The Annals of Statistics
1995, Vol. 23, No. 2, 525–550

FINITE-SAMPLE CONFIDENCE ENVELOPES FOR SHAPE-RESTRICTED DENSITIES[1]

BY NICOLAS W. HENGARTNER AND PHILIP B. STARK

University of California, Berkeley

A conservative finite-sample simultaneous confidence envelope for a density can be found by solving a finite set of finite-dimensional linear programming problems if the density is known to be monotonic or to have at most k modes relative to a positive weight function. The dimension of the problems is at most $(n/\log n)^{1/3}$, where n is the number of observations. The linear programs find densities attaining the largest and smallest values at a point among cumulative distribution functions in a confidence set defined using the assumed shape restriction and differences between the empirical cumulative distribution function evaluated at a subset of the observed points. Bounds at any finite set of points can be extrapolated conservatively using the shape restriction. The optima are attained by densities piecewise proportional to the weight function with discontinuities at a subset of the observations and at most five other points. If the weight function is constant and the density satisfies a local Lipschitz condition with exponent ϱ, the width of the bounds converges to zero at the optimal rate $(\log n/n)^{\varrho/(1+2\varrho)}$ outside every neighborhood of the set of modes, if a "bandwidth" parameter is chosen correctly. The integrated width of the bounds converges at the same rate on intervals where the density satisfies a Lipschitz condition if the intervals are strictly within the support of the density. The approach also gives algorithms to compute confidence intervals for the support of monotonic densities and for the mode of unimodal densities, lower confidence intervals on the number of modes of a distribution and conservative tests of the hypothesis of k-modality. We use the method to compute confidence bounds for the probability density of aftershocks of the 1984 Morgan Hill, CA, earthquake, assuming aftershock times are an inhomogeneous Poisson point process with decreasing intensity.

1. Introduction. Articles on density estimation abound, but most results on the uncertainty of density estimates are asymptotic and rely on assumptions about the density that are difficult to establish or justify (e.g., $\|f^{(2)}\|_2 \le C$); for example, see [1, 7, 22, 24]. Without some regularity condition, any density estimate may suffer from unbounded bias. An assumption we sometimes find compelling is that the density is monotone, possibly relative to some strictly positive weight function. For example, it is usually assumed that the probability of earthquake aftershocks decreases with time

Received March 1994; revised June 1994.
[1]Supported by the NSF PYI DMS-89-57573 and Grant DMS-88-10192, and NASA Grant NAGW-2515.
AMS 1991 *subject classifications*. Primary 62G15; secondary 65U05, 62G05.
Key words and phrases. Simultaneous confidence intervals, density estimation, monotone densities, unimodal densities, shape restrictions, confidence intervals for modes, nonparametric tests, linear programming, seismology.

after the "main event." We apply the method presented here to that problem in Section 6. Grenander [9] originated nonparametric density estimation with monotonicity constraints; see also Birgé [2], Groeneboom [10], Prakasa Rao [20] and Wang [28]. The uncertainty results in these papers are asymptotic, except [2], which computes the nonasymptotic L_1 risk of the Grenander estimate.

We show here that under the assumption of monotonicity or unimodality, one can compute conservative finite-sample confidence regions for the entire density by linear programming. The results extend to the less restrictive assumption that the density has at most k modes; see Section 7.2. One can also compute a lower confidence bound for the number of modes of an arbitrary distribution, a confidence interval for the support of a monotone decreasing density, a confidence interval for the mode of a unimodal density and test hypotheses of k-modality and monotonicity. The procedure produces a data-dependent confidence region—the width of the region (as a function of x) depends nonlinearly on the observations. The technique is computationally intensive, but manageable for large sets of data (millions). The rate of convergence of the method is optimal if a "bandwidth" parameter is chosen correctly.

Suppose $\{X_j\}_{j=1}^n$ are iid F. We present a way to construct a $1 - \alpha$ confidence region for the density f of F from the observations $X_j = x_j$, $j = 1, \dots, n$; that is, a pair of random functions $\gamma^-(x)$, $\gamma^+(x)$ such that

$$\mathbf{P}_F\{[\gamma^-(x), \gamma^+(x)] \ni f(x), \forall\, x \in \mathbf{R}\} \geq 1 - \alpha. \tag{1}$$

The coverage probability is *conservative and simultaneous* for all x.

Assumptions and conditions.

A1. $w(x)$ is nonnegative.

A2. F has density f with respect to Lebesgue measure.

A3. The support of f is connected and contained in the interval $[a, b]$, $-\infty \leq a < b \leq \infty$, a subset of the support of w.

In addition, either of the following restrictions holds:

U. $f(x)/w(x)$ is unimodal with mode $\mu \in \mathscr{I} = [\mu^-, \mu^+]$.

M. $f(x)/w(x)$ is monotone with x in the support of f, which may be known or unknown.

Without loss of generality we assume that $x_1 \leq x_2 \leq \cdots \leq x_n$.

2. A confidence region for *F*. The approach derives from the *strict bounds* technique used in geophysical inverse theory (see, e.g., [25]), which is similar to the *neighborhood procedure* of Donoho [8]. The key idea is to define a nonparametric confidence region for F in the set of distributions satisfying the shape restriction M or U. We solve optimization problems to find the largest and smallest value densities of distribution functions in the confidence region attain at a fixed point. Our main result is that it is sufficient to hunt among densities piecewise proportional to w with discontinuities at a subset of the data and at most three additional points, reducing the computations to finite-dimensional linear programs.

One might expect that the Kolmogorov–Smirnov (K-S) distance would provide a good confidence region for F; indeed all the following problems can be worked out for the K-S norm [12]. However, the rate of convergence using the K-S distance is slower. This is related to current work by Z. Landsmann and M. Rom (personal communication, 1993) on the (in)efficiency of the K-S test against multimodal alternatives.

Instead, we base our confidence region on the distribution of differences $F(X_{(j)}) - F(X_{(k)})$ for j and k in a subset of $\{1, \dots, n\}$. Let $K = K(n)$ be an integer less than n, and for fixed K define

$$M' \equiv \lfloor n/K \rfloor \tag{2}$$

and

$$M \equiv \lceil n/K \rceil. \tag{3}$$

Define

$$k_i = \begin{cases} (i-1)K + 1, & i = 1, \dots, M', \\ n, & i = M \text{ if } M \neq M'. \end{cases} \tag{4}$$

If F is the true cumulative distribution function, the differences

$$\Delta_i \equiv F(X_{(k_i)}) - F(X_{(k_{i-1})})$$
$$\sim \begin{cases} \dfrac{\Gamma(K,1)}{\sum_{l=1}^{M'} \Gamma(K,1) + \Gamma(M - M' + 1, 1)}, & i = 1, \dots, M', \\ \dfrac{\Gamma(n - KM', 1)}{\sum_{l=1}^{M'} \Gamma(K,1) + \Gamma(M - M' + 1, 1)}, & i = M \neq M', \end{cases} \tag{5}$$

where $\Gamma(\cdot, \cdot)$ is the gamma distribution. Note that $\{\Delta_i\}$ are not independent. Let $\{c_i^-(\alpha)\}_{i=1}^M$ and $\{c_i^+(\alpha)\}_{i=1}^M$ satisfy

$$\mathbf{P}_F\{c_i^-(\alpha) \leq \Delta_i \leq c_i^+(\alpha), i = 1, \dots, M\} \geq 1 - \alpha, \tag{6}$$

and denote by $\mathbf{c}$ the $2M$-vector $((c_i^-)_{i=1}^M, (c_i^+)_{i=1}^M)$. We will usually take all the c_i^- to be equal and all the c_i^+ to be equal, and denote their common values by c^- and c^+. Approximate values of c^- and c^+ can be found by simulation using (5) (see Section 6.2); the Appendix bounds the asymptotic behavior of c^- and c^+.

Let $\mathscr{P}$ be the set of cumulative distribution functions of probability measures on $\mathbf{R}$ and let $\mathscr{Q}$ be the set of cumulative distribution functions of subprobability measures on $\mathbf{R}$. Define

$$\mathscr{D} = \mathscr{D}_c \equiv \{G \in \mathscr{Q}: c_i^-(\alpha) \leq G(X_{(k_i)}) - G(X_{(k_{i-1})}) \leq c_i^+(\alpha)\} \tag{7}$$

($\mathscr{D}$ stands for data). Then

$$\mathbf{P}_F\{\mathscr{D} \ni F\} \geq 1 - \alpha. \tag{8}$$

Let w be nonnegative (assumption A1). Define $\mathscr{M} \equiv \mathscr{M}_w$ to be the set of cumulative distribution functions of subprobabilities on $\mathbf{R}$ satisfying A2, A3 and M, with monotone decreasing densities; monotone increasing densities can be handled by reflecting the x axis. Let $\mathscr{U}_{\mathscr{I}} \equiv \mathscr{U}_{w,\mathscr{I}}$ be the set of cumulative distribution functions of subprobabilities on $\mathbf{R}$ satisfying A2, A3 and U: densities unimodal relative to w with mode $\mu \in \mathscr{I}$, where $\mathscr{I} = [\mu^-, \mu^+]$ is a known, possibly unbounded interval. Note that $\mathscr{M} \subset \mathscr{U}_{[a, x_1]}$.

Let $\mathscr{F}$ denote $\mathscr{M}$ or $\mathscr{U}_{\mathscr{I}}$. If we know a priori that $F \in \mathscr{F}$, it follows that

$$\mathbf{P}_F\{\mathscr{F} \cap \mathscr{D} \ni F\} \geq 1 - \alpha. \tag{9}$$

Any inequality satisfied by the entire set $\mathscr{F} \cap \mathscr{D}$ holds for F as well, with probability at least $1 - \alpha$. In particular, for any fixed $y \in \mathbf{R}$, define

$$\gamma_{\mathscr{F}}^-(y) \equiv \inf_{G \in \mathscr{F} \cap \mathscr{D} \cap \mathscr{P}} g(y) \tag{10}$$

and

$$\gamma_{\mathscr{F}}^+(y) \equiv \sup_{G \in \mathscr{F} \cap \mathscr{D} \cap \mathscr{P}} g(y), \tag{11}$$

where g is the density of G. Then

$$\mathbf{P}_F\{[\gamma_{\mathscr{F}}^-(y), \gamma_{\mathscr{F}}^+(y)] \ni f(y)\ \forall y \in \mathbf{R}\} \geq 1 - \alpha; \tag{12}$$

that is, $[\gamma_{\mathscr{F}}^-(y), \gamma_{\mathscr{F}}^+(y)]$ is a conservative $1 - \alpha$ confidence region for $f(y)$. Knowing that $F \in \mathscr{F}$ allows us to interpolate and extrapolate the confidence intervals found at finitely many points y to obtain a confidence "envelope" with simultaneous $1 - \alpha$ coverage probability for all $y \in \mathbf{R}$ (see Section 5).

3. Bounds from finite-dimensional problems. The infinite-dimensional problems (10) and (11) can be reduced to finite-dimensional optimization problems. Given $z_1 < z_2 < \cdots < z_m$, $m \geq 2$, with $\mu \in \{z_k\}$, define

$$\begin{aligned}\mathscr{C}_\mu^- \equiv \Bigg\{ G \in \mathscr{G}\colon g(x) &= w(x) \\ &\times \Bigg(\sum_{\{j:\, z_{j+1} < \mu\}} \beta_j 1_{(z_j, z_{j+1}]}(x) + \max_{\{j:\, [z_j, z_{j+1}] \ni \mu\}} \beta_j 1_\mu(x) \\ &\quad + \sum_{\{j:\, [z_j, z_{j+1}] \ni \mu\}} \beta_j 1_{(z_j, z_{j+1})}(x) + \sum_{\{j:\, z_j > \mu\}} \beta_j 1_{[z_j, z_{j+1})}(x) \Bigg) \Bigg\},\end{aligned} \tag{13}$$

where g is the density of G (cumulative distribution functions of subprobability measures G with densities g piecewise proportional to w, left-continuous to the left of μ and right-continuous to the right of μ). Similarly, define

$$\mathscr{C}_\mu^+ \equiv \Bigg\{G \in \mathscr{Q}\colon g(x) = w(x)\Bigg(\sum_{\{j:\, z_{j+1} \le \mu\}} \beta_j \mathbf{1}_{[z_j, z_{j+1})}(x) + \infty \cdot \mathbf{1}_\mu(x) + \sum_{\{j:\, z_j > \mu\}} \beta_j \mathbf{1}_{(z_j, z_{j+1}]}(x)\Bigg)\Bigg\} \tag{14}$$

(cumulative distribution functions of subprobabilities G with densities g right-continuous to the left of μ and left-continuous to the right of μ).

Assume that $\{x_j\}_{j=1}^n$ contains n distinct elements (with probability 1, it will; this assumption can be relaxed at the expense of some bookkeeping). Let z_j be the jth smallest element of the set $\{a, \mu^-, \mu^+, y, x_{k_1}, x_{k_2}, \ldots, x_{k_M}, b\}$. Let N denote the number of distinct elements in the set $\{z_j\}$ ($M \le N \le M + 5$). Our basic observation is that if we define the constants $\{\beta_j\}$ in definitions (13) and (14) to be w-weighted averages of a unimodal density g over the intervals (z_j, z_{j+1}), the resulting densities are unimodal with the same mode, have cumulative distribution functions that match $G(x_{k_i}) - G(x_{k_{i-1}})$ (so they agree with the data if G does) and bracket $g(z_j)$. We use the convention that $\inf_\varnothing(\cdot) = \infty$ and $\sup_\varnothing(\cdot) = 0$. When the mode is known ($\mu^- = \mu^+ = \mu$), we have the following theorem:

THEOREM 3.1. *Define*

$$\tilde{\gamma}_\mu^-(y) \equiv \inf_{G \in \mathscr{C}_\mu^- \cap \mathscr{U}_\mu \cap \mathscr{D}} g(y), \tag{15}$$

and

$$\tilde{\gamma}_\mu^+(y) \equiv \sup_{G \in \mathscr{C}_\mu^+ \cap \mathscr{U}_\mu \cap \mathscr{D}} g(y). \tag{16}$$

Then

$$\gamma_{\mathscr{U}_\mu}^-(y) \ge \tilde{\gamma}_\mu^-(y) \tag{17}$$

and

$$\gamma_{\mathscr{U}_\mu}^+(y) \le \tilde{\gamma}_\mu^+(y). \tag{18}$$

We defined $\mathscr{C}_\mu^+$ so that the upper bound $\gamma_\mu^+(y)$ is infinite if $y = \mu$. Since the closure of the set of measures whose densities have mode μ contains measures with point masses at μ, $\gamma_{\mathscr{U}_\mu}^+(\mu) = \infty$ if $\mathscr{U}_\mu \cap \mathscr{D}$ contains interior points. If the mode is known to be attained at every point in the interval $[\mu^-, \mu^+]$, nontrivial upper bounds at $y \in [\mu^-, \mu^+]$ are possible; the definitions of $\mathscr{C}_\mu^\pm$ need to be modified.

When μ is only known to lie in the interval $\mathscr{I}$ (e.g., when f is monotonic with unknown support, $\mathscr{I} = [a, x_1]$), the situation is only slightly more complicated. Since $\{x_{k_i}\} \subset \{z_j\}$, the data cannot distinguish between $\mu \in [z_j, z_{j+1}]$ and $\mu \in \{z_j, z_{j+1}\}$. Thus it suffices to consider $\mu \in \{z_j\}_{j=1}^N$.

THEOREM 3.2. *We have that*

$$\gamma^-_{\mathscr{U}_\mathscr{F}}(y) \geq \tilde{\gamma}^-_{\mathscr{F}}(y) \equiv \min_{\mu \in \{z_j\} \cap \mathscr{F}} \tilde{\gamma}^-_\mu(y) \tag{19}$$

and

$$\gamma^+_{\mathscr{U}_\mathscr{F}}(y) \leq \tilde{\gamma}^+_{\mathscr{F}}(y) \equiv \max_{\mu \in \{z_j\} \cap \mathscr{F}} \tilde{\gamma}^+_\mu(y). \tag{20}$$

Theorem 3.1 is an immediate consequence of the fact that for any $G \in \mathscr{U}_\mu \cap \mathscr{D}$, we can construct $G^\pm \in \mathscr{C}^\pm_\mu \cap \mathscr{U}_\mu \cap \mathscr{D}$ whose densities $g^\pm$ bracket $g(y)$. Given $G \in \mathscr{U}_\mu$, define

$$\beta_j \equiv \frac{\int_{z_j}^{z_{j+1}} g\,dx}{\int_{z_j}^{z_{j+1}} w\,dx}, \qquad j = 1, \dots, N-1. \tag{21}$$

It is easy to verify that the elements $G^- \in \mathscr{C}^-_\mu$ and $G^+ \in \mathscr{C}^+_\mu$ with densities g^- and g^+ defined by taking these coefficients in (13) and (14) are also in $\mathscr{U}_\mu$: suppose $z_j, z_{j+1} \leq \mu$. Then for any $s \in (z_j, z_{j+1})$,

$$\begin{aligned} \frac{g^\pm(s)}{w(s)} &= \frac{\int_{z_j}^{z_{j+1}} g\,dx}{\int_{z_j}^{z_{j+1}} w\,dx} \frac{w(s)}{w(s)} \\ &\leq \frac{g(z_{j+1})}{w(z_{j+1})} \frac{\int_{z_j}^{z_{j+1}} w\,dx}{\int_{z_j}^{z_{j+1}} w\,dx} \\ &= \frac{g(z_{j+1})}{w(z_{j+1})}. \end{aligned} \tag{22}$$

Similarly,

$$\frac{g^\pm(s)}{w(s)} \geq \frac{g(z_j)}{w(z_j)}. \tag{23}$$

If $z_j, z_{j+1} \geq \mu$,

$$\frac{g(z_j)}{w(z_j)} \geq \frac{g^\pm(s)}{w(s)} \geq \frac{g(z_{j+1})}{w(z_{j+1})}. \tag{24}$$

The unimodality of $g^\pm$ thus follows from that of g.

Now suppose that $G \in \mathscr{D}$. By construction,

$$G(x_{k_i}) - G(x_{k_{i-1}}) = G^\pm(x_{k_i}) - G^\pm(x_{k_{i-1}}), \qquad i = 2, \dots, M. \tag{25}$$

4. The linear programs. Observing that $G^\pm \in \mathscr{F} \cap \mathscr{C}^\pm_\mu \cap \mathscr{D}$ if and only if $\{\beta_j\}$ satisfy a finite set of linear inequalities shows that the finite-dimensional optimization problems are in fact linear programs. Define

$$\omega_j \equiv \int_{z_j}^{z_{j+1}} w\,dx. \tag{26}$$

For $G \in \mathscr{C}_\mu^{\pm}$,

$$G(x_{k_i}) - G(x_{k_{i-1}}) = \sum_{\{j:\, x_{k_{i-1}} < z_j \le x_{k_i}\}} \beta_j \omega_j. \tag{27}$$

LEMMA 4.1. *Suppose G has density g of the form in* (13) *or* (14). $G \in \mathscr{C}$ *if and only if*:

B1. $\beta_j \ge 0$, $j = 1, \dots, N-1$.

B2. $\sum_{j=1}^{N-1} \omega_j \beta_j \le 1$.

$G \in \mathscr{D}$ *if and only if* B1 *and* B2 *hold and*:

B3. $c_i^- \le \sum_{\{j:\, x_{k_{i-1}} < z_j \le x_{k_i}\}} \beta_j \omega_j \le c_i^+$, $i = 2, \dots, M$.

$G \in \mathscr{U}_\mu$ *if and only if* B1 *and* B2 *hold and*:

B4. $\beta_j \le \beta_{j+1}$ *for all j such that* $z_{j+1} < \mu$ *and* $\beta_j \ge \beta_{j+1}$ *for all j such that* $z_j \ge \mu$.

B1 ensures that G is the distribution function of a positive measure; B2 ensures that $G(\infty) \le 1$; B4 ensures that g is unimodal with mode μ. B1–B4 are linear inequalities on $\{\beta_j\}$. The proof of the lemma is trivial.

The value of the density of an element of $\mathscr{C}_\mu^-$ at the point $y = z_j$ is

$$g(y) = \beta_{j-1} w(y), \qquad y < \mu, \tag{28}$$

and

$$g(y) = \beta_j w(y), \qquad y > \mu. \tag{29}$$

For elements of $\mathscr{C}_\mu^+$, the definitions are reversed. Maximizing or minimizing (28) or (29) subject to B1–B4 are $N-1$-dimensional linear programs.

When $y = \mu$, the upper bound is infinite if the program is feasible, and there are two linear programs to solve for the lower bound—we take the larger of min $\beta_{j-1} w(y)$ and min $\beta_j w(y)$ subject to B1–B4.

REMARK. If the linear programs are infeasible, the data are not consistent with the hypothesis of unimodality with mode μ at significance level α. The confidence interval is then empty. The range of mode locations for which the linear programs are feasible is a conservative confidence interval for the mode of a unimodal distribution. This is similar to the approach of Bogomolov [3], but produces shorter intervals. In contrast to methods for estimating the mode using the number of points in intervals of a given length (or the length of intervals containing a number of points) [4, 5, 27], the linear programming method can reject the hypothesis of unimodality when it is severely violated, and gives conservative coverage probability for finite n.

REMARK. If $f(\mu)$ is known to be attained on the entire interval $[\mu^-, \mu^+]$ we can find nontrivial upper and lower bounds by constraining $\beta_j = \beta_{j+1}$ $\forall j$ such that $\mu^- \le z_j$ and $z_{j+1} \le \mu^+$. This decreases the dimension of the linear programs.

5. From confidence intervals to confidence envelopes. For any fixed set $\{y_j\}$, we can find conservative $1-\alpha$ confidence intervals for $f(y_j)$ by solving linear programs. Since the extremal densities are piecewise proportional to $w(y)$ between the selected order statistics, it makes sense to compute the bounds at the subset of the order statistics $\{X_{(k_j)}\}_{j=1}^{M}$, where k_j and M are defined in (4) and (3). Each (feasible) linear program produces a cumulative distribution function $G_{y_j} \in \mathscr{F} \cap \mathscr{D}$, so the confidence level for the intervals at $\{y_j\}$ is $1-\alpha$ *simultaneously*. The constraint $F \in \mathscr{F}$ allows us to interpolate and extrapolate bounds at $\{y_j\}_{j=1}^{M}$ to get a conservative confidence envelope for $f(y)$ $\forall y \in \mathbf{R}$, as described in the following sections.

5.1. *Known mode.* If $G \in \mathscr{U}_\mu \cap \mathscr{D}$, then

$$\tilde{\gamma}_\mu^-(y_m) \le g(y_m) \le \frac{w(y_m)}{w(y)} g(y), \tag{30}$$

for $y \in [y_m, \mu]$ if $y_m \le \mu$ or for $y \in [\mu, y_m]$ if $y_m \ge \mu$. Similarly

$$\tilde{\gamma}_\mu^+(y_m) \ge g(y_m) \ge \frac{w(y_m)}{w(y)} g(y), \tag{31}$$

for $y \in [a, y_m]$ if $y_m < \mu$ or for $y \in [y_m, b]$ if $y_m > \mu$. This yields upper and lower confidence curves for $f(y)$ when $\mu^- = \mu^+ = \mu$:

$$l_\mu(y) \equiv \begin{cases} \max\limits_{\{m:\, y_m \le y \le \mu\}} \dfrac{w(y)}{w(y_m)} \tilde{\gamma}_\mu^-(y_m), & y \le \mu, \\ \max\limits_{\{m:\, y_m \ge y \ge \mu\}} \dfrac{w(y)}{w(y_m)} \tilde{\gamma}_\mu^-(y_m), & y > \mu, \end{cases} \tag{32}$$

and

$$u_\mu(y) \equiv \begin{cases} \min\limits_{\{m:\, y \le y_m \le \mu\}} \dfrac{w(y)}{w(y_m)} \tilde{\gamma}_\mu^+(y_m), & y \le \mu, \\ \min\limits_{\{m:\, y \ge y_m \ge \mu\}} \dfrac{w(y)}{w(y_m)} \tilde{\gamma}_\mu^+(y_m), & y > \mu, \end{cases} \tag{33}$$

and $P_F\{[l_\mu(y), u_\mu(y)] \ni f(y), \forall y\} \ge 1-\alpha$.

5.2. *Unknown mode.* Define

$$\xi^- \equiv \inf\{\mu \in [\mu^-, \mu^+]: \mathscr{U}_\mu \cap \mathscr{D} \ne \varnothing\} \tag{34}$$

and

$$\xi^+ \equiv \sup\{\mu \in [\mu^-, \mu^+]: \mathscr{U}_\mu \cap \mathscr{D} \ne \varnothing\}. \tag{35}$$

The upper confidence bound on $f(y)$ is infinite for all $y \in [\xi^-, \xi^+]$. All densities of cumulative distribution functions $G \in \mathscr{D} \cap \mathscr{U}_{\mathscr{I}}$ are monotone increasing relative to w left of ξ^- and monotone decreasing relative to w

right of ξ^+. Thus the interpolation scheme in the previous section works outside $[\xi^-, \xi^+]$:

$$(36)\qquad l_{\mathscr{I}}(y) \equiv \begin{cases} w(y) \max_{\{m:\, y_m \le y \le \xi^-\}} \dfrac{\tilde{\gamma}_{\mathscr{I}}^-(y_m)}{w(y_m)}, & y \le \xi^-, \\ w(y) \max_{\{m:\, y_m \ge y \ge \xi^+\}} \dfrac{\tilde{\gamma}_{\mathscr{I}}^-(y_m)}{w(y_m)}, & y > \xi^+, \\ w(y) \min_{\{m:\, [y_m, y_{m+1}] \ni y\}} \left\{ \dfrac{\tilde{\gamma}_{\mathscr{I}}^-(y_m)}{w(y_m)}, \dfrac{\tilde{\gamma}_{\mathscr{I}}^-(y_{m+1})}{w(y_{m+1})} \right\}, & \\ \qquad y \in [y_m, y_{m+1}] \cap [\xi^-, \xi^+], & \end{cases}$$

and

$$(37)\qquad u_{\mathscr{I}}(y) \equiv \begin{cases} \min_{\{m:\, y \le y_m \le \xi^-\}} \dfrac{w(y)}{w(y_m)} \tilde{\gamma}_{\mathscr{I}}^+(y_m), & y \le \xi^-, \\ \min_{\{m:\, y \ge y_m \ge \xi^+\}} \dfrac{w(y)}{w(y_m)} \tilde{\gamma}_{\mathscr{I}}^+(y_m), & y > \xi^+, \\ \infty, & y \in [\xi^-, \xi^+], \end{cases}$$

and we have

$$(38)\qquad \mathbf{P}_F\{[l_{\mathscr{I}}(y), u_{\mathscr{I}}(y)] \ni f(y), \forall y\} \ge 1 - \alpha.$$

6. Earthquake aftershocks.

6.1. *Inhomogeneous Poisson process models for aftershocks.* It is generally believed that immediately following a large earthquake, the chance of aftershocks is largest, decreasing as time goes by (at least for a while until stresses build up again). The number of aftershocks by time t, $S(t)$, is often modeled as an inhomogeneous Poisson point process with intensity $\lambda(t)$. [However, see Ogata [18].] Omori's law [19] and its modification (e.g., [26]) are parametric Poisson intensity functions often fitted to aftershock sequences. In the modified Omori law, the intensity is

$$(39)\qquad \lambda(t) = \frac{\text{const}}{(c + t)^q}.$$

Davis and Frohlich [6] verified that the modified Omori law with $q \approx 0.87$ provides a probabilistically adequate fit to many small aftershock sequences; other investigators have generally (though not invariably) found $q > 1$ for large earthquakes with many aftershocks (e.g., [26]). A number of theoretical studies using different physical models of earthquakes predict a modified Omori law with $q > 1$ for some ranges of time [26, 17, 13, 29]. Kagan and Knopoff [13] argue that a modified Omori law should hold initially, followed by a transition to an exponentially decaying intensity function. Given the pervasive use of Omori's law to model aftershock sequences, it has become a touchstone for theoretical physical models, so the agreement of theory and observation is more suggestive than conclusive (see, e.g., [15]).

Assuming a particular parametric form for the intensity of aftershocks is extremely restrictive. Furthermore, researchers are wont to draw conclusions about physical differences between events using uncertainty estimates for the parameters in the modified Omori law and to project earthquake hazard after main shocks using the law (see, e.g., [6, 21]). The uncertainty estimates are suspect, since they are conditional on the truth of the parametric model. Estimates of q are also sensitive to the time after the main shock at which the aftershock observations begin (C. Frohlich, personal communication, 1992).

Suppose aftershock times $\{X_j\}$ are an inhomogeneous Poisson point process with monotone decreasing intensity $\lambda(t)$ and define

$$f(t) = \frac{\lambda(t)}{\Lambda(T)}, \qquad 0 \le t \le T, \tag{40}$$

where

$$\Lambda(T) \equiv \int_0^T \lambda(t)\, dt. \tag{41}$$

Then $f(t)$ is the conditional density of the (iid) times $\{X_j\}_{j=1}^n$ given $S(T) = n$. The density $f(t)$ is just the intensity $\lambda(t)$ normalized to unit area over the observation interval, so monotonicity of $\lambda(t)$ implies that $f(t)$ is monotonic too. In the next section, we find confidence bounds for $f(t)$, $t \in (0, T]$, assuming that λ is monotonic.

6.2. *Data and results.* Robert Uhrhammer (personal communication, 1992) provided us with USGS-identified aftershocks of the 21:15:18 24 April 1984 magnitude 5.9 Morgan Hill, California, event, located near 37°18.58′N latitude, 121°40.60′W longitude, for 24 April through 31 December 1984. Typically, earthquakes are identified by seismologists as aftershocks of a "main shock" if they are smaller and later than the main shock and are sufficiently close to the main shock in space and time. The smallest events reliably detected by the seismographic network are magnitude 2.0 (R. Uhrhammer, personal communication, 1992). The data we used include all events of magnitude 2.0 and larger located between 37.0 and 37.5°N latitude and between 121.5 and 121.83°W longitude. There were 766 such events.

To apply the method, we need to select a "bandwidth" K. The optimal value of K depends on the number of data n, as well as unknown properties of f (such as its smoothness and, indeed, its value at y; see Appendix). However, the simultaneous coverage probability of the bounds is conservative, regardless of how K is chosen or the true smoothness of f. To illustrate the method, we plot confidence bounds for two choices of K (10 and 50) and for the linear programming method based on the Kolmogorov–Smirnov distance of the empirical cumulative distribution function from the true [12].

To approximate conservative critical values to use for c^+ and c^-, we simulated $766/K$ independent gamma random variables with parameter K. We calculated the maximum M and minimum W of

$$U_j = \frac{X_j}{\sum_{j=1}^{766/K} X_j}. \tag{42}$$

We simulated sets of $766/K$ gamma random variables 5000 times to obtain empirical distribution functions for M and W. Critical values c^- and c^+ of these "Monte Carlo" distribution functions are given in Table 1.

Using these critical values we can obtain conservative, joint one-sided confidence intervals for M and W:

$$\mathbf{P}\{M \le c^+ \cap W \ge c^-\} \ge 1 - \mathbf{P}\{M \le c^+\} - \mathbf{P}\{W \ge c^-\}. \tag{43}$$

Thus the $K = 10$ 97.5% critical values $c^- = 3.34 \times 10^{-3}$ and $c^+ = 31.2 \times 10^{-3}$ give us conservative 95% joint confidence intervals for M and W. Using these critical values, we solved the linear programming problems to find confidence bounds on the probability density of aftershocks. Figure 1 shows the resulting 95% confidence bounds and the maximum likelihood estimate of the modified Omori density, which had $\hat{q} = 0.554$ and $\hat{c} = 0.0018$. (If this value of q held for all time, the expected number of aftershocks would be infinite.) The Omori density estimate is the solid curve, the $K = 10$ confidence bounds are the long-dashed curves and the $K = 50$ confidence bounds are the short-dashed curves. The dotted curves are a different set of nonparametric 95% confidence bounds also found using linear programming [12], but using a confidence set based on the Kolmogorov–Smirnov distance between the empirical and true cumulative distribution functions, calibrated using the result of Massart [16]. The rate of convergence of the second method is suboptimal and the bounds

TABLE 1

Critical values differences of 77 uniform order statistics ($K = 10$) and 16 uniform order statistics ($K = 50$), found from 5000 simulations

K	$1 - \alpha$*	$10^3 c^-$†	$10^3 c^+$‡
10	0.95	3.66	29.9
	0.975	3.34	31.2
	0.995	2.72	34.7
50	0.95	36.7	88.0
	0.975	40.2	90.4
	0.995	41.7	96.3

*Coverage probability.
†Lower confidence bound for $\min_{j \le \lceil 766/K \rceil}(F(X_{(K(j+1))}) - F(X_{(Kj)}))$.
‡Upper confidence bound for $\max_{j \le \lceil 766/K \rceil}(F(X_{(K(j+1))}) - F(X_{(Kj)}))$.

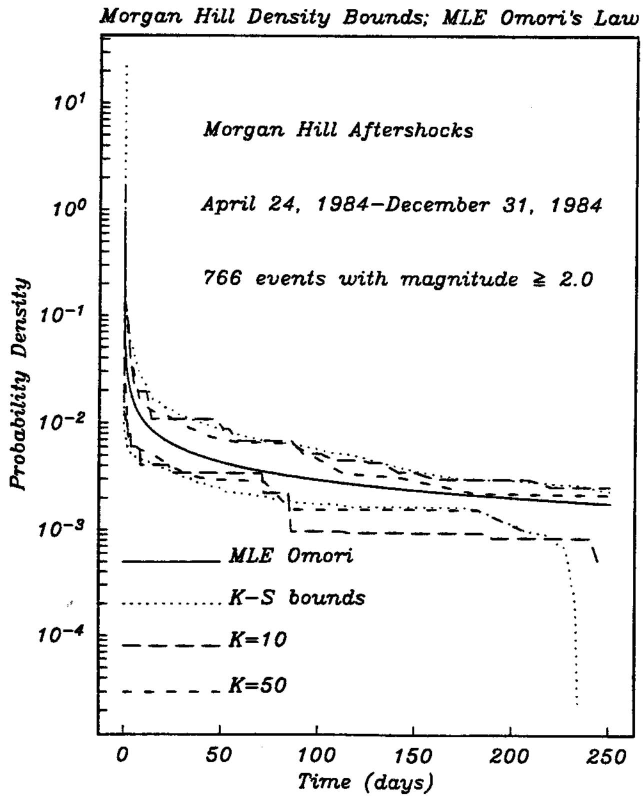

FIG. 1. *95% confidence bounds on the probability density function of aftershocks of the* 24 *April* 1984 *Morgan Hill event based on* 766 *aftershocks with magnitudes at least* 2, *identified by the U.S. Geological Survey between* 24 *April and* 31 *December* 1984. *The long-dashed curves are simultaneous* 95% *confidence bounds derived by the linear programming method introduced in this paper, using a bandwidth* $K = 10$. *The short-dashed curves are simultaneous* 95% *confidence bounds with* $K = 50$. *The solid curve is the pdf implied by the MLE-modified Omori law, which had* $\hat{q} = 0.554$, $\hat{c} = 0.0018$. *The dotted curves are* 95% *confidence bounds derived by a different conservative, finite-sample technique also based on linear programming, using the Kolmogorov–Smirnov distance instead of differences of order statistics to define the confidence region for the cumulative distribution function* F.

are more expensive to compute: the dimension of the linear programs is about n, rather than $n^{1/3}$ (see [12]). On the other hand, the method that uses the Kolmogorov–Smirnov distance is adaptive (one need not specify a bandwidth K) and can have faster convergence than the present method if the assumed smoothness of f is erroneous. Furthermore, for finite samples, it is not possible to predict which method will have narrower bounds. See also Section 7.3.

Computing these bounds took roughly 1 cpu minute for the $K = 10$ bounds and a few seconds for the $K = 50$ bounds using Numerical Algorithms Group

(NAG) routines to solve the linear programming problems on a Sun SparcStation 2.

7. Discussion.

7.1. *Testing.* With the linear programming formulation we can test the hypothesis that $\{x_j\}_{j=1}^n$ are iid samples from a distribution $F \in \mathscr{F}$: Finding the smallest $c_i^+ - c_i^-$ for which the linear program is feasible is a linear program. Simulations then allow one to compute a p-value. This approach is similar to the "dip test" of Hartigan and Hartigan [11], but linear programming yields a conservative, finite-sample test without appealing to asymptotics or specifying a particular null distribution ([11] use the uniform).

7.2. *Extension to k or fewer modes.* The linear programming formulation also gives a procedure to compute conservative, finite-sample lower confidence intervals for the number of modes of a distribution. (Compare this with the kernel approach to estimating the number of modes; see, e.g., Silverman [23].) The approach is related to the "neighborhood procedure" analyzed by Donoho [8] (he does not suggest an algorithm). Define the number of modes of a distribution to be the number of local maxima (relative to w) of the density. The fundamental observation justifying the linear programming approach is that the density piecewise proportional to w obtained by averaging any density on intervals (z_j, z_{j+1}) (as in the proof of Theorem 1) cannot have more modes than the original density, and the cumulative distribution function derived from the averaged density matches $G(x_{k_i}) - G(x_{k_{i-1}})$. That is, suppose g is the density of some $G \in \mathscr{D}$ and that g has k modes relative to the weight function w. Then the cumulative distribution function $\tilde{G}$ whose density $\tilde{g}$ is given by sums of the form in (14) and (13), with β_j defined by (21), has at most k modes relative to w (by the mean value theorem) and, as we have already shown, is in $\mathscr{D}$.

We may impose the restriction that the density of an element of $\mathscr{C}_\mu^{\pm}$ has modes on the intervals $(z_j, z_{j+1})_{j \in \mathscr{I}^+}$ and antimodes on the intervals $(z_j, z_{j+1})_{j \in \mathscr{I}^-}$ by a suitable set of linear inequalities among the coefficients $\{\beta_j\}$. We may then sequentially check whether there exists a density with mode between z_1 and z_2, between z_2 and z_3, between z_3 and z_4 and so forth, whose cumulative distribution function satisfies $c_i^- \le G(x_{j_i}) - G(x_{j_{i-1}}) \le c_i^+$, $i = 1, \dots, M$, using these inequalities and the constraints B1–B3. Testing the consistency of the set of linear inequalities for each postulated location of the mode is a linear programming feasibility problem. If none of these problems is feasible, we then check whether the data are consistent with modes on (z_1, z_2) and (z_3, z_4) and an antimode on (z_2, z_3), with modes on (z_1, z_2) and (z_4, z_5) with an antimode on either (z_2, z_3) or (z_3, z_4) and so forth. The smallest number of modes for which it is possible to construct a density piecewise proportional to w and consistent with the data is a lower $1 - \alpha$ confidence bound for the number of modes. (Any truly nonparametric confidence interval for the number of modes must have an infinite upper endpoint [8].) To find the lower endpoint we must solve at most

$$\sum_{k=1}^{\lfloor N/2 \rfloor} \binom{N-1}{2k-1} = 2^{N-2} - 1 \tag{44}$$

linear programming feasibility problems. In practice, fewer problems will be required, since for reasonable confidence levels, $\mathscr{D}$ will contain cumulative distribution functions whose densities have far fewer than $N/2$ modes, even if f has many more.

Similarly, the linear programming approach can be used to find simultaneous confidence bounds on a density assumed to have k or fewer modes by imposing the shape restriction for a particular set of assumed locations for the modes (as described in the previous paragraph) and solving linear programs to find the smallest and largest values the density can have at the point y. By taking the largest linear programming upper bound and smallest linear programming lower bound as the assumed mode locations range over all $\binom{N+1}{k}$ possibilities, we can find a conservative confidence interval for $f(y)$.

7.3. *Rate of convergence.* We wish to emphasize that the proposed technique does not require conditions on f other than $F \in \mathscr{F}$; however, the rate at which the distance between the upper and lower confidence bounds converges does depend on smoothness, and details of the procedure can be tailored to speed convergence if the degree of smoothness is known. On the other hand, if f has k or fewer modes, f is differentiable almost everywhere, which is enough to guarantee L_1 convergence of the confidence bounds on "most" bounded intervals except for some sets of arbitrarily small measure (Corollary 7.3). The point-by-point rate of convergence does not apply at discontinuities and modes of the density, where the bounds do not converge to single points (otherwise the coverage probability could not be conservative).

For f locally Lipschitz with exponent ϱ, Khas'minskii [14] gives a lower bound of $(\log n/n)^{\varrho/(1+2\varrho)}$ for the minimax rate at which the uniform norm of the error of any estimator of the density goes to zero. Our method attains this rate (which is therefore optimal) if the "bandwidth" K grows with n in a way that depends on ϱ. For $\varrho = 1$ ($K = \text{const} \times n^{2/3}(\log n)^{1/3}$), the rate is $(\log n/n)^{1/3}$, which is essentially the pointwise rate of the Grenander estimate for the same Lipschitz condition ($n^{-1/3}$), modified by a $(\log n)^{1/3}$ term needed for simultaneous coverage probability. The following theorems are all for $w = 1$, but can be extended to general w.

THEOREM 7.1. *Let y be an interior point of the support of f, but not a mode of f. Suppose that f is unimodal and that at y, f satisfies the local Lipschitz condition*

$$|f(x) - f(y)| \le C|x - y|^{\varrho} \tag{45}$$

for some constants C and $\varrho > 0$. Suppose K goes to infinity as n does, $K < n$, *and that for some constant* $\tau > 0$, *K satisfies*

$$\frac{\log K}{\log(n/K)} + \frac{2\tau^3}{3}\sqrt{\frac{\log(n/K)}{K}} \le \tau^2 - 2. \tag{46}$$

Then

$$\liminf_{n\to\infty} \mathbf{P}_F\left\{|\gamma^+(y) - \gamma^-(y)| \le 2f(y)\tau\sqrt{\frac{\log n}{K_n}} + 2Cf^{-\varrho}(y)\left(\frac{K_n}{n}\right)^{\varrho}\right\} > 0. \tag{47}$$

Theorem 7.1 and the next two corollaries are proved in the Appendix.

COROLLARY 7.2. *Let f and y be as in Theorem* 7.1. *Let the "bandwidth"*

$$K = \left(\frac{\tau}{2C\varrho}\right)^{2/(1+2\varrho)}(1+2\varrho)^{-1/(1+2\varrho)} f^{2(1+\varrho)/(1+2\varrho)}(y) \times (n^{2\varrho}\log n)^{1/(1+2\varrho)}. \tag{48}$$

Then

$$\liminf_{n\to\infty} \mathbf{P}\left\{|\gamma^+(y) - \gamma^-(y)| \le \left(\frac{f(y)}{1+2\varrho}\right)^{\varrho/(1+2\varrho)} (4C\varrho(1+\varrho))^{1/(1+2\varrho)} \times (2+\varrho^{-1})\left(\frac{\log n}{n}\right)^{\varrho/(1+2\varrho)}\right\} > 0. \tag{49}$$

For $\varrho = 1$, this gives the rate $(\log n/n)^{1/3}$.

The integrated width of the bounds also converges at this rate:

COROLLARY 7.3. *Suppose f is a density with at most k modes and let* $K = \text{const}(n^2 \log n)^{1/3}$. *Suppose that the interval* $T = [c, d]$ $(-\infty < c < d < \infty)$ *contains no mode of f, that*

$$\mathbf{P}\{X_j < c\} > 0 \tag{50}$$

and

$$\mathbf{P}\{X_j > d\} > 0. \tag{51}$$

Then for any $\varepsilon > 0$, *there exists a set* $T_\varepsilon \subset T$ *such that* $\text{meas}\, T \setminus T_\varepsilon \le \varepsilon$ *and*

$$\liminf_{n\to\infty} \mathbf{P}\left\{\left(\int_{T_\varepsilon} |\gamma^+(x) - \gamma^-(x)|^p\, dx\right)^{1/p} \le C\left(\frac{\log n}{n}\right)^{1/3}\right\} > 0 \tag{52}$$

for some constant C.

The proof of Corollary 7.3 is basically an application of the following lemma.

LEMMA 7.4. *If f is a monotone function of bounded variation on a bounded interval T, then for any $\varepsilon > 0$ there exists a finite C such that*

$$\operatorname{meas}\{x \in T: |f(x) - f(y)| > C|x - y| \text{ for some } y \in T\} < \varepsilon. \tag{53}$$

Lemma 7.4 follows from an argument similar to the proof of Lebesgue's theorem that the measure of the set of discontinuities of a monotone function is zero.

Conditions (50) and (51), together with the fact that $K/n \to 0$, imply that

$$\liminf_{n \to \infty} \mathbf{P}\{[X_{(K)}, X_{(n-K)}] \supset [c, d]\} = 1, \tag{54}$$

which in turn implies that the pointwise rate argument of Corollary 7.2 ultimately applies throughout T except on sets of arbitrarily small measure.

The optimality of the rate of convergence depends on selecting K optimally, which in turn requires information about the smoothness of f. What if one errs in the smoothness assumption? It is better to err conservatively by assuming f is not necessarily very smooth than to be overly optimistic:

COROLLARY 7.5. *Suppose K is chosen to be optimal for Lipschitz exponent ϱ, but in fact the Lipschitz exponent of f is ϱ'. Then for some constants C_1 and C_2,*

$$\liminf_{n \to \infty} \mathbf{P}\left\{|\gamma^+(y) - \gamma^-(y)| \le C_1\left(\frac{\log n}{n}\right)^{\varrho/(1+2\varrho)} + C_2\left(\frac{\log n}{n}\right)^{\varrho'/(1+2\varrho)}\right\} > 0. \tag{55}$$

That is, the rate of convergence is

$$\left(\frac{\log n}{n}\right)^{\min(\varrho, \varrho')/(1+2\varrho)} \tag{56}$$

This follows algebraically from Theorem 7.1 and Corollary 7.2.

Recall that the coverage probability of the confidence bounds is conservative regardless of how smooth or rough f is and how K is chosen. Corollary 7.5 shows that if we assume only that f is monotone (almost everywhere Lipschitz 1) but f is in fact smoother, the additional smoothness is not reflected in the rate [it is still $(\log n/n)^{1/3}$]. On the other hand, if we assume that f is smoother than it really is, the rate of convergence will suffer still more. It is less damaging to make the more conservative assumption that f is not very smooth than to assume incorrectly that f is smoother than in fact it is. Note that this is in contrast to the performance of the linear programming approach that uses the Kolmogorov–Smirnov distance to form the confidence region [12]; there, the rate of convergence is always $n^{\varrho/(2+2\varrho)}$, regardless of ϱ, and one need not specify any "bandwidth" or other tuning parameter.

7.4. *Probabilities versus subprobabilities.* If $\int_a^{x_1} w\,dx < \infty$ and $\int_{x_n}^b w\,dx < \infty$, we can restrict $\mathscr{C}_\mu^\pm$ to contain only cumulative distribution functions of probability measures, not subprobability measures. The construction of $\{\beta_j\}$ in (21) will produce distribution functions of probability measures as well, and we can replace condition B2 by the following:

B2′. $\sum_{j=1}^{N-1} \omega_j \beta_j = 1.$

Suppose $\int_{x_n}^b w\,dx = \infty$ and consider the case f is monotone decreasing. Any mass needed to bring $G(\infty) = 1$ for $G \in \mathscr{F} \cap \mathscr{D}$ can be accommodated between z_{N-1} and b without violating the monotonicity in the tail or changing the fit to the data. By symmetry, for monotone increasing or unimodal f, it is sufficient that $\int_a^{x_1} w\,dx = \infty$ for this to work also. In these cases, the values of the optimization problems posed over sets of cumulative distribution functions of subprobabilities and over cumulative distribution functions of probabilities are equal.

APPENDIX

Proof of Theorem 7.1 and Corollaries 7.2 and 7.3. In this section, please keep in mind the implicit dependence of c^-, c^+ and K on n. We assume implicitly that K divides n; when K does not divide n, the results still hold since the relevant quantities are bounded stochastically from below and above by problems where K does divide n.

A.1. *Analysis.* Without loss of generality, we assume f decreases monotonically on a neighborhood of y. Let

$$\zeta = F(y), \tag{57}$$

define l to satisfy

$$\frac{(l-1)K}{n} < \zeta \le \frac{lK}{n} \tag{58}$$

and define

$$l_j = (l+j)K. \tag{59}$$

We can upper-bound the width of the confidence band at y by ignoring the restriction to subprobabilities: for any $G \in \mathscr{D}$ with monotone decreasing density g,

$$g(y) \ge \frac{c^-}{\omega_{l+1}} \tag{60}$$

and

$$g(y) \leq \frac{c^+}{\omega_l}. \tag{61}$$

Thus

$$\gamma^+(y) - \gamma^-(y) \leq D \equiv \left| \frac{c^+}{\omega_l} - \frac{c^-}{\omega_{l+1}} \right|. \tag{62}$$

By the mean value theorem, there exists $\eta_l \in (X_{(l_{-1})}, X_{(l_0)})$ such that

$$\begin{aligned} \Delta_l &= F(X_{(l_0)}) - F(X_{(l_{-1})}) \\ &= (X_{(l_0)} - X_{(l_{-1})}) f(\eta_l) \\ &= \omega_l f(\eta_l) \end{aligned} \tag{63}$$

and $\eta_{l+1} \in (X_{(l_0)}, X_{(l_1)})$ such that

$$\Delta_{l+1} = \omega_{l+1} f(\eta_{l+1}). \tag{64}$$

As $n \to \infty$, with probability tending to 1, f decreases monotonically on the interval $[X_{(l_{-1})}, X_{(l_1)}]$. Thus, asymptotically,

$$\frac{\Delta_l}{f(X_{(l_{-1})})} \leq \omega_l \leq \frac{\Delta_l}{f(X_{(l_0)})} \tag{65}$$

and

$$\frac{\Delta_{l+1}}{f(X_{(l_0)})} \leq \omega_{l+1} \leq \frac{\Delta_{l+1}}{f(X_{(l_1)})}, \tag{66}$$

so

$$\begin{aligned} D &\leq \left| \frac{f(X_{(l_{-1})}) c^+}{\Delta_l} - \frac{f(X_{(l_1)}) c^-}{\Delta_{l+1}} \right| \\ &\leq f(X_{(l_1)}) \left| \frac{c^+}{\Delta_l} - \frac{c^-}{\Delta_{l+1}} \right| + \frac{c^+}{\Delta_l} \left| f(X_{(l_{-1})}) - f(X_{(l_1)}) \right| \\ &\leq f(X_{(l_1)}) \left| \frac{c^+}{\Delta_l} - \frac{c^-}{\Delta_{l+1}} \right| + \frac{Cc^+}{\Delta_l} \left| X_{(l_{-1})} - X_{(l_1)} \right|^{\varrho} \\ &\leq f(X_{(l_1)}) \left| \frac{c^+}{\Delta_{l-1}} - \frac{c^-}{\Delta_{l+1}} \right| + \frac{Cc^+}{\Delta_l} \left(\frac{\Delta_l + \Delta_{l+1}}{f(X_{(l_1)})} \right)^{\varrho}. \end{aligned} \tag{67}$$

A.2. *Asymptotic behavior of c^+ and c^-.* Let $\{Y_j\}_{j=1}^{n+1}$ be iid exponential random variables with mean 1. Define

$$M_n \equiv \max_{j=1,\ldots,n/K} \sum_{m=(j-1)K+1}^{jK} Y_m \tag{68}$$

and

$$W_n \equiv \min_{j=1,\dots,n/K} \sum_{m=(j-1)K+1}^{jK} Y_m \tag{69}$$

and let

$$S_n \equiv \sum_{m=1}^{n} Y_m. \tag{70}$$

It is well known that

$$\Delta_l =_{\mathscr{D}} \frac{S_{l_0} - S_{l_{-1}}}{S_{n+1}}, \tag{71}$$

so c^+ and c^- must satisfy

$$\mathbf{P}\left\{c^- \le \frac{W_n}{S_{n+1}} \text{ and } \frac{M_n}{S_{n+1}} \le c^+\right\} \ge 1 - \alpha. \tag{72}$$

We seek a bound $\delta = \delta(n)$ on D satisfying

$$\liminf_{n\to\infty} \mathbf{P}\{D \le \delta\} > 0. \tag{73}$$

To find such a δ, we need to study the asymptotic behavior of M_n and W_n. If $c^{\pm}(n)$ satisfy

$$\mathbf{P}\left\{\frac{W_n}{S_{n+1}} \le c^-\right\} \le \frac{\alpha}{2} \tag{74}$$

and

$$\mathbf{P}\left\{\frac{M_n}{S_{n+1}} \ge c^+\right\} \le \frac{\alpha}{2}, \tag{75}$$

then

$$\begin{aligned} \mathbf{P}\{c^- \le W_n \text{ and } M_n \le c^+\} &= 1 - \mathbf{P}\left\{\frac{W_n}{S_{n+1}} \le c^- \text{ or } \frac{M_n}{S_n} \ge c^+\right\} \\ &\ge 1 - \mathbf{P}\left\{\frac{W_n}{S_{n+1}} \le c^-\right\} - \mathbf{P}\left\{\frac{M_n}{S_{n+1}} \ge c^+\right\} \\ &\ge 1 - \alpha, \end{aligned} \tag{76}$$

so (72) will hold. We may therefore study the behavior of M_n/S_n and W_n/S_n separately.

Consider the triangular array $\{G_{n,1}, G_{n,2}, \dots, G_{n,n}\}_{n=1}^{\infty}$ of iid $\Gamma(K,1)$ random variables. Then

$$M_n =_{\mathscr{D}} \max_{j=1,\dots,n/K} G_{n/K,j}. \tag{77}$$

Now

$$\begin{aligned} \mathbf{P}\{M_n \le g\} &= \mathbf{P}\{G_{n/K,1} \le g\}^{n/K} \\ &= \left(1 - \mathbf{P}\{G_{n/K,1} > g\}\right)^{n/K}. \end{aligned} \tag{78}$$

Integrating the gamma density by parts, we find

$$\mathbf{P}\{G_{n/K,1} > g\} = \int_g^{\infty} \frac{s^{K-1}}{\Gamma(K)} e^{-s}\, ds \tag{79}$$

$$= \sum_{j=0}^{K-1} \frac{g^j}{j!} e^{-g} \tag{80}$$

$$= \frac{g^K}{\Gamma(K)} e^{-g} \left[\frac{1}{g} + \frac{K-1}{g^2} + \cdots + \frac{(K-1)!}{g^K} \right]. \tag{81}$$

Suppose

$$g = K + \tau\sqrt{K \log(n/K)}\,. \tag{82}$$

Since $g > K$,

$$\frac{1}{g} + \frac{K-1}{g^2} + \cdots + \frac{(K-1)!}{g^K} \le K\frac{1}{g} \le 1, \tag{83}$$

and so

$$\mathbf{P}\{G_{n/K,1} > g\} \le \frac{g^K e^{-g}}{\Gamma(K)}. \tag{84}$$

Stirling's approximation to $\Gamma(K)$ yields

$$\begin{aligned} \log\left(\frac{g^K e^{-g}}{\Gamma(K)} \right) &= K \log g + K - g - \left(K - \frac{1}{2} \right) \log K - \frac{1}{2} \log 2\pi + O(K^{-1}) \\ &= K \log\left(1 + \tau\sqrt{\frac{\log(n/K)}{K}} \right) - \tau\sqrt{K \log\left(\frac{n}{K} \right)} + \frac{1}{2} \log K \\ &\quad - \frac{1}{2} \log 2\pi + O(K^{-1}). \end{aligned} \tag{85}$$

The Taylor series for $\log(1 + x)$ is an alternating absolutely convergent power series, so

$$\begin{aligned} &\log\left(\frac{g^K}{\Gamma(K)} e^{-g} \right) \\ &\quad \le K\tau\sqrt{\frac{\log(n/K)}{K}} - \frac{1}{2} K\tau^2 \frac{\log(n/K)}{K} + \frac{1}{3} K\tau^3 \left(\frac{\log(n/K)}{K} \right)^{3/2} \\ &\qquad - \tau\sqrt{K \log(n/K)} + \frac{1}{2} \log K - \frac{1}{2} \log 2\pi + O(K^{-1}) \end{aligned}$$

$$\begin{aligned} &= -\frac{1}{2}\tau^2 \log\left(\frac{n}{K} \right) + \frac{\tau^3 \log(n/K)}{3} \sqrt{\frac{\log(n/K)}{K}} \\ &\quad + \frac{1}{2} \log K - \frac{1}{2} \log 2\pi + O(K^{-1}) \end{aligned} \tag{86}$$

$$= \log\left(\frac{n}{K}\right)\left(-\frac{\tau^2}{2} + \frac{\log K}{2\log(n/K)} + \frac{\tau^3}{3}\sqrt{\frac{\log(n/K)}{K}}\right)$$
$$-\frac{1}{2}\log 2\pi + O(K^{-1}).$$

For this to yield a useful bound, we need to guarantee that

$$-\frac{1}{2}\tau^2 + \frac{\log K}{2\log(n/K)} + \frac{\tau^3}{3}\sqrt{\frac{\log(n/K)}{K}} \le -1, \tag{87}$$

that is,

$$\frac{\log K}{\log(n/K)} + \frac{2\tau^3}{3}\sqrt{\frac{\log(n/K)}{K}} \le \tau^2 - 2. \tag{88}$$

If inequality (88) holds asymptotically, we have

$$\begin{aligned}\liminf_{n\to\infty} \mathbf{P}\left\{M_n \le K + \tau\sqrt{K\log(n/K)}\right\} &\ge \lim_{n\to\infty}\left(1 - \frac{K}{n}\frac{1}{\sqrt{2\pi}}\right)^{n/K} \\ &= \exp\left(-\frac{1}{\sqrt{2\pi}}\right).\end{aligned} \tag{89}$$

A similar argument shows that

$$\liminf_{n\to\infty} \mathbf{P}\left\{W_n \ge K - \tau\sqrt{K\log(n/K)}\right\} \ge \exp(-1/\sqrt{2\pi}). \tag{90}$$

Note that if τ is chosen so that inequality (88) is strict, the probabilities in (89) and (90) will tend to 1, which is sufficient to establish a rate of convergence. The normal approximation to S_n gives

$$\mathbf{P}\left\{|S_n - n| \ge \sqrt{n}\, z_{1-\theta/2}\right\} = 1 - \theta + O(n^{-1/2}). \tag{91}$$

Thus if (88) holds strictly,

$$\lim_{n\to\infty}\left(c^-(n) - \left(\frac{K - \sqrt{K}\,\tau\sqrt{\log(n/K)}}{n + \sqrt{n}\, z_{1-\alpha/2}}\right)\right) \ge 0 \tag{92}$$

and

$$\lim_{n\to\infty}\left(c^+(n) - \left(\frac{K + \sqrt{K}\,\tau\sqrt{\log(n/K)}}{n - \sqrt{n}\, z_{1-\alpha/2}}\right)\right) \le 0. \tag{93}$$

A.3. *Asymptotic behavior of Δ_l and D.* From the normal approximation to the denominator and numerator of (71) we find

$$\Delta_l =_{\mathscr{D}} \frac{K + \sqrt{K}Z_1 + O_P(1)}{n + 1 + \sqrt{n+1}\,Z_2 + O_P(1)}, \tag{94}$$

where Z_1 and Z_2 are correlated standard normal random variables, so

$$\mathbf{P}\left\{\left|\Delta_l - \frac{K}{n}\right| \le \frac{\sqrt{K}}{n} z_{1-\theta}\right\} \ge 1 - 2\theta + O\left(\max\left(\sqrt{\frac{K}{n}}, \frac{1}{\sqrt{K}}\right)\right). \tag{95}$$

Combining (95) with (92) and (93) and simplifying the result, we find

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{c^-}{\Delta_{l+1}} \ge 1 - \tau\sqrt{\frac{\log(n/K)}{K}}\right\} \ge 1 - 2\theta \tag{96}$$

and

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{c^+}{\Delta_l} \le 1 + \tau\sqrt{\frac{\log(n/K)}{K}}\right\} \ge 1 - 2\theta. \tag{97}$$

Note that the coverage probabilities of (96) and (97) are simultaneous. Thus, in particular,

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{c^+}{\Delta_l} - \frac{c^-}{\Delta_{l+1}}\right| \le 2\tau\sqrt{\frac{\log(n/K)}{K}}\right\} \ge 1 - 2\theta, \tag{98}$$

and since f is monotone

$$\lim_{n\to\infty} \mathbf{P}\left\{f(X_{l_1})\left|\frac{c^+}{\Delta_l} - \frac{c^-}{\Delta_{l+1}}\right| \le 2\tau f(y)\sqrt{\frac{\log(n/K)}{K}}\right\} \ge 1 - 2\theta. \tag{99}$$

We also estimate (67):

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{c^+}{\Delta_l}(\Delta_l + \Delta_{l+1})^\varrho \le 2\left(1 + \tau\sqrt{\frac{\log(n/K)}{K}}\right)\left(\frac{K}{n} + \frac{\sqrt{K}}{n} z_{1-\theta}\right)^\varrho\right\} \ge 1 - 2\theta. \tag{100}$$

Keeping the leading term in the Taylor expansion gives

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{c^+}{\Delta_l}(\Delta_l + \Delta_{l+1})^\varrho \le 2\left(\frac{K}{n}\right)^\varrho\right\} \ge 1 - 2\theta. \tag{101}$$

Since f is continuous and strictly positive at y, $\lim_{n\to\infty} X_{l_1} = y$ a.s. and hence

$$\lim_{n\to\infty} \mathbf{P}\left\{f^{-\varrho}(X_{l_1})\frac{c^+}{\Delta_l}(\Delta_l + \Delta_{l+1})^\varrho \le 2f^{-\varrho}(y)\left(\frac{K}{n}\right)^\varrho\right\} \ge 1 - 2\theta. \tag{102}$$

Thus

$$\liminf_{n\to\infty} \mathbf{P}\left\{|\gamma^+(y) - \gamma^-(y)| \le 2f(y)\tau\sqrt{\frac{\log(n/K)}{K}} + 2Cf^{-\varrho}(y)\left(\frac{K}{n}\right)^\varrho\right\} > 0, \tag{103}$$

which proves Theorem 7.1.

A.4. *Proving Corollary* 7.2. To prove Corollary 7.2, we need to find the optimal dependence of K on n. The choice $K = B(n^{2\varrho} \log n)^{1/(1+2\varrho)}$, for some $B > 0$, leads to

$$\liminf_{n \to \infty} \mathbf{P}\Bigg\{ |\gamma^+(y) - \gamma^-(y)| \le 2f(y)\left(\frac{\tau}{\sqrt{1+2\varrho}} B^{-1/2} + Cf^{-(\varrho+1)}(y) B^{\varrho}\right) \times \left(\frac{\log n}{n}\right)^{\varrho/(1+2\varrho)} \Bigg\} > 0. \tag{104}$$

The "rate" $(\log n/n)^{\varrho/(1+2\varrho)}$ is the same as the lower bound found by Khas'minskii [14], and is therefore optimal. We can find the optimal constant B in K as well; stationarity yields

$$B = \left(\frac{\tau}{2C\varrho}\right)^{2/(1+2\varrho)} (1 + 2\varrho)^{-1/(1+2\varrho)} f^{2(1+\varrho)/(1+2\varrho)}(y). \tag{105}$$

It remains to verify that a constant τ satisfying (88) exists, and to choose the smallest possible τ to optimize the constant B. For K of the chosen form, (88) requires that asymptotically

$$\tau^2 \ge 2 + 2\varrho, \tag{106}$$

so the smallest constant B results from $\tau^2 = 2 + 2\varrho$. This yields

$$K = \left(\frac{\sqrt{1+\varrho}}{\sqrt{2}\, C\varrho}\right)^{2/(1+2\varrho)} (1 + 2\varrho)^{-1/(1+2\varrho)} f^{2(1+\varrho)/(1+2\varrho)}(y) \times (n^{2\varrho} \log n)^{1/(1+2\varrho)}. \tag{107}$$

Combining all the results, we find that for K defined by (107),

$$\liminf_{n \to \infty} \mathbf{P}\Bigg\{ |\gamma^+(y) - \gamma^-(y)| \le \left(\frac{f(y)}{1+2\varrho}\right)^{\varrho/(1+2\varrho)} (4C\varrho(1+\varrho))^{1/(1+2\varrho)} \times (2 + \varrho^{-1})\left(\frac{\log n}{n}\right)^{\varrho/(1+2\varrho)} \Bigg\} > 0, \tag{108}$$

which proves Corollary 7.2.

A.5. *Proving Corollary* 7.3. By Lemma 7.4, for any $\varepsilon > 0$, there exists a $C(\varepsilon)$ such that f is Lipschitz with exponent 1 and constant $C(\varepsilon)$ on a set $T_\varepsilon \subset T$ whose measure differs from that of T by at most ε. By the assumptions of the corollary, there is a positive probability of increasing numbers of observations below c and above d.

We integrate (103):

$$\liminf_{n\to\infty} \mathbf{P}\left\{ \int_{T_\varepsilon} |\gamma^+(y) - \gamma^-(y)|\, dy \right.$$

$$(109) \qquad \le 2\left(\int_{T_\varepsilon} f(y)\, dy\right)\tau\sqrt{\frac{\log(n/K)}{K}}$$

$$\left. +2C(\varepsilon)\left(\int_{T_\varepsilon} f^{-\varrho}(y)\, dy\right)\left(\frac{K}{n}\right)^{\varrho}\right\} > 0.$$

Using the same *choice* for K,

$$(110) \qquad K = B(n^{2\varrho}\log n)^{1/(1+2\varrho)},$$

we find that

$$\liminf_{n\to\infty} \mathbf{P}\left\{ \int_{T_\varepsilon} |\gamma^+(y) - \gamma^-(y)|\, dy \right.$$

$$(111) \qquad \le \left[2\left(\int_{T_\varepsilon} f(y)\, dy\right)\tau\frac{1}{\sqrt{(1+2\varrho)B}} + 2C(\varepsilon)B^{\varrho}\left(\int_{T_\varepsilon} f^{-\varrho}(y)\, dy\right)\right]$$

$$\left. \times\left(\frac{\log n}{n}\right)^{\varrho/(1+2\varrho)}\right\} > 0.$$

Incorporating the constraint (88) on τ and solving for the optimal B as before, we find

$$(112) \qquad K = \left(\frac{\sqrt{1+\varrho}\,(\int_{T_\varepsilon} f)}{\sqrt{2(1+2\varrho)}\,C(\varepsilon)\varrho\int_{T_\varepsilon} f^{-\varrho}}\right)^{2/(1+2\varrho)} (n^{2\varrho}\log n)^{1/(1+2\varrho)}.$$

With this choice of K, we find

$$\liminf_{n\to\infty} \mathbf{P}\left\{ \int_{T_\varepsilon} |\gamma^+(y) - \gamma^-(y)|\, dy \right.$$

$$(113) \qquad \le \left(\frac{\sqrt{2(1+\varrho)}\,(\int_{T_\varepsilon} f)}{\sqrt{1+2\varrho}}\right)^{2\varrho/(1+2\varrho)}$$

$$\left. \times\left(2\varrho C(\varepsilon)\int_{T_\varepsilon} f^{-\varrho}\right)^{1/(1+2\varrho)}\left(2+\frac{1}{\varrho}\right)\right\} > 0.$$

Acknowledgments. We are grateful to Peter Bickel, David Donoho, Lutz Dumbgen, Cliff Frohlich, Christopher Genovese, Lucien Le Cam, Mark Low and Charles Stone for discussions, suggestions and comments on previous drafts. David Donoho and Mark Low suggested looking at order statistics. We are grateful to R. Uhrhammer for USGS data for the 1984 Morgan Hill, CA event.

REFERENCES

[1] BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

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

[3] BOGOMOLOV, N. A. (1979). Confidence regions for unimodal and symmetric distribution functions. *Theory Probab. Appl.* **24** 419–423.

[4] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41.

[5] DALENIUS, T. (1965). The mode—a neglected statistical parameter. *J. Roy Statist. Soc. Ser. A* **128** 110–117.

[6] DAVIS, S. D. and FROHLICH, C. (1991). Single-link cluster analysis of earthquake aftershocks: Decay laws and regional variations. *Journal of Geophysical Research* **96** 6335–6350.

[7] DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

[8] DONOHO, D. L. (1988). One-sided inference about functionals of a density. *Ann. Statist.* **16** 1390–1420.

[9] GRENANDER, U. (1956). On the theory of mortality measurement. *Skand. Aktuarietidskr.* **39** 125–153.

[10] 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** 529–555. Wadsworth, Belmont, CA.

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

[12] HENGARTNER, N. W. and STARK, P. B. (1992). Conservative finite-sample confidence envelopes for monotone and unimodal densities. Technical Report 341, Dept. Statistics, Univ. California, Berkeley.

[13] KAGAN, Y. Y. and KNOPOFF, L. (1987). Random stress and earthquake statistics: Time dependence. *Geophysical Journal of the Royal Astronomical Society* **88** 723–731.

[14] KHAS'MINSKII, R. Z. (1978). A lower bound on the risks of non-parametric estimates of densities in the uniform metric. *Theory Probab. Appl.* **23** 794–798.

[15] KISSLINGER, C. (1993). The stretched exponential function as an alternative model for aftershock decay rate. *Journal of Geophysical Research* **98** 1913–1921.

[16] MASSART, P. (1990). The tight constant in the Dvoretsky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283.

[17] MIKUMO, T. and MIYATAKE, T. (1979). Earthquake sequences on a frictional fault model with nonuniform strengths and relaxation times. *Geophysical Journal of the Royal Astronomical Society* **59** 497–522.

[18] OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83** 9–27.

[19] OMORI, F. (1895). On the aftershocks of earthquakes. *Journal of the College of Science, Imperial University of Tokyo* **7** 111–200.

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

[21] REASENBERG, P. A. and JONES, L. M. (1989). Earthquake hazard after a main shock in southern California. *Science* **243** 1173–1176.

[22] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference.* Wiley, Chichester.

[23] SILVERMAN, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B* **43** 97–99.

[24] SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

[25] STARK, P. B. (1992). Inference in infinite-dimensional inverse problems: Discretization and duality. *Journal of Geophysical Research* **97** 14,055–14,082.

[26] UTSU, T. (1961). A statistical study on the occurrence of aftershocks. *Geophysical Magazine* **30** 521–605.

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

[28] WANG, Y. (1991). Contributions to the theory of estimation of monotone and unimodal densities. Technical Report 325, Dept. Statistics, Univ. California, Berkeley.

[29] YAMASHITA, T. and KNOPOFF, L. (1987). Models of aftershock occurrence. *Geophys. J. Roy. Astron. Soc.* **91** 13–26.

DEPARTMENT OF STATISTICS
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06510

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720