitivity (sample size $n = 100$, cut-off parameter $\mu = .8/\sqrt{n}$, $\mu = .6/\sqrt{n}$, $\mu = .4/\sqrt{n}$).

4. TESTS FOR MULTIMODALITY BASED ON EXCESS MASS ESTIMATES

The excess difference test of unimodality versus M-modality rejects when $\Delta_{n,M} = \max_\lambda D_{n,M}(\lambda)$ is too large. For $M = 2$ the test statistic $\Delta_{n,2}$ is equal to twice Hartigan's dip (cf. Hartigan and Hartigan 1985). This can be shown by a refinement of Hartigan's "string argument." A test statistic similar to $\Delta_{n,2}$, but not taking into account the splitting possibility of (3), has been suggested independently by Hartigan (1987) in the context of density contour estimation with two-dimensional data.

The determination of critical values encounters the difficulty that the asymptotic distribution of $\Delta_{n,M} = \max_\lambda D_{n,M}(\lambda)$ is not known. Moreover, Theorem 2 (Sec. 5) implies that the large-sample distribution of $\Delta_{n,M}$ depends heavily on differential properties of the underlying unimodal distributions (in particular, the second derivative of the density f at the mode). There is even a change in rate of the test statistic $\Delta_{n,M}$ from $O_P(n^{-3/5}(\log n)^{3/5})$ for a unimodal distribution fulfilling natural regularity conditions (which in particular exclude flat parts) to $O_P(n^{-1/2})$ for a uniform distribution U (more generally, for every distribution containing flat parts; cf. the remark after Th. 2). In this situation we use finite-sample bounds (in the sense of stochastic ordering) for $\Delta_{n,M}$ to determine critical values.

The Oscillation Bound. A general bound in the case of unimodality is provided by the following result [cf. (6)]:

$$\Pr[\Delta_{n,2} \geq \kappa] < \Pr[\max_C |(U_n - U)(C)| \geq \kappa], \qquad (4)$$

where U is the standard uniform distribution and U_n is the empirical distribution of a sample drawn from it. Therefore, if $\Delta_{n,2}$ is used with critical value κ_α chosen to make the right-hand side equal to α, the probability of false rejection will be smaller than or equal to α. Asymptotically, as $n \to \infty$, the probability on the right-hand side of (4) is well known from the distribution of the range of a standard Brownian bridge [as given in Kuiper (1960), together with an $n^{-1/2}$ correction term]. Unfortunately the bound (4) appears to be very conservative as is demonstrated by simulation studies (cf. Figure 9).

Using Critical Values from the Uniform Distribution. From a more pragmatic point of view, one uses the standard uniform distribution $U(0, 1)$ to calculate critical values κ_α^* for $\Delta_{n,2}$ from $\mathscr{L}(\Delta_{n,2} \mid U(0, 1))$ (see Tab. 1). We know that $\Pr[\Delta_{n,2} \geq \kappa_\alpha^* \mid F] \leq \alpha$ does not hold for general unimodal distributions F; in fact, this inequality may be violated for small n (see Hartigan and Hartigan 1985). However, such an inequality is suggested to hold for large n by our theoretical results (i.e., the difference in asymptotic rates revealed by Th. 2 and the remark following it). It is also supported by extensive simulations for moderate values of n, examples of which are illustrated by Figures 9 and 10.

For the chondrite data, the excess difference statistic $\max_\lambda(\mathbf{E}_{n,2}(\lambda) - \mathbf{E}_{n,1}(\lambda))$ takes the value .161, which corresponds to a significance probability of .226 (based on 10,000 simulation runs). Using the test statistic $\max_\lambda(\mathbf{E}_{n,3}(\lambda) - \mathbf{E}_{n,1}(\lambda))$, however, would give a value of .291, corresponding to a significance probability of .072. Taking into account additional modes thus may increase the power.

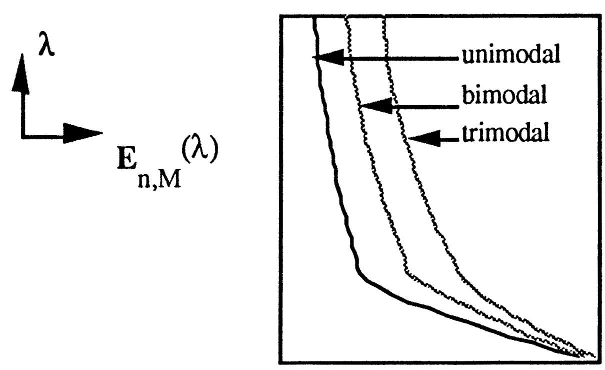


Figure 8. Excess Mass Function for the Chondrite Data. The values of the test statistics are $\max_\lambda(E_{n,2}(\lambda) - E_{n,1}(\lambda)) = .1614$, $\max_\lambda(E_{n,3}(\lambda) - E_{n,2}(\lambda)) = .1367$, and $\max_\lambda(E_{n,3}(\lambda) - E_{n,1}(\lambda)) = .2912$.

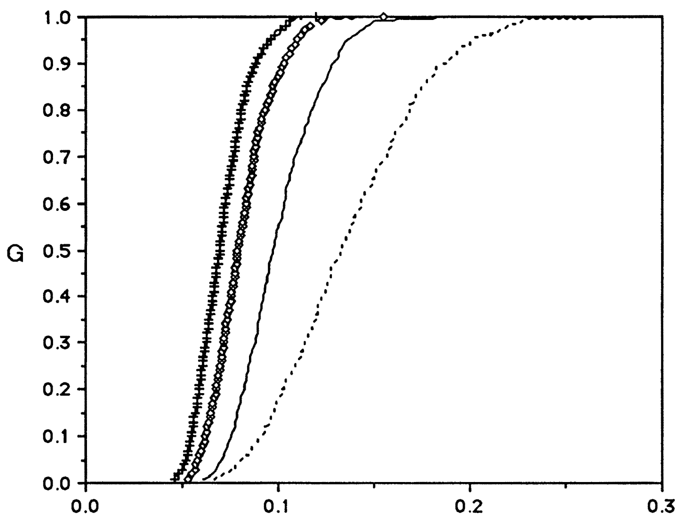


Figure 9. Simulated Distribution Function G of the Test Statistic $\Delta_{n,2}$ for Bimodality Under Uniform, Gaussian and Cauchy Distribution, and Simulated Oscillation Bound (denoted by MaxVar(Uniform): maximum absolute variation for uniform distribution). Sample size n = 50. Under each condition, 1,000 runs were done using the SANE random generator of the MPW Pascal Library on a Macintosh II. MaxVar(Unif) is represented by a dashed line, the uniform distribution by a solid line, the Gaussian by the diamonds, and the Cauchy by the pluses.

For testing unimodality against bimodality, simulations were performed with samples taken from a uniform distribution, a standard Gaussian distribution, and a test distribution putting uniform mass on the intervals [0, .25], (.25, .75), [.75, 1.00] in the proportions 3:2:3 (as in Hartigan and Hartigan 1985). The simulations clearly show that the excess-difference test discriminates between the unimodal distributions and Hartigan's alternative, with the uniform giving a conservative bound on the hypothesis (Figure 10).

A test for bimodality versus higher order multimodality can be based on the statistic $\max_\lambda(\mathbf{E}_{n,3}(\lambda) - \mathbf{E}_{n,2}(\lambda))$, with critical value κ_α derived from (4).

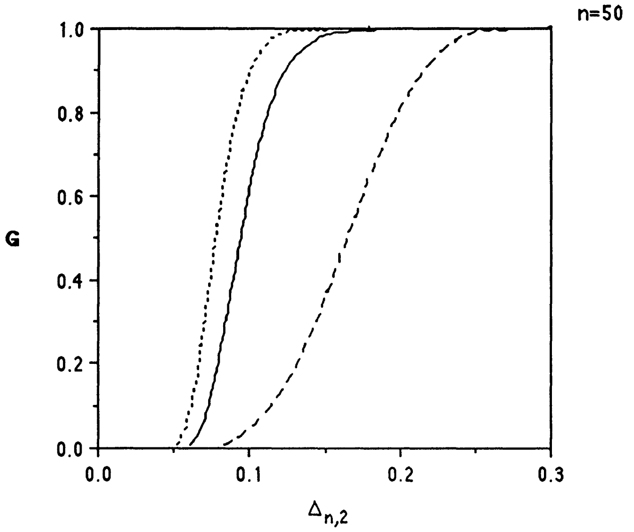
5. ASYMPTOTICS OF EXCESS MASS ESTIMATES

In this section we study the large-sample behavior of the excess mass estimators $\mathbf{E}_{n,M}(\lambda)$. Proofs are given in Appendix A. Since $H_{n,\lambda}(x)$ converges to $H_\lambda(x)$ uniformly in

Table 1. Percentage Points of the Maximal Excess Mass Difference Bi-Unimodal in Uniform Samples

p	n = 10	n = 50	n = 75	n = 100
.01	.12092	.06227	.05182	.04478
.05	.14344	.06984	.05807	.05067
.10	.15580	.07483	.06216	.05422
.25	.17606	.08428	.07004	.06111
.50	.19552	.09713	.08072	.07042
.75	.22810	.11272	.09354	.08188
.90	.26009	.12903	.10721	.09382
.95	.27871	.13971	.11622	.10162
.99	.31858	.16085	.13452	.11760

NOTE: Based on 30,000 simulations using Apple's SANE generator.



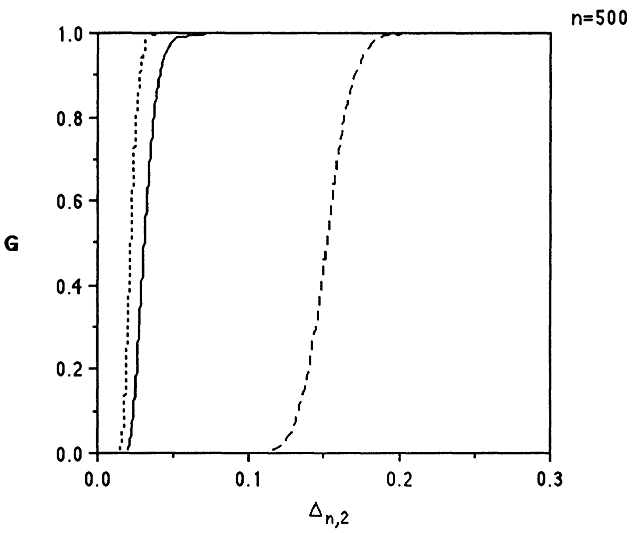


Figure 10. Simulated Distribution Function of the Test Statistic $\Delta_{n,2}$ for Bimodality. Under each condition, 10,000 runs were done using the random generator of the NAg Pascal Library on a Macintosh II. The Gaussian distribution is represented by the dashed line, the uniform by the solid line, and the test mixture by the broken line.

$x \in \mathbf{R}^1$ almost surely, we have

Proposition 1. For given $\lambda > 0$ let the function $x \to H_\lambda(x)$ have exactly $2m_\lambda$ local extrema (i.e., m_λ local minima and m_λ local maxima). Then for all $M \geq m_\lambda$,

$$\lim_{n\to\infty} \mathbf{E}_{n,M}(\lambda) = \mathbf{E}(\lambda) \text{ almost surely.}$$

If $M \geq m_\lambda$ for all $\lambda > 0$, then this convergence is uniform in $\lambda > 0$.

For convergence of the empirical λ-clusters to the true λ-clusters, some condition excluding flat parts of the distribution will be necessary. We consider the following condition:

For all $\lambda > 0$ the function $x \to H_\lambda(x)$ has exactly

$$2m_\lambda < +\infty \text{ local extrema.} \quad (5)$$

In particular, under (5), there are no intervals of positive F mass on which f is constant, that is, $F[f(x) = \lambda] = 0$ ($\lambda \geq 0$) (note that local maxima (minima) of the function $x \to H_\lambda(x)$ correspond to down (up) crossings of level λ by the density f); hence, also, H_λ does not have flat parts of positive F mass. We denote by $\Gamma(\lambda)$ the union of all λ-clusters and by $\Gamma_{M,n}(\lambda)$ the union of all M empirical λ-clusters. The following proposition is intuitive and easy to prove.

Proposition 2. Let condition (5) be satisfied. If $M \geq m_\lambda$ for all $\lambda > 0$ then $\sup_{\lambda\geq 0} F(\Gamma_{M,n}(\lambda)\Delta\Gamma(\lambda)) \to 0$ almost surely as $n \to \infty$.

The following theorem identifies the limit of the excess mass process $\lambda \to \mathbf{E}_{n,M}(\lambda)$.

Theorem 1. Let condition (5) be satisfied. Let $M \geq m_\lambda$ for all $\lambda > 0$. Then for every Λ, $0 < \Lambda \leq +\infty$, as $n \to \infty$, the excess mass process $\lambda \to \sqrt{n}(\mathbf{E}_{n,M}(\lambda) - \mathbf{E}(\lambda))$ converges in distribution on the space of continuous functions on $[0, \Lambda]$ to the process $\lambda \to B(a_\lambda)$, where B is a standard Brownian bridge and $a_\lambda = F[f(x) \geq \lambda]$. If $N \geq M$ then $\max_{0\leq\lambda\leq\Lambda} \sqrt{n}(\mathbf{E}_{n,N}(\lambda) - \mathbf{E}_{n,M}(\lambda)) \to 0$ in probability.

More detailed information about the rate of $\Delta_{n,M} \equiv \max_{0\leq\lambda\leq\Lambda} (\mathbf{E}_{n,M}(\lambda) - \mathbf{E}_{n,1}(\lambda))$ in the unimodal case is given by Theorem 2.

Theorem 2. Let f be a density, unimodal in the strong sense that there is an x_0 such that $f'(x) = 0$ holds only for $f(x) = 0$ or $x = x_0$. Let the following conditions be satisfied:

(i) f has a continuous derivative f', ultimately monotone in the tails;

(ii) f has a bounded second derivative in a neighborhood of x_0 such that $f''(x_0) < 0$;

(iii) f has a third derivative bounded in a neighborhood of x_0.

Then

(1) for $\lambda = f(x_0)$ the quantity $\mathbf{D}_{n,M}(\lambda)$ is of the exact order $O_P(n^{-3/5})$;

(2) for each $\epsilon > 0$ the quantity $\max_{\lambda\leq f(x_0)-\epsilon} \mathbf{D}_{n,M}(\lambda)$ is of the order $O_P(n^{-2/3} \log^{2/3} n)$;

(3) $\Delta_{n,M} \equiv \max_\lambda \mathbf{D}_{n,M}(\lambda)$ is of the order $O_P(n^{-3/5} \log^{3/5} n)$.

The proof shows that under unimodality the higher order empirical λ-clusters tend to concentrate near the endpoints of $C_1(\lambda)$ $[0 < \lambda < f(x_0)]$. In contrast to the result of Theorem 2, in the case of the uniform distribution $F = U(0, 1)$ the quantity $\max_{\lambda\leq\Lambda} \mathbf{D}_{n,2}(\lambda)$ is of the order $O_P(n^{-1/2})$. This follows already from (4). For $\Lambda \geq 1$ this order is exact. For $\Lambda > 1$ the quantity $\sqrt{n} \max_{\lambda\leq\Lambda} \mathbf{D}_{n,2}(\lambda)$ converges in distribution to $\max_{-\infty<\lambda<+\infty}(\mathbf{E}_{\infty,2}(\lambda) - \mathbf{E}_{\infty,1}(\lambda))$. Here $\mathbf{E}_{\infty,j}(\lambda) = \sup_C(B(C) - \lambda\cdot\text{Leb}(C))$, where for $j = 1$, C is an interval and for $j = 2$, C is a union of two disjoint intervals ($C \subset [0, 1]$); B is a standard Brownian bridge indexed by sets. This follows from $\sqrt{n}H_{n,\lambda}(C) = B_n(C) + \sqrt{n}(1 - \lambda)\text{Leb}(C)$, where B_n converges in law to B.

The proof of Theorem 2 makes use of a finite-sample result that is of independent interest. According to the splitting property, $\mathbf{D}_{n,2}(\lambda) \leq H_{n,\lambda}(C_{n,j}(\lambda))$ for $j = 1, 2$, and $\mathbf{D}_{n,2}(\lambda) \leq -H_{n,\lambda}(C_n)$, where C_n is the open interval in between $C_{n,1}(\lambda)$ and $C_{n,2}(\lambda)$. Now for any interval C, either $C_n \subset C$ or one of the $C_{n,j}(\lambda) \subset C^c$. Specializing $C = C_1(\lambda)$ yields

Proposition 3.

$$\mathbf{D}_{n,2}(\lambda) \le \max\{H_{n,\lambda}(C)\colon C \subset C_1(\lambda)^c\} \vee \max\{-H_{n,\lambda}(C)\colon C \subset C_1(\lambda)\}.$$

For a unimodal distribution F this leads to a stochastic bound for $\Delta_{n,2}$: Since

$$H_{n,\lambda}(C) = H_\lambda(C) + (F_n - F)(C) \le (F_n - F)(C) \quad \text{if } C \subset C_1(\lambda)^c,$$

$$-H_{n,\lambda}(C) = -H_\lambda(C) - (F_n - F)(C) \le -(F_n - F)(C) \quad \text{if } C \subset C_1(\lambda),$$

one has

$$\mathbf{D}_{n,2}(\lambda) \le \max\{(F_n - F)(C)\colon C \subset C_1(\lambda)^c\} \vee \max\{-(F_n - F)(C)\colon C \subset C_1(\lambda)\}.$$

Hence

$$\Delta_{n,2} \le \max_C |(F_n - F)(C)|. \tag{6}$$

Since the distribution of the right-hand side does not depend on F, this also proves (4). By the same arguments, for distributions with m modes one obtains the following generalization of (4):

$$\Pr[\max_\lambda(\mathbf{E}_{n,M+1}(\lambda) - \mathbf{E}_{n,M}(\lambda)) \ge \kappa] \le \Pr[\max_C|(U_n - U)(C)| \ge \kappa].$$

APPENDIX A: PROOFS

Proof of Theorem 1. Let $\mathbf{E}^*_{n,M}(\lambda) = \sum_{j=1}^M H_{n,\lambda}(C_j(\lambda))$. We show that $\max_{0\le\lambda\le\Lambda}\sqrt{n}(\mathbf{E}^*_{n,M}(\lambda) - \mathbf{E}_{n,M}(\lambda))$ converges to 0 in probability as $n \to \infty$. Then $\mathbf{E}_{n,M}$ and $\mathbf{E}^*_{n,M}$ have the same limit distribution whose form can be derived from $\mathbf{E}^*_{n,M}(\lambda) - \mathbf{E}(\lambda) = \sum_{j=1}^M (F_n(C_j(\lambda)) - F(C_j(\lambda))) = (F_n - F)\,[f(x) \ge \lambda]$. Using the definitions of $C_j(\lambda)$ and $C_{n,j}(\lambda)$ and remembering that $H_{n,\lambda} - H_\lambda = F_n - F$ one arrives at the inequality

$$\begin{aligned} 0 &\ge \mathbf{E}^*_{n,M}(\lambda) - \mathbf{E}_{n,M}(\lambda) \\ &= \sum_{j=1}^M ((F_n - F)(C_j(\lambda)) - (F_n - F)(C_{n,j}(\lambda))) \\ &\quad + \sum_{j=1}^M (H_\lambda(C_j(\lambda)) - H_\lambda(C_{n,j}(\lambda))) \\ &= (F_n - F)(\Gamma_M(\lambda)) - (F_n - F)(\Gamma_{M,n}(\lambda)) \\ &\quad + H_\lambda(\Gamma_M(\lambda)) - H_\lambda(\Gamma_{M,n}(\lambda)). \end{aligned}$$

Here the second term on the right is nonnegative such that $\mathbf{E}^*_{n,M} - \mathbf{E}_{n,M}$ inherits the order of the first term. Now $B_n(\Gamma) = \sqrt{n} \times (F_n - F)(\Gamma)$ is a process indexed by sets Γ that are unions of at most M intervals and converges in law to a Brownian bridge process B (expectation $B(\Gamma_1)B(\Gamma_2) = F(\Gamma_1 \cap \Gamma_2) - F(\Gamma_1)F(\Gamma_2)$). Hence it follows from the asymptotic continuity of B_n and Proposition 2 that the first term is of the uniform order $o_P(n^{-1/2})$, as asserted. Since $\lambda \to S_n(\lambda) = \sqrt{n}(\mathbf{E}_{n,M}(\lambda) - \mathbf{E}(\lambda))$ and $\lambda \to T_n(\lambda) = \sqrt{n}(\mathbf{E}_{n,N}(\lambda) - \mathbf{E}(\lambda))$ converge to the same limit, and S_n being $\le T_n$, it follows that $\max_{0\le\lambda\le\Lambda}(T_n(\lambda) - S_n(\lambda)) \to 0$ in probability.

Proof of Theorem 2. We use a Hungarian embedding and the scaling properties of Brownian motion in much the same way as Groeneboom (1985). For an $\epsilon > 0$ to be determined later, the full proof treats the following four cases

$$\lambda \le \epsilon; \quad \epsilon \le \lambda \le f(x_0) - \epsilon; \quad f(x_0) - \epsilon \le \lambda \le f(x_0); \quad \lambda \ge f(x_0).$$

The main arguments being essentially similar, we only outline the proof in the most interesting case $f(x_0) - \epsilon \le \lambda \le f(x_0)$. It suffices to prove the assertion for $\mathbf{D}_{n,2}(\lambda)$, since $\max_\lambda \mathbf{D}_{n,2}(\lambda) \le \max_\lambda \mathbf{D}_{n,M}(\lambda) \le (M - 1)\max_\lambda \mathbf{D}_{n,2}(\lambda)$. We are going to show that

$$\max_{f(x_0)-\epsilon\le\lambda\le f(x_0)} \mathbf{D}_{n,2}(\lambda) = O_P(n^{-3/5}\log^{3/5} n).$$

After Proposition 3, the term $\mathbf{D}_{n,2}(\lambda)$ is bounded by one of the two similar quantities

$$\max\{H_{n,\lambda}(C)\colon C \subset C_1(\lambda)^c\}, \qquad \max\{-H_{n,\lambda}(C)\colon C \subset C_1(\lambda)\}.$$

We just show that the first term is of the stated order uniformly in $f(x_0) - \epsilon \le \lambda \le f(x_0)$. Let $\mathcal{U}$ be a convex open neighborhood of x_0, in which (ii) and (iii) are satisfied, such that $f''(x)$ is bounded away from 0 ($x \in \mathcal{U}$).

Let $C_1(\lambda) = [a_\lambda, b_\lambda]$. Then for small $\epsilon > 0$ we can assume $b_\lambda \in \mathcal{U}$. Without loss of generality we consider the contribution $H_{n,\lambda}(C)$ for intervals $C \equiv (x', x''] \subset (b_\lambda, +\infty)$. For $b_\lambda \le x \in \mathcal{U}$ we have

$$H'_\lambda(x) \le f''((b_\lambda)/2!)\,(x - b_\lambda)^2 + o((x - b_\lambda)^2),$$

where the order of the remainder term is uniform in b_λ because of (iii). Hence

$$H'_\lambda(x) \le d/dx(H_\lambda(b_\lambda) - \gamma(x - b_\lambda)^3)$$

for some constant $\gamma > 0$. This yields

$$\begin{aligned} H_\lambda(x'') - H_\lambda(x') &\le -\gamma[(x'' - b_\lambda)^3 - (x' - b_\lambda)^3] \\ &\le -(\gamma/4)(x'' - x')^3 \end{aligned} \tag{A.1}$$

for $b_\lambda \le x' \le x'' \in \mathcal{U}$. For $b_\lambda \le x' \le x''$ both not in $\mathcal{U}$ one obtains a similar inequality with linear terms replacing the cubic terms and then completes the argument along the following lines.

In order to describe the behavior of the empirical quantities $H_{n,\lambda}(x'') - H_{n,\lambda}(x')$ for $x', x'' \in \mathcal{U}$ ($x' \le x''$), we utilize the embedding of F_n in a standard Brownian bridge B (cf. Komlós, Major, and Tusnády 1975):

$$\Pr[d_n \equiv \sup_t|(F_n(t) - F(t)) - n^{-1/2}B(F(t))| \ge n^{-1+\beta}] \to 0$$

as $n \to \infty$ for every $\beta > 0$.

Hence

$$\begin{aligned} H_{n,\lambda}(x'') - H_{n,\lambda}(x') &= n^{-1/2}(B(F(x'')) - B(F(x'))) \\ &\quad + H_\lambda(x'') - H_\lambda(x') + e_n, \end{aligned}$$

where e_n is a random variable bounded by d_n for all x'', x'.

Now one applies Lévy's modulus of continuity for B

$$\sup|B(v) - B(u)| \cdot (|v - u| \log|v - u|^{-1})^{-1/2} = O_P(1)$$

together with (A.1) to establish the bound

$$\begin{aligned} \sup_{b_\lambda\le x'\le x''} (H_{n,\lambda}(x'') - H_{n,\lambda}(x')) &\le \sup n^{-1/2}(F(x'') - F(x'))^{1/2} \\ &\quad \times |\log(F(x'') - F(x'))|^{1/2} O_P(1) - (\gamma/4)(x'' - x')^3 + e_n \\ &\le \sup n^{-1/2}(F(x'') - F(x'))^{1/2}|\log(F(x'') - F(x'))|^{1/2} O_P(1) \\ &\quad - (\gamma/4) f(x_0)^{-3}(F(x'') - F(x'))^3 + e_n. \end{aligned}$$

This quantity has the order of the maximum of $n^{-1/2}\xi^{1/2} \times (-\log \xi)^{1/2} - \xi^3$, that is, $O_P(n^{-3/5}\log^{3/5} n)$.

The monotonicity stated in Section 3 is an immediate consequence of the following.

Lemma. Let the following assumptions (a) and (b) be satisfied:

(a) for each $0 \le \lambda \le \Lambda$ let $H_{n,\lambda}(C_{n,1}(\lambda)) > H_{n,\lambda}(C)$ for all intervals C disjoint to $C_{n,1}(\lambda)$;

(b) all the difference quotients of order statistics $x_{(1)} < \cdots <$

$x_{(n)}: (x_{(j)} - x_{(i)})/(j - i)$, where $i < j$ are pairwise different numbers. Then $C_{n,1}(\lambda') \subset C_{n,1}(\lambda)$ for all $0 \leq \lambda < \lambda' \leq \Lambda$.

Note that the intervals $C_{n,1}(\lambda)$ may not be uniquely determined; the lemma holds no matter how the maximizing intervals are chosen.

Proof of the Lemma. $C_{n,1}(\lambda) = [x_{(i)}, x_{(j)}]$ for some $i < j$ maximizing

$$H_{n,\lambda}([x_{(i)}, x_{(j)}]) = n^{-1}(j - i + 1) - \lambda(x_{(j)} - x_{(i)})$$

or, equivalently, maximizing

$$n^{-1}(j - i) - \lambda(x_{(j)} - x_{(i)}) = H_{n,\lambda}(x_{(j)}) - H_{n,\lambda}(x_{(i)}). \quad \text{(A.2)}$$

There may be several such pairs $i < j$. Any two pairs $i' < j'$, $i'' < j''$ both maximizing (A.2) must have one index in common: Either $i' = i''$ or $j' = j''$. The reason is that they overlap after assumption (a): Without restriction of generality it suffices to consider the case $i' \leq i'' \leq j' \leq j''$. If $i' < i''$ and $j' < j''$ then

$$H_{n,\lambda}(x_{(i')}) = H_{n,\lambda}(x_{(i'')}) \quad \text{and} \quad H_{n,\lambda}(x_{(j')}) = H_{n,\lambda}(x_{(j'')}),$$

which would contradict assumption (b). Therefore $i' = i''$ or $j' = j''$. By the same argument it follows that there can be at most two different intervals maximizing (A.2). Now consider $\lambda < \lambda'$. Let $i < j$, $i' < j'$ be maximizing pairs, respectively:

$$n^{-1}(j - i) - \lambda(x_{(j)} - x_{(i)}) = \max,$$

$$n^{-1}(j' - i') - \lambda'(x_{(j')} - x_{(i')}) = \max.$$

This gives the following pair of inequalities:

$$n^{-1}[j - j' - (i - i')] - \lambda[x_{(j)} - x_{(j')} - (x_{(i)} - x_{(i')})] \geq 0, \quad \text{(A.3)}$$

$$n^{-1}[j - j' - (i - i')] - \lambda'[x_{(j)} - x_{(j')} - (x_{(i)} - x_{(i')})] \leq 0. \quad \text{(A.4)}$$

Subtracting (A.4) from (A.3), one gets $x_{(j)} - x_{(j')} - (x_{(i)} - x_{(i')}) \geq 0$, so that, by (A.3),

$$j - j' \geq i - i'. \quad \text{(A.5)}$$

Now the mapping $\lambda \to C_{n,1}(\lambda)$ is piecewise constant. Thus the monotonicity has to be shown at the jumps only. Let λ' be the position of a jump. There will be exactly two different maximizing pairs $i < j$ and $i' < j'$:

$$n^{-1}(j - i) - \lambda'(x_{(j)} - x_{(i)}) = \max,$$

$$n^{-1}(j' - i') - \lambda'(x_{(j')} - x_{(i')}) = \max.$$

One of them, say $i < j$, also maximizes $n^{-1}(j - i) - \lambda(x_{(j)} - x_{(i)})$ for $\lambda < \lambda'$ and λ sufficiently close to λ'. Then $i = i'$ or $j = j'$. Without restriction of generality, let $i = i'$. Using (A.5) one concludes $j' \leq j$, hence $[x_{(i')}, x_{(j')}] \subset [x_{(i)}, x_{(j)}]$.

APPENDIX B: ALGORITHM

An algorithm is given that calculates excess mass estimates $\mathbf{E}_{n,M}(\lambda)$ and finds empirical λ-clusters $C^M_{n,1}, \ldots, C^M_{n,M}$ $(M \leq n)$. The calculation is done recursively in M, using a direct search algorithm at every step. For $M = 1$, the search for ranks i, j maximizing $H_{n,\lambda}[x_{(i)}, x_{(j)}]$ produces an interval $C^1_{n,1} = [x_{(i)}, x_{(j)}]$, with $\mathbf{E}_{n,1}(\lambda) = H_{n,\lambda}(C^1_{n,1})$. For the general step from M to $M + 1$ the following relation is used [a generalization of the splitting property (3)]:

If $C^M_{n,1}, \ldots, C^M_{n,M}$ are disjoint intervals such that $\mathbf{E}_{n,M}(\lambda) = \Sigma_j H_{n,\lambda}(C^M_{n,j})$, then $\mathbf{E}_{n,M+1}(\lambda) = \Sigma_j H_{n,\lambda}(C^{M+1}_{n,j})$ for a set of disjoint intervals $C^{M+1}_{n,1}, \ldots, C^{M+1}_{n,M+1}$ which satisfies either

$$\{C^{M+1}_{n,1}, \ldots, C^{M+1}_{n,M+1}\} = \{C^M_{n,1}, \ldots, C^M_{n,M}\} \cup \{C\}$$

with $C \cap C^M_{n,j} = \emptyset$ for all j, or

$$\{C^{M+1}_{n,1}, \ldots, C^{M+1}_{n,M+1}\} = \{C^M_{n,1}, \ldots, C^M_{n,i-1}, C^M_{n,i+1}, \ldots, C^M_{n,M}\} \cup \{C', C''\},$$

where C', C'' are disjoint with convex hull $C^M_{n,i}$ for some i.

Hence for the step from M to $M + 1$ an interval C with $C \cap C^M_{n,j} = \emptyset$ for all j has to be found which maximizes $H_{n,\lambda}$ and an interval $\tilde{C}$ with $\tilde{C} \subset C^M_{n,i}$ for some i which maximizes $-H_{n,\lambda}$. If $H_{n,\lambda}(C) > |H_{n,\lambda}(\tilde{C})|$ the new interval C is added, else $C^M_{n,i}$ is replaced by the two intervals forming the complement $C^M_{n,i} \setminus \tilde{C}$. Since $H_{n,\lambda}(C) < 0$ for any open interval $C \neq \emptyset$ not containing data points, C is a closed and $\tilde{C}$ an open interval with data points as endpoints. C is found by exhaustive search with search ranges restricted to the intervals forming $(\cup C^M_{n,j})^c$ and $\tilde{C}$ by search restricted to the intervals $C^M_{n,i}$.

The direct search for ranks i and j maximizing $H_{n,\lambda}(C)$ for $C = [x_{(i)}, x_{(j)}]$ is reduced to a problem of complexity $O(n)$ by successively keeping the record minimum and ascent (hiker's record keeping method):

$$\max_{i,j} H_{n,\lambda}[x_{(i)}, x_{(j)}] = \max_j (H_{n,\lambda}(x_{(j)}) - \min_{i \leq j} H_{n,\lambda}(x_{(i)}\text{-})).$$

A complete PASCAL procedure performing this maximization within a specified search range is given. For a given range of ranks `par_from. . . par_to`, it returns the endpoint ranks of the interval $C = [x_{(\text{max\_from})}, x_{(\text{max\_to})}]$ maximizing the empirical excess mass $H_{n,\lambda}(C)$ over this range. The result $H_{n,\lambda}(C)$ is returned as `variation`. Tabulated values of $nH_{n,\lambda}(x_{(\cdot)})$ must be supplied in a global array `nHLambda`.

```
procedure localmax {finds an interval}
                   {of maximum or minimum variation}
  (count: integer; {total number of observations}
  par_from, par_to: integer; {search index range}
  var max_from: integer; {position of min when}
                         {variation is maximal}
  var max_to: integer; {position of max when}
                       {variation is maximal}
  var variation: real); {maximal excess mass}
                        {over interval}
var
  max_variation: real;
     {current rescaled estimate of max variation}
     {the internal algorithm operates on left open,}
     {right closed intervals}
{variation=max_variation/count+endpointcorrection}
  curval: real; {current value of function under test}
                {function under test is}
                {n_obs * H(.;n_obs, lambda)}
  minval: real; {min(i ≤ j) of function under test}
  minindex: integer;
  maxkrit: real; {precalculated critical value}
                 {to speed up}
                 {curval > maxkrit iff}
                 {curval - minval > max_variation}
  j: integer; {loop index}
begin
  max_variation := 0; {initial values}
  max_to := par_from;
  max_from := par_from;
  minindex := par_from;
  minval := nHLambda [par_from];
  maxkrit := minval; {define critical value}
                     {to speed up}
  for j := par_from + 1 to par_to do
```

```
  begin
    curval := nHLambda[j];
                           {this is the function to test}
    if curval < minval then {do we have a new minimum?}
    begin
      minval := curval; {note minimum value}
      minindex := j; {and position}
      maxkrit := max_variation + minval;
                                  {adapt critical value}
    end
    else if curval > maxkrit then
                                {new maximum variation?}
    begin
      max_variation := curval - minval;
                                  {note variation value}
      max_to := j; {note interval limits}
      max_from := minindex;
      maxkrit := curval; {adapt critical value}
    end;
  end; {loop}
  variation := (max_variation + 1) / count;
              {endpoint correction for closed interval}
        {(max_variation - 1) / count for open intervals}
end; {localmax}
```

A typical computation time of the complete algorithm is less than 6 seconds per 1,000 data points for $M \leq 3$ on a Macintosh II.

[*Received November 1988. Revised November 1990.*]

REFERENCES

Ahrens, L. H. (1965), "Observations on the Fe-Si-Mg Relationship in Chondrites," *Geochimica et Cosmochimica Acta*, 29, 801–806.

Donoho, D. L. (1988), "One-Sided Inference About Functionals of a Density," *The Annals of Statistics*, 16, 1390–1420.

Good, I. J., and Gaskins, R. A. (1980), "Density Estimation and Bump-Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data," *Journal of the American Statistical Association*, 75, 42–73.

Groeneboom, P. (1985), "Estimating a Monotone Density," in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (Vol. II), eds. L. LeCam and R. Olshen, Monterey, CA: Wadsworth, pp. 539–555.

Hartigan, J. A. (1975), *Clustering Algorithms*, New York: John Wiley.

——— (1977), "Distribution Problems in Clustering," in *Classification and Clustering*, ed. J. Van Ryzin, New York: Academic Press.

——— (1987), "Estimation of a Convex Density Contour in Two Dimensions," *Journal of the American Statistical Association*, 82, 267–270.

Hartigan, J. A., and Hartigan, P. M. (1985), "The DIP Test of Unimodality," *The Annals of Statistics*, 13, 70–84.

Komlós, J., Major, P., and Tusnády, G. (1975), "An Approximation of Partial Sums of Independent rv'-s, and the Sample df.I," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32, 111–131.

Kuiper, N. H. (1960), "Tests Concerning Random Points on a Circle," *Nederlandse Akademie van Wetenschappen: Proceedings A*, 63, 38–47.

Müller, D. W., and Sawitzki, G. (1987), "Using Excess Mass Estimates to Investigate the Modality of a Distribution," Preprint 398, January 1987, Universität Heidelberg, Sonderforschungsbereich 123 Stochastische Mathematische Modelle.

Silverman, B. W. (1981), "Using Kernel Density Estimates to Investigate Multimodality," *Journal of the Royal Statistical Society*, Ser. B, 43, 97–99.

Wong, M. A. (1985), "A Bootstrap Testing Procedure for Investigating the Number of Subpopulations," *Journal of Statistical Computation and Simulation*, 22, 99–112.