The Annals of Statistics
2002, Vol. 30, No. 2, 528–555

RISK BOUNDS IN ISOTONIC REGRESSION

BY CUN-HUI ZHANG

Rutgers University

Nonasymptotic risk bounds are provided for maximum likelihood-type isotonic estimators of an unknown nondecreasing regression function, with general average loss at design points. These bounds are optimal up to scale constants, and they imply uniform $n^{-1/3}$-consistency of the ℓ_p risk for unknown regression functions of uniformly bounded variation, under mild assumptions on the joint probability distribution of the data, with possibly dependent observations.

1. Introduction. In this paper, we provide nonasymptotic risk bounds for maximum likelihood-type isotonic estimators of an unknown nondecreasing regression function, with general average loss at design points, for possibly dependent observations.

In the simplest model under consideration here, the relationship between the response variables y_i and covariates t_i is specified by

$$y_i \equiv f(t_i) + \varepsilon_i, \qquad 1 \le i \le n, \tag{1.1}$$

where ε_i are i.i.d. errors with $E\varepsilon_i = 0$ and $E\varepsilon_i^2 = \sigma^2$, t_i are deterministic design points and $f(t)$ is a nondecreasing regression function. The least squares estimator (LSE) of the unknown f is a left-continuous step function $\hat{f}_n$ with jumps only at t_i, defined by

$$\hat{f}_n \equiv \arg\min\left\{\sum_{i=1}^{n}\big(y_i - f(t_i)\big)^2 : f \text{ is nondecreasing}\right\}. \tag{1.2}$$

Let $V(f)$ be the total variation of f. In Sections 2 and 3, we develop uniform upper bounds, in terms of $(n, V(f), \sigma)$, for the ℓ_p risk

$$R_{n,p}(f) \equiv \left(\frac{1}{n}\sum_{i=1}^{n} E\big|\hat{f}_n(t_i) - f(t_i)\big|^p\right)^{1/p}. \tag{1.3}$$

Our risk bounds are quite sharp. For $1 \le p < 3$, they imply the uniform cube-root convergence with tight constants:

$$0.64 + o(1) \le \frac{n^{1/3}}{\sigma^{2/3}V^{1/3}} \sup_{V(f)\le V} R_{n,p}(f) \le M_p + o(1), \tag{1.4}$$

Received February 1999; revised September 2001.

AMS 2000 subject classifications. Primary 62G08, 62G05; secondary 62J02, 62G20.

Key words and phrases. Nonparametric regression, isotonic regression, risk bounds, least squares estimator, maximum likelihood estimator.

where M_p, depending on p only, are the constants in Theorem 2.3, for example, $M_2 < 2.75$.

The LSE (1.2) and related methods for estimating a monotone regression or density function $f(\cdot)$ were proposed by Ayer, Brunk, Ewing, Reid and Silverman (1955), van Eeden (1956) and Grenander (1956). The convergence of $n^{1/3}\{\hat{f}_n(x_0) - f(x_0)\}$ in distribution at a fixed x_0 was established by Prakasa Rao (1969) and Brunk (1970). Groeneboom (1985) obtained asymptotic distributions of the L_1 loss and $\hat{f}_n(x_0)$ for the Grenander estimator. van de Geer (1990, 1993) obtained rates of convergence in probability for the LSE and median regression estimators, including the $n^{-1/3}$-consistency in probability of the ℓ_2 loss of (1.2) for independent errors with $\max_{i \le n} E \exp(b_0 \varepsilon_i^2) = O(1)$ for some $b_0 > 0$. Donoho (1991) obtained $n^{1/3} \sup_{V(f) \le V} R_{n,2}(f) = O(1)$ for i.i.d. normal errors. Birgé and Massart (1993) weakened Donoho's assumption on i.i.d. errors to the finiteness of some exponential moment. Wang (1996) considered nonasymptotic bounds of $R_{n,2}(f)$ and the boundedness of $(n/\log n)^{1/3} \sup_{V(f) \le V} R_{n,2}(f)$ for i.i.d. errors with finite second moment. Recently, Meyer and Woodroofe (2000) obtained bounds for $R_{n,2}^2(f)$ for i.i.d. normal errors based on Stein's (1981) unbiased estimation of mean squared errors. For estimating monotone densities, Birgé (1987, 1989) obtained nonasymptotic bounds for the L_1 risk of the Grenander estimator which imply the $n^{-1/3}$-consistency. For a general discussion of statistical methods with order restrictions, see Barlow, Bartholomew, Bremner and Brunk (1972), Grenander (1980), Robertson, Wright and Dykstra (1988) and Groeneboom and Wellner (1992).

Our risk bounds are derived through an inequality for the number of terms greater than $\{\sigma^2/m\}^{p/2}$ in the sum in (1.3). As a result, we use relatively "light" probabilistic tools, for example, Doob's inequality for martingales and "good-λ inequality," instead of entropy-type arguments, as used, for example, in van de Geer (1990). Our methods are applicable to general loss functions and dependent observations and allow model misspecification for nonmonotone regression functions. These extensions are given in Sections 4 and 5. In Sections 6–8, we consider general isotonic estimators of the form

$$\hat{f}_n \equiv \arg\max\left\{\sum_{i=1}^{n} \phi_i\big(f(t_i); y_i\big) : f \text{ is nondecreasing}\right\},$$

for example, log-likelihood $\phi_i(\theta; y) \equiv \log\{g_i(y|\theta)\}$ for certain densities g_i. To simplify the notation, we assume throughout that $t_1 \le t_2 \le \cdots \le t_n$. Let $x_+ \equiv x \vee 0$ and $x_- \equiv (-x)_+$.

2. Risk bounds for the LSE.

For $p \ge 1$, define

$$r_{+,p}(m, v) \equiv \max_{n_1 < j \le n_2 - m} E\left(v + \min_{j \le \ell \le j+m} \max_{1 \le k \le j} \frac{\sum_{i=k}^{\ell} \varepsilon_i}{\ell - k + 1}\right)_+^p,$$

$$r_{-,p}(m, v) \equiv \max_{n_1 + m < j \le n_2} E\left(-v + \min_{j \le \ell \le n} \max_{j - m \le k \le j} \frac{\sum_{i=k}^{\ell} \varepsilon_i}{\ell - k + 1}\right)_-^p,$$

$0 \le n_1 \le n_2 \le n$, $m = 0, 1, 2, \dots$, $v \ge 0$, and define

$$r_p(m, v) \equiv r_{p,n_1,n_2}(m, v) \equiv r_{+,p}(m, v) + r_{-,p}(m, v). \tag{2.1}$$

THEOREM 2.1. *Let $\hat{f}_n$ be the LSE in* (1.2) *based on observations* (y_i, t_i), $i = 1, \dots, n$, *from the regression model* (1.1) *with a nondecreasing $f(t)$ and arbitrary errors $\{\varepsilon_i\}$. Let $p \ge 1$, $0 \le n_1 \le n_2 \le n$ and $r_p(m, v)$ be as in* (2.1). *Then*

$$\frac{1}{n_*} \sum_{j=n_1+1}^{n_2} E\big|\hat{f}_n(t_j) - f(t_j)\big|^p \le \int_{0<x<\infty} r_p\big(\lfloor x \rfloor, v(\lfloor x \rfloor)\big)\, dH_v\left(x; n_*, \frac{V_*}{2}\right) \tag{2.2}$$

for all nonincreasing, nonnegative continuous functions $v(x)$, where $V_* \equiv f(t_{n_2}) - f(t_{n_1+1})$, $n_* \equiv n_2 - n_1$, $\lfloor x \rfloor$ is the integer part of x and $H_v(x; n, V)$ is a continuous increasing function of x with

$$H_v(x; n, V) \equiv \min\big[1, x\{1 + V/v(x)\}/n\big].$$

Theorem 2.1, proved in Section 3, imposes no assumption on the stochastic structure of the errors $\{\varepsilon_i\}$. Since (2.1) depends only on moment-type properties of the familiar partial-sum processes of the errors, ℓ_p risk bounds for the LSE can be easily derived from (2.2); cf. (3.5) and (3.7) below. The risk bound in (2.2) can be viewed as a weighted sum of $r_p(m, v(m))$ with total weight $H_v(\infty; n_*, V_*/2) = 1$; that is, $E|\hat{f}_n(t_j) - f(t_j)|^p \le r_p(m_j, v(m_j))$ for certain m_j, with the worst distribution of $\{m_j : n_1 < j \le n_2\}$ being dominated by the discrete version of $H_v(dx; n_*, V_*/2)$. The power of (2.2) rests in its validity for all nonincreasing functions $v(\cdot)$, for example, allowing optimization over a parametric family of such functions. Moreover, (2.2) is localized since the bound for the interval $\{j : n_1 < j \le n_2\}$ depends only on the total variation of f in the interval $[t_{n_1+1}, t_{n_2}]$.

In the rest of this section, we shall focus on independent errors with zero mean and bounded variance. Since the order of $r_p(m, v)$ is $v^p + m^{-p/2}$ in the i.i.d. case, it is natural to consider $v(m) \equiv c/\sqrt{m+1}$. We shall provide risk bounds for (1.3) only, as their local versions can be generated from Theorem 2.1 in the same manner with $n \to n_*$ and $V(f) \to V_*$. Let

$$J_p(n, V) \equiv \int_{0<x<\infty} (x \vee 1)^{-p/2} d \min\left[1, n^{-1} \int_0^x \{1 + (3/2)V\sqrt{t \vee 1}\}\, dt\right]. \tag{2.3}$$

By calculus, $J_p(n, V) \sim n^{-(p \wedge 3)/3} (\log n)^{I\{p=3\}}$ for fixed $V > 0$ and $p \ge 1$; cf. (3.7) and Lemma A.1. Let $r_{p,n_1,n_2}(m, v)$ be as in (2.1) and define

$$K^*_{p,c} \equiv \left\{\sup_{m\ge 0} \frac{r_{p,0,n}(m, c/\sqrt{m+1})}{(m+1)^{-p/2}}\right\}^{1/p}, \qquad p \ge 1,\ c > 0. \tag{2.4}$$

For nonnegative random variables X, $c > 0$ and $1 \le p < 3$, define

$$M_{p,c}(X) \equiv \left\{ \frac{6E(c+X)^p}{(3-p)(2c)^{p/3}} \right\}^{1/p}, \qquad M_p(X) \equiv \inf_{c>0} M_{p,c}(X). \tag{2.5}$$

THEOREM 2.2. (i) *Let* $R_{n,p}(f)$, $J_p(n,V)$ *and* $K^*_{p,c}$ *be as in* (1.3), (2.3) *and* (2.4), *respectively*, *and let* $V(f)$ *be the total variation of* f. *Then*

$$R_{n,p}(f) \le \inf_{c>0} K^*_{p,c}\{J_p(n, V(f)/(2c))\}^{1/p}, \qquad p \ge 1. \tag{2.6}$$

(ii) *If* ε_i *are independent with* $E\varepsilon_i = 0$ *and* $E|\varepsilon_i|^{p\vee 2} \le \sigma_p^{p\vee 2}$, $p \ge 1$, *then*

$$\begin{aligned} &R_{n,p}(f) \\ &\quad \le 2^{1/p}\sigma_p \inf_{c>0}\big[(c/2 + C_p)\{J_p(n, V(f)/(c\sigma_p))\}^{1/p}\big] \\ &\quad \le 2^{1/p}\sigma_p C_p \min\left[1, \frac{3}{2}\left\{\frac{3}{(3-p)_+}\left(\frac{V(f)}{n\sigma_p C_p}\right)^{p/3} + \frac{1}{n}\int_0^n \frac{dx}{(x\vee 1)^{p/2}}\right\}^{1/p}\right], \end{aligned} \tag{2.7}$$

where C_p *are constants depending on* p *only in general*, *and* $C_p = \sqrt{2}$ *for i.i.d.* ε_i *with* $p \le 2$.

(iii) *If* ε_i *are i.i.d.* $N(0, \sigma^2)$ *with* $\sigma \le \sigma_p$, *then* (2.7) *holds with* $C_p = 1$ *for* $1 \le p \le 2$, *and for* $1 \le p < 3$

$$R_{n,p}(f) \le \sigma M_p(Z_0)\left\{\left(\frac{V(f)}{n\sigma}\right)^{p/3} + \frac{1}{n}\int_0^n (x\vee 1)^{-p/2}\,dx\right\}^{1/p}, \tag{2.8}$$

where $Z_0 \sim |N(0,1)|$ *and* $M_p(X)$ *is as in* (2.5). *In particular*, $M_2(Z_0) \approx 3.50$.

For $1 \le p \le 2$ with $\sigma_p^2 \equiv \sigma^2 = E\varepsilon_i^2$, the statistical content of the right-hand side of (2.7) is clear: (a) the lower bound $\sigma\{\int_0^n (x\vee 1)^{-p/2}\,dx/n\}^{1/p}$ is due to the spikes of the LSE near the large jumps of f and the endpoints t_1 and t_n [cf. (2.11)]; (b) the upper bound σ represents the minimax error for estimating $f(t_i)$ by y_i for each i when $V(f)$ is of larger order than n and $f(t_i)$ are widely spread; (c) between these two extreme cases, $\sigma\{V(f)/(n\sigma)\}^{1/3}$ provides the cube-root consistency of the LSE when $V(f) = O(1)$. None of these three factors can be removed from (2.7). In this sense, (2.7) is sharp up to a scale constant, and the conditions cannot be weakened.

For i.i.d. normal errors and $p = 2$, Meyer and Woodroofe (2000) proved that

$$R^2_{n,2}(f) \le \frac{\sigma^2 ED_n}{n} \le \frac{\sigma^2}{n}\left[\kappa_0\left\{\frac{V(f)}{\sigma} + \log n\right\} + \kappa_1\left\{\frac{V(f)}{\sigma}\right\}^{2/3} n^{1/3}\right],$$

where $D_n \equiv \{1 < j \le n : \hat{f}_n(t_j) > \hat{f}_n(t_{j-1})\}$. Since $D_n \le n$, their results imply (2.7) up to a constant factor in this special case. The constants $\sqrt{\kappa_0}$ and $\sqrt{\kappa_1}$, comparable to our $M_2(Z_0) \approx 3.50$ in (2.8), were not explicitly given.

Next, we consider asymptotic bounds. Let $\varphi_1(x) \equiv 4\{\varphi(x) - x\int_x^\infty \varphi(y)\,dy\} \times I_{\{x>0\}}$, with $\varphi(x) \equiv e^{-x^2/2}/\sqrt{2\pi}$. By calculus, we have $\int_0^\infty x^p \varphi_1(x)\,dx = 4\int_0^\infty x^p \varphi(x)\,dx/(p+2)$ for $p > -1$; for example, $(4/3)/\sqrt{2\pi}$ for $p = 1$ and $1/2$ for $p = 2$. Groeneboom (1983) identified φ_1 as the density of the slope, at $t = 1$, of the concave majorant of the standard Brownian motion. We shall consider a double array of errors $\varepsilon_i \equiv \varepsilon_{n,i}$, $1 \le i \le n$, in (1.1).

THEOREM 2.3. *Let Z_1 be a variable with density φ_1 and let Z be the location of the maximum of $W(t) - t^2$ for a two-sided standard Brownian motion W. Suppose $\{\varepsilon_i \equiv \varepsilon_{n,i},\ i \le n\}$ are independent variables with $E\varepsilon_{n,i} = 0$ and $E\varepsilon_{n,i}^2 = \sigma^2$, $\{\varepsilon_{n,i}^2,\ i \le n,\ n \ge 1\}$ is uniformly integrable and $\sup_n \max_{i\le n} E|\varepsilon_{n,i}|^p < \infty$. Let $1 \le p < 3$. Then, for $V > 0$ and large n,*

$$\begin{aligned} &2^{2/3}\{E|Z|^p\}^{1/p} + o(1) \\ &\qquad \le M_{n,p} \equiv \frac{n^{1/3}}{\sigma^{2/3}V^{1/3}} \sup_{V(f)\le V} R_{n,p}(f) \le M_p + o(1), \end{aligned} \tag{2.9}$$

where $M_p \equiv M_p(Z_1)$ are as in (2.5); for example, $M_2 < 2.75$. If the empirical $\sum_{i=1}^n I\{t_i \le t\}/n$ converges in distribution to a continuous $G(t)$, then

$$n^{1/3} R_{n,p}(f) \le M_p \sigma^{2/3}\left[\int \{df(t)/dG(t)\}^{p/3}\,dG(t)\right]^{1/p} + o(1), \tag{2.10}$$

where df/dG is the Radon–Nikodym derivative of the absolutely continuous part of f with respect to G. If $f(\cdot)$ is a constant and $1 \le p \le 2$, then $R_{n,p}^p(f) \le \sum_{m=0}^n r_p(m,0)/n$ and

$$\begin{aligned} R_{n,p}^p(f) &= \bigl(1 + o(1)\bigr)\sum_{m=0}^n \frac{r_p(m,0)}{n} \\ &= \bigl(1 + o(1)\bigr)\frac{\sigma^p}{n}\left\{2\int x^p \varphi_1(x)\,dx\right\}\int_0^n (x \vee 1)^{-p/2}\,dx. \end{aligned} \tag{2.11}$$

REMARK 2.1. By Groeneboom (1985), $E|Z| \approx 0.41$, so that the lower bound on the left-hand side of (2.9) is no less than $2^{2/3}E|Z| > 0.64$ and (1.4) holds. The proof of Theorem 2.3 indicates that the lower bound in (2.9) is sharp and that (2.10) should hold with equality for $M_p = 2^{2/3}\{E|Z|^p\}^{1/p}$.

REMARK 2.2. If df is singular to dG, then $n^{1/3}R_{n,p}(f) \to 0$ by (2.10) for $p < 3$.

3. Proofs of Theorems 2.1–2.3. We provide a mathematical description of our basic ideas here by proving our risk bounds in the simplest model (1.1).

PROOF OF THEOREM 2.1. Let $f_i \equiv f(t_i)$ and $\bar{f}_{k,\ell} \equiv \sum_{i=k}^{\ell} f_i/(\ell-k+1)$. The proof is based on the well-known minimax formula for (1.2):

$$\hat{f}_n(t_j) = \min_{\ell \geq j} \max_{k \leq j} \frac{\sum_{i=k}^{\ell} y_i}{\ell - k + 1}; \tag{3.1}$$

cf. page 23 of Robertson, Wright and Dykstra (1988) and Proposition 6.1. Define $m_j \equiv \max\{m \geq 0 : \bar{f}_{j,j+m} \leq f_j + v(m),\ \ j+m \leq n_2\}$. The minimax formula implies

$$\begin{aligned} \hat{f}_n(t_j) &\leq \min_{j \leq \ell \leq j+m_j} \max_{k \leq j} \left(\frac{\sum_{i=k}^{\ell} \varepsilon_i}{\ell - k + 1} + \bar{f}_{k,\ell} \right) \\ &\leq f_j + v(m_j) + \min_{j \leq \ell \leq j+m_j} \max_{k \leq j} \frac{\sum_{i=k}^{\ell} \varepsilon_i}{\ell - k + 1}, \end{aligned} \tag{3.2}$$

as $\bar{f}_{k,\ell}$ is nondecreasing in both k and ℓ. Thus, by the definition of $r_{+,p}(m, v)$ above (2.1), $E(\hat{f}_n(t_j) - f_j)_+^p \leq r_{+,p}(m_j, v(m_j))$. Set $\ell(m) \equiv \#\{j : m_j < m,\ n_1 < j \leq n_2\}$. We have

$$\sum_{j=n_1+1}^{n_2} E\big(\hat{f}_n(t_j) - f_j\big)_+^p \leq \sum_{m=0}^{\infty} r_{+,p}\big(m, v(m)\big)\big\{\ell(m+1) - \ell(m)\big\}. \tag{3.3}$$

Since $\bar{f}_{j,j+m}$ is nondecreasing in m, $m_j \leq m$ and $n_1 + 1 \leq j \leq n_2 - (m+1)$ imply $\bar{f}_{j,j+m+1} - f_j \geq v(m+1)$, so that $\ell(m+1)$ is bounded by the sum of $m+1$ and

$$\begin{aligned} \sum_{j=n_1+1}^{n_2-(m+1)} \frac{\bar{f}_{j,j+m+1} - f_j}{v(m+1)} &= \sum_{j=n_1+1}^{n_2-(m+1)} \sum_{i=j}^{j+m+1} \frac{\sum_{k=j}^{i-1}(f_{k+1} - f_k)}{(m+2)v(m+1)} \\ &\leq \sum_{k=n_1+1}^{n_2-1} \sum_{j=k-m}^{k} \sum_{i=k+1}^{j+m+1} \frac{f_{k+1} - f_k}{(m+2)v(m+1)} \\ &= \sum_{k=n_1+1}^{n_2-1} \frac{f_{k+1} - f_k}{(m+2)v(m+1)} \frac{(m+1)(m+2)}{2}. \end{aligned} \tag{3.4}$$

Thus,

$$\ell(m+1) \leq \min\big(n_*, (m+1)\big[1 + V_*/\{2v(m+1)\}\big]\big) = n_* H_v(m+1; n_*, V_*/2).$$

Since $r_{p,+}(m, v(m))$ is nonincreasing in m, we are allowed to replace $\ell(m)$ by its upper bound $n_* H_v(m; n_*, V_*/2)$ in (3.3). The proof is completed by applying the same method to the negative part and then summing the two parts together. □

PROOF OF THEOREM 2.2. (i) Set $v(x) \equiv c/\sqrt{x+1}$. By (1.2), Theorem 2.1 and (2.4),

$$R_{n,p}^p(f) \le \int_0^\infty (K_{p,c}^*)^p (1+\lfloor x\rfloor)^{-p/2}\, dH_0\big(x; n, V(f)/(2c)\big),$$

where $H_0(x; n, V) \equiv \min\{1, x(1+V\sqrt{x+1})/n\}$. Thus, (2.6) follows from

$$\int_0^\infty (1+\lfloor x\rfloor)^{-p/2}\, dH_0(x; n, V) \le J_p(n, V). \tag{3.5}$$

Inequality (3.5) is part of Lemma A.1.

(ii) Let $h_p(t) \equiv \{v + t^{1/(p\vee 2)}\}^p$ and

$$Y_{+,j,m} \equiv \max_{1\le k\le j} \frac{(\sum_{i=k}^{j+m} \varepsilon_i)_+}{j+m-k+1}, \qquad Y_{-,j,m} \equiv \max_{\ell\ge j} \frac{(\sum_{i=j-m}^{\ell} \varepsilon_i)_-}{\ell - j + m + 1}.$$

Since $h_p(t)$ is concave for $t > 0$, by (2.1) and the Jensen inequality,

$$\begin{aligned} r_p(m, v) &\le \sup_j E h_p\big(Y_{+,j,m}^{p\vee 2}\big) + \sup_j E h_p\big(Y_{-,j,m}^{p\vee 2}\big) \\ &\le 2h_p\bigg(\frac{1}{2}\Big(\sup_j EY_{+,j,m}^{p\vee 2} + \sup_j EY_{-,j,m}^{p\vee 2}\Big)\bigg). \end{aligned}$$

Since ε_i are independent, it follows from (A.7) of Lemma A.2, with $b_i = \max(i, m+1)$, that $\sup_j EY_{\pm,j,m}^{p\vee 2} \le C_p^{p\vee 2}\sigma_p^{p\vee 2}/(m+1)^{(p\vee 2)/2}$ for certain universal constants C_p. Thus,

$$(K_{p,c}^*)^p \le \sup_{m\ge 0} \frac{r_p(m, c/\sqrt{m+1})}{(m+1)^{-p/2}} \le 2(c + C_p\sigma_p)^p. \tag{3.6}$$

For i.i.d. ε_i and $1 \le p \le 2$, the exchangeability of ε_i and an application of Doob's inequality for the reverse submartingales $(\sum_{i=1}^{\ell} \varepsilon_i/\ell)_\pm$ yield

$$\begin{aligned} \sup_j EY_{+,j,m}^2 + \sup_j EY_{-,j,m}^2 &\le E \sup_{\ell\ge m+1} \bigg(\sum_{i=1}^{\ell} \frac{\varepsilon_i}{\ell}\bigg)_+^2 + E \sup_{\ell\ge m+1} \bigg(\sum_{i=1}^{\ell} \frac{\varepsilon_i}{\ell}\bigg)_-^2 \\ &\le 4E\bigg(\sum_{i=1}^{m+1} \frac{\varepsilon_i}{m+1}\bigg)_+^2 + 4E\bigg(\sum_{i=1}^{m+1} \frac{\varepsilon_i}{m+1}\bigg)_-^2 \\ &= \frac{4\sigma^2}{m+1}, \end{aligned}$$

so that (3.6) holds with $C_p = \sqrt{2}$. Thus, in either the general or the i.i.d. cases, (2.6) and (3.6) imply the first inequality of (2.7), with the C_p stated, after a change of variable $c \to c\sigma_p/2$.

The second inequality of (2.7) follows from

$$J_p(n,V) \le \min\left\{1, \frac{3}{(3-p)_+}\left(\frac{V}{n}\right)^{p/3} + \frac{1}{n}\int_0^n (x\vee 1)^{-p/2}\,dx\right\}, \tag{3.7}$$

which is part of Lemma A.1. Note that $c=0$ and $c=C_p$ are used in the infimum in (2.7) respectively for the first and second bounds in the minimum.

(iii) For normal ε_i, $\ell^{-1}\sum_{i=1}^{\ell}\varepsilon_i/\sigma = \tilde{W}(\ell)/\ell = W(1/\ell)$ for some Brownian motion processes $\tilde{W}(\cdot)$ and $W(\cdot)$, so that

$$\sqrt{m}\sup_{\ell\ge m}\ell^{-1}\left(\sum_{i=1}^{\ell}\varepsilon_i\right)_{\pm}\Big/\sigma \le \sqrt{m}\max_{t\le 1/m}\{\pm W(t)\}\sim Z_0.$$

Thus, $(K^*_{p,c})^p \le 2E\bigl(c+\sigma Z_0\bigr)^p$ by (2.1) and (2.4). This implies $(K^*_{p,c})^p \le 2(c+\sigma)$ for $p\le 2$ by the concavity of $(1+\sqrt{x})^p$, so that (2.7) holds with $C_p=1$.

Finally, let us prove (2.8). Assume $\sigma=1$ by scale invariance. By (2.6) and (3.7),

$$R^p_{n,p} \le \frac{6E(c+Z_0)^p}{(3-p)(2c)^{p/3}}\left\{\left(\frac{V}{n}\right)^{p/3} + \bigl\{(2c)^{p/3}(3-p)/3\bigr\}\frac{1}{n}\int_0^n (x\vee 1)^{-p/2}\,dx\right\},$$

since $(K^*_{p,c})^p \le 2E(c+Z_0)^p$. By (2.5), the rest follows from $(2c_p)^{p/3}(3-p)/3\le 1$, proved in Lemma A.3, where $c_p \equiv \arg\min\{M_{p,c}(Z_0): c>0\}$. □

PROOF OF THEOREM 2.3. By the uniform integrability of $\varepsilon^2_{n,i}$, the Lindeberg condition holds uniformly for $\{\varepsilon_{n,i}, k\le i\le \ell\}$ as $\ell-k\to\infty$. Thus, by the invariance principle,

$$\frac{\sqrt{m+1}}{\sigma}\min_{j\le\ell\le j+m}\max_{1\le k\le j}\frac{\sum_{i=k}^{\ell}\varepsilon_{n,i}}{\ell-k+1} \approx_D \min_{0<s<1}\max_{t>1}\frac{W(s)-W(t)}{t-s}\sim\varphi_1,$$

where $t\approx (j+m-k+1)/(m+1)$, $s\approx (j+m-\ell)/(m+1)$ and $W(\cdot)$ is a standard Brownian motion. Let $v(x)\equiv c/\sqrt{x+1}$. By (2.1) and as in the proof of Theorem 2.2,

$$\lim_{m\to\infty}\frac{r_p\bigl(m,v(m)\bigr)}{(m+1)^{-p/2}} = \lim_{m\to\infty}\frac{2r_{\pm,p}\bigl(m,v(m)\bigr)}{(m+1)^{-p/2}} = 2E(c+\sigma Z_1)^p.$$

Since $H_v(x;n,V/2)=O(1/n)$ for all $x>0$ and $V>0$, by (2.2),

$$\sup_{V(f)\le V} R^p_{n,p}(f) \le \bigl(1+o(1)\bigr)2E(c+\sigma Z_1)^p\int_0^\infty \bigl(1+\lfloor x\rfloor\bigr)^{-p/2}H_v(x;n,V/2),$$

with c being the minimizer of $M_{p,c}(Z_1)$. Thus, $M_{n,p}\le M_p+o(1)$ by the proof of (2.8). If G is continuous and $f(t)=VG(t)$, then, by Brunk (1970),

$$\lim_{n\to\infty} n^{1/3}R_{n,p}(f) = \bigl\{E|2Z|^p\bigr\}^{1/p}\sigma^{2/3}2^{-1/3}V^{1/3}.$$

This gives the lower bound for $M_{n,p}$ and so (2.9) holds. The proof of (2.11) is simpler and omitted.

Finally, we prove (2.10) by dividing the real line into several intervals and using local versions of (2.9). Let $\mathbf{s} \equiv \{-\infty \equiv s_0 < s_1 < \cdots < s_{k-1} < s_k = \infty\}$. The local version of (2.9) implies

$$\begin{aligned}\lim_{n\to\infty} n^{p/3} R^p_{n,p}(f) &\le \inf_{k,\mathbf{s}} \sum_{j=1}^k M_p^p \sigma^{2p/3} \Delta_j^{p/3}(f) \Delta_j^{1-p/3}(G) \\ &= M_p^p \sigma^{2p/3} \int (df/dG)^{p/3}\, dG,\end{aligned}$$

where $\Delta_j(h) \equiv \Delta_{j,\mathbf{s}}(h) \equiv h(s_j) - h(s_{j-1})$ for all h. Note that the infimum involves only the absolutely continuous part of f with respect to G, since the sum of $\Delta_j^{p/3}(f)\Delta_j^{1-p/3}(G)$ over $\{j : \Delta_j(f) > M\Delta_j(G)\}$ is bounded by $M^{p/3-1}V(f) = o(1)$ for large M. $\square$

4. Nonmonotone regression functions and general loss.

Let

$$y_i \equiv \mu(t_i) + \varepsilon_i, \qquad 1 \le i \le n, \tag{4.1}$$

where μ is an arbitrary function and the errors ε_i are possibly dependent. Although (1.2) is derived for the purpose of estimating nondecreasing regression functions, the true $\mu(\cdot)$ may not be monotone. Most results in the literature concern the case of monotone $\mu(\cdot)$. Birgé (1989) showed that the Grenander estimator performs reasonably well when the true density is nearly monotone.

Define the population version of (1.2) by

$$f_{(n)} \equiv \arg\min\left\{\sum_{i=1}^n \big(\mu(t_i) - f(t_i)\big)^2 : f \text{ is nondecreasing}\right\}. \tag{4.2}$$

If $\mu(t_1) \le \cdots \le \mu(t_n)$, then $f_{(n)} = \mu$ at design points. If $\tilde{f}_n$ is an isotonic estimator, then (4.2) implies, without condition on $\mu(\cdot)$,

$$\|f_{(n)} - \mu\|_n \le \|\tilde{f}_n - \mu\|_n \le \|f_{(n)} - \mu\|_n + \|\tilde{f}_n - f_{(n)}\|_n, \tag{4.3}$$

where $\|h\|_n \equiv \{\sum_{i=1}^n h^2(t_i)/n\}^{1/2}$. We argue that isotonic estimators should be used when we have reason to believe that $\mu(\cdot)$ is isotonic or nearly so, and by (4.3) we should look for estimators close to $f_{(n)}$. We may also view $\|\tilde{f}_n - f_{(n)}\|_n$ as the estimation error and $\|f_{(n)} - \mu\|_n$ the model approximation error. Thus, we consider in this section risks of the form $\sum_{i=n_1+1}^{n_2} EL(\hat{f}_n(t_i) - f_{(n)}(t_i))$ for a general loss function $L(x)$ and the $f_{(n)}$ in (4.2) without assuming the monotonicity of $\mu(\cdot)$.

Let $L_+(x) \equiv L(x) I_{\{x \ge 0\}}$ and $L_-(x) \equiv L(x) I_{\{x<0\}}$ and define

$$r_L(m, v) \equiv \max_{1<j\le n-m} EL_+\left(v + \max_{1\le k\le j} \frac{\sum_{i=k}^{j+m} \varepsilon_i}{j+m-k+1}\right) + \max_{1+m<j\le n} EL_-\left(-v + \min_{\ell \ge j} \frac{\sum_{i=j-m}^{\ell} \varepsilon_i}{\ell - j + m + 1}\right). \tag{4.4}$$

THEOREM 4.1. *Let $\hat f_n$ be the LSE in* (1.2) *based on* $\{(y_i, t_i), i \le n\}$ *from the regression model* (4.1) *with arbitrary* $\mu(\cdot)$ *and* $\{\varepsilon_i\}$. *Let* $L(x) \ge 0$ *be a loss function that is nonincreasing in* $(-\infty, 0)$ *and nondecreasing in* $[0, \infty)$. *Then, for* $0 \le n_1 \le n_2 \le n$,

$$\frac{1}{n_*} \sum_{j=n_1+1}^{n_2} EL\big(\hat f_n(t_j) - f_{(n)}(t_j)\big) \le \int_{0<x<\infty} r_L\big(\lfloor x\rfloor, v(\lfloor x\rfloor)\big)\, dH_v(x; n_*, V_*) \tag{4.5}$$

for all nonincreasing, nonnegative continuous $v(x)$, *where* $V_*(f_{(n)}) \equiv f_{(n)}(t_{n_2}) - f_{(n)}(t_{n_1+1})$, $f_{(n)}$ *is as in* (4.2) *and* n_* *and* $H_v(x; n, V)$ *are as in Theorem* 2.1.

REMARK 4.1. By (4.2), $V_*(f_{(n)}) \le V(f_{(n)}) \le \max_{1\le i\le j\le n}\{\mu(t_j) - \mu(t_i)\}$.

REMARK 4.2. For $L(x) = |x|^p$, the differences between (4.5) and (2.2) are the replacement of r_p by the slightly larger (4.4) and the loss of factor $1/2$ in $V_*/2$. Thus, Theorems 2.2 and 2.3 can be easily extended, for the ℓ_p risk of $\hat f_n - f_{(n)}$, to the case of general nonmonotone μ in (4.1).

PROOF OF THEOREM 4.1. Let $f_{(n),j} \equiv f_{(n)}(t_j)$. Define

$$m_j \equiv \max\big\{m \ge 0 \colon f_{(n),j+m} \le f_{(n),j} + v(m),\ \ j+m \le n_2\big\}.$$

Let $\ell(j)$ be the largest $\ell \le n$ satisfying $f_{(n),\ell} = f_{(n),j}$ and let $\ell_j^* \equiv \ell(j + m_j)$. By (3.1),

$$\hat f_n(t_j) = \min_{\ell \ge j} \max_{k\le j} \frac{\sum_{i=k}^{\ell} y_i}{\ell - k + 1} \le \max_{k \le j} \frac{\sum_{i=k}^{\ell_j^*} \varepsilon_i}{\ell_j^* - k + 1} + \max_{k\le j} \frac{\sum_{i=k}^{\ell_j^*} \mu(t_i)}{\ell_j^* - k + 1}$$

and $\max_{k\le \ell_j^*} \sum_{i=k}^{\ell_j^*} \mu(t_i)/(\ell_j^* - k + 1) = f_{(n),\ell_j^*} = f_{(n),j+m_j} \le f_{(n),j} + v(m_j)$. Thus,

$$L_+\big(\hat f_n(t_j) - f_{(n),j}\big) \le L_+\left(v(m_j) + \max_{k \le \ell_j^* - m} \frac{\sum_{i=k}^{\ell_j^*} \varepsilon_i}{\ell_j^* - k + 1}\right) \tag{4.6}$$

holds almost surely, and certainly in expectation, by simple algebra and the monotonicity of L_+. This leads to a slightly different version of (3.3). Since (3.2) is no longer valid for the current m_j, the upper bound for $\ell(m+1)-(m+1)$ in (3.4) is replaced by

$$\sum_{j=n_1+1}^{n_2-(m+1)} \frac{f_{(n),j+m+1}-f_{(n),j}}{v(m+1)} = \sum_{j=n_1+1}^{n_2-(m+1)} \frac{\sum_{k=j}^{j+m}(f_{(n),k+1}-f_{(n),k})}{v(m+1)} \le (m+1)\frac{V_*(f_{(n)})}{v(m+1)}.$$

The rest of the proof is identical to the parallel parts of the proof of Theorem 2.1 and is omitted. □

5. Dependent errors. In this section, we apply Theorem 4.1 to the ℓ_p risk

$$R^*_{n,p} \equiv \left(\frac{1}{n}\sum_{i=1}^{n} E\big|\hat{f}_n(t_i) - f_{(n)}(t_i)\big|^p\right)^{1/p}, \tag{5.1}$$

with dependent errors in (4.1) satisfying the moment condition

$$\left(E\left|\sum_{i=k}^{\ell}\frac{\varepsilon_i}{\ell-k+1}\right|^{p'}\right)^{1/p'} \le \frac{\sigma}{(\ell-k+1)^\alpha} \quad \forall k \le \ell, \tag{5.2}$$

for some $0<\alpha<1$, $p' \ge 1/(1-\alpha)$ and $\sigma<\infty$. Consider the case of $E\varepsilon_i = 0$. For $p'=2$, (5.2) holds for $\alpha = 1/2$ if the errors are uncorrelated with $E\varepsilon_i^2 \le \sigma^2$, or for $\alpha<1/2$ if the errors are stationary with $E\varepsilon_1\varepsilon_k = O(k^{-2\alpha})$. For independent ε_i and $p' \ge 1$, (5.2) for any α implies $\max_{i\le n} E|\varepsilon_i|^{p'} = O(1)$, which then implies (5.2) for $\alpha = 1-1/(2\wedge p')$; cf. Lemma A.2.

THEOREM 5.1. *Let $0<\alpha<1\le p<\infty$ and $c \equiv c_\alpha \equiv \alpha e^{1/\alpha - 1}$. Let $\hat{f}_n$, $f_{(n)}$ and $R^*_{n,p}$ be as in* (1.2), (4.2) *and* (5.1). *Suppose that* (5.2) *holds for $p' \ge \max\{p, 1/(1-\alpha)\}$. Then*

$$R^*_{n,p} \le M^*\sigma\left\{\int_0^\infty \frac{\log^{\beta p}(x+1+c)}{\max(x^\alpha, x^{p\alpha})}\, dH_{\alpha,\beta,c}(x; n, V(f_{(n)})/\sigma)\right\}^{1/p}, \tag{5.3}$$

where $H_{\alpha,\beta,c}(x;n,V) \equiv \min[1, (x/n)\{1 + Vx^\alpha/\log^\beta(x+1+c)\}]$, $\beta \equiv I_{\{p'=1/(1-\alpha)\}}$ and $M^*<\infty$ depending on (p,p',α) only. Consequently,

$$\limsup_{n\to\infty} \frac{n^{\alpha/(1+\alpha)}}{(\log n)^{\beta/(1+\alpha)}} \sup_{V(f_{(n)})\le V} R^*_{n,p} \le \frac{M^*\sigma^{1/(1+\alpha)}V^{\alpha/(1+\alpha)}}{1-p\alpha/(1+\alpha)} \tag{5.4}$$

for $p<1+1/\alpha$.

REMARK 5.1. Under (5.2), Theorem 5.1 describes the connection between the convergence rate $n^{-\alpha}$ for the estimation of a common mean by the sample mean of $\{y_i\}$ and the convergence rate $n^{-\alpha/(1+\alpha)}$ for (1.2); for example, $\alpha/(1+\alpha) = 1/3$ for $\alpha = 1/2$.

REMARK 5.2. The LSE $\hat{f}_n$ is a local average of $\{y_i\}$ over a data-driven partition of $\{1, \dots, n\}$. For $\mu(\cdot)\uparrow$ and $\beta = 0$, the order of $|\hat{f}_n(t_i) - \mu(t_i)|^p$ is $(\sigma/m^\alpha)^p$ under (5.2) if $\hat{f}_n(t_i)$ is roughly the average of a block of m of the y_i's and $\mu(t_i)$ does not change much in the block. As in Theorem 2.1, (5.3) is obtained by finding upper bounds on the number of such blocks of size m, and thus the pth power of its right-hand side is a weighted average of $(\sigma/m^\alpha)^p$.

PROOF OF THEOREM 5.1. Let $v(x) \equiv \sigma\{\log(x+1+c_\alpha)\}^\beta/(x+1)^\alpha$. It follows from Lemma A.2(i) and condition (5.2) that, for the loss $L(x) = |x|^p$,

$$2^{1-p} E L_+\left(v + \max_{1\le k\le j} \frac{\sum_{i=k}^{j+m} \varepsilon_i}{j+m-k+1}\right) \le v^p + \left[K_{p',\alpha}^{p'} \frac{\{\log(m+2)\}^{\beta p'}}{(m+2)^{\alpha p'}} \sigma^{p'}\right]^{p/p'},$$

so that $r_L(m, v(m)) \le (M^*_{p',\alpha})^p v^p(m)$ by (4.4). Thus, by Theorem 4.1,

$$(R^*_{n,p})^p \le (M^*_{p',\alpha})^p \sigma^p \int_0^\infty v_1^p(\lfloor x \rfloor)\, dH_{v_1}(x; n, V(f_{(n)})/\sigma),$$

with $v_1(x) \equiv v(x)/\sigma$, and (5.3) follows from (A.1) of Lemma A.1 with $h(x) \equiv \{\log(x+c_\alpha)\}^\beta$. Note that $x^\alpha/\log(x+c)$ is increasing in $[0,\infty)$ iff $c \ge c_\alpha$. The asymptotic bound (5.4) follows from (5.3) by straightforward calculus. □

6. General isotonic regression methods. Let $-\infty \le a_* < a^* \le \infty$ and let $\phi_i(\theta; y_i)$ be observable continuous functions of θ from $[a_*, a^*]$ into $[-\infty, \infty]$, given the response variables y_i. Here, the topology in the extended real line allows $x_n \to \pm\infty$ in the usual sense. In this section, we consider restricted MLE-type general isotonic estimators

$$\hat{f}_n \equiv \arg\max\left\{\sum_{i=1}^n \phi_i\big(f(t_i); y_i\big) : f \text{ is nondecreasing and } a_* \le f \le a^*\right\}. \tag{6.1}$$

Estimators of the form (6.1) have been considered by van Eeden (1957a, b), Robertson and Waltman (1968), Brunk and Johansen (1970) and Barlow and Ubhaya (1971), among others. van de Geer (1990, 1993) obtained the $n^{-1/3}$-consistency in probability of the ℓ_2 loss functions for the median regression with $\phi_i(\theta; y_i) = |y_i| - |y_i - \theta|$.

The estimator (6.1) is quite general. In location models, $\phi_i(\theta; y_i) = w_i \tilde{\phi}(\{y_i - g(\theta)\}/\sigma_i)$ is often used, where w_i and σ_i are constants, $g(\theta)$ is a link function and $\tilde{\phi}(\cdot)$ is a reward function or a log-likelihood function. For example, $\phi_i(\theta; y_i) = \log\{\phi_0((y_i - \theta)/\sigma_i)/\sigma_i\}$ for some known density function ϕ_0. For weighted

ℓ_p regression, (6.1) is used with $\phi_i(\theta; y_i) = w_i\{|y_i|^p - |y_i - \theta|^p\}/p$ for certain $p \ge 1$. The LSE (1.2) is a special case of (6.1) with $\phi_i(\theta; y_i) = \{y_i^2 - (y_i - \theta)^2\}/2$. For estimators based on quasi-likelihood, $\phi_i(\theta; y_i) = \{g(\theta)y_i - \psi(g(\theta))\}/\sigma_i^2$ for a convex function ψ. For example, $\phi_i(\theta; y_i) = \theta y_i - \log(1 + e^\theta)$ for Bernoulli y_i and $\phi_i(\theta; y_i) = y_i \log(\theta) - \theta$ for Poisson y_i. In Section 8, we shall consider median regression with $\phi_i(\theta; y_i) = |y_i| - |y_i - \theta|$.

Suppose there exist $a_* \le \hat\theta^{\pm}_{k,\ell} \le a^*$ for $1 \le k \le \ell \le n$ such that

$$\sum_{i=k}^{\ell} \phi_i(\theta'; y_i) - \sum_{i=k}^{\ell} \phi_i(\theta''; y_i) \begin{cases} < 0, & \text{if } a_* \le \theta' < \theta'' \le \hat\theta^-_{k,\ell}, \\ = 0, & \text{if } \hat\theta^-_{k,\ell} \le \theta' < \theta'' \le \hat\theta^+_{k,\ell}, \\ > 0, & \text{if } \hat\theta^+_{k,\ell} \le \theta' < \theta'' \le a^*. \end{cases} \tag{6.2}$$

This unimodality condition implies that for the estimation of a common parameter θ based on $y_k, \dots, y_\ell$, the "MLE," that is, the set of maximizers of the "log-likelihood" $\sum_{i=k}^{\ell} \phi_i(\theta; y_i)$, is a closed interval $[\hat\theta^-_{k,\ell}, \hat\theta^+_{k,\ell}]$. The estimator (6.1) can be easily computed using the *pool-adjacent-violators algorithm* under (6.2), as both families of modes $\{\hat\theta^+_{k,\ell}\}$ and $\{\hat\theta^-_{k,\ell}\}$ satisfy the *Cauchy-mean* condition; cf. (6.6) and Robertson and Waltman (1968).

Let $L(\cdot)$ be a loss function, with $L(0) = 0$, such that $L\uparrow$ in $[0, \infty]$ and $L\downarrow$ in $[-\infty, 0]$ [$\lim_{x\to\pm\infty} L(x) = \infty$ allowed]. Let f_0 be a nondecreasing function. We consider upper bounds for sums of $EL(\hat f(t_i) - f_0(t_i))$. Define, via integrating by parts if necessary,

$$r_+(m, v; f_0) \equiv \max_{n_1 < j \le n_2 - m} \int_0^\infty P\Big\{\min_{\ell \ge j} \max_{k \le j} \hat\theta^+_{k,\ell} - f_0(t_{j+m}) > x - v\Big\}\, dL(x),$$

$$r_-(m, v; f_0) \equiv \max_{n_1 + m < j \le n_2} \bigg|\int_{-\infty}^0 P\Big\{\min_{\ell \ge j} \max_{k \le j} \hat\theta^-_{k,\ell} - f_0(t_{j-m}) < x + v\Big\}\, dL(x)\bigg|$$

for $v \ge 0$, $m = 0, 1, \dots$ and integers $0 \le n_1 \le n_2 \le n$, and define

$$r(t, v; f_0) \equiv r_+(m, v; f_0) + r_-(m, v; f_0). \tag{6.3}$$

THEOREM 6.1. *Let $f_0\uparrow$ and let $r(t, v; f_0)$ be as in* (6.3). *Then, for $0 \le n_1 \le n_2 \le n$,*

$$\frac{1}{n_*} \sum_{j=n_1+1}^{n_2} EL\big(\hat f_n(t_j) - f_0(t_j)\big) \le \int_0^\infty r\big(\lfloor x\rfloor, v(\lfloor x\rfloor); f_0\big)\, dH_v(x; n_*, V_*) \tag{6.4}$$

for all nonincreasing, nonnegative continuous $v(x)$, where $V_* \equiv f_0(t_{n_2}) - f_0(t_{n_1+1})$ and $n_* \equiv n_2 - n_1$, $H_v(x; n, V) \equiv \min[1, x\{1 + V/v(x)\}/n]$ and $\lfloor x\rfloor$ are as in Theorem 2.1.

A remarkable aspect of Theorem 6.1 is that (6.4) holds for *all* nondecreasing continuous functions v. This is probably related to the insensitivity of the norm that is used to find the isotonic estimator; cf. Section 1.5 of Robertson, Wright and Dykstra (1988). If $r(m, v(m); f_0) \le \{Mv(m)\}^p$ for $v(m) \equiv h(m+1)/(m+1)^\alpha$ and a suitable h, then (6.4) and Lemma A.1 can be used to derive risk bounds as in Theorem 5.1. Explicit risk bounds for more specific loss functions will be derived from (6.4) in Sections 7 and 8.

The proof of Theorem 6.1 is based on the minimax bounds in the following proposition. Minimax formulas of slightly different form were obtained by van Eeden (1957a, b) and Robertson and Waltman (1968), among others.

PROPOSITION 6.1. *Let* $\hat\theta^{\pm}_{k,\ell}$ *be as in* (6.2) *and let* $\hat f_n$ *be a solution of* (6.1). *Then*

$$\min_{\ell\ge j}\max_{k\le j}\hat\theta^-_{k,\ell} \le \hat f_n(t_j) \le \min_{\ell\ge j}\max_{k\le j}\hat\theta^+_{k,\ell}, \qquad 1\le j\le n. \tag{6.5}$$

In particular, $\hat f_n(t_j) = \min_{\ell\ge j}\max_{k\le j}\sum_{i=k}^{\ell} y_i/(\ell-k+1)$ *for the LSE* (1.2).

PROOF OF THEOREM 6.1. Set $m_j \equiv \max\{m\ge 0 : f_{j+m} \le f_j + v(m),\ j+m\le n_2\}$ as in the proof of Theorem 4.1, where $f_j \equiv f_0(t_j)$. Then, by (6.5) and (6.3),

$$\int_0^\infty P\{\hat f_n(t_j) - f_j > x\}\,dL(x) \le \int_0^\infty P\{\hat f_n(t_j) - f_{j+m_j} > x - v(m_j)\}\,dL(x)$$

is bounded by $r_+(m_j, v(m_j); f_0)$. The rest of the proof is the same as that of Theorem 4.1 and is omitted. □

PROOF OF PROPOSITION 6.1. First, let us verify the Cauchy-mean property for $\hat\theta^+_{k,\ell}$:

$$\min(\hat\theta^+_{k,j}, \hat\theta^+_{j+1,\ell}) \le \hat\theta^+_{k,\ell} \le \max(\hat\theta^+_{k,j}, \hat\theta^+_{j+1,\ell}), \qquad k\le j<\ell. \tag{6.6}$$

Since both $S_1(\theta) \equiv \sum_{i=k}^{j}\phi_i(\theta; y_i)$ and $S_2(\theta) \equiv \sum_{i=j+1}^{\ell}\phi_i(\theta; y_i)$ are nondecreasing in $\theta \le \min(\hat\theta^+_{k,j}, \hat\theta^+_{j+1,\ell})$, the sum $S(\theta) \equiv \sum_{i=k}^{\ell}\phi_i(\theta; y_i)$ is nondecreasing in θ in the same interval, which implies the first inequality of (6.6) by (6.2). Likewise, the second inequality of (6.6) holds, since both $S_1(\theta)$ and $S_2(\theta)$ are strictly decreasing in $\theta > \max(\hat\theta^+_{k,j}, \hat\theta^+_{j+1,\ell})$.

By symmetry, we shall only prove the second inequality of (6.5) for a fixed $j = j_0$. Since the minimax formula is nondecreasing in j, we assume $\hat f_n(t_{j_0-1}) < \hat f_n(t_{j_0})$, with the convention $\hat f_n(t_0) \equiv -\infty$. It suffices to show $\hat f_n(t_{j_0}) \le \hat\theta^+_{j_0,\ell_0}$ for every fixed $\ell_0 \ge j_0$.

Let $j_0 < j_1 < \cdots < j_m$ be the jump points of $\hat f_n$ in $[j_0, \ell_0]$ and let $j_{m+1} = n+1$. Let k be fixed, $1 \le k \le m+1$, and set $\tilde f(t_i) \equiv \hat f_n(t_i) - a$ for $i \in [j_{k-1}, \ell_0 \wedge (j_k - 1)]$

and $\tilde f(t_i) \equiv \hat f_n(t_i)$ otherwise. Since $\hat f_n(t_{j_{k-1}-1}) < \hat f_n(t_{j_{k-1}})$, $\tilde f(t_i)$ is nondecreasing in i for sufficiently small $a > 0$, so that, by the optimality of $\hat f_n$,

$$\sum_{i=j_{k-1}}^{\ell_0\wedge(j_k-1)} \phi_i\big(\hat f_n(t_{j_{k-1}}); y_i\big) - \sum_{i=j_{k-1}}^{\ell_0\wedge(j_k-1)} \phi_i\big(\hat f_n(t_{j_{k-1}}) - a; y_i\big)$$

$$= \sum_{i=1}^{n} \phi_i\big(\hat f_n(t_i); y_i\big) - \sum_{i=1}^{n} \phi_i\big(\tilde f(t_i); y_i\big) \ge 0.$$

This and the unimodality (6.2) imply $\hat f_n(t_{j_{k-1}}) \le \hat\theta^+_{j_{k-1},\ell_0\wedge(j_k-1)}$. Since $\hat f_n$ is nondecreasing and $1 \le k \le m+1$ is arbitrary, by the Cauchy-mean property (6.6),

$$\hat f_n(t_{j_0}) \le \min_{1\le k\le m+1} \hat f_n(t_{j_{k-1}}) \le \min_{1\le k\le m+1} \hat\theta^+_{j_{k-1},\ell_0\wedge(j_k-1)} \le \hat\theta^+_{j_0,\ell_0}.$$

This completes the proof. □

7. Truncated $\boldsymbol{\ell_p}$ and zero–one losses. We shall apply Theorem 6.1 to loss functions $L(x) = (|x| \wedge \delta_0)^p$ and $L(x) = I_{\{|x|>\delta_0\}}$. Let $\psi_i(\theta) \equiv E\phi_i(\theta; y_i)$. As in Section 4, we consider (6.4) with $f_0 = f_{(n)}$, the population version of (6.1), given by

$$f_{(n)} \equiv \arg\max\left\{\sum_{i=1}^{n} \psi_i\big(f(t_i)\big) : f \text{ is nondecreasing and } a_* \le f \le a^*\right\}. \tag{7.1}$$

Assume throughout this section that

$$\phi_i(\theta; y_i) = \tilde\phi_i\big(g(\theta); y_i\big) \tag{7.2}$$

for certain random concave functions $\tilde\phi_i(\cdot; y_i)$ and an increasing continuous g. Define

$$\rho_i^{\pm}(\theta) \equiv \lim_{\varepsilon\to 0\pm} \frac{\tilde\phi_i\big(g(\theta)+\varepsilon; y_i\big) - \tilde\phi_i\big(g(\theta); y_i\big)}{\varepsilon}, \tag{7.3}$$

that is, the right- and left-continuous versions of $\tilde\phi_i(g(\theta) + dx; y_i)/dx$. For $[a_*, a^*] \ne [-\infty, \infty]$, the domain of g is assumed to be $[-\infty, \infty]$, through natural extension of g if necessary, so that (7.3) is meaningful for all $-\infty < \theta < \infty$. For (4.1) with $\phi(\theta; y_i) = \{y_i^2 - (y_i - \theta)^2\}/2$, $\rho_i^{\pm}(\theta) = y_i - \theta = (\partial/\partial\theta)\phi_i(\theta; y_i)$. Define

$$\begin{aligned} k(j) &\equiv \min\{k : f_{(n)}(t_k) = f_{(n)}(t_j)\},\\ \ell(j) &\equiv \max\{\ell : f_{(n)}(t_\ell) = f_{(n)}(t_j)\}. \end{aligned} \tag{7.4}$$

By the concavity of $\tilde\phi_i(\cdot; y_i)$ and the monotonicity of g, $\rho_i^{\pm}(\theta; y_i)$ are nonincreasing in θ. Since $\hat\theta^{\pm}_{k,\ell}$ are modes of $\sum_{i=k}^{\ell} \phi_i(\theta; y_i)$, $(\theta - \hat\theta^{\pm}_{k,\ell}) \sum_{i=k}^{\ell} \rho_i^{\pm}(\theta; y_i) \le 0$

for all $a_* < \theta < a^*$. Thus, for $k \le \ell \le \ell_{(j+m)}$, $\hat{\theta}^+_{k,\ell} > f_{(n)}(t_{j+m}) + x$ implies

$$0 \le \sum_{i=k}^{\ell} \rho_i^+\big(f_{(n)}(t_{j+m}) + x; y_i\big) \le \sum_{i=k}^{\ell} \rho_i^+\big(f_{(n)}(t_i) + x; y_i\big)$$

by the monotonicity of $\rho_i^+(\cdot; y_i)$ and (7.4). Consequently,

$$\begin{aligned} &\Big\{\min_{\ell \ge j} \max_{k \le j} \hat{\theta}^+_{k,\ell} - f_{(n)}(t_{j+m}) > x\Big\} \\ &\qquad \subseteq \Bigg\{\min_{j \le \ell \le \ell(j+m)} \max_{k \le j} \sum_{i=k}^{\ell} \rho_i^+\big(f_{(n)}(t_i) + x; y_i\big) \ge 0\Bigg\}. \end{aligned} \tag{7.5}$$

Let $\varepsilon_i^+(x)$ be nonincreasing [$\varepsilon_i^-(x)$ nondecreasing] random functions of x such that

$$\begin{aligned} \sum_{i=k}^{\ell_{(j)}} \rho_i^+\big(f_{(n)}(t_i) + x\big) &\le \sum_{i=k}^{\ell_{(j)}} \varepsilon_i^+(x), \\ \sum_{i=k_{(j)}}^{\ell} \rho_i^-\big(f_{(n)}(t_i) - x\big) &\ge \sum_{i=k_{(j)}}^{\ell} \varepsilon_i^-(x) \end{aligned} \tag{7.6}$$

for all $1 \le k \le j \le \ell \le n$, for example, $\varepsilon_i^\pm(x) = \rho_i^\pm(f_{(n)}(t_i) \pm x)$, where $k(j)$ and $\ell(j)$ are as in (7.4). We shall derive risk bounds based on moment conditions on $\varepsilon_i^\pm(x)$ and the relationship

$$\begin{aligned} &\Big\{\min_{\ell \ge j} \max_{k \le j} \hat{\theta}^+_{k,\ell} - f_{(n)}(t_{j+m}) > x\Big\} \\ &\qquad \subseteq \Bigg\{\max_{k \le j} \sum_{i=k}^{\ell(j+m)} \varepsilon_i^+(x) \ge 0,\ f_{(n)}(t_{j+m}) + x < a^*\Bigg\} \end{aligned} \tag{7.7}$$

from (7.5) and its counterpart for $\varepsilon_i^-(x)$, in view of (6.3). Note that $a_* \le \hat{\theta}^\pm_{k,\ell} \le a^*$.

Let $0 < \alpha < 1 \le p < \infty$, $p' \ge 1/(1-\alpha)$, $x > 0$ and $d_0 > 0$. Consider conditions

$$\sum_{i=k}^{\ell_{(j)}} \frac{E\varepsilon_i^+(x)}{\ell_{(j)} - k + 1} \le -d_0 x, \qquad \sum_{i=k_{(j)}}^{\ell} \frac{E\varepsilon_i^-(x)}{\ell - k_{(j)} + 1} \ge d_0 x \qquad \forall k \le j \le \ell \tag{7.8}$$

for $0 < x < a^* - f_{(n)}(t_j)$ in the first inequality and $0 < x < f_{(n)}(t_j) - a_*$ in the second,

$$\Bigg\{E\Bigg|\sum_{i=k}^{\ell} \frac{\varepsilon_i^\pm(x) - E\varepsilon_i^\pm(x)}{\ell - k + 1}\Bigg|^{p'}\Bigg\}^{1/p'} \le \frac{\sigma d_0}{(\ell - k + 1)^\alpha} \qquad \forall k \le \ell, \tag{7.9}$$

cf. the discussion after (5.2), and for all $j \geq 1$ and $m \geq 1$,

$$\int_0^{\delta_0} P\{\pm\{\varepsilon_j^{\pm}(x) - E\varepsilon_j^{\pm}(x)\} \geq m\, d_0 x/2\}\, dx^p \leq (\sigma/m)^p, \tag{7.10}$$

where $\varepsilon_i^{\pm}(x)$ are as in (7.6) and $k_{(j)}$ and $\ell_{(j)}$ are as in (7.4).

Let $\psi_i(\theta) \equiv E\phi_i(\theta; y_i)$ as in (7.1) and $\dot\psi_i^{\pm}(\theta)$ be their left and right derivatives. If the limit in (7.3) is exchangeable with the expectation and $\varepsilon_i^{\pm}(x) = \rho_i^{\pm}(f(t_i) \pm x)$ is chosen for (7.6), then $E\varepsilon_i^{\pm}(x) = \dot\psi_i^{\pm}(f_{(n)}(t_i) \pm x)$. In this case,

$$\sum_{i=k}^{\ell_{(j)}} E\varepsilon_i^{+}(x) \leq \sum_{i=k}^{\ell_{(j)}} E\varepsilon_i^{+}(0) \leq 0, \qquad 0 \leq \sum_{i=k_{(j)}}^{\ell} E\varepsilon_i^{-}(0) \leq \sum_{i=k_{(j)}}^{\ell} E\varepsilon_i^{-}(x) \tag{7.11}$$

for all the (x, j, k, ℓ) considered in (7.8), by the monotonicity of $\varepsilon_i^{\pm}(x)$ and the optimality of $f_{(n)}$. Note that, by (7.1), $\sum_{i=k}^{\ell_{(j)}} \psi_i(f_{(n)}(t_i) + x) \leq \sum_{i=k}^{\ell_{(j)}} \psi_i(f_{(n)}(t_i))$ for $x > 0$ and $k(j) \leq k \leq \ell(j)$. Consequently, (7.8) holds if $\dot\psi_i^{\pm}(\theta + x) - \dot\psi_i^{\pm}(\theta) \geq d_0 x$ for all $a_* < \theta < \theta + x < a^*$.

In the location model (4.1) with $\phi(\theta; y_i) = \{y_i^2 - (y_i - \theta)^2\}/2$,

$$\begin{aligned}\sum_{i=k}^{\ell} \rho_i^{\pm}(f_{(n)}(t_i) \pm x) &= \sum_{i=k}^{\ell}\{y_i - f_{(n)}(t_i) \mp x\} \\ &= \sum_{i=k}^{\ell}(\varepsilon_i \mp x) + \sum_{i=k}^{\ell}\{\mu(t_i) - f_{(n)}(t_i)\}.\end{aligned}$$

Since $\sum_{i=1}^{\ell} f_{(n)}(t_i)$ is the convex minorant of $\sum_{i=k}^{\ell} \mu(t_i)$, by (7.4),

$$\sum_{i=k}^{\ell_{(j)}}\{\mu(t_i) - f_{(n)}(t_i)\} \leq 0 \leq \sum_{i=k_{(j)}}^{\ell}\{\mu(t_i) - f_{(n)}(t_i)\}, \qquad k \leq j \leq \ell.$$

Thus, (7.6) holds for $\varepsilon_i^{\pm}(x) = \varepsilon_i \mp x$. Furthermore, for either choices $\varepsilon_i^{\pm}(x) = \varepsilon_i \mp x$ and $\varepsilon_i^{\pm}(x) = \rho_i^{\pm}(f_{(n)}(t_i) \pm x)$, (7.8) holds with $d_0 = 1$ and (7.9) and (7.10) follow from (5.2). In fact, $r_L(m, v) \leq r(m, v; f_{(n)})$ by (4.4), (7.7) and (6.3). It is clear that (7.8) may not hold if $k_{(j)}$ and $\ell_{(j)}$ are replaced by general $1 \leq k \leq \ell \leq n$.

With $\hat f_n$ and $f_{(n)}$ in (6.1) and (7.1), respectively, let

$$R_{n,p}^* \equiv \left(\frac{1}{n}\sum_{i=1}^{n} E\min\{|\hat f_n(t_i) - f_{(n)}(t_i)|^p, \delta^p\}\right)^{1/p}. \tag{7.12}$$

THEOREM 7.1. *Let* $0 < \alpha < 1 \le p < \infty$, $0 < \delta_0 \le \infty$ *and* $p' \ge \max\{p, 1/(1-\alpha)\}$. *Suppose* (7.8) *and* (7.9) *hold for all* $0 < x < \delta_0$. *Suppose that either* (a) $\delta_0 < \infty$ *for* $p' = p$ *or* (b) *condition* (7.10) *holds and* $\alpha \le \min(1/2, 1 - 1/p)$ *and both sequences* $\{\varepsilon_i^+(x),\ i \le n\}$ *and* $\{\varepsilon_i^-(x),\ i \le n\}$ *are independent for each* x. *Let* $\beta \equiv I_{\{p'=1/(1-\alpha)\}} + p^{-1} I_{\{p'=p\}}$ *under* (a) *and* $\beta \equiv 0$ *under* (b). *Then* (5.3) *holds with the* $R^*_{n,p}$ *in* (7.12), $M^* \equiv M^*_{p,p',\alpha}$ *and* $c \equiv (\alpha/\beta)e^{\beta/\alpha - 1}$. *Consequently,* (5.4) *holds.*

THEOREM 7.2. *Suppose* (7.8) *and* (7.9) *hold for* $x = \delta_0/2$ *and* $p' \ge 1/(1-\alpha)$. *Then*

$$\sum_{i=1}^n P\{|\hat f_n(t_i) - f_{(n)}(t_i)| > \delta_0\} \le M_{p',\alpha} \frac{\sigma^{p'} \log^{\beta'}(n+1)}{\delta_0^{p'} n^{-(1-p'\alpha)_+}} \biggl(1 + \frac{V(f_{(n)})}{\delta_0}\biggr), \tag{7.13}$$

where $\beta' \equiv p' I_{\{p'=1/(1-\alpha) \le 1/\alpha\}} + I_{\{p'=1/\alpha\}}$ *without additional conditions,* $\beta' \equiv I_{\{p'=1/\alpha\}}$ *if both* $\{\varepsilon_i^+(\delta_0/2),\ j \le n\}$ *and* $\{\varepsilon_i^-(\delta_0/2),\ j \le n\}$ *are independent sequences and* $\beta' \equiv 0$ *if* $\alpha = 1 - 1/p' > 0$ *and* $1 < p' \le 2$ *and* $\{\varepsilon_i^\pm(\delta_0/2),\ j \le n\}$ *are both i.i.d. sequences.*

The proofs of Theorems 7.1 and 7.2 are based on the moment and tail probability inequalities provided in Lemma A.2.

PROOF OF THEOREM 7.1. First, consider the case $p' > \max\{p, 1/(1-\alpha)\}$. It follows from the definition of $r_+(m, v; \delta_0)$ in (6.3), (7.7), the monotonicity of $\varepsilon_i^+(\cdot)$ and (7.8) that

$$\begin{aligned} &2^{1-p} r_+(m, v; f_{(n)}) \\ &\quad \le \max_j 2^{1-p} \int_0^{\delta_0} P\Biggl\{\max_{k \le j-m} \sum_{i=k}^{\ell_{(j)}} \varepsilon_i^+(x - v) \ge 0\Biggr\} dx^p \\ &\quad \le v^p + \max_j \int_0^{\delta_0} P\Biggl\{\max_{k \le j-m} \sum_{i=k}^{\ell_{(j)}} \frac{\varepsilon_i^+(x) - E\varepsilon_i^+(x)}{\ell_{(j)} - k + 1} \ge d_0 x\Biggr\} dx^p. \end{aligned} \tag{7.14}$$

By the Markov inequality, Lemma A.2(i) and (7.9), the integration on the right-hand side is bounded by

$$\int_0^{\delta_0} \min\biggl\{1, \frac{K_{p',\alpha}^{p'} (\sigma d_0)^{p'}}{(m+1)^{p'\alpha} (d_0 x)^{p'}}\biggr\} dx^p \le \frac{K_{p',\alpha}^p \sigma^p}{(m+1)^{\alpha p}} \int_0^\infty \min\{1, x^{-p'}\} dx^p.$$

These and the same for $r_-(m, v; f_{(n)})$ imply $r(m, v(m); f_{(n)}) \le K^* v^p(m)$ for $v(m) \equiv \sigma/(m+1)^\alpha$. Thus, with $\beta = 0$, (5.3) follows from Lemma A.1 as in the proof of Theorem 5.1.

For $p' = 1/(1-\alpha)$, the probability inside the integration is bounded by the smaller of one and $K_{p',\alpha}^{p'}(\sigma d_0)^{p'} \log^{p'}(m+1+c)/\{(m+1)^{p'\alpha}(d_0 x)^{p'}\}$ by Lemma A.2(i). For $p' = p$, $\int_0^\infty \min(1, x^{-p'})\, dx^p$ should be replaced by

$$\int_0^{\delta_0 (m+1)^\alpha/(K\sigma)} \min\{1, x^{-p'}\}\, dx^p = \alpha \log(m+1+c) + O(1).$$

These two modifications of the calculation yield $r(m, v(m); f_{(n)}) \le K^* v^p(m)$ for $v(m) \equiv \sigma \log^\beta(1+c+m)/(m+1)^\alpha$. Again, Lemma A.1 can be used to prove (5.3).

Now, consider the independence case. Set $\ell \equiv \ell_{(j)}$ and $X_i \equiv \varepsilon_{\ell-i+1}^+(x) - E\varepsilon_{\ell-i+1}^+(x)$. By (A.6) of Lemma A.2 with $(k_0, c, t) = (2, 2/3, (3/4)d_0 x)$ and $b_i = (m+1) \vee i$, the integration on the right-hand side of (7.14) is bounded by

$$\begin{aligned}
&\int_0^{\delta_0} P\left\{\max_{1\le i\le \ell} \frac{S_i}{(m+1)\vee i} > x d_0\right\} dx^p \\
&\quad \le \int_0^{\delta_0} P\left\{\max_{1\le i\le \ell} \frac{X_i}{(m+1)\vee i} > \frac{x d_0}{2}\right\} dx^p \\
&\qquad + \int_0^{\delta_0} \left(\min\left\{1, \frac{4^p}{(x d_0)^p} E \max_{1\le i\le \ell}\left|\frac{S_i}{(m+1)\vee i}\right|^p\right\}\right)^2 dx^p.
\end{aligned}$$

By (7.10), the first integral above is on the order of $\sum_{i=1}^{\ell}\{(m+1)\vee i\}^{-p} \sim (m+1)^{1-p} \le (m+1)^{-\alpha p}$, while the second one is on the order of $(m+1)^{-\alpha p}$ by (A.7) of Lemma A.2 and (7.9). The rest is the same as the proof of Theorem 5.1. □

PROOF OF THEOREM 7.2. Let $v(x) \equiv v_0 \equiv \delta_0/2$ (a constant). By Theorem 6.1 with $L(x) \equiv I_{\{|x|>\delta_0\}}$ and Lemma A.2, the left-hand side of (7.13) is bounded by

$$\begin{aligned}
&\sum_{m=0}^{n} r(m, v_0; f_{(n)})\left(1 + \frac{V(f_{(n)})}{v_0}\right) \\
&\qquad \le K_{p'\alpha}^{p'}\left(\frac{2\sigma}{\delta_0}\right)^{p'}\left(1 + \frac{V(f_{(n)})}{v_0}\right)\sum_{m=0}^{n} \frac{\log^{\beta_1 p'}(m+1)}{(m+1)^{p'\alpha}}
\end{aligned}$$

as in the proof of Theorem 7.1, where $\beta_1 \equiv I_{\{p'=1/(1-\alpha)\}}$ in general and $\beta_1 \equiv 0$ for independent $\varepsilon_i^\pm(v_0)$. Thus, (7.13) holds for $\beta' = \beta_1 p' I_{\{p'\alpha\le 1\}} + I_{\{p'\alpha=1\}}$.

In the i.i.d. case with $1 < p' \le 2$, (7.9) implies $E|X_i|^{p'} \le (\sigma d_0)^{p'}$, where $X_i \equiv \varepsilon_i^+(v_0) - E\varepsilon_i^+(v_0)$. Let $S_k \equiv \sum_{i=1}^k X_i$. By Chow and Lai (1978),

$$\sum_{m=0}^n r(m, v_0; f_{(n)}) \le \sum_{m=0}^n \frac{n^{2-p'}}{m^{2-p'}} P\left\{\sup_{k \ge m+1} \frac{S_k}{k} > d_0 v_0\right\}$$
$$\le \frac{K_{p'}^{p'} n^{2-p'}}{(d_0 v_0)^{p'}} E|X_1|^{p'}. \qquad \square$$

8. Rates of convergence in probability. Although Theorems 7.1 and 7.2 deal with truncated ℓ_p and zero–one losses, they imply convergence in probability of the ℓ_p losses without truncation under a mild additional condition (8.1). We shall also consider here median regression as an example.

THEOREM 8.1. *Suppose* (5.4) *and* (7.13) *hold for certain* $(\alpha, p, p', \beta, \beta', \delta_0)$, *with* $p < 1 + 1/\alpha$, $p' > p$ *and* $0 < \delta_0 < \infty$ *as in Theorems* 7.1 *and* 7.2. *Let* $\gamma \equiv 1/p - \alpha/(1+\alpha) - (1-p'\alpha)_+/p$ *and* $\beta'' \equiv \beta/(1+\alpha) - \beta'/p$. *Define*

$$b_n \equiv \frac{n^{\alpha/(1+\alpha)}}{(\log n)^{\beta/(1+\alpha)}}, \qquad x_n \equiv \frac{n^{1/p}/b_n}{\{n^{(1-p'\alpha)_+}(\log n)^{\beta'}\}^{1/p}} = n^{\gamma}(\log n)^{\beta''}.$$

Let $\hat\theta_{k,\ell}^{\pm}$ *be as in* (6.2) *and let* $k_{n,\varepsilon} \equiv \lfloor n^{(1-p'\alpha)_+}(\log n)^{\beta'}/\varepsilon\rfloor$. *Suppose that, for all* $\varepsilon > 0$,

$$\begin{aligned} &P\left\{\max_{0\le k<k_{n,\varepsilon}} \hat\theta_{n-k,n}^{+} \ge f_{(n)}(t_n) + Mx_n\right\} \\ &\qquad + P\left\{\min_{1\le k\le k_{n,\varepsilon}} \hat\theta_{1,k}^{-} \le f_{(n)}(t_1) - Mx_n\right\} \to 0 \end{aligned} \tag{8.1}$$

as $n \to \infty$ *and then* $M \to \infty$, *where* $f_{(n)}$ *are as in* (7.1). *Then, for* $\hat f_n$ *in* (6.1),

$$\lim_{M\to\infty} \limsup_{n\to\infty} P\left\{b_n\left[\frac{1}{n}\sum_{i=1}^n |\hat f_n(t_i) - f_{(n)}(t_i)|^p\right]^{1/p} \ge M\right\} = 0. \tag{8.2}$$

REMARK 8.1. Since $p < 1 + 1/\alpha$ and $p' > p$, $\gamma > 0$ and $x_n \to \infty$ in (8.1). If $p' > 1/\alpha$ and $\beta' = 0$, then $k_{n,\epsilon} = \lfloor 1/\epsilon\rfloor$ does not depend on n.

EXAMPLE 8.1 (Median regression). For median regression,

$$\hat f_n \equiv \arg\min\left\{\sum_{i=1}^n |y_i - f(t_i)| : f \text{ is nondecreasing and } a_* \le f \le a^*\right\},$$

which is a special case of (6.1) with $\phi_i(\theta; y_i) = |y_i| - |y_i - \theta|$. Let $f_{(n)}$ be as in (7.1) and assume $f_{(n)}(t_n) - f_{(n)}(t_1) \equiv V(f_{(n)}) \le V_0$ for some fixed $V_0 < \infty$. Suppose that y_i are independent variables.

By (7.3), $\rho_i^+(\theta) = 2I\{\theta \le y_i\} - 1$ and $\rho_i^-(\theta) = 2I\{\theta < y_i\} - 1$, and (7.6) holds for

$$\varepsilon_i^+(x) \equiv 2I\{y_i \ge f_{(n)}(t_i) + x\} - 1, \qquad \varepsilon_i^-(x) \equiv 2I\{y_i > f_{(n)}(t_i) - x\} - 1.$$

Let δ_0 be a (small) positive number. Since $|\varepsilon_i^\pm(x)| \le 1$, (7.11) holds, so that (7.8) holds if

$$\begin{aligned} &P\{f_{(n)}(t_i) \le y_i < f_{(n)}(t_i) + x\} \ge \frac{d_0 x}{2}, \\ &P\{f_{(n)}(t_i) - x < y_i \le f_{(n)}(t_i)\} \ge \frac{d_0 x}{2} \end{aligned} \tag{8.3}$$

for $0 < x \le \delta_0$. Since $|\varepsilon_i^\pm(x)| \le 1$, (7.9) holds for $(\alpha, p', \sigma) = (1/2, 4, 4/d_0)$, and (7.10) holds for $\sigma = 4\delta_0/d_0$. Let $2 \le p < 3$. By Theorems 7.1 and 7.2, (5.4) and (7.13) hold with $(\alpha, p', \beta, \beta') = (1/2, 4, 0, 0)$ under (8.3), so that $\gamma = 1/p - 1/3 > 0$, $\beta'' = 0$, $b_n = n^{1/3}$, $x_n = n^\gamma$ and $k_{n,\varepsilon} = \lfloor 1/\varepsilon \rfloor$ in Theorem 8.1. Furthermore, since $k_{n,\varepsilon}$ do not depend on n and $\hat\theta_{k,\ell}^\pm$ are the medians of $\{y_k, \dots, y_\ell\}$, (8.1) holds if either $a^* - a_* < \infty$ or

$$\lim_{M\to\infty} \limsup_{n\to\infty} \sup_{1\le i\le n} P\{|y_i - f_{(n)}(t_i)| > Mn^{1/p-1/3}\} = 0. \tag{8.4}$$

Consequently, by Theorem 8.1, if (8.3) holds with $a^* - a_* < \infty$ or (8.3) and (8.4) both hold, then

$$\lim_{M\to\infty} \limsup_{n\to\infty} P\left\{ n^{1/3} \left[\frac{1}{n} \sum_{i=1}^n |\hat f_n(t_i) - f_{(n)}(t_i)|^p \right]^{1/p} \ge M \right\} = 0. \tag{8.5}$$

In the special case where the medians of y_i are nondecreasing, that is, $\mathrm{median}(y_i) = f_{(n)}(t_i)$, van de Geer (1990) obtained (8.5) under condition (8.3) for the estimator

$$\hat f_n \equiv \arg\min\left\{ \sum_{i=1}^n |y_i - f(t_i)| : f \text{ is nondecreasing and } V(f) \le V_0 \right\}. \tag{8.6}$$

The estimator (8.6) is similar to (7.1) for $a^* - a_* < \infty$, so that our results are comparable to hers in this case. We also allow here $[a_*, a^*] = [-\infty, \infty]$ with the extra condition (8.4) to control the contribution of the spikes of $\hat f_n$ at t_1 and t_n to the ℓ_p loss. Condition (8.4) holds if the errors $y_i - f_{(n)}(t_i)$ are uniformly stochastically bounded.

PROOF OF THEOREM 8.1. Let $I_n(x) \equiv \sum_{i=1}^n I\{|\hat f_n(t_i) - f_{(n)}(t_i)| > x\}$ and $L_{n,p}(x) \equiv n^{-1} \sum_{i=1}^n |\hat f_n(t_i) - f_{(n)}(t_i)|^p \wedge x^p$. By (5.4) and (7.13),

$$b_n^p E L_{n,p}(2Mx_n) \le b_n^p E L_{n,p}(\delta_0) + b_n^p \big((2M)^p x_n^p / n\big) E I_n(\delta_0) = O(1)$$

for each fixed $0 < M < \infty$. Since $L_{n,p}(\infty) = L_{n,p}(2Mx_n)$ in the event of $I_n(2Mx_n) = 0$ and $EI_n(\delta_0)/k_{n,\varepsilon} = o(1)$ as $n \to \infty$ and then $\varepsilon \to 0+$, it suffices to show

$$\limsup_{M\to\infty} \limsup_{n\to\infty} P\{I_n(2Mx_n) > 0,\ I_n(\delta_0) \le k_{n,\varepsilon}\} = 0, \qquad \varepsilon > 0.$$

By the monotonicity of both $\hat{f}_n$ and $f_{(n)}$, $I_n(2Mx_n) > 0$ implies either $\hat{f}_n(t_n) - f_{(n)}(t_1) \ge 2Mx_n$ or $\hat{f}_n(t_1) - f_{(n)}(t_n) \le -2Mx_n$. By (8.1) and symmetry, we shall only prove

$$\begin{aligned}&\{\hat{f}_n(t_n) - f_{(n)}(t_1) \ge 2Mx_n,\ I_n(\delta_0) \le k_{n,\varepsilon}\}\\ &\quad\subseteq \Big\{\max_{0\le k<k_{n,\varepsilon}} \hat{\theta}^+_{n-k,n} \ge f_{(n)}(t_n) + Mx_n\Big\}.\end{aligned} \tag{8.7}$$

Let $M \ge 1$ and let n be large enough such that $x_n \ge V(f_{(n)}) + \delta_0$. Suppose $\hat{f}_n(t_n) \ge f_{(n)}(t_1) + 2Mx_n$ and $I_n(\delta_0) \le k_{n,\varepsilon}$. Since $f_{(n)}(t_j) \le f_{(n)}(t_1) + V(f_{(n)})$, $\hat{f}_n(t_n) \ge f_{(n)}(t_j) + Mx_n + \delta_0$ for all j. Moreover, since there exist at most $k_{n,\varepsilon}$ of $j \le n$ for which $\hat{f}_n(t_j) \ge f_{(n)}(t_j) + \delta_0$, $\hat{f}_n(t_n) = \hat{f}_n(t_{n-k}) > \hat{f}_n(t_{n-k-1})$ for some random $k < k_{n,\varepsilon}$, so that, for small $x > 0$, $\sum_{i=n-k}^n \phi_i(\hat{f}_n(t_n) - x; y_i) \le \sum_{i=n-k}^n \phi_i(\hat{f}_n(t_n); y_i)$. It follows that $\hat{\theta}^+_{n-k,n} \ge \hat{f}_n(t_n)$ for certain $k < k_{n,\varepsilon}$ in view of (6.2) and (7.2). This completes the proof of (8.7) and therefore the theorem. □

APPENDIX

We provide three lemmas here.

LEMMA A.1. *Let $0 < \alpha < 1 \le p < \infty$ and let $h(x)$ be a continuous function such that $x^\alpha/h(x)\uparrow$ for $x \ge 0$ and $x^{1-\alpha}h(x)\uparrow$ for $x \ge 1$. Let $v(m) \equiv h(m+1)/(m+1)^\alpha$ and let $H_v(x; n, V) \equiv \min[1, x\{1 + V/v(x)\}/n]$ be as in* (2.2). *If $h(x)\uparrow$, then, with $\delta = 1$,*

$$\begin{aligned}&\int_{0<x<\infty} v^p(\lfloor x\rfloor)\, dH_v(x; n, V)\\ &\quad\le \int_{0<x<\infty} \frac{h^p(x+\delta)}{\max(x^\alpha, x^{p\alpha})}\, d\min\Big\{1, \frac{x}{n} + \frac{x^{1+\alpha}V}{h(x+\delta)n}\Big\}.\end{aligned} \tag{A.1}$$

If $h(1) \le h(x)$ for $0 < x < 1$ and $xh'(x)/h(x)\uparrow$ for $x > 0$, then (A.1) *holds with $\delta = 0$. If $h(x) = 1$ and $\alpha = 1/2$, then* (3.5) *holds for $p \ge 1$ and* (3.7) *holds for $1 \le p < 3$.*

PROOF. Let $H(x) \equiv x\{1 + V/v(x)\}/n$. By definition, the left-hand side of (A.1) equals

$$v^p(m_0)\{H(t_0) - H(m_0)\} + \sum_{m=0}^{m_0-1} v^p(m)\{H(m+1) - H(m)\}, \tag{A.2}$$

where $m_0 \le t_0 < m_0 + 1$ and $H(t_0) = 1$. Since $h(x)/x^{\alpha}\downarrow$ for $0 \le m \le m_0$,

$$
\begin{aligned}
&nv(m)\{H(m+c) - H(m)\} \\
&\quad = \frac{h(m+1)}{(m+1)^{\alpha}}\left[c + (m+c)\frac{(m+1+c)^{\alpha}V}{h(m+1+c)} - m\frac{(m+1)^{\alpha}V}{h(m+1)}\right] \\
&\quad \le \int_m^{m+c} \frac{h(x)}{x^{\alpha}}\,dx + V\left\{\frac{h(m+1)}{(m+1)^{\alpha-1}}\int_{m+1}^{m+1+c} d\frac{x^{\alpha}}{h(x)} + c\right\},
\end{aligned}
$$

where $c \equiv 1$ for $m < m_0$ and $c \equiv t_0 - m_0$ for $m = m_0$. Since $h(x)x^{1-\alpha}\uparrow$ for $x \ge 1$,

$$
\begin{aligned}
&\frac{h(m+1)}{(m+1)^{\alpha-1}}\int_{m+1}^{m+1+c} d\frac{x^{\alpha}}{h(x)} + c \\
&\quad \le \int_{m+1}^{m+1+c} \frac{h(x)}{x^{\alpha-1}}\,d\frac{x^{\alpha}}{h(x)} + c = \int_{m+1}^{m+1+c} \frac{h(x)}{x^{\alpha}}\,d\frac{x^{\alpha+1}}{h(x)}.
\end{aligned}
$$

Now, the above two inequalities imply

$$
\begin{aligned}
&nv(m)\{H(m+c) - H(m)\} \\
&\quad \le \int_m^{m+c} \frac{h(x)}{x^{\alpha}}\,dx + V\int_{m+1}^{m+1+c} \frac{h(x)}{x^{\alpha}}\,d\frac{x^{\alpha+1}}{h(x)}.
\end{aligned}
\tag{A.3}
$$

If $h(x)\uparrow$, then $\int_m^{m+c} x^{-\alpha}h(x)\,dx \le \int_m^{m+c} x^{-\alpha}h(x+1)\,dx$ and

$$
\begin{aligned}
&\int_{m+1}^{m+1+c} \frac{h(x)}{x^{\alpha}}\,d\frac{x^{\alpha+1}}{h(x)} \\
&\quad \le \int_m^{m+c} xh(x+1)\,d\frac{1}{h(x+1)} + c(1+\alpha) \\
&\quad = \int_m^{m+c} \frac{h(x+1)}{x^{\alpha}}\,d\frac{x^{1+\alpha}}{h(x+1)}.
\end{aligned}
$$

Thus, (A.3) is bounded by $\int_m^{m+c} x^{-\alpha}h(x+1)\,d\{x + Vx^{1+\alpha}/h(x+1)\}$, and, by (A.2),

$$
\int_{0<x<\infty} v(\lfloor x\rfloor)\,dH_v(x;n,V) \le \int_0^{t_0} x^{-\alpha}h(x+\delta)\,d\left\{\frac{x}{n} + \frac{x^{1+\alpha}V}{h(x+\delta)n}\right\},
\tag{A.4}
$$

with $\delta = 1$. This implies (A.1) with $\delta = 1$, since $v^{p-1}(m) \le \{h(x+1)/(x \vee 1)^{\alpha}\}^{p-1}$ for $m \le x < m+1$ and $t_0/n + t_0^{1+\alpha}V/\{h(t_0+1)n\} \le H(t_0) = 1$.

If $xh'(x)/h(x)\uparrow$, then the second integration in (A.3) is decreasing in m, so (A.3) leads to (A.4) with $\delta = 0$ and then to (A.1) with $\delta = 0$. We omit the details here.

Now consider $h(x) = 1$ and $\alpha = 1/2$. The difference between the proofs of (3.5) and (A.1) is the treatment of the first term in (A.2). Similar to (A.4), we obtain

$$\int_{0<x<\infty} v(\lfloor x \rfloor)\, dH_0(x; n, V) \leq \int_0^{t_0} (x \vee 1)^{-1/2}\, dH_1(x)$$

for $t_0 \geq 1$, where $H_1'(x) \equiv \max\{(1 + V\sqrt{2})/n, 1/n + (3/2)V\sqrt{x}/n\}$ and $H_1(0) \equiv 0$. This implies (3.5) for $t_0 \geq 1$, since $H_1(t_0) \leq 1$ and the measure in the integration in (2.3) puts more mass in $[0, 1)$ than $H_1(dx)$ does. Inequality (3.5) for the case of $t_0 < 1$ is trivial.

Finally, let us prove (3.7). Let $t > 1$ satisfy $\int_0^t \{1 + (3/2)V\sqrt{x \vee 1}\}\, dx = n$. By (2.3),

$$\begin{aligned} J_p(n, V) &= \frac{1}{n} \int_{0<x<t} (x \vee 1)^{-p/2} \{1 + (3/2)V\sqrt{x \vee 1}\}\, dx \\ &= \frac{1}{n} \int_0^t (x \vee 1)^{-p/2}\, dx + \frac{3V}{2n} \int_0^t (x \vee 1)^{(1-p)/2}\, dx. \end{aligned}$$

If $t \geq 1$, then $t + V/2 + Vt^{3/2} = n$, so that (3.7) follows from $t \leq \min\{n, (n/V)^{2/3}\}$ for $p \geq 1$. If $t < 1$, then $n/(1 + 3V/2) = t < 1 \leq n$ and, for $1 \leq p < 3$, (3.7) follows from

$$\begin{aligned} &\frac{3V}{2n} \int_0^t (x \vee 1)^{(1-p)/2}\, dx \\ &\quad = \frac{3V/2}{1 + 3V/2} \leq \left(\frac{3V/2}{1 + 3V/2}\right)^{p/3} \leq \left(\frac{3V}{2n}\right)^{p/3} \leq \frac{3}{3-p} \left(\frac{V}{n}\right)^{p/3}. \qquad \square \end{aligned}$$

LEMMA A.2. *Let $\{X_i\}$ be a sequence of random variables and let $\{b_n\}$ be a nondecreasing sequence of positive constants. Set $S_n \equiv \sum_{i=1}^n X_i$ with $S_0 \equiv 0$. Let $0 < \alpha < 1$.*

(i) *Let $p \geq 1/(1 - \alpha)$. Then, for $\beta \equiv I_{\{p=1/(1-\alpha)\}}$ and all $m \geq 1$,*

$$\begin{aligned} &\sup_k E \sup_{\ell \geq k+m} \left| \frac{S_\ell - S_k}{\ell - k} \right|^p \\ &\quad \leq K_{p,\alpha}^p \frac{\{\log(m+1)\}^{p\beta}}{(m+1)^{p\alpha}} \sup_{\ell > k} \frac{E|S_\ell - S_k|^p}{(\ell - k)^{p(1-\alpha)}}, \end{aligned} \tag{A.5}$$

where $K_{p,\alpha} < \infty$ are universal constants depending on (p, α) only.

(ii) *Suppose* $\{X_i\}$ *are independent variables. Then, for* $0 < c < 1$ *and* $k_0 = 2, 3, \dots,$

$$
\begin{aligned}
&P\left\{\max_{1\le i\le n}\frac{S_i}{b_i} > (k_0-c)t\right\}\\
&\qquad\le P\left\{\max_{1\le i\le n}\frac{X_i}{b_i} > ct\right\}\\
&\qquad\quad+\max_{0\le j<n}P^{k_0}\left[\bigcup_{j<i\le n}\left\{\frac{S_i-S_j}{b_i} > (1-c)t,\ \max_{j<\ell\le i}\frac{X_\ell}{b_\ell}\le ct\right\}\right],
\end{aligned}
\tag{A.6}
$$

and under the additional condition $EX_i = 0$ *there exist universal* $K_p < \infty$ *such that*

$$
E\max_{1\le i\le n}\left|\frac{S_i}{b_i}\right|^p \le K_p^p\left\{\sum_{i=1}^n E\left|\frac{X_i}{b_i}\right|^p + \left(\sum_{i=1}^n\frac{E|X_i|^2}{b_i^2}\right)^{p/2} I_{\{p>2\}}\right\}.
\tag{A.7}
$$

PROOF. (i) Let $k = 0$. By Serfling (1970), we have $E\max_{\ell\le 2^j m}|S_\ell|^p \le K_{p,\alpha}^p(2^j m)^{(1-\alpha)p}$ for $p > 1/(1-\alpha)$. This implies (A.5) after dividing $\sup_{\ell\ge m}$ into $\sum_{j=1}^\infty \max_{2^{j-1}m\le\ell<2^j m}$. The proof for $p = 1/(1-\alpha)$ is the same, using Rademacher–Mensov [cf. Serfling (1970)].

(ii) Inequality (A.6) is a version of the *good*-λ inequality [cf., e.g., Hoffmann-Jorgensen (1974) and Chow and Lai (1975, 1978)]. Set $\tau(j) \equiv \inf\{i > j : (S_i - S_j)/b_i \ge (1-c)t\}$, $T_{j+1} \equiv \tau(T_j)$, $T_0 \equiv 0$ and $A_{j,m} \equiv \{X_i/b_i \le ct,\ j < i \le m\}$. Since the left-hand side of (A.6) is bounded by $P\{A_{0,n}^c\} + P\{T_{k_0} \le n, A_{0,n}\}$, (A.6) follows from induction via

$$
\begin{aligned}
&P\{T_{m+1}\le n, A_{j,T_{m+1}} | T_m = j, A_{0,j}\}\\
&\qquad\le P\left[\bigcup_{j<i\le n}\left\{\frac{S_i-S_j}{b_i} > (1-c)t,\ \max_{j<\ell\le i}\frac{X_\ell}{b_\ell}\le ct\right\}\right].
\end{aligned}
$$

For $EX_i = 0$, we find, with truncation at level $x > 0$,

$$
P\left\{\max_{j<i\le n}(S_i - S_j) > x\right\} \le \left\{1 + \left(\frac{3}{2}\right)^2\right\}\sum_{i=j+1}^n\frac{E|X_i|^p}{x^p},
\tag{A.8}
$$

$1 \le p \le 2$, by the Kolmogorov inequality, since $|\sum_{i=j+1}^\ell EX_i I\{|X_i|\le x\}| \le x^{1-p}\sum_{i=j+1}^n E|X_i|^p \le x/3$ for $\ell \le n$ and $\sum_{i=j+1}^n E|X_i|^p/x^p \le 4/13$. Let $b_{n_k} \le 2^k < b_{n_k+1}$. By (A.8),

$$
\frac{4}{13}P\left\{\max_{j<i\le n}\frac{S_i-S_j}{b_i} > x\right\} \le \sum_k\frac{\sum_{i=j+1}^{n_k}E|X_i|^p}{x^p b_{n_{k-1}+1}^p} \le \sum_{i=j+1}^n\frac{4^p E|X_i|^p}{b_i^p x^p(2^p-1)}.
$$

This and (A.7) provide, with $c = 1/2$ and $k_0 > \max(1, p/2)$,

$$(k_0 - 1/2)^{-p} E \max_{1\le i\le n}\left|\frac{S_i}{b_i}\right|^p \le \int_0^\infty P\left\{\max_{1\le i\le n}\frac{|X_i|}{b_i} > \frac{x}{2}\right\} dx^p$$

$$+ (K'_p)^p \int_0^\infty \left[\min\left\{1, \sum_{i=1}^n \frac{E|X_i|^{p\wedge 2}}{(b_i x)^{p\wedge 2}}\right\}\right]^{k_0} dx^p$$

$$\le K_p^p \left\{\sum_{i=1}^n E\left|\frac{X_i}{b_i}\right|^p + \left(\sum_{i=1}^n \frac{E|X_i|^2}{b_i^2}\right)^{p/2} I_{\{p>2\}}\right\}. \qquad \square$$

LEMMA A.3. *Let* $c_p \equiv \arg\min\{E(c + X)^p / c^{p/3} : c > 0\}$ *for a nonnegative random variable* X *with* $EX^p < \infty$. *Then* c_p *is increasing in* p *for* $p \ge 1$. *Moreover, for* $X \sim |N(0, 1)|$, $c_3 \le 2/3$, *so that* $(2c_p)^{p/3}(3 - p)_+/3 \le 8/9$ *for* $1 \le p < 3$.

PROOF. Let $h(a; p) \equiv E(a^2 + X/a)^p$ and $a(p) \equiv \arg\min\{h(a; p) : a > 0\}$. Then $c_p = a^3(p)$ is the minimizer of $h(c^{1/3}; p)$. Define $h_k(a; p) \equiv (\partial/\partial a)^k h(a; p)$. Since $h(a; p)$ is strictly convex in a, $h_1(a(p); p) = 0$ and $h_2(a(p); p) > 0$. Moreover,

$$\frac{\partial}{\partial p} h_1\big(a(p); p\big) = a'(p) h_2\big(a(p); p\big) + h_1\big(a(p); p\big)/p$$

$$+ pE(a^2 + X/a)^{p-1}(2a - X/a^2)\log(a^2 + X/a)\big|_{a=a(p)} = 0,$$

with the expectation being negative, since $\log(a^2 + X/a)$ is increasing and $2a - X/a^2$ is decreasing in X. Thus, $a'(p) > 0$.

If $X \sim |N(0, 1)|$, then $h(a; 3) = a^6 + 3a^3\sqrt{2/\pi} + 3 + 2\sqrt{2/\pi}/a^3$ is minimized at $a(3) \approx 0.87$, so that $c_3 = a^3(3) \le 2/3$ and $(2c_p)^{p/3}(3 - p)_+/3 \le (4/3)(3 - 1)/3 = 8/9$. $\square$

Acknowledgment. This research was partially supported by the National Science Foundation. The author thanks a referee for his/her helpful suggestions.

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.

BARLOW, R. E. and UBHAYA, V. A. (1971). Isotonic approximation. In *Optimizing Methods in Statistics* (J. S. Rustagi, ed.) 77–86. Academic Press, New York.

BIRGÉ, L. (1987). Estimating a density under order restrictions: Nonasymptotic minimax risk. *Ann. Statist.* **15** 995–1012.

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

BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150.

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

BRUNK, H. D. and JOHANSEN, S. (1970). A generalized Radon–Nikodym derivative. *Pacific J. Math.* **34** 585–617.

CHOW, Y. S. and LAI, T. L. (1975). Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings. *Trans. Amer. Math. Soc.* **208** 51–72.

CHOW, Y. S. and LAI, T. L. (1978). Paley-type inequalities and convergence rates related to the law of large numbers and extended renewal theory. *Z. Wahrsch. Verw. Gebiete* **45** 1–19.

DONOHO, D. (1991). Gel'fand n-widths and the method of least squares. Technical report, Dept. Statist., Univ. California, Berkeley.

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

GRENANDER, U. (1980). *Abstract Inference.* Wiley, New York.

GROENEBOOM, P. (1983). The concave majorant of Brownian motion. *Ann. Probab.* **11** 1016–1027.

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

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser, New York.

HOFFMANN–JORGENSEN, J. (1974). Sums of independent Banach space valued random variables. *Studia Math.* **52** 159–189.

MEYER, M. and WOODROOFE, M. (2000). On the degrees of freedom in shape-restricted regression. *Ann. Statist.* **28** 1083–1104.

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

ROBERTSON, T. and WALTMAN, P. (1968). On estimating monotone parameters. *Ann. Math. Statist.* **39** 1030–1039.

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

SERFLING, R. J. (1970). Moment inequalities for the maximum cumulative sum. *Ann. Math. Statist.* **41** 1227–1234.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

VAN DE GEER, S. (1990). Estimating a regression function. *Ann. Statist.* **18** 907–924.

VAN DE GEER, S. (1993). Hellinger consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14–44.

VAN EEDEN, C. (1956). Maximum likelihood estimation of ordered probabilities. *Indag. Math.* **18** 444–455.

VAN EEDEN, C. (1957a). Maximum likelihood estimation of partially or completely ordered parameters I. *Indag. Math.* **19** 128–136.

VAN EEDEN, C. (1957b). Maximum likelihood estimation of partially or completely ordered parameters II. *Indag. Math.* **19** 201–211.

WANG, Y. (1996). The L_2 risk of an isotonic estimate. *Comm. Statist. Theory Methods* **25** 281–294.

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