ELSEVIER

Statistics & Probability Letters 39 (1998) 245–254

STATISTICS & PROBABILITY LETTERS

On mode testing and empirical approximations to distributions

Ming-Yen Cheng *, Peter Hall

Centre for Mathematics and its Applications, Australian National University, Canberra, ACT0200, Australia

Received December 1996; received in revised form January 1998

Abstract

There are close connections between the theory of statistical inference under order restrictions and the theory of tests for unimodality. In particular, a result of Kiefer and Wolfowitz, on the error of convex approximations to empirical distribution functions, is basic to limit theory for the dip test for unimodality. We develop a version of Kiefer and Wolfowitz' result in the context of distributions that are strongly unimodal, and apply it and related limit theory to compare the powers, against local alternatives, of three different tests of unimodality. In this context it is shown that the dip, excess mass and bandwidth tests are all able to detect departures of size $n^{-3/5}$ (measured in terms of the distribution function) from the null hypothesis, where n denotes sample size; but are not able to detect departures of smaller order. Thus, they have similar powers. © 1998 Elsevier Science B.V. All rights reserved

AMS-classification: primary 62G05; secondary 62G09

Keywords: Bandwidth test; Concave majorant; Convex minorant; Density estimation; Dip test; Excess mass test; Local alternative hypothesis; Order constraints; Unimodal distribution

1. Introduction

There is a close relationship between the accuracy of empirical approximations to distributions and the performance of tests for modality. Sometimes this is expressed through confidence bands. For example, if no plausible unimodal density lies within a 95% confidence band for the density, then a test of the hypothesis that the density is unimodal would be rejected at the 5% level. See for example Hall and Titterington (1988). Alternatively, confidence bands could be constructed for the distribution function. For example, a distribution function corresponding to a density that has a unique mode at a point x_0 is convex to the left of x_0 and concave to the right, and so we could reject the null hypothesis of unimodality if no distribution with these properties were sufficiently close to the empirical distribution function. This is the basis of the dip test of Hartigan and Hartigan (1985); see also Hartigan (1985).

The problem of estimating a distribution under convexity or concavity assumptions is also of interest in the theory of statistical inference under order constraints. The literature there has its roots in work of Brunk (1956), van Eeden (1956) and Grenander (1956) on the greatest convex minorant of a distribution function,

* Corresponding author.

0167-7152/98/$19.00
PII S0167-7152(98)00058-3

and has been surveyed by Barlow et al. (1972). Hartigan and Hartigan's (1985) dip test statistic uses a "string" construction adapted from the idea of a minorant. Specifically, the dip test employs that distribution function $\widehat{U}$ which is nearest to $\widehat{F}$ in an L_∞ sense, subject to being convex to the left and concave to the right of some point x. Kiefer and Wolfowitz (1976) showed that if the true distribution function, F, is convex then the greatest convex minorant, $\widehat{C}$, of the empirical distribution function of a random sample drawn from F, is an order of magnitude closer to $\widehat{F}$ than it is to F. This result implies that first-order empirical process theory for $\widehat{C}$ is identical to that for $\widehat{F}$, and that fact was employed in Hartigan and Hartigan's (1985) development of limit theory for the dip test statistic.

Kiefer and Wolfowitz (1976) actually proved, under regularity conditions that imply convexity of F, a result that is marginally stronger than the following:

$$\sup_{-\infty<x<\infty} |\widehat{C} - \widehat{F}| = O_p\{n^{-2/3}(\log n)^{5/6}\}, \tag{1.1}$$

where n denotes sample size. It is clear from their work that (1.1) fails if the distribution F is unimodal, rather than convex, and if $\widehat{C}$ is replaced by $\widehat{U}$. We shall prove that in the latter case, assuming mild additional conditions, the exact rate of convergence is $O_p(n^{-3/5})$, a little slower than the rate at (1.1). In fact,

$$n^{3/5} \sup_{-\infty<x<\infty} |\widehat{U} - \widehat{F}| \to C(F)Z \tag{1.2}$$

in distribution, where the constant $C(F)$ depends on F only through the second and third derivatives of F at its mode, and where the random variable Z has a distribution that does not depend on F. This result will be presented in Section 2.

Using the result at (1.2) we are able to give a concise account of asymptotic properties of the dip test statistic under the null hypothesis of unimodality. Since the dip statistic is closely related to the excess mass statistic of Müller and Sawitzki (1991), our results also apply in that context. Indeed, the relationship between the two statistics is employed to derive (1.2).

A formula similar to (1.2), but for Silverman's (1981) bandwidth test of unimodality, has already been given by Mammen et al. (1992). By generalizing all these results we are able to develop theory describing performance of the bandwidth, dip and excess mass tests under local alternatives. In this way we show that all three tests have approximately equal power, in the sense that each is just able to distinguish alternatives that are distant $n^{-3/5}$, but no less, from the null hypothesis of unimodality. Here, distance is measured between distribution functions, and in any L_p metric for $1 \leqslant p \leqslant \infty$. This work will be described in Section 3. Technical arguments will be summarised in Section 4.

2. The dip test, and result (1.2)

Let $\mathscr{U}$ be the class of all unimodal distribution functions. The *dip* of a distribution function F is given by

$$D(F) = \inf_{G\in\mathscr{U}} \sup_{-\infty<x<\infty} |F(x) - G(x)|.$$

To test the null hypothesis H_0 that the distribution F has a unimodal density f, against the alternative H_1 that it has more than one mode, Hartigan and Hartigan (1985) proposed the statistic $D(\widehat{F})$, where $\widehat{F}$ denotes the empirical distribution function of a random n-sample. They suggested a conservative testing procedure, based on comparing the distribution of $D(\widehat{F})$ with that which arises if F is the Uniform distribution function, H, on $[0,1]$; and rejecting H_0 in an α-level test if $D(\widehat{F})$ exceeds the α-level critical point under the assumption $F=H$. That this approach is asymptotically conservative follows from the fact that $n^{1/2}D(\widehat{F})$ has a strictly positive limiting distribution as $n \to \infty$, provided $F=H$; whereas $D(\widehat{F}) = o_p(n^{-1/2})$ for a wide variety of strictly unimodal distributions. See Theorems 3 and 5 of Hartigan and Hartigan (1985).

The simplicity and elegance of the assumption of uniformity make it attractive, but its high degree of conservatism gives it poor performance against, for example, the bandwidth test, if both use the same nominal level. We claim that, in order for the actual level of the test to be bounded away from 0 and 1 under the null hypothesis, the critical point should be of size $n^{-3/5}$ rather than $n^{-1/2}$. Moreover, if a critical point of this size is employed then the test is only barely able to detect local alternative hypotheses distant $n^{-3/5}$ from the null.

To make this claim explicit we introduce a class of local alternatives to F. Let ψ be an antisymmetric function with support equal to the compact interval $[-v, v]$, having two continuous derivatives on the real line, within $(-v, v)$ having the property that ψ'' vanishes at only a finite number of points, and such that $\Psi(x) = \int_{u \leqslant x} \psi(u)\,du$ crosses the horizontal axis at least twice in $(-v, v)$. Let $\eta = n^{-1/5}$, let $c > 0$ be a constant, let F be the distribution function corresponding to a fixed unimodal density with its mode at x_0, and define

$$\Delta_n(x) = (c\eta)^3\, \Psi\left(\frac{x - x_0}{c\eta}\right) \quad \text{and} \quad F_n(x) = F(x) + \Delta_n(x). \tag{2.1}$$

The antisymmetric nature of ψ, and its boundedness and compact support, ensure that for each choice of c the function F_n is a proper distribution function for all sufficiently large n. Furthermore, if F is strongly unimodal (e.g. satisfies (2.2) below) then for all sufficiently large c, the density corresponding to F_n has at least two modes near x_0 for all sufficiently large n. (This follows from a Taylor expansion of F_n''.) We could also incorporate a location change into the definition of Δ_n, for example replacing $\Psi\{(x - x_0)/c\eta\}$ there by $\Psi\{(x - x_0 + d\eta)/c\eta\}$ for an arbitrary constant d, but for the sake of simplicity have not done so.

Assume that

$$f' \text{ exists on } \mathbb{R}, \text{ and is continuous and ultimately monotone in each tail; and } f'' \text{ exists within a neighbourhood of the unique mode } x_0, \text{ and is Hölder continuous there, with } f''(x_0) < 0. \tag{2.2}$$

We also need an extra condition on ψ: defining $\Theta(y) = \Psi(y) - \frac{1}{6}|f''(x_0)|\, y^3$ we ask that

$$\sup_{y_1 < \cdots < y_4} \{\Theta(y_2) - \Theta(y_1) + \Theta(y_4) - \Theta(y_3)\} - \sup_{y_1 < y_2} \{\Theta(y_2) - \Theta(y_1)\} > 0. \tag{2.3}$$

This holds if $\psi = \text{const.}\ \psi_0$ for a fixed, antisymmetric function ψ_0 satisfying the conditions imposed on ψ in Section 2, provided the constant is sufficiently large. To appreciate why, consider a graph of $\Theta(y)$. It diverges to $+\infty$ as $y \to -\infty$, and to $-\infty$ as $y \to +\infty$. From this property, and the fact that Ψ crosses the horizontal axis at least twice, we see that if $\psi = \text{const.}\ \psi_0$ then Θ has at least two local maxima and two local minima for all sufficiently large values of the constant. This implies (2.3).

The theorem below shows that, under this condition, (a) the dip test statistic, after rescaling by a factor $n^{3/5}$, converges in distribution under the null hypothesis; (b) the power of the test against F_n, defined by (2.1), may be made arbitrarily close to 1 by choosing c sufficiently large; and (c) for any fixed $c > 0$ the power is strictly less than 1.

Define $C = C(F) = \{f(x_0)^3/|f''(x_0)|\}^{1/5}$, and let Z denote a random variable whose distribution does not depend on F and which we shall define in Section 4.

Theorem 2.1 (Dip test). *Assume* (2.2), *and let* $t_n = \eta^3 u$ *for some* $u > 0$. *Then,*

$$P_F\{D(\widehat{F}) > t_n\} \to P(CZ > u) \tag{2.4}$$

as $n \to \infty$. *If in addition* (2.3) *holds then the limit*

$$p(c) \equiv \liminf_{n \to \infty} P_{F_n}\{D(\widehat{F}) > t_n\} \tag{2.5}$$

exists and satisfies $0<p(c)<1$ *for each* $0\leqslant c<\infty$, *and*

$$\lim_{c\to\infty} p(c)=1. \tag{2.6}$$

Since $\sup|\widehat{U}-\widehat{F}|=D(\widehat{F})$ then (2.4) is equivalent to (1.2).

3. Power of excess mass and bandwidth tests

We begin by describing the excess mass test statistic. Define

$$E_{nm}(\lambda)=\sup_{I_1,\ldots,I_m}\sum_{j=1}^{m}\{\widehat{F}(I_j)-\lambda\,\|I_j\|\},$$

the "empirical measure for m modes", with the supremum being taken over all sequences $\{I_1,\ldots,I_m\}$ of intervals, and $\|I\|$ denoting the length of the interval I. Then the excess mass statistic for testing the null hypothesis of unimodality against the alternative of two or more modes is

$$\Delta_{n2}=\sup_{\lambda>0}\{E_{n2}(\lambda)-E_{n1}(\lambda)\}.$$

See Müller and Sawitzki (1991). The null hypothesis of unimodality is rejected if Δ_{n2} is too large.

Theorem 3.1. *For every* n,

$$\Delta_{n2}=2D(\widehat{F}). \tag{3.1}$$

The version of (3.1) for the distribution function F also holds; i.e. $\Delta_{n2}(F)=2D(F)$. Of course, (3.1) is an identity in the data $\{X_i\}$, and so is valid regardless of the correctness of the null hypothesis. It implies that, except for the obvious alteration (i.e. multiplying either C or Z by 2), Theorem 2.1 holds if $D(\widehat{F})$ is replaced by Δ_{n2}. The validity of Theorem 3.1 was first noted by Müller and Sawitzki (1991), but apparently no proof has been published before.

The bandwidth test, introduced by Silverman (1981), is based on a standard kernel-type density estimator,

$$\hat{f}(x\,|\,h)=(nh)^{-1}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right),$$

where K is taken to be the *standard normal* density (so that the number of modes of kernel estimates is monotone in h), and h is the bandwidth. Define $\hat{h}_{\text{crit}}$ to be the infimum of values of h such that $\hat{f}(\cdot\,|\,h)$ is unimodal, and reject H_0 in favour of H_1 if $\hat{h}_{\text{crit}}$ exceeds a critical value, t_n say.

Mammen et al. (1992) showed that under H_0, $\hat{h}_{\text{crit}}$ is of size η, and in fact $\hat{h}_{\text{crit}}/\eta\to C'Z'$ in distribution, where $C'=\{f(x_0)/f''(x_0)^2\}^{1/5}$ is a constant and the random variable Z' has a continuous distribution which does not depend on f. A sufficient regularity condition is the following:

the density f corresponding to F has a unique turning point at x_0 (say), and has support equal to a compact interval $\mathscr{I}=[a,b]$ on which it has two continuous derivatives, with $f''(x_0)<0$, $f'(a+)>0$ and $f'(b-)<0$. (3.2)

Under (3.2) the function

$$\beta_F(u)\equiv\lim_{n\to\infty}P_F(\hat{h}_{\text{crit}}>\eta u)$$

is well-defined, satisfies $0<\beta_F(u)<1$ for all $u>0$, and $\beta_F(u)\uparrow 1$ as $u\downarrow 0$, $\beta_F(u)\downarrow 0$ as $u\uparrow\infty$. This suggests that the role of the critical point $t_n=\eta^3 u$ in Theorem 2.1 should here be played by ηu.

Let F_n be as in (2.1), with the functions ψ, Ψ and Δ_n having the definitions given in Section 2. Recall that F_n depends on a positive constant c, through Δ_n. Our next result is the analogue of Theorem 2.1 for the bandwidth test, demonstrating that, like the dip and excess mass approaches, it is just able to distinguish alternatives that are distant $n^{-3/5}$ from the null hypothesis. Result (3.3) is due to Mammen et al. (1992).

Theorem 3.2 (Bandwidth test). *Assume condition* (3.2), *and let* $t_n=\eta u$ *for some* $u>0$. *Then,*

$$P_F(\hat{h}_{\text{crit}}>t_n)\to P(C'Z'>u) \tag{3.3}$$

as $n\to\infty$; *the limit*

$$p(c)\equiv\liminf_{n\to\infty} P_{F_n}(\hat{h}_{\text{crit}}>t_n)$$

exists and satisfies $0<p(c)<1$ *for each* $0\leqslant c<\infty$; *and*

$$\lim_{c\to\infty} p(c)=1. \tag{3.4}$$

4. Outline of technical arguments

We begin with two lemmas, pertaining to the excess mass and bandwidth tests, respectively. Recall that $\eta=n^{-1/5}$.

Let $a=f(x_0)$, $b=-\frac{1}{6}f''(x_0)$ and W be a standard Wiener process, and given $u\in\mathbb{R}$ define

$$w(y_1,y_2,u\,|\,a,b,c)=a^{1/2}\{W(y_2)-W(y_1)\}-b(y_2^3-y_1^3)+u(y_2-y_1)+c^3\{\Psi(y_2/c)-\Psi(y_1/c)\},$$

$$U(a,b,c)=\sup_{-\infty<u<\infty}\left[\sup_{-\infty<y_1<\cdots<y_4<\infty}\{w(y_1,y_2,u\,|\,a,b,c)+w(y_3,y_4,u\,|\,a,b,c)\}\right.$$
$$\left.-\sup_{-\infty<y_1<y_2<\infty} w(y_1,y_2,u\,|\,a,b,c)\right].$$

For the most part we shall suppress a and b, writing $U(a,b,c)$ as $U(c)$. It may be proved that $U(c)$ is finite and positive with probability 1, and that its distribution has no atoms.

Lemma 4.1. *Let* F *satisfy* (2.2). *Then, for each* $c\geqslant 0$ *and* $u>0$, $P_{F_n}(\Delta_{n2}\leqslant\eta^3 u)\to P\{U(c)\leqslant u\}$ *as* $n\to\infty$. *If in addition* (2.3) *holds then* $P\{U(c)\leqslant u\}\to 0$ *as* $c\to\infty$.

Proof. Using the embedding of Komlós et al. (1975), and modifying the argument of Müller and Sawitzki (1991), we may show that $\Delta_{n2}/\eta^3\to U(c)$ in distribution. This establishes the first part of the lemma. To prove the second part, observe that with $\Theta(y)=\Psi(y)-by^3$,

$$U(c)\geqslant\sup_{-\infty<y_1<\cdots<y_4<\infty}\{w(cy_1,cy_2,0\,|\,a,b,c)+w(cy_3,cy_4,0\,|\,a,b,c)\}$$
$$-\sup_{-\infty<y_1<y_2<\infty} w(cy_1,cy_2,0\,|\,a,b,c)$$

$$= c^3 \left[\sup_{-\infty < y_1 < \cdots < y_4 < \infty} \{\Theta(y_2) - \Theta(y_1) + \Theta(y_4) - \Theta(y_3)\} \right.$$

$$\left. - \sup_{-\infty < y_1 < y_2 < \infty} \{\Theta(y_2) - \Theta(y_1)\} \right] + o_p(c^3)$$

$$\to \infty$$

in probability as $c \to \infty$, by (2.3). This completes the proof of the lemma. □

Put $b_1 = -f''(x_0) > 0$, let W be a standard Wiener process, and define

$$V(r,y) = V(r,y \mid a, b_1, c) = -b_1 y + c \int \psi'\left(\frac{y - rz}{c}\right) K(z)\,dz + a^{1/2} r^{-3} \int K''(y+z)\, W(rz)\,dz.$$

Let $Z(c)$ denote the infimum of all values of r such that $V(r,y)$, as a function of y, changes sign exactly once on the real line. (We suppress a, b_1 and d from the notation for $Z(c)$, since only c will be varied.) Note that, by differentiation with respect to y, with probability one $V(r,y)$ (as a function of y) does not have any turning points on any given interval $[-y_0, y_0]$, for all sufficiently large r; that (since $V(y,r) \to -b_1 y$ as $r \to \infty$) it must have at least one zero-crossing point there, and so have exactly one, for all sufficiently large r; and that (since $V(y,r) \to -b_1 y$ *uniformly* in y as $r \to \infty$) it has no zero-crossing points outside $[-y_0, y_0]$, for large r. Hence, with probability one, for all sufficiently large r the function $V(r,\cdot)$ has exactly one zero-crossing point. More simply, with probability 1 it has at least two zero-crossing points for all sufficiently small r. Therefore, $Z(c)$ is well-defined.

Lemma 4.2. *Let F satisfy* (3.2). *Then, for each $u > 0$, $P_{F_n}(\hat{h}_{\text{crit}} \leqslant \eta u) \to P\{Z(c) \leqslant u\}$ as $n \to \infty$, and $P\{Z(c) \leqslant u\} \to 0$ as $c \to \infty$.*

Proof. Arguing as in Mammen et al. (1992) we may show that, sampling from F_n rather than F, $\eta^{-1} \hat{f}'(x_0 + \eta y \mid \eta r)$ is closely approximated by $V(r,y)$ for an appropriate choice of W, and thence that $\hat{h}_{\text{crit}}/\eta$ converges in distribution to $Z(c)$. (Mammen et al. (1992) treated the case $c = 0$.) This proves the first part of the lemma. To establish the second part, note that the term involving c in the definition of $V(r,y)$ is

$$c \int \psi'\left(\frac{y - rz}{c}\right) K(z)\,dz = c\,\{\psi'(y/c) + o(1)\}$$

as $c \to \infty$. The function ψ' changes sign at least four times on its support, and so for each fixed $r_0 > 0$, and all sufficiently large c, $V(r,y)$ must change sign at least four times in the interval $(-c^{-1}v, c^{-1}v)$ for all $0 < r \leqslant r_0$. Therefore, $Z(c) \to \infty$ as $c \to \infty$, completing the proof of Lemma 4.2. □

Proof of Theorem 2.1 (*The case of the excess mass test*). Taking $c = 0$ we see from the first part of Lemma 4.1 that (2.4), with Δ_{n2} replacing $D(\widehat{F})$ and $U(0) = U(a,b,0)$ replacing CZ, holds. Note that $U(a,b,0)$ has the same distribution as $(a^3/b)^{1/5}\ U(1,1,0)$; and that if $C = \{f(x_0)^3/|f''(x_0)|\}^{1/5}$ then, in the case of the excess mass test, the variable Z in the theorem has the distribution of $6^{1/5}\,U(1,1,0)$. (In the context of the dip test, for which the theorem was originally stated; and in the case of (1.2); it has the distribution of half this variable.) Lemma 4.1 implies (2.5) with $p(c) = P\{U(c) > c\}$, and also implies (2.6) if (2.3) holds. □

Proof of Theorem 3.2. Taking $c = 0$ and using the first part of Lemma 4.2 we see that (3.3) holds with $C'Z'$ replaced by $Z(0)$. The latter random variable is a function only of a and b_1. Making a scale transform

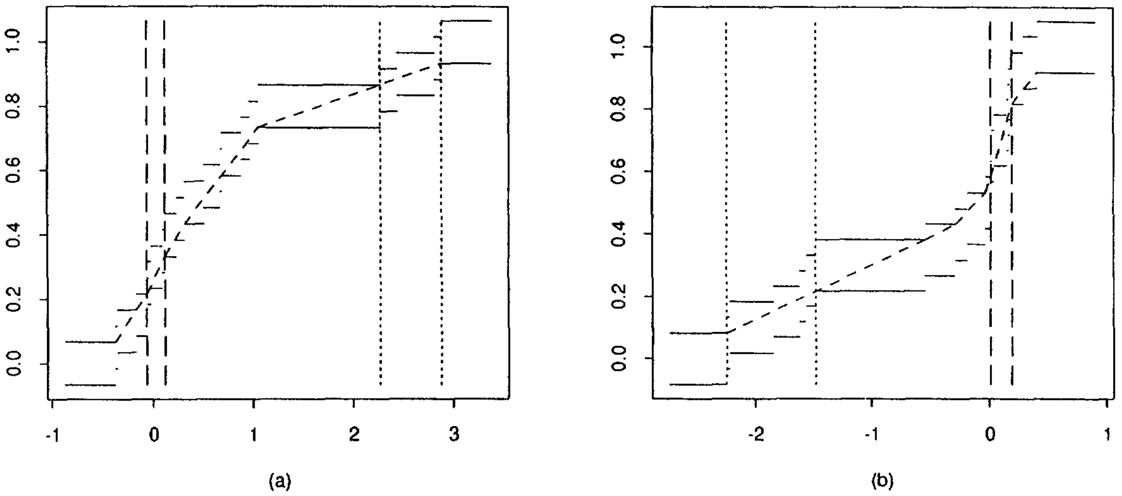

Fig. 1. Typical plots of $\widehat{F} \pm D(\widehat{F})$ and $\widehat{U}$. Panels (a) and (b) correspond to different samples of size $n = 20$. The horizontal lines represent $\widehat{F} \pm D(\widehat{F})$, the steadily increasing short-dashed line depicts $\widehat{U}$, the vertical dashed lines define the interval J_0, and the vertical dotted lines indicate the interval J_2. Note that $x_0 > X_{(U)}$ in panel (a), and $x_0 < X_{(L)}$ in panel (b).

of the the process W we may show that the bivariate stochastic process $U_1(r,y) = V\{(a/b_1^2)^{1/5}r, y \mid a, b_1, 0\}$ has the same distribution as $U_2(r,y) = b_1\, V(r,y \mid 1,1,0)$, and hence that $Z(0)$ may be written as $C'Z'$, where $C' = (a/b_1^2)^{1/5}$ and the distribution of Z' does not depend on unknowns. Result (3.4) follows from the second part of Lemma 4.2. □

Proof of Theorem 3.1. Let $H_\lambda(x) = \widehat{F}(x) - \lambda x$, with the analogous definition when x is replaced by an interval. Note that

$$\Delta_{n2} = -H_{\lambda_0}(b,c) = H_{\lambda_0}[c,d]$$

for some $\lambda_0 > 0$. Here $b < c < d$ are data points, b, d are local maxima of H_{λ_0}, and $H_{\lambda_0}(b) = H_{\lambda_0}(d)$. First we prove that $\Delta_{n2} \leqslant 2D(\widehat{F})$. If to the contrary $\Delta_{n2} > 2D(\widehat{F})$, then

$$H_{\lambda_0}(c-) + D(\widehat{F}) < H_{\lambda_0}(d) - D(\widehat{F}) = H_{\lambda_0}(b) - D(\widehat{F}).$$

Hence, if $\mathscr{G}$ denotes the set of distribution functions that are convex to the left of some point, and concave to the right of that point, then no element of $\mathscr{G}$ has a graph that is contained within the band $\{(x,y) : H_{\lambda_0}(x) - D(\widehat{F}) \leqslant y \leqslant H_{\lambda_0}(x) + D(\widehat{F})\}$. This contradicts the identity

$$D(\widehat{F}) = \inf_{G \in \mathscr{U}} \sup_{-\infty < x < \infty} |\widehat{F}(x) - G(x)| = \inf_{g \in \mathscr{G}} \sup_{-\infty < x < \infty} |H_{\lambda_0}(x) - g(x)|.$$

Therefore, $\Delta_{n2} \leqslant 2D(\widehat{F})$.

To obtain the reverse inequality, let $J_0 = [X_{(L)}, X_{(U)}]$ be the interval over which $\widehat{U}$ has greatest constant slope. From Theorem 6 of Hartigan and Hartigan (1985), $|\widehat{U}(x_0) - \widehat{F}(x_0)| = D(\widehat{F})$ for some $x_0 \notin J_0$, where $\widehat{U}$ is the greatest convex minorant of $\widehat{F} + D(\widehat{F})$ in $(-\infty, X_{(L)})$ and is the least concave majorant of $\widehat{F} - D(\widehat{F})$ in $(X_{(U)}, +\infty)$. Let $J_1 \equiv [X_{(A)}, X_{(B)}]$ be the interval containing x_0, where $X_{(A)}$, $X_{(B)}$ are two successive vertices of $\widehat{U}$. Denote the slopes of $\widehat{U}$ in J_0 and J_1 by λ^* and $\lambda^\dagger$, respectively; then $\lambda^* > \lambda^\dagger$.

First consider the case $x_0 > X_{(U)}$. Let $J_2 = [X_{(C)}, X_{(B)}]$, where $X_{(C)}$ is given by $D(\widehat{F}) = \widehat{U}(X_{(C)}-) - \widehat{F}(X_{(C)}-)$. It follows that x_0 may be taken as any point strictly less than $X_{(C)}$ and sufficiently close to $X_{(C)}$. Denote the vertices of $\widehat{U}$ in $[X_{(U)}, +\infty)$ by $X_{(U)} = u_1 < \cdots < u_I = X_{(A)} < u_{I+1} = X_{(B)} < \cdots < u_m < u_{m+1} = +\infty$, where $u_1, \ldots, u_m$ are data points, and let λ_i be the slope of $\widehat{U}$ in $[u_i, u_{i+1})$, for $i = 1, \ldots, m$. Note that $\lambda_1 > \cdots > \lambda_m$. For each $i = 1, \ldots, m$, the graph of $\widehat{U}$ restricted to $[u_i, u_{i+1}]$ is just the line segment joining the points $(u_i, \widehat{F}(u_i) - D(\widehat{F}))$ and $(u_{i+1}, \widehat{F}(u_{i+1}) - D(\widehat{F}))$:

$$\widehat{U}(x) = \widehat{F}(u_i) - D(\widehat{F}) + \lambda_i(x - u_i), x \in [u_i, u_{i+1}], \tag{4.1}$$

where $\lambda_i = \{\widehat{F}(u_{i+1}) - \widehat{F}(u_i)\}/(u_{i+1} - u_i)$. Since $|\widehat{F}(x) - \widehat{U}(x)| \leqslant D(\widehat{F})$ then by (4.1),

$$H_{\lambda_i}(u_i) = H_{\lambda_i}(u_{i+1}), \tag{4.2}$$

$$-2D(\widehat{F}) \leqslant H_{\lambda_i}(x) - H_{\lambda_i}(u_i) \leqslant 0, \quad x \in [u_i, u_{i+1}]. \tag{4.3}$$

For any $y > X_{(A)} = u_I$, let $u_J \leqslant y$ be the vertex of $\widehat{U}$ that is closest to y. Using (4.2), (4.3) and the fact that $\lambda^\dagger = \lambda_I > \cdots > \lambda_J$, we may show that

$$\begin{aligned} H_{\lambda^\dagger}[X_{(L)}, y] &= \{H_{\lambda^\dagger}(y) - H_{\lambda^\dagger}(u_J)\} + \{H_{\lambda^\dagger}(u_J) - H_{\lambda^\dagger}(u_I)\} \\ &\quad + \{H_{\lambda^\dagger}(u_I) - H_{\lambda^\dagger}(X_{(L)}-)\} \\ &= \{H_{\lambda_J}(y) - H_{\lambda_J}(u_J) + (\lambda_J - \lambda^\dagger)(y - u_J)\} + \sum_{l=I}^{J-1}(\lambda^\dagger - \lambda_l)(u_l - u_{l+1}) \\ &\quad + H_{\lambda^\dagger}[X_{(L)}, X_{(A)}] \\ &\leqslant H_{\lambda^\dagger}[X_{(L)}, X_{(A)}]. \end{aligned} \tag{4.4}$$

Furthermore,

$$\begin{aligned} H_{\lambda^\dagger}[X_{(L)}, X_{(A)}] &= \{H_{\lambda^\dagger}(X_{(A)}) - H_{\lambda^\dagger}(X_{(U)})\} + \{H_{\lambda^\dagger}(X_{(U)}) - H_{\lambda^\dagger}(X_{(L)}-)\} \\ &= \sum_{l=1}^{I-1}(\lambda^\dagger - \lambda_l)(u_l - u_{l+1}) + \{H_{\lambda^\dagger}(X_{(U)}) - H_{\lambda^\dagger}(X_{(L)}-)\} \\ &> H_{\lambda^\dagger}(X_{(U)}) - H_{\lambda^\dagger}(X_{(L)}-) \\ &= H_{\lambda^*}(X_{(U)}) - H_{\lambda^*}(X_{(L)}-) + (\lambda^* - \lambda^\dagger)(X_{(U)} - X_{(L)}) \\ &= 2D(\widehat{F}) + (\lambda^* - \lambda^\dagger)(X_{(U)} - X_{(L)}) \\ &> 2D(\widehat{F}), \end{aligned} \tag{4.5}$$

where the last equality follows from the fact that $\widehat{U}$ restricted to J_0 is the line segment joining $(X_{(L)}, \widehat{F}(X_{(L)}-) + D(\widehat{F}))$ to $(X_{(U)}, \widehat{F}(X_{(U)}) - D(\widehat{F}))$.

Next we show that for any $y > x \geqslant X_{(A)}$,

$$H_{\lambda^\dagger}[x, y] \leqslant 2D(\widehat{F}). \tag{4.6}$$

Let u_K, u_L (where $u_K \leqslant x < u_L \leqslant y$) be the vertices of $\widehat{U}$ that are closest to x and y. If $u_K \leqslant x < y \leqslant u_{K+1}$, then from (4.3) and since $\lambda_K \leqslant \lambda^\dagger$,

$$H_{\lambda^\dagger}[x,y] = H_{\lambda^\dagger}(y) - H_{\lambda^\dagger}(x-) = H_{\lambda_K}(y) - H_{\lambda_K}(x-) + (\lambda_K - \lambda^\dagger)(y-x) \leqslant 2D(\widehat{F}).$$

Moreover, if $u_K \leqslant x \leqslant u_{K+1} \leqslant u_L \leqslant y$ then (4.1)–(4.3) imply

$$\begin{aligned} H_{\lambda^\dagger}[x,y] &= \{H_{\lambda^\dagger}(y) - H_{\lambda^\dagger}(u_L)\} + \{H_{\lambda^\dagger}(u_L) - H_{\lambda^\dagger}(u_{K+1})\} \\ &\quad + \{H_{\lambda^\dagger}(u_{K+1}) - H_{\lambda^\dagger}(x-)\} \\ &= \{H_{\lambda_L}(y) - H_{\lambda_L}(u_L) + (\lambda_L - \lambda^\dagger)(y - u_L)\} \\ &\quad + \sum_{l=K+1}^{L-1} (\lambda^\dagger - \lambda_l)(u_l - u_{l+1}) \\ &\quad + \{H_{\lambda_K}(u_{K+1}) - H_{\lambda_K}(x-) + (\lambda_K - \lambda^\dagger)(u_{K+1} - x)\} \\ &\leqslant 2D(\widehat{F}). \end{aligned}$$

It follows from (4.4)–(4.6) that

$$C_{n,1}(\lambda^\dagger) \subset (-\infty, X_{(A)}],$$

where $C_{n,1}(\lambda^\dagger)$ denotes the interval maximizing $\widehat{F}(I_1) - \lambda^\dagger \|I_1\|$ over all intervals I_1. Using (a) formula (3) of Müller and Sawitzki (1991), (b) our formulae (4.1) and (4.2), and (c) the results $J_2 \subset C_{n,1}(\lambda^\dagger)^c$ and $D(\widehat{F}) = \widehat{U}(X_{(C)}-) - \widehat{F}(X_{(C)}-)$, we may prove that

$$\begin{aligned} \Delta_{n2} &= \sup_{\lambda>0} \{E_{n2}(\lambda) - E_{n1}(\lambda)\} \geqslant E_{n2}(\lambda^\dagger) - E_{n1}(\lambda^\dagger) \\ &\geqslant H_{\lambda^\dagger}(J_2) = 2D(\widehat{F}). \end{aligned}$$

When $x_0 < X_{(L)}$ we have $D(\widehat{F}) = \widehat{F}(X_{(C)}) - \widehat{U}(X_{(C)})$, and so we may take $x_0 = X_{(C)}$. Let $J_2 = [X_{(A)}, X_{(C)}]$. As in the case $x_0 > X_{(U)}$ we may prove that $C_{n,1}(\lambda^\dagger) \subset [X_{(A)}, +\infty)$ and that $H_{\lambda^\dagger}(J_2) = 2D(\widehat{F})$ holds. The desired inequality follows from these properties.

Acknowledgements

We are grateful to D.W. Müller for helpful comments.

References

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D., 1972. Statistical Inference Under Order Restrictions. Wiley, New York.
Brunk, H.D., 1956. On an inequality for convex functions. Proc. Amer. Math. Soc. 7, 817–824.
Eeden, C. van, 1956. Maximum likelihood estimation of ordered probabilities. Indag. Math. 18, 444–455.
Grenander, U., 1956. On the theory of mortality measurement, Part II. Skand. Act. 39, 125–153.
Hall, P., Titterington, D.M., 1988. On confidence bands in nonparametric density estimation and regression. J. Multivariate. Anal. 27, 228–254.
Hartigan, J.A. and Hartigan, P.M., 1985. The dip test of unimodality. Ann. Statist. 13, 70–84.

Hartigan, P.M., 1985. Algorithm AS217: Computation of the dip statistic to test for unimodality. Appl. Statist. 34, 320–325.

Kiefer, J., Wolfowitz, J., 1976. Asymptotically minimax estimation of concave and convex distribution functions. Z. Wahrscheinlichkeitstheorie verw. Gebiete 34, 73–85.

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

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

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

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