The Annals of Statistics
2001, Vol. 29, No. 3, 648–665

LEAST SQUARES ESTIMATORS OF THE MODE OF A UNIMODAL REGRESSION FUNCTION

BY JYH-MING SHOUNG AND CUN-HUI ZHANG

Rutgers University

In this paper, we consider nonparametric least squares estimators of the mode of an unknown unimodal regression function. We establish almost sure convergence of these estimators with nearly optimal convergence rates, under the assumption of the exponential tail for the error distributions.

1. Introduction. Consider the regression model

$$Y_i = f_0(X_i) + \varepsilon_i, \qquad i = 1, ..., n, \tag{1.1}$$

with independent mean-zero errors independent of the design points $\{X_i\}$, where $f_0(x)$ is an unknown unimodal function with unknown mode m_0. The nonparametric least squares estimator (NPLSE) of the mode m_0 is defined by

$$\widehat{m}_n \equiv X_{\widehat{j}}, \quad \widehat{j} \equiv \arg\min_{1\le j\le n} \left[\min_{f\in\mathscr{F}(X_j)} \sum_{i=1}^{n} \{Y_i - f(X_i)\}^2 \right], \tag{1.2}$$

where $\mathscr{F}(m) = \{f : f \text{ is a unimodal function with mode } m\}$. In this paper, we establish almost sure convergence of (1.2) with optimal convergence rates up to a logarithmic factor, under the assumption of the exponential tail for the error distributions.

Suppose throughout the sequel that $\{X_i\}$ are iid random variables from a continuous distribution G_X such that the density G'_X exists in a neighborhood of the unknown mode m_0 and G'_X is positive and continuous at m_0. Consider smoothness condition

$$|f_0(x) - f_0(m_0)| = (B + o(1))|x - m_0|^s \qquad \text{as } x \to m_0 \tag{1.3}$$

at the mode m_0, with smoothness index $s > 0$ and certain constant $B > 0$.

THEOREM 1.1. *Let $\widehat{m}_n$ be given by* (1.2) *with the data from* (1.1). *Suppose $f_0(x)$ is a unimodal function satisfying* (1.3) *at its mode m_0. Suppose the errors in* (1.1) *are independent with $E\varepsilon_i = 0$ and that $P\{|\varepsilon_i| > t\} \le \mathscr{M} \exp\big(-ct^\alpha\big)$ for all $t > 0$ and all i, where α, $\mathscr{M}$ and c are positive numbers. Then, with $\gamma = \max\{1, 1/2 + 1/\alpha\}$*

$$\limsup_{n\to\infty} \big\{n/(\log n)^{2\gamma}\big\}^{1/(2s+1)} \big|\widehat{m}_n - m_0\big| < \infty \qquad a.s. \tag{1.4}$$

Received June 1999; revised Jauary 2001.
AMS 2000 *subject classifications*. Primary 62G08; secondary 62G05.
Key words and phrases. Least squares estimation, mode, nonparametric regression, convergence rate.

The convergence rate in (1.4) is optimal up to a logarithmic factor [cf. Donoho and Liu (1991) and Hasminskii (1979)]. Let the smoothness index s be a positive integer and $f^{(k)}$ be the kth derivative of f. Let $\mathscr{C}_{\epsilon,M} \equiv \{f : \|f^{(s)} - f_0^{(s)}\|_\infty \le \epsilon, \|f^{(s-1)} - f_0^{(s-1)}\|_\infty \le Mn^{-1/(2s+1)}\}$ with sufficiently small $\epsilon > 0$ and large $M < \infty$. If the regression function f_0 is s-times continuously differentiable, (1.3) implies $f_0^{(s-1)}(m_0) = 0 > f_0^{(s)}(m_0)$, so that $f^{(s-1)}(m_0) \approx f_0^{(s)}(m_0)\{m_0 - \mathrm{mode}(f)\}$ for $f \in \mathscr{C}_{\epsilon,M}$ and $\epsilon << f_0^{(s)}(m_0)$. Thus, for iid $N(0,1)$ errors, the minimax convergence rate for the estimation of the mode of f over $C_{\epsilon,M}$ is the same as the minimax convergence rate $n^{-1/(2s+1)}$ for the estimation of $f^{(s-1)}(m_0)$ over $\mathscr{C}_{\epsilon,M}$ [cf. Stone (1980)]. Here, $m_0 \equiv \mathrm{mode}(f_0) \not\equiv \mathrm{mode}(f)$ in general. It seems that $(n/\log n)^{-1/(2s+1)}$ is a lower bound for the convergence rates of the NPLSE (1.2) and all other estimators adaptive to different values of the smoothness index s in (1.3), in view of Lepskii (1992), Gill and Levit (1995) and Brown and Low (1996).

Nonparametric regression (1.1) with unimodal regression function is closely related to nonparametric density problems based on iid observations from an unknown unimodal probability density function. Both models have been considered by many authors in the literature. We shall provide a brief review of three types of relevant previous results: isotonic estimation with known mode, rate specific estimation of the mode m_0 under a given smoothness condition, and rate adaptive estimation of m_0.

When the mode m_0 is known, the NPLSE of f_0 is given by two isotonic regression estimators, one on each side of m_0. The NPLSE and related isotonic methods for estimating a monotone regression or density function were proposed by Ayer et al. (1955), Eeden (1956) and Grenander (1956). The asymptotic distributions of these estimates at a fixed x_0 were established by Prakasa Rao (1969) and Brunk (1970). Groeneboom (1985) obtained asymptotic distribution of the L_1 loss for the Grenander estimator. Comprehensive account of the subject can be found in Barlow et al. (1972) and Robertson et al. (1988).

In the case of unknown mode m_0 and for classes with known smoothness index s, Venter (1967) used clustering methods to estimate the mode of a density and proved the following convergence rates for his estimators: $o(1)n^{-1/(2s+1)}(\log n)^{1/s}$ for $s \ge 1/2$, and $o(1)n^{-1/2}(\log n)^{1/s}$ for $0 < s < 1/2$. Under stronger smoothness conditions on f_0 in a neighborhood of m_0; Parzen (1962), Chernoff (1964), Eddy (1980) and Müller (1989) provided certain kernel estimators of mode with faster convergence rates.

For unspecified smoothness index, Wegman (1970) proved strong consistency for an MLE of the mode of a unimodal density. Grund and Hall (1995) proved the L^p consistency of an estimate of the mode based on kernel methods. Birgé (1997) provided L_1 risk bounds for certain minimum distance estimators of a density with unknown mode. Bickel and Fan (1996) proved that certain modified MLE of the mode of a unimodal density function converges at the rate of $o(1)n^{-1/(4s+2)}(\log n)^{2/(2s+1)}$ with the smoothness index $s \ge 1$ in (1.3).

In Section 2 we discuss in detail NPLSE and give an outline of the proof of Theorem 1.1. The full proof is provided in Section 3.

2. NPLSE of mode and an outline of proof. In this section, we provide an outline of our proof of Theorem 1.1 bsaed on an alternative expression of (1.2) and a deviation inequality due to Ledoux and Talagrand.

The NPLSE of an unknown unimodal function with known mode m is defined by

$$\widehat{f}_n(\cdot; m) = \operatorname*{arg\,min}_{f \in \mathscr{F}(m)} \sum_{i=1}^{n} \{Y_i - f(X_i)\}^2, \tag{2.1}$$

where $\mathscr{F}(m)$ is as in (1.2) and for definiteness $\widehat{f}_n(\cdot; m)$ is taken to be the version which is a left-continuous step function with jumps only at design points X_i. Thus, $\widehat{f}_n(\cdot; m') = \widehat{f}_n(\cdot; m'')$ if $[m', m''] \cap \{X_i\} = \varnothing$. The minimization problem in (2.1) can be solved by separately minimizing the sum of squares for $\{i : X_i \le m\}$ and $\{i : X_i > m\}$ with standard isotonic regression methods. The solution can be easily computed with the pool-adjacent-violators algorithm. See Barlow *et al* (1972) or Robertson *et al* (1988). Here, a function f defined on an interval (a, b), $-\infty \le a < b \le \infty$, is unimodal if there exists $m \in (a, b)$ such that f is nondecreasing in (a, m) and nonincreasing in (m, b) and that $f(m) \ge \min\{f(m-), f(m+)\}$. Thus, $\sup_x f(x)$ is reached either at m or $m\pm$. In this case, m is called a mode of f.

Let $t_1 < t_2 < \cdots < t_k$ be the location of jumps of $\widehat{f}_n(\cdot; m)$ in (2.1). Within each interval $I_\ell = (t_{\ell-1}, t_\ell]$, $t_0 = -\infty$, the value of $\widehat{f}_n(\cdot; m)$ is the average of $\{Y_i : X_i \in I_\ell\}$, as it minimizes the sum $\sum_{X_i \in I_\ell} \{Y_i - c\}^2$ over real c, $\ell = 1, \ldots, k$. Thus,

$$\begin{aligned} \sum_{i=1}^{n} \{Y_i - \widehat{f}_n(X_i; m)\}^2 &= \sum_{\ell=1}^{k} \sum_{X_i \in I_\ell} \{Y_i^2 - \widehat{f}_n^{\,2}(t_\ell; m)\} \\ &= \sum_{i=1}^{n} Y_i^2 - \sum_{i=1}^{n} \widehat{f}_n^{\,2}(X_i; m). \end{aligned} \tag{2.2}$$

Since the first term on the right-hand side of (2.2) is fixed for the given data, $\widehat{f}_n(\cdot; m)$ is actually the maximizer of $\sum_{i=1}^{n} f^2(X_i)$ for all $f \in \mathscr{F}(m)$ which are piecewise averages of $\{Y_i\}$. Furthermore, the NPLSE of the mode defined in (1.2) is equivalent to

$$\widehat{m}_n = X_{\hat{j}}, \quad \sum_{i=1}^{n} \widehat{f}_n^{\,2}(X_i; X_{\hat{j}}) = \max_m \sum_{i=1}^{n} \widehat{f}_n^{\,2}(X_i; m). \tag{2.3}$$

The proof of Theorem 1.1, in Section 3, is based on our investigation of the sum of squares in (2.3) and an application of the following deviation inequality [Ledoux (1996)]

$$P\{\tilde{h}(\tilde{\eta}_1, \ldots, \tilde{\eta}_n) > \operatorname{median}\left(\tilde{h}(\tilde{\eta}_1, \ldots, \tilde{\eta}_n)\right) + t\} \le 2\exp\left(-t^2/4\right), \tag{2.4}$$

for independent variables $\tilde{\eta}_i$ living in a unit cube and convex $\tilde{h}$ with $\|\tilde{h}\|_{Lip} \le 1$. Here $\|\tilde{h}\|_{Lip} = \sup_{\vec{x} \ne \vec{y}} \|\tilde{h}(\vec{x}) - \tilde{h}(\vec{y})\| / \|\vec{x} - \vec{y}\|$ is the Lipschitz norm.

A crucial step in the proof is to control the contribution of the spike of $\widehat{f}_n(\cdot;\widehat{m}_n)$ around m_0 to the sum of squares in (2.3); that is, to control the contribution of the positive part $\{\widehat{f}_n(x;\widehat{m}_n)-f_0(m_0)\}^+$. This is done, in subsection 3.2, by invoking (2.4) for

$$T'_j = \sqrt{\sum_{k=1}^n \max_{k\le\ell\le n}\left(\sum_{i=1}^{\ell}\eta_{i+j}\varepsilon_{i+j}/\ell\right)^2}, \qquad j\ge 0, \tag{2.5}$$

conditionally on $\{\varepsilon_i\}$, where $\{\eta_i\}$ is a Rademacher sequence independent of $\{\varepsilon_i\}$. Here is the connection between (2.3) and (2.5). By the minimax formula [cf., e.g., Barlow et al. (1972)]

$$\begin{aligned}\widehat{f}_n(X_j;m) &= \max_{s\le X_j}\min_{X_j\le t\le m}\frac{\sum_{s\le X_i\le t}Y_i}{\#\{i: s\le X_i\le t\}}\\ &\le f_0(m_0)+\max_{s\le X_j}\frac{\sum_{s\le X_i\le m}\varepsilon_i}{\#\{i: s\le X_i\le m\}},\end{aligned} \tag{2.6}$$

for $X_j\le m$, with $\#\{A\}$ being the size of set A, and

$$\begin{aligned}\widehat{f}_n(X_j;m) &= \min_{m<s\le X_j}\max_{X_j\le t}\frac{\sum_{s\le X_i\le t}Y_i}{\#\{i: s\le X_i\le t\}}\\ &\le f_0(m_0)+\max_{X_j\le t}\frac{\sum_{m<X_i\le t}\varepsilon_i}{\#\{i: m< X_i\le t\}},\end{aligned} \tag{2.7}$$

for $X_j>m$. These minimax formulas imply

$$\begin{aligned}&\sup_m\sum_{j=1}^n\left[\left\{\widehat{f}_n(X_j;m)-f_0(m_0)\right\}^+\right]^2\\ &\quad\le \max_{1\le j^*\le n}\sum_{j=1}^{j^*}\max_{1\le j_1\le j}\left\{\frac{\left(\sum_{i=j_1}^{j^*}\varepsilon'_i\right)^+}{j^*-j_1+1}\right\}^2\\ &\qquad+\max_{1\le j^*\le n}\sum_{j=j^*+1}^{n}\max_{j\le j_2\le n}\left\{\frac{\left(\sum_{i=j^*+1}^{j_2}\varepsilon'_i\right)^+}{j_2-j^*}\right\}^2,\end{aligned} \tag{2.8}$$

where $\{\varepsilon'_i\}$ is the permutation of $\{\varepsilon_i\}$ according to the ranks of $\{X_i\}$. This inequality and standard symmetrization methods provide the connection between T'_j in (2.5) and the contribution of the spike of $\widehat{f}_n(x;\widehat{m}_n)$ to the sum of squares in (2.3).

We conclude the section with two remarks about the conditions of Theorem 1.1.

REMARK 2.1. The conclusion of Theorem 1.1 remains valid without the continuity assumption on G_X on $\mathbb{R}$, as long as G'_X is positive and continuous at m_0. We impose this extra continuity condition in order to focus on the main ideas in the proofs.

REMARK 2.2. The condition on the tail probability of the errors is used to control the contribution of the spike of $\widehat{f}_n(\cdot;\widehat{m}_n)$ to the sum of squares in (2.3). It is not clear if the condition can be weakened to $\sup_i E|\varepsilon_i|^p < \infty$ for certain $2 \le p < \infty$ for the NPLSE (1.2) or its modifications adaptive to smoothness classes of different index s.

3. Proof of Theorem 1.1. We divide the proof into 3 subsections. We shall discuss the continuity of the NPLSE of unimodal functions and the magnitude of the spike of the NPLSE at the mode in the first two subsections to prepare for the proof of Theorem 1.1 in the third subsection. Since our conditions and conclusions are invariant under monotone transformation of the design points $\{X_i\}$, we assume without loss of generality in this section that $\{X_i\}$ are iid uniform random variables in an interval $[a,b]$ of unit length with $a < m_0 < b$.

3.1. *Continuity of NPLSE of unimodal functions.* Let $[a,b]$ be an interval with $b = a+1$. For $a \le m \le b$, right-continuous functions H and right-continuous nondecreasing functions G on $[a,b]$, define $D_m(H|G) = \widehat{h}(\cdot;m)$ by

$$D_m(H|G)(x)=\widehat{h}(x;m)=\begin{cases} \sup\limits_{G(s)<G(x)} \inf\limits_{x\le t\le m} \dfrac{H(t)-H(s)}{G(t)-G(s)}, & a\le x\le m,\\ \inf\limits_{\substack{G(s)<G(x)\\ m<s}} \sup\limits_{x\le t} \dfrac{H(t)-H(s)}{G(t)-G(s)}, & m<x\le b, \end{cases} \tag{3.1}$$

with the convention $\sup\varnothing = -\infty$ and $\inf\varnothing = \infty$. See (2.6) and (2.7).

REMARK 3.1. By the minimax formula, $\widehat{h}(G^{-1}(x);m)$ is the left-derivative of the convex minorant (concave majorant) of $H(G^{-1}(x))$ on $[a,m]$ (on $(m,b]$ respectively), where G^{-1} is the inverse function of G. Thus, (3.1) is a local estimator; e.g. If $\widehat{h}(x;m)$ has a jump at a point $x_0 \in [a,m]$, then $\widehat{h}(x;m')$ for $x \le x_0 < m' \le m$ does not depend on the behavior of the functions H and G in $(x_0, b]$.

Consider a left-continuous unimodal function h_0 on $[a,b]$ with a mode $m_0 \in (a,b)$ and $\sup_x h_0(x) = 0$. Let $H_0(x) = \int_a^x h_0(t)dt$ and G_0 be the uniform distribution on $[a,b]$. In Lemma 3.1 below, we provide certain continuity properties of the mapping $D_m(H|G)$ in (3.1), as $(H,G) \to (H_0,G_0)$. These results are used in the proof of Theorem 1.1.

LEMMA 3.1. *Let $h_0(\cdot;m) = D_m(H_0|G_0)$ with the D_m in* (3.1). *Let $C_1 = 4(1+\|h_0\|_\infty)$ and $0<\delta<\delta_0/4$. Then, for $x \in A_{m,\delta_0} = (a+\delta_0, b-\delta_0]\setminus(m-\delta_0, m+\delta_0]$,*

$$\left|\widehat{h}(x;m) - h_0(x;m)\right| \le C_1\delta/\delta_0 + \max_{|t-x|\le\delta_0} \left|h_0(t) - h_0(x-\delta_0)\right| \tag{3.2}$$

for all (H, G) *with* $\|H - H_0\|_\infty \vee \|G - G_0\|_\infty \le \delta$. *Moreover, for all* $\epsilon > 0$ *and* $a < m^* < b$ *there exist* $\delta = \delta_{\epsilon, m^*, h_0} > 0$ *such that*

$$\begin{aligned}\int_a^b \Big\{\widehat{h}^2(x;m) - \widehat{h}^2(x;m^*)\Big\} dG(x) - \int_a^b \Big\{h_0^2(x;m) - h_0^2(x;m^*)\Big\} dG_0(x)\\ \le \epsilon + \int_a^b \Big\{\widehat{h}^+(x;m)\Big\}^2 dG(x),\end{aligned} \tag{3.3}$$

uniformly for all real numbers $a \le m \le b$ *and functions* H *and distribution functions* G *on* $[a, b]$ *satisfying* $\|G - G_0\|_\infty \vee \|H - H_0\|_\infty \le \delta$, *where* $\widehat{h}^+ = \max(\widehat{h}^+, 0)$.

REMARK 3.2. Since h_0 is increasing in $[a, m_0]$ and decreasing in $(m_0, b]$, by (3.1),

$$h_0(x;m) = h_0(x) I_{\{x \notin (m', m'']\}} + \frac{H_0(m'') - H_0(m')}{m'' - m'} I_{\{x \in (m', m'']\}}, \tag{3.4}$$

where $m' = m$ and $m'' = \inf\{x > m_0 : (H_0(x) - H_0(m))/(x - m) \ge h_0(x)\}$ for $m \le m_0$, and $m'' = m$ and $m' = \sup\{x < m_0 : (H_0(m) - H_0(x))/(m - x) \le h_0(x)\}$ for $m > m_0$, that is, $H_0(x;m) \equiv \int_a^x h_0(t;m)dt$ is continuous, linear in $(m', m'']$ and identical to H_0 outside $(m', m'']$. It follows that $h_0(m_0;m)$ is the average of $h_0(\cdot)$ over $(m', m'']$ and

$$\begin{aligned}\int_a^b \Big\{h_0^2(x;m) - h_0^2(x;m_0)\Big\} dG_0(x)\\ = -\int_{m'}^{m''} \big\{h_0(x) - h_0(m_0;m)\big\}^2 dG_0(x).\end{aligned} \tag{3.5}$$

REMARK 3.3. Let $[a, b] = [-1/2, 1/2]$ and $h_0(x) = -B|x|^s$ for some positive numbers B and s. Then $m_0 = 0$, $m'' = \lambda_s |m|$ for $-1/2 \le m < 0$ and $m' = -\lambda_s m$ for $0 < m \le 1/2$ in (3.4), where λ_s satisfies $\lambda_s^s(s\lambda_s + s + 1) = 1$ and $0 < \lambda_s < 1$. Moreover, by (3.5),

$$\int_{-1/2}^{1/2} \Big\{h_0^2(x;m)dx - h_0^2(x;0)\Big\} dx = -\Lambda_s B^2 |m|^{2s+1} < 0 \tag{3.6}$$

for all $0 < |m| \le 1/2$, where $\Lambda_s = (\lambda_s^s - 1)^2/(2s + 1) > 0$.

REMARK 3.4. Let $[a, b]$ and h_0 be as in Remark 3.3. If $|m| \le \beta' < \beta < 1/2$ and $\|H - H_0\|_\infty \le \delta_{\beta', \beta}$, for sufficiently small $\delta_{\beta', \beta} > 0$, then $\widehat{h}(x;m)$ has increments in both intervals $(-\beta, -\beta')$ and (β', β). This is due to (3.2) and the fact that $h_0(x;m)$ is strictly monotone in both intervals $\pm(\beta', \beta)$ for $|m| \le \beta' < \beta$ (as $\lambda_s < 1$).

PROOF OF LEMMA 3.1. Assume $a < m_0 = 0 < b$ without loss of generality.

Step 1. *Proof of* (3.2). Suppose $\|H - H_0\|_\infty \le \delta$ and $\|G - G_0\|_\infty \le \delta$ with $0 < \delta \le \delta_0/4$. We shall first show

$$\widehat{h}(x;m)\begin{cases} \le h_0(x+\delta_0;m) + C_1\delta/\delta_0 & \text{if } a \le x \le m-\delta_0, \\ \ge h_0(x-\delta_0;m) - C_1\delta/\delta_0 & \text{if } a+\delta_0 < x \le m, \\ \le h_0(x-\delta_0;m) + C_1\delta/\delta_0 & \text{if } m+\delta_0 < x \le b, \\ \ge h_0(x+\delta_0;m) - C_1\delta/\delta_0 & \text{if } m < x \le b-\delta_0. \end{cases} \tag{3.7}$$

For $a \le s < x$ and $x + \delta_0 \le t \le m$, $G(t) - G(s) \ge (t - s) - 2\delta \ge \delta_0/2$, so that

$$\begin{aligned} \left|\frac{H(t)-H(s)}{G(t)-G(s)} - \frac{H_0(t)-H_0(s)}{t-s}\right| &\le \frac{2\|H-H_0\|_\infty}{G(t)-G(s)} + \frac{2\|G-G_0\|_\infty}{G(t)-G(s)}\frac{|H_0(t)-H_0(s)|}{t-s} \\ &\le \frac{2\delta}{\delta_0/2} + \frac{2\delta}{\delta_0/2}\|h_0\|_\infty \le C_1\delta/\delta_0, \end{aligned}$$

which implies by (3.1) that

$$\begin{aligned} \widehat{h}(x;m) &\le \sup_{s<x} \inf_{x+\delta_0\le t\le m} \frac{H(t)-H(s)}{G(t)-G(s)} \\ &\le \sup_{s<x} \inf_{x+\delta_0\le t\le m} \left\{\frac{C_1\delta}{\delta_0} + \frac{H_0(t)-H_0(s)}{t-s}\right\} \le \frac{C_1\delta}{\delta_0} + h_0(x+\delta_0;m). \end{aligned}$$

This gives (3.7) for the case of $a \le x \le m - \delta_0$. The proofs for the three other cases are nearly identical and omitted.

It follows from (3.7) and the monotonicity of $h_0(\cdot;m)$ in both intervals $[a, m]$ and $(m, b]$ that $|\widehat{h}(x;m) - h_0(x;m)| \le |h_0(x+\delta_0;m) - h_0(x-\delta_0;m)| + C_1\delta/\delta_0$ for $x \in A_{m,\delta_0} = (a+\delta_0, b-\delta_0] \setminus (m-\delta_0, m+\delta_0]$. Thus, (3.2) is a consequence of

$$|h_0(x+\delta_0;m) - h_0(x-\delta_0;m)| \le h'_{0,\delta_0}(x) \qquad \forall\ x \in A_{m,\delta_0}. \tag{3.8}$$

where $h'_{0,\delta_0}(x) = \max_{|t-x|\le\delta_0} |h_0(t) - h_0(x-\delta_0)|$. Let us verify (3.8) for $a + \delta_0 < x \le m - \delta_0$. Let $(m', m'']$ be as in (3.4). If $x + \delta_0 \le m'$, then $h_0(\cdot;m) = h_0(\cdot)$ on $[a, m']$ by (3.4) and (3.8) holds automatically. If $m' < x - \delta_0$, then $h_0(x-\delta_0;m) = h_0(x+\delta_0;m)$ by (3.4) and (3.8) holds automatically. Finally, if $x - \delta_0 \le m' < x + \delta_0$, then by (3.4) $m' < m = m''$, $h_0(x-\delta_0;m) = h_0(x-\delta_0)$ and $h_0(m') \le h_0(x+\delta_0;m) \le h_0(m'+)$, so that (3.8) holds. Therefore, (3.8) holds in all the cases with $a + \delta_0 < x \le m - \delta_0$. By symmetry, (3.8) also holds for $m + \delta_0 < x \le b - \delta_0$. Thus, the proof of (3.2) is complete.

Step 2. *Proof of* (3.3). Let $0 < \delta < 1$ and $\delta_0 = \max(\sqrt{\delta}, 4\delta)$ be small constants with $a + 4\delta_0 < m^* < b - 4\delta_0$. Suppose $\|H - H_0\|_\infty \vee \|G - G_0\|_\infty \le \delta$. It suffices to show

$$\begin{aligned}(3.9)\qquad &\int_a^b \left\{ \left(\widehat{h}^-(x;m)\right)^2 - \left(\widehat{h}^-(x;m^*)\right)^2 \right\} dG(x) \\ &\quad - \int_a^b \left\{ h_0^2(x;m) - h_0^2(x;m^*) \right\} dG_0(x) \le \epsilon,\end{aligned}$$

where $\widehat{h}^-(x;m) = \max\{-\widehat{h}(x;m), 0\}$. This will be done by splitting the integrations over two regions, $I_{\delta_0} = (a+\delta_0, b-\delta_0]$ and $[a,b] \setminus I_{\delta_0}$.

Let us first consider the integrations over I_{δ_0}. Set $C_2 = \|h_0(\cdot)\|_\infty + C_1\delta/\delta_0 \le \|h_0(\cdot)\|_\infty + C_1$. By (3.1) $-\|h_0(\cdot)\|_\infty \le h_0(x;m) \le 0$. If $a+\delta_0 < x \le m$, then $\widehat{h}(x;m) \ge h_0(a;m) - C_1\delta/\delta_0$ by (3.7). If $m < x \le b-\delta_0$, then $\widehat{h}(x;m) \ge h_0(b;m) - C_1\delta/\delta_0$ by (3.7). Thus, $\widehat{h}(x;m) \ge -C_2$ in I_{δ_0}. Furthermore, since $-\{\widehat{h}^-(\cdot;m)\}^2$ is unimodal, its total variation in I_{δ_0} is bounded by $2C_2^2$. It follows from these facts that, via integrating by parts,

$$\left| \int_{I_{\delta_0}} \{\widehat{h}^-(x;m)\}^2 d(G-G_0)(x) \right| \le \|G-G_0\|_\infty (2C_2^2 + 2C_2^2) \le 4\delta C_2^2.$$

Also in I_{δ_0}, $\widehat{h}(x;m) \ge -C_2$ and (3.2) imply $\left|\{\widehat{h}^-(x;m)\}^2 - h_0^2(x;m)\right| \le 2C_2\{h'_{0,\delta_0}(x) + C_1\sqrt{\delta}\} + 2C_2^2 I\{|x-m| \le \delta_0\}$ for all m, where h'_{0,δ_0} is as in (3.8) and $\delta_0 = \max(\sqrt{\delta}, 4\delta)$. This implies

$$\left| \int_{I_{\delta_0}} \{(\widehat{h}^-(x;m))^2 - h_0^2(x;m)\} dG_0(x) \right| \le 2C_2 \int_{I_{\delta_0}} h'_{0,\delta_0} dG_0 + 2C_1C_2\sqrt{\delta} + 4C_2^2\delta_0 \to 0$$

as $\delta \to 0$ and $\delta_0 = \max(\sqrt{\delta}, 4\delta) \to 0$, by $G_0([m-\delta_0, m+\delta_0]) = 2\delta_0$ and the dominated convergence theorem. Note that $\|h'_{0,\delta_0}\|_\infty \le 2\|h_0(\cdot)\|_\infty$ and $h'_{0,\delta_0}(x) \to 0$ almost everywhere as $\delta_0 \to 0$. The above two inequalities with the integrations over I_{δ_0} give

$$(3.10)\qquad \sup_{a \le m \le b} \left| \int_{I_{\delta_0}} \{\widehat{h}^-(x;m)\}^2 dG(x) - \int_{I_{\delta_0}} h_0^2(x;m) dG_0(x) \right| \le \epsilon/4$$

for sufficiently small δ and $\delta_0 = \max(\sqrt{\delta}, 4\delta)$.

For integrations over $[a,b] \setminus I_{\delta_0} = [a, a+\delta_0] \cup (b-\delta_0, b]$, we shall first prove

$$\begin{aligned}(3.11)\qquad &\left\{\widehat{h}^-(x;m)\right\}^2 - \left\{\widehat{h}^-(x;m^*)\right\}^2 \le C_3^2, \\ &C_3 = \frac{2(\|H_0\| + \delta)}{\min(b - 2\delta_0 - m^*, m^* - a - 2\delta_0)},\end{aligned}$$

uniformly for $a \le m \le b$ and $x \in [a,b] \setminus I_{\delta_0}$. Note that $a + 4\delta_0 < m^* < b - 4\delta_0$ implies $C_3 \le 4(\|H_0\| + \delta)/\min(b - m^*, m^* - a) < \infty$. By symmetry, we shall only consider the case $a \le m \le m^*$. Let x be a fixed point in $[a,b] \setminus I_{\delta_0}$.

For $a \le x \le m \le m^*$, $\widehat{h}(x;m) \ge \widehat{h}(x;m^*)$ by (3.1), so that (3.11) holds. For $a \le m < x \le m^*$, we have $x < a + \delta_0$ and (3.1) implies

$$\widehat{h}(x;m) \ge \inf_{m<s<x} \frac{H(b)-H(s)}{G(b)-G(s)} \ge -\frac{2(\|H_0\|+\delta)}{b-a-\delta_0-2\delta} \ge -C_3.$$

Finally, for $m \le m^* < x \le b$, we have $x \ge b - \delta_0$ and (3.1) implies

$$\begin{aligned}\widehat{h}(x;m) &= \min\left\{\widehat{h}(x;m^*), \inf_{m<s\le m^*} \sup_{x\le t\le b} \frac{H(t)-H(s)}{G(t)-G(s)}\right\}\\ &\ge \min\left\{\widehat{h}(x;m^*), -\frac{2(\|H_0\|+\delta)}{b-\delta_0-m^*-2\delta}\right\}.\end{aligned}$$

Thus, (3.11) holds in all the cases. Now, (3.11) implies

$$\begin{aligned}&\int_{[a,b]\setminus I_{\delta_0}} \left\{\left(\widehat{h}^-(x;m)\right)^2 - (\widehat{h}^-(x;m^*))^2\right\} dG(x)\\ &\qquad \le C_3^2 G([a, a+\delta_0] \cup (b-\delta_0, b]) \le C_3^2\{4\delta + 2\delta_0\} \le \epsilon/4,\end{aligned}$$

for sufficiently small δ with $\delta_0 = \max(\sqrt{\delta}, 4\delta)$. This and (3.10) imply (3.9). □

3.2. *Upper bound for the magnitude of the spike.* In this section we provide upper bounds for the tail probability of (2.8) via (2.4).

LEMMA 3.2. *Suppose the errors ε_i are independent random variables with $E\varepsilon_i = 0$ and $P\{|\varepsilon_i| > t\} \le \mathscr{M} \exp(-ct^\alpha)$ for all $t \ge 0$, where α, $\mathscr{M}$ and c are positive numbers. Let $\gamma = \max\{1, 1/2 + 1/\alpha\}$. Then, there exists a constant $\omega < \infty$ such that for all $\xi > 0$,*

$$\begin{aligned}P\left\{\sup_m \sum_{j=1}^n \left[\{\widehat{f}_n(X_j;m) - f_0(m_0)\}^+\right]^2 \ge [\omega\{1+\log n + \xi\}]^{2\gamma}\right\}\\ \le \exp(-\xi).\end{aligned} \tag{3.12}$$

Let $\{a_i\}$ be a sequence of real numbers. For $\vec{x}_n = (x_1, \ldots, x_n)$ and $\vec{\ell}_n = (\ell_1, \ldots, \ell_n)$ define

$$g_{\vec{\ell}_n}(\vec{x}_n) = \langle g_{\ell_1}(\vec{x}_n), \ldots, g_{\ell_n}(\vec{x}_n)\rangle \in \mathbb{R}^n$$

with $g_\ell(\vec{x}_n) = g_\ell(x_1, \ldots, x_n) = \sum_{i=1}^\ell x_i a_i/\ell$. Furthermore, define

$$h_n(\vec{x}_n) = \frac{\max\{\|g_{\vec{\ell}_n}(\vec{x}_n)\| : \vec{\ell}_n \in \Omega_n\}}{\sqrt{2}S_n}, \qquad S_n = \left\{\sum_{k=1}^n a_k^2/k\right\}^{1/2}, \tag{3.13}$$

where $\|\vec{x}_n\|$ is the Euclidean distance and $\Omega_n = \{(\ell_1, \ldots, \ell_n) : k \le \ell_k \le n, k = 1, \ldots, n\}$. The following lemma is used in the proof of Lemma 3.2.

LEMMA 3.3 (i) *Let $\{\eta_i\}$ be a Rademacher sequence, iid variables with $P\{\eta_i = \pm 1\} = 1/2$. Then, for all $t \geq 0$ and $\vec{a}_n$*

$$P\big\{h_n(\eta_1, \dots, \eta_n) > 16/\sqrt{3} + t\big\} \leq 2\exp\big(-t^2/16\big). \tag{3.14}$$

(ii) *Let T'_j be as in* (2.5). *Then, there exists $M_0 < \infty$ such that for all $t \geq 0$,*

$$\begin{aligned} P\bigg\{ & T'_j > \big(16/\sqrt{3} + t\big)\sqrt{2M_0 v^{(2-\alpha)^+}[t^2/16 + 1 + \log n]},\ \max_{1\leq i\leq n} |\varepsilon_{i+j}| \leq v\bigg\} \\ & \leq 3\exp\big(-t^2/16\big). \end{aligned} \tag{3.15}$$

PROOF. (i) Let $\vec{a}_n$ be fixed and $\tilde{\eta}_i = \eta_i/2$. Since $h_n(\eta_1, \dots, \eta_n) = 2h_n(\tilde{\eta}_1, \dots, \tilde{\eta}_n)$, (3.14) is an immediate consequence of

$$P\big\{h_n(\tilde{\eta}_1, \dots, \tilde{\eta}_n) > M + t\big\} \leq 2\exp\big(-t^2/4\big) \tag{3.16}$$

with $M = 8/\sqrt{3}$. Since $\tilde{\eta}_i$ are iid variables living in a unit cube, (3.16) follows from (2.4) under the following conditions:

(i) $h_n(\vec{x}_n)$ is convex in $\vec{x}_n$;
(ii) $\|h_n\|_{Lip} \leq 1$; and
(iii) $\operatorname{median}(h_n(\tilde{\eta}_1, \dots, \tilde{\eta}_n)) \leq M$. We check these three conditions below.

First, since $g_{\vec{\ell}_n}(\vec{x}_n)$ is a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$, $\|g_{\vec{\ell}_n}(\vec{x}_n)\|/S_n$ is convex. Since maximum of convex functions is convex, $h_n(\vec{x}_n)$ is convex.

Second, we show $\|h_n\|_{Lip} \leq 1$. For $\ell_k \geq k$, $1 \leq k \leq n$,

$$\sum_{k=1}^{n}\sum_{i=1}^{\ell_k} \frac{a_i^2}{\ell_k^2} \leq \sum_{i=1}^{n}\sum_{k=1}^{n} \frac{a_i^2}{i^2 \vee k^2} \leq \sum_{i=1}^{n}\left\{\frac{a_i^2}{i} + a_i^2 \sum_{k=i+1}^{n} \frac{1}{k^2}\right\} \leq \sum_{i=1}^{n} \frac{2a_i^2}{i}. \tag{3.17}$$

Since $|g_\ell(\vec{x}_n)|^2 \leq \|\vec{x}_n\|^2 \sum_{i=1}^{\ell} a_i^2/\ell^2$ and $\ell_k \geq k$ in the vector $\vec{\ell}_n \in \Omega_n$, by (3.17),

$$\|g_{\vec{\ell}_n}(\vec{x}_n)\|^2 = \sum_{k=1}^{n} |g_{\ell_k}(\vec{x}_n)|^2 \leq \|\vec{x}_n\|^2 \sum_{k=1}^{n}\sum_{i=1}^{\ell_k} \frac{a_i^2}{\ell_k^2} \leq \|\vec{x}_n\|^2 \sum_{i=1}^{n} \frac{2a_i^2}{i}.$$

Thus, the norm of the linear mapping $g_{\vec{\ell}_n}(\vec{x}_n)/(\sqrt{2}S_n)$ is bounded by one, which implies that the Lipschitz norm of $h^*_{\vec{\ell}_n}(\vec{x}_n) = \|g_{\vec{\ell}_n}(\vec{x}_n)\|/(\sqrt{2}S_n)$ is bounded by one. Since h_n is the maximum of $h^*_{\vec{\ell}_n}$ over Ω_n, $\|h_n\|_{Lip} \leq \max_{\vec{\ell}_n\in\Omega_n} \|h^*_{\vec{\ell}_n}\|_{Lip} \leq 1$.

Third, we prove $\operatorname{median}(h_n(\tilde{\eta}_1, \dots, \tilde{\eta}_n)) \leq M = 8/\sqrt{3}$. Since $h_n(\tilde{\eta}_1, \dots, \tilde{\eta}_n) \geq 0$, we have $P\{h_n(\tilde{\eta}_1,\dots,\tilde{\eta}_n) \geq 2Eh_n(\tilde{\eta}_1,\dots,\tilde{\eta}_n)\} \leq 1/2$. Since $h_n(\eta_1,\dots,\eta_n) = 2h_n(\tilde{\eta}_1, \dots, \tilde{\eta}_n)$, it suffices to prove $Eh_n(\eta_1, \dots, \eta_n) \leq M = 8/\sqrt{3}$. By definition, $Eh_n^2(\eta_1, \dots, \eta_n) = E_\eta/(2S_n^2)$, where

$$E_\eta = E \max_{(\ell_1,\dots,\ell_k)\in\Omega_n} \sum_{k=1}^{n}\left(\sum_{i=1}^{\ell_k} \frac{\eta_i a_i}{\ell_k}\right)^2 = \sum_{k=1}^{n} E \max_{k\leq\ell\leq n}\left(\sum_{i=1}^{\ell} \frac{\eta_i a_i}{\ell}\right)^2.$$

By the Doob inequality we have

$$
\begin{aligned}
E \max_{k\le\ell\le n}\left(\sum_{i=1}^{\ell}\frac{\eta_i a_i}{\ell}\right)^2 &\le \sum_{m=0}^{\infty} E \max_{2^m k\le\ell\le(2^{m+1}k)\wedge n}\left(\frac{\sum_{i=1}^{\ell}\eta_i a_i}{2^m k}\right)^2\\
&\le 4\sum_{m=0}^{\infty}\sum_{i=1}^{(2^{m+1}k)\wedge n}\frac{a_i^2}{(2^m k)^2}\\
&\le 4\sum_{i=1}^{n} a_i^2\sum_{m=0}^{\infty}(2^m k)^{-2} I_{\{2^{m+1}k\ge i\}} \le \frac{64}{3}\sum_{i=1}^{n}\frac{a_i^2}{(i\vee 2k)^2},
\end{aligned}
$$

which implies by (3.17) that

$$
E_\eta \le \sum_{k=1}^{n}\frac{64}{3}\sum_{i=1}^{n}\frac{a_i^2}{(i\vee k)^2} \le \frac{64}{3}\sum_{i=1}^{n}\frac{2a_i^2}{i} = 2M^2 S_n^2.
$$

Hence, $\text{median}(h_n(\tilde\eta_1,\dots,\tilde\eta_n)) \le 2\sqrt{Eh_n^2(\tilde\eta_1,\dots,\tilde\eta_n)} = \sqrt{Eh_n^2(\eta_1,\dots,\eta_n)} = \sqrt{E_\eta/(2S_n^2)} \le M$, and the proof of (3.14) is complete.

(ii) We shall only prove (3.15) for $j=0$. Set $S_n^2 = \sum_{k=1}^{n}\varepsilon_k^2/k$. By (2.5) and (3.13) $T_0' = h_n(\eta_1,\dots,\eta_n)\sqrt{2}S_n$ given $\vec\varepsilon_n = \vec a_n$. By (3.14),

$$
\begin{aligned}
&P\left\{T_0' > (16/\sqrt{3}+t)\sqrt{2M_0 v^{(2-\alpha)^+}(t^2/16+1+\log n)},\ \max_{1\le i\le n}|\varepsilon_i|\le v\right\}\\
&\quad\le \left\| P\left\{h_n(\eta_1,\dots,\eta_n) > 16/\sqrt{3}+t \;\middle|\; \vec\varepsilon_n=\vec a_n\right\}\right\|_\infty\\
&\qquad + P\left\{2S_n^2 > 2M_0 v^{(2-\alpha)^+}(t^2/16+1+\log n),\ \max_{1\le i\le n}|\varepsilon_i|\le v\right\}\\
&\quad\le 2\exp\left(-t^2/16\right) + P\left\{\sum_{i=1}^{n}|\varepsilon_i|^{\alpha\wedge 2}/(M_0 i) > t^2/16+1+\log n\right\}.
\end{aligned}
\tag{3.18}
$$

Note that $S_n^2 \le v^{(2-\alpha)^+}\sum_{i=1}^{n}|\varepsilon_i|^{\alpha\wedge 2}/i$ on the event $\{\max_{1\le i\le n}|\varepsilon_i|\le v\}$. Since our condition on the tail probability of ε_i is uniform over all $\{\varepsilon_i\}$, there exists a large $M_0<\infty$ such that $\sup_i E\exp\{|\varepsilon_i|^{\alpha\wedge 2}/(M_0 k)\} \le 1+1/k \le e^{1/k}$ for all k. Thus,

$$
\begin{aligned}
&P\left\{\sum_{i=1}^{n}|\varepsilon_i|^{\alpha\wedge 2}/(M_0 i) > t^2/16+1+\log n\right\}\\
&\qquad\le \exp\left(-t^2/16-1-\log n\right)\exp\left(\sum_{k=1}^{n}1/k\right) \le \exp\{-t^2/16\}.
\end{aligned}
$$

Inserting the above inequality into (3.18), we obtain (3.15). □

Proof of Lemma 3.2. Let $T_j = \left[\sum_{k=1}^{n-j}\max_{k\le l\le n-j}\left\{\left(\sum_{i=1}^{l}\varepsilon_{i+j}/l\right)^+\right\}^2\right]^{1/2}$. The second term on the right-hand side of (2.8) is bounded by $\max_{0\le j<n}T_j^2$

with $\{\varepsilon_i\}$ replaced by $\{\varepsilon'_i\}$ and the first term is bounded by $\max_{0\le j<n} T_j^2$ with $\{\varepsilon_i\}$ replaced by $\{\varepsilon'_{n+1-i}\}$. Since $\{X_i\}$ and $\{\varepsilon_i\}$ are independent and the conditions on ε_i are uniform in i, it suffices to prove

$$P\Big\{\max_j T_j^2 > 4C_{n,\xi}^2 + 1\Big\} \le E\Big(\max_j T_j^2 - 4C_{n,\xi}^2\Big)^+ \le \exp\big(-\xi\big), \tag{3.19}$$

where $C_{n,\xi} = [\omega\{1 + \log n + \xi\}]^\gamma$ for some $\omega \in [1, \infty)$. Since the maximum and sum of convex functions are convex, T_j^2 and $\max\big\{\max_j T_j^2 - 4C_{n,\xi}^2, 0\big\}$ are convex functions of $\{\varepsilon_i\}$, so that by standard symmetrization methods

$$\begin{aligned} &E\Big\{\max_{0\le j<n} T_j^2 - 4C_{n,\xi}^2\Big\}^+ \\ &\quad\le E\left[\max_{0\le j<n}\sum_{k=1}^{n-j}\max_{k\le \ell\le n-j}\left\{\left(\sum_{i=1}^{\ell}\eta_{i+j}(\varepsilon_{i+j} - \tilde\varepsilon_{i+j})/\ell\right)^+\right\}^2 - 4C_{n,\xi}^2\right]^+ \\ &\quad\le E\left[\max_{0\le j<n}(T'_j + \tilde T'_j)^2 - 4C_{n,\xi}^2\right]^+ \\ &\quad\le 4E\left[\max_{0\le j<n}(T'_j)^2 - C_{n,\xi}^2\right]^+ \end{aligned} \tag{3.20}$$

where $\{\tilde\varepsilon_i\}$ is an independent copy of $\{\varepsilon_i\}$, $\{\eta_i\}$ is a Rademacher sequence independent of $(\{\varepsilon_i\}, \{\tilde\varepsilon_i\})$, T'_j are the $\{\varepsilon_i\}$ version of (2.5) and $\tilde T'_j$ are the $\{\tilde\varepsilon_i\}$ version of (2.5).

Let $\beta = 1/(\alpha\gamma)$. We split the expectation on the right-hand side of (3.20) into two parts:

$$\begin{aligned} E\Big[\max_{0\le j<n}(T'_j)^2 - C_{n,\xi}^2\Big]^+ &= \int_{C_{n,\xi}}^\infty P\Big\{\max_{0\le j<n} T'_j > u\Big\}du^2 \\ &\le \int_{C_{n,\xi}}^\infty P\Big\{\max_{1\le i\le 2n-1}|\varepsilon_i| > u^\beta\Big\}du^2 \\ &\quad + \int_{C_{n,\xi}}^\infty \sum_{j=0}^{n-1} P\Big\{T'_j > u, \max_{1\le i\le n}|\varepsilon_{i+j}| \le u^\beta\Big\}du^2. \end{aligned} \tag{3.21}$$

It follows from our assumption on the tail probability of ε_i that

$$\begin{aligned} &\int_{C_{n,\xi}}^\infty P\Big\{\max_{1\le i\le 2n-1}|\varepsilon_i| > u^\beta\Big\}du^2 \\ &\quad\le 2n\int_{C_{n,\xi}}^\infty \mathscr{M}\exp\big\{-cu^{\alpha\beta}\big\}du^2 \\ &\quad\le 2n\mathscr{M}\exp\big\{-(c/2)C_{n,\xi}^{\alpha\beta}\big\}\int_1^\infty \exp\big\{-(c/2)u^{\alpha\beta}\big\}du^2 \\ &\quad= 2n\mathscr{M}\exp\big\{-(c/2)\omega(1 + \log n + \xi)\big\}\int_1^\infty \exp\big(-(c/2)u^{1/\gamma}\big)du^2 \\ &\quad\le e^{-\xi}/8 \end{aligned} \tag{3.22}$$

for sufficiently large $\omega < \infty$. Let $(t, v) = (c_0 u^{1/(2\gamma)}, u^\beta)$ in (3.15) with a small $c_0 > 0$. Let $\omega > 16/c_0^2$. Since $C_{n,\xi}^{1/\gamma} \ge \omega(1 + \log n) \ge 1$, $t^2/16 + 1 + \log n \le c_0^2 u^{1/\gamma}/16 + C_{n,\xi}^{1/\gamma}/\omega \le c_0^2 u^{1/\gamma}/8$ for $u \ge C_{n,\xi}$. Thus, for $u \ge C_{n,\xi}$ the quantity in (3.15) is bounded by

$$\begin{aligned}(16/\sqrt{3} + t)&\sqrt{2M_0 v^{(2-\alpha)^+}(t^2/16 + 1 + \log n)}\\ &\le \bigl(16/\sqrt{3} + c_0\bigr)u^{1/(2\gamma)}\sqrt{2M_0 u^{\beta(2-\alpha)^+} c_0^2 u^{1/\gamma}/8}\\ &\le u^{1/\gamma + \beta(2-\alpha)^+/2} = u\end{aligned}$$

for sufficiently small $c_0 > 0$ and $\omega = \omega_{c_0} > 16/c_0^2$, since $\beta = 1/(\alpha\gamma)$ and $\gamma = \max(1, 1/2 + 1/\alpha)$. It follows from Lemma 3.3(ii) that

$$\begin{aligned}\int_{C_{n,\xi}}^\infty &\sum_{j=0}^{n-1} P\Bigl\{T'_j > u,\ \max_{1\le i\le n}|\varepsilon_{i+j}| \le u^\beta\Bigr\}du^2\\ &\le \int_{C_{n,\xi}}^\infty 3n\exp\biggl(-\frac{c_0^2 u^{1/\gamma}}{16}\biggr)du^2 \le \frac{e^{-\xi}}{8}\end{aligned} \tag{3.23}$$

for sufficiently large ω. The second inequality above holds from the argument in (3.22) after the first inequality of (3.22). Inserting (3.22) and (3.23) into (3.21), we find that the right-hand side of (3.20) is bounded by $e^{-\xi}$. This gives (3.19) and ends the proof. $\square$

PROOF OF THEOREM 1.1. As mentioned in the beginning of the section, we assume that $\{X_i\}$ are iid uniform variables from (a, b). We shall further assume without loss of generality that $f_0(m_0) = 0$, $m_0 = 0$, $a < 0 < b$ and $b - a = 1$. Let $k_0 \ge 1$ be the smallest positive integer satisfying $a \le -1/2^{(k_0+1)} < 1/2^{(k_0+1)} \le b$. For $k \ge k_0$, define

$$F_{n,k}(x) = \frac{1}{n_k \vee 1}\sum_{i=1}^n 2^{ks} Y_i I\{-1/2 \le 2^k X_i \le x\},\ -1/2 \le x \le 1/2, \tag{3.24}$$

$$G_{n,k}(x) = \frac{1}{n_k \vee 1}\sum_{i=1}^n I\{-1/2 \le 2^k X_i \le x\},\ -1/2 \le x \le 1/2, \tag{3.25}$$

with $n_k = \sum_{i=1}^n I\{|X_i| \le 1/2^{(k+1)}\}$, and define

$$\widehat{m}_{n,k} = \arg\max_m \int_{-1/2}^{1/2} \widehat{f}_{n,k}^{\,2}(x;m)\,dG_{n,k}(x),\quad \widehat{f}_{n,k}(\cdot;m) = D_m\bigl(F_{n,k}\big|G_{n,k}\bigr), \tag{3.26}$$

where the "arg max" is taken over all m in the set $\{2^k X_i : |X_i| \le 2^{-k-1}\}$. Thus, $|\widehat{m}_{n,k}| \le 1/2$. Let $F_\infty(x) = \int_{-1/2}^x f_\infty(t)\,dt$ with $f_\infty(t) = -B|t|^s$ [the special $h_0(t)$ in Remarks 3.2 and 3.3 of Subsection 3.1]. Let $G_0(x)$ be the uniform distribution on $[-1/2, 1/2]$.

To obtain the conclusion (1.4), it suffices to prove

$$P\Big\{2^{p_n+1}|\widehat{m}_n|>1,\ \text{i.o.}\Big\}=0,\quad 2^{p_n}\le\left(\frac{n}{M(1+\log n)^{2\gamma}}\right)^{1/(2s+1)}<2^{p_n+1}\tag{3.27}$$

for some large $M<\infty$. This is done by proving

$$P\Big\{\widehat{m}_n\ne 2^{-k_1}\widehat{m}_{n,k_1}\ \text{i.o.}\Big\}=0\tag{3.28}$$

for some large $k_1\ge k_0$ and that

$$\sum_{n=1}^{\infty}P\left\{\bigcup_{k=k_1}^{p_n}(D^c_{n,k,\Delta}\cup E^c_{n,k,\delta})\right\}<\infty\tag{3.29}$$

where

$$E_{n,k,\delta}=\big\{\|F_{n,k}-F_\infty\|_\infty\vee\|G_{n,k}-G_0\|_\infty\le\delta\big\}\tag{3.30}$$

for some small $\delta>0$, and for some small $\Delta>0$,

$$D_{n,k,\Delta}=\left\{\int_{-1/2}^{1/2}\big\{\widehat{f}^{+}_{n,k}(x;\widehat{m}_{n,k})\big\}^2 dG_{n,k}(x)<\Delta\right\}.\tag{3.31}$$

The details are given in two steps.

Step 1. *Proof of* (3.27) *based on* (3.28) *and* (3.29). It suffices to show

$$\big\{E_{n,k,\delta}\cap D_{n,k,\Delta}\big\}\cap\big\{E_{n,k+1,\delta}\cap D_{n,k+1,\Delta}\big\}\subseteq\big\{\widehat{m}_{n,k}=\widehat{m}_{n,k+1}/2\big\}\tag{3.32}$$

for $k_1\le k\le p_n$. This implies (3.27) based on (3.28) and (3.29), since it implies

$$\big\{\widehat{m}_n\ne 2^{-p_n}\widehat{m}_{n,p_n}\big\}\subseteq\big\{\widehat{m}_n\ne 2^{-k_1}\widehat{m}_{n,k_1}\big\}\cup\left[\bigcup_{k=k_1}^{p_n}(E^c_{n,k,\delta}\cup D^c_{n,k,\Delta})\right].$$

Note that $|\widehat{m}_{n,k}|\le 1/2$ by (3.26) and $P\big\{2^{p_n}\widehat{m}_n\ne\widehat{m}_{n,p_n}\ \text{i.o.}\big\}=0$ by (3.28), (3.29) and the above argument.

Let us prove (3.32). Let $h_0=f_\infty$ in Lemma 3.1 and $f_\infty(\cdot;m)=D_m(F_\infty|G_0)$. Let $\delta>0$ be sufficiently small so that by Remark 3.4 $\widehat{f}^{\,2}_{n,k}(x;m)$ has jumps in all four intervals $\pm(1/16,1/8)$ and $\pm(1/8,1/4)$ for all $|m|\le 1/16$ on $E_{n,k,\delta}$, so that by (3.1) and Remark 3.1 the following two identities hold: (a) $\widehat{f}_{n,k}(x;m)=\widehat{f}_{n,k}(x;0)$ for all $|x|\ge 1/8$ and $|m|\le 1/16$; and (b) $\widehat{f}_{n,k+1}(2x;2m)=\widehat{f}_{n,k}(x;m)$ for all $|x|\le 1/8$ and $|m|\le 1/32$. These imply

$$\begin{aligned}\operatorname*{arg\,max}_{|m|\le 1/32}\int_{-1/2}^{1/2}\widehat{f}^{\,2}_{n,k}(x;m)dG_{n,k}&=\operatorname*{arg\,max}_{|m|\le 1/32}\int_{-1/8}^{1/8}\widehat{f}^{\,2}_{n,k}(x;m)dG_{n,k}\\&=\tfrac12\operatorname*{arg\,max}_{|m|\le 1/16}\int_{-1/4}^{1/4}\widehat{f}^{\,2}_{n,k+1}(x;m)dG_{n,k+1}\\&=\tfrac12\operatorname*{arg\,max}_{|m|\le 1/16}\int_{-1/2}^{1/2}\widehat{f}^{\,2}_{n,k+1}(x;m)dG_{n,k+1}\end{aligned}\tag{3.33}$$

on $E_{n,k,\delta} \cap E_{n,k+1,\delta}$. The last equation above follows from the first identity (a) for $k+1$. Note that by (3.24) and (3.25),

$$\frac{F_{n,k+1}(2t) - F_{n,k+1}(2s)}{G_{n,k+1}(2t) - G_{n,k+1}(2s)} = 2^s \frac{F_{n,k}(t) - F_{n,k}(s)}{G_{n,k}(t) - G_{n,k}(s)}$$

for $|s| \vee |t| \le 1/4$, so that the second identity (b) above follows from (3.1).

Let $\Delta > 0$ in (3.31) and $\epsilon > 0$ be sufficiently small such that $\epsilon + \Delta < \Lambda_s B^2 (1/32)^{2s+1}$. It follows from (3.26), (3.3), (3.31) and (3.6) that, for $\delta \le \delta_{\epsilon,0,f_\infty}$ as in Lemma 3.1,

$$\begin{aligned} 0 &\le \int_{-1/2}^{1/2} \Big\{ \widehat{f}^{\,2}_{n,k}(x; \widehat{m}_{n,k}) - \widehat{f}^{\,2}_{n,k}(x;0) \Big\} dG_{n,k}(x) \\ &\le \epsilon + \Delta + \int_{-1/2}^{1/2} \Big\{ f^2_\infty(x; \widehat{m}_{n,k}) - f^2_\infty(x;0) \Big\} dG_0(x) = \epsilon + \Delta - \Lambda_s B^2 \big|\widehat{m}_{n,k}\big|^{2s+1} \end{aligned}$$

on $D_{n,k,\Delta} \cap E_{n,k,\delta}$. Since $\epsilon + \Delta < \Lambda_s B^2 (1/32)^{2s+1}$, $\big|\widehat{m}_{n,k}\big| < 1/32$ on $D_{n,k,\Delta} \cap E_{n,k,\delta}$. This and (3.33) imply $\widehat{m}_{n,k} = 2^{-1} \widehat{m}_{n,k+1}$ on $E_{n,k,\delta} \cap D_{n,k,\Delta} \cap E_{n,k+1,\delta} \cap D_{n,k+1,\Delta}$ for sufficiently small $\delta > 0$ and $\Delta > 0$. Thus, the proof of (3.32) is complete.

Step 2. *Proofs of* (3.28) *and* (3.29). In addition to Lemma 3.2, the exponential inequality in Lemma 3.4 below is used in the proof. The proof of Lemma 3.4 is omitted since it is simpler than that of Lemma 3.2 and it follows from standard methods [cf., e.g., Pollard (1984)].

LEMMA 3.4. *Let X_i, ε_i be as in* (1.1) *satisfying the tail probability condition on ε_i in Theorem* 1.1. *Let $N = \#\{i \le n : X_i \in A\}$ with a Borel set A. Let ϕ_i be Borel functions such that $\big|\phi_i(x,t)\big| \le 1 + |t|$ and $E\big[\phi_i(X_i, \varepsilon_i)\big|X_i \in A\big] = 0$. Let $\gamma = \max(1, 1/2 + 1/\alpha)$ be as in Theorem* 1.1. *Then*,

$$P\bigg\{ \sup_x \bigg| \sum_{X_i \in A} \phi_i(X_i, \varepsilon_i) I\{X_i \le x\} \bigg| > \sqrt{N} \omega \big\{1 + \log N + \xi\big\}^\gamma \bigg| N \bigg\} \le e^{-\xi}$$

for all $\xi > 0$, provided that ω is a sufficiently large constant.

Let us first prove (3.29). Consider the events in (3.30). Let n_k be as in (3.25). Since n_k are binomial variables with parameters $(n, 1/2^k)$,

$$P\big\{n_k \le n/2^{k+1}\big\} \le E \exp\big(n/2^{k+1} - n_k\big) \le \exp\big\{ - n(1/2 - 1/e)/2^k \big\},$$

so that by the definition of p_n in (3.27)

$$\sum_{n=1}^\infty \sum_{k=1}^{p_n} P\big\{n_k \le n/2^{k+1}\big\} \le \sum_{n=1}^\infty p_n \exp\big\{ - n(1/2 - 1/e)/2^{p_n} \big\} < \infty \tag{3.34}$$

for all choices of $M < \infty$ in (3.27). Let $\mu_{n,k}(x) = E\big[F_{n,k}(x)\big|n_k\big]$. Set $A = [-1/2^{k+1}, 1/2^{k+1}]$, $N = n_k$ and $\phi_i(X_i, \varepsilon_i) = Y_i - \mu_{n,k}(X_i)$ in Lemma 3.4. Then, conditions of Lemma 3.4 follows from those of Theorem 1.1 and $(n_k/2^{ks})$

$\{F_{n,k}(x)-\mu_{n,k}(x)\}=\sum_{X_i\in A}\phi_i(X_i,\varepsilon_i)I\{X_i\le x\}$ by (3.24). Thus, by Lemma 3.4,

$$P\big\{\|F_{n,k}-\mu_{n,k}\|_\infty\ge\delta/2,\ n_k>n/2^{k+1}\big\}\le\exp(-2\log n)=1/n^2, \tag{3.35}$$

provided that $\sqrt{n_k}\,\omega\{1+\log n_k+\xi\}^\gamma\le(n_k/2^{ks})\delta/2$ for $n_k>n/2^{k+1}$ and $\xi=2\log n$. By (3.27), $2^{k(2s+1)}\le n(1+\log n)^{-2\gamma}/M$ for $k\le p_n$, so that $\omega\{1+3\log n\}^\gamma\le\sqrt{n/2^{k+1}}2^{-ks}\delta/2$ for $M>(2\omega 3^\gamma/\delta)^2$. Thus, (3.35) holds. It follows from (3.24) that for $n_k>0$,

$$\mu_{n,k}=E\Big[2^{ks}f(X_i)I\{2^kX_i\le x\}\big||X_i|\le 1/2^{k+1}\Big]=2^{ks}\int_{-1/2}^{x}f(t/2^k)dt,$$

which converge uniformly to $F_\infty(x)=\int_{-1/2}^{x}(-Bt^s)dt=\int_{-1/2}^{x}f_\infty(t)dt$ by our conditions on f. Choosing k_1 large enough to ensure $\|\mu_{n,k}-F_\infty\|_\infty\le\delta/2$ for the given δ, we find by (3.35),

$$\sum_{n=1}^{\infty}\sum_{k=k_1}^{p_n}P\big\{\|F_{n,k}-F_\infty\|_\infty\ge\delta,\ n_k>n/2^{k+1}\big\}\le\sum_{n=1}^{\infty}p_n/n^2<\infty.$$

This and (3.34) give $\sum_{n=1}^{\infty}\sum_{k=k_1}^{p_n}P\big\{\|F_{n,k}-F_\infty\|_\infty\ge\delta\big\}<\infty$. We omit the proof of $\sum_{n=1}^{\infty}\sum_{k=k_1}^{p_n}P\big\{\|G_{n,k}-G_0\|_\infty\ge\delta\big\}<\infty$, as it is similar. Thus, we obtain

$$\sum_{n=1}^{\infty}\sum_{k=k_1}^{p_n}P\big\{E^c_{n,k,\delta}\big\}<\infty. \tag{3.36}$$

Let us consider $D_{n,k,\Delta}$. Let $F_n(x)=\sum_{i=1}^{n}Y_iI\{X_i\le x\}$ and G_n be the empirical distribution function of $\{X_i\}$. It follows from (3.1), (3.26) and (2.3) that

$$\widehat{f}_{n,k}(x;m)\le 2^{ks}D_{m/2^k}(F_n|G_n)(x/2^k)=2^{ks}\widehat{f}_n(x/2^k;m/2^k)$$

for all $|x|\le 1/2$ and $|m|\le 1/2$, so that by (3.25),

$$\int_{-1/2}^{1/2}\big\{\widehat{f}^+_{n,k}(x;m)\big\}^2dG_{n,k}\le\frac{2^{2ks}}{n_k\vee 1}\sup_m\sum_{j=1}^{n}\big\{\widehat{f}^+_n(X_j;m)\big\}^2.$$

This and the definition of p_n in (3.26) imply

$$\begin{aligned}\bigcup_{k=1}^{p_n}\Big(D^c_{n,k,\Delta}\cap\{n_k>n/2^{k+1}\}\Big)&\subseteq\left\{\sup_{k\le p_n}\frac{2^{k(2s+1)}}{n/2}\sup_m\sum_{j=1}^{n}\big\{\widehat{f}^+_n(x_j;m)\big\}^2>\Delta\right\}\\&\subseteq\left\{\sup_m\sum_{j=1}^{n}\big\{\widehat{f}^+_n(x_j;m)\big\}^2>2M\Delta(1+\log n)^{2\gamma}\right\}.\end{aligned}$$

It then follows from (3.34) and Lemma 3.2 with $\xi=2\log n$ that for $M>(2\Delta)^{-1}(3\omega)^{2\gamma}$,

$$\sum_{n=1}^{\infty}P\left\{\bigcup_{k=k_1}^{p_n}D^c_{n,k,\Delta}\right\}\le\sum_{n=1}^{\infty}\exp(-2\log n)+\sum_{n=1}^{\infty}\sum_{k=k_1}^{p_n}P\{n_k\le n/2^{k+1}\}<\infty.$$

This and (3.36) imply (3.29).

Finally let us prove (3.28). Since $\widehat{f}_n(\cdot; m) = D_m(F_n|G_n)$, the proof is nearly identical to Step 1 and we shall omit detains. By the standard results in empirical process theory [cf., e.g., Pollard (1984)], $\|F_n - F\|_\infty \to 0$ and $\|G_n - G_X\|_\infty \to 0$, where $F(x) = \int_a^x f(t)dt$. Thus, Lemma 3.1 applies with $h_0 = f$ and $G_0 = G_X \sim$ Unif[a,b]. The conclusion follows since the contribution of the spike to the sum of squares are controlled by Lemma 3.2. □

REFERENCES

AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York.

BICKEL, P. J. and FAN, J. (1996). Some problems on the estimation of unimodal densities. *Statist. Sinica* **3** 23–45.

BIRGÉ, L. (1989). The Grenander estimator: A nonasymptotic approach. *Ann. Statist.* **17** 1532–1549.

BIRGÉ, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25** 970–981.

BROWN, L. D. and LOW, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* **24** 2524–2535.

BRUNK, H. D. (1970). Estimation of isotonic regression (with discussion by Ronald Pyke). In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 177–197. Cambridge Univ. Press.

CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Stat. Math.* **16** 31–41.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence, II. *Ann. Statist.* **19** 633–667.

EDDY, W. F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.* **8** 870–882.

GILL, R. D. and LEVIT, B. Y. (1995). Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli* **1** 59–79.

GRENANDER, U. (1956). On the theory of mortality measurement, Part II. *Skand. Akt.* **39** 125–153.

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2** 539-555. Wadsworth, Belmont, CA.

GRUND, B. and HALL, P. (1995). On the minimisation of L^p error in mode estimation. *Ann. Statist.* **23** 2264–2284.

HASMINSKII, R. Z. (1979). Lower bound for the risks of nonparametric estimates of the mode. In *Contribution to Statistics: Hajek Memorial Volume* (J. Jureckova, ed.) 91–97. Academia, Prague.

LEDOUX, M. (1996). On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics* **1** 63–87.

LEPSKII, O. V. (1992). On problems of adaptive estimation in white Gaussian noise. In *Advances in Soviet Mathematics* **12**. Amer. Math. Soc., Providence, RI.

MÜLLER, H.-G. (1989). Adaptive nonparametric peak estimation. *Ann. Statist.* **17** 1053–1069.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

VAN EEDEN, C. (1956). Maximum likelihood estimation of ordered probabilities. *Proc. K. ned. Akad. Wet A* **59**/*Indag. Math.* **18** 444–455.

VENTER, J. H. (1967). On estimation of the mode. *Ann. Math. Statist.* **38** 1446-1455.

WEGMAN, E. J. (1970). Maximum likelihood estimation of a unimodal density function. *Ann. Math. Statist.* **41** 457–471.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
HILL CENTER
BUSCH CAMPUS
PISCATAWAY, NEW JERSEY 08854
E-MAIL: cunhui@stat.rutgers.edu