Ann. Inst. Statist. Math.
Vol. 44, No. 4, 737–744 (1992)

UPPER BOUNDS FOR THE L_1-RISK OF THE MINIMUM L_1-DISTANCE REGRESSION ESTIMATOR

L. GAJEK AND M. KAŁUSZKA

Institute of Mathematics, Polish Academy of Sciences,
Institute of Mathematics, Technical University of Lódz,
Al. Politechniki 11, 90-924 Lódz, Poland

(Received April 30, 1991; revised February 27, 1992)

Abstract. A new estimator of a regression function is introduced via minimizing the L_1-distance between some empirical function and its theoretical counterpart plus penalty for the roughness. The L_1-risk of the estimator is bounded from above for every sample size no matter what the dependence structure of the observed random variables is. In the case of independent errors of measurement with a common variance the estimator is shown to achieve the optimal L_1-rate of convergence within the class of m-times differentiable functions with bounded derivatives.

Key words and phrases: Nonlinear regression, minimum distance estimation, rates of convergence.

1. Introduction

Let us consider the problem of identifying a real valued trajectory $r(t)$, $t \in (\underline{t}, \bar{t})$, of a physical object under investigation. We assume that the function $r(\cdot)$ is smooth, say m-times differentiable, and that the limit values of $r^{(i)}(t)$, $i = 0, 1, \dots, m$, are fixed and known at both boundary points $\underline{t}$ and $\bar{t}$, which may be finite or not.

One can give many practical examples corresponding to the above model. Consider, for instance, an electric circuit, where the charge $r(\underline{t})$ is concentrated inside a condenser at time $\underline{t} = 0$ while at $\bar{t} = \infty$ the state of stability with $r(\bar{t}) = r'(\bar{t}) = 0$ is achieved. Other examples one can find in mechanics, astronomy and so on.

The problem of our interest is how to identify the function $r(\cdot)$ on the basis of a finite number of observations of $r(t_i) + \epsilon_i$, $i = 1, \dots, n$, where ϵ_i is a random error of measuring. In this paper we treat the case where t_i's are independent random variables with a common absolutely continuous distribution function. Knowing the distribution of t_i's, one can transform them to the random variables uniformly distributed on $(0, 1)$. Also the boundary conditions for r can be assumed, without loss of generality, to be of the form $r^{(i)}(\underline{t}) = r^{(i)}(\bar{t}) = 0$, $i = 0, 1, \dots, m$.

So let us consider independent uniformly distributed on $(0,1)$ random variables $T_1,\dots,T_n$. Let r be an unknown regression function on $(0,1)$. Consider random variables $Y_1,\dots,Y_n$ which are assumed to satisfy

$$Y_i = r(T_i) + \epsilon_i, \qquad i = 1,\dots,n,$$

where $\epsilon_1,\dots,\epsilon_n$ are independent of $T_1,\dots,T_n$ random errors with $E\epsilon_i = 0$ and $\mathrm{cov}(\epsilon_i,\epsilon_j) = \sigma_{ij}$, $i,j = 1,\dots,n$. The problem is to estimate r given that r is from the class $\mathcal{R}^m$ of functions with a bounded m-th derivative on $[0,1]$ and such that $r^{(i)}(0) = r^{(i)}(1) = 0$, $i = 0,1,\dots,m$. The least squares estimation is meaningless in the present context unless we modify the method e.g. by adding $\int |r^{(m)}|^2$, the penalty for roughness, to the sum of squares. This leads to smoothing noisy data by spline functions (see Silverman (1985)). On the other hand, the spline smoothing is asymptotically equivalent to using variable kernel method (see Silverman (1984)). We refer the reader to the paper of Jennen-Steinmetz and Gasser (1988) for references concerning the kernel methods.

Another possibility of smooth estimation of r is to minimize the sum of absolute deviations $\sum_{i=1}^n |Y_i - r(T_i)|$ penalized, if wanted, by adding $\int |r^{(m)}|$. For general motivation, historical background and characteristic properties of least absolute deviation (LAD) methods, we refer the reader to the book of Dodge (1987). Asymptotic results concerning LAD estimators of regression function from a finite dimensional space one can find in Pollard (1990); a nonparametric case was partly treated by van de Geer (1990). Here we investigate a different approach to the LAD regression function estimation. As is well known the accuracy of estimating r is limited by the fact that there does not exist an unbiased estimator of r from the class $\mathcal{R}^m$. Still, for the function

$$R(t) = \int_0^t r(s)ds \tag{1.1}$$

an unbiased estimator is available. Since (1.1) defines a one-to-one map on $\mathcal{R}^m$, we shall use the notation $R \in \mathcal{R}^m$ whenever the corresponding $r = R'$ belongs to $\mathcal{R}^m$. We start our investigations with constructing an unbiased estimator of R. To this end let us consider an empirical counterpart of R,

$$R_n(t) = \frac{1}{n}\sum_{i=1}^n Y_i \mathbf{1}_{(0,t)}(T_i). \tag{1.2}$$

It can be easily shown that R_n is an unbiased estimator of R.

Let $\|\ \|_1$ denote L_1-norm in the space of functions which are integrable on $[0,1]$. Though $E\|R_n - R\|_1 = O(n^{-1/2})$, R_n is not a smooth function. The minimum L_1-distance smoother $\hat{R}_n$ of R_n is defined via minimizing the functional

$$J_n(R) \equiv \|R_n - R\|_1 + \beta\|R^{(m+1)}\|_1$$

over the class of $(m+1)$-times differentiable on $\mathbb{R}$ functions with a bounded support, where $\beta = \beta(n) > 0$ are prescribed smoothing parameters. The minimum L_1-penalized distance (MPD) estimator of $R^{(i)}$, $i = 1,\dots,m$ is defined as $\hat{R}_n^{(i)}$.

The main idea of the above definition follows the idea of the MPD estimation of a density and its derivatives, which is investigated in Gajek (1990) and earlier, in the case of L_2 and L_∞ distances, in Gajek (1989). In the present paper we show that a similar approach leads to the regression estimator which achieves the optimal rate of L_1-convergence.

In fact we have proven somewhat stronger result; Theorem 2.1 gives an upper bound for the L_1-risk of the MPD regression estimator which is valid for each sample size n no matter what the dependence structure of the sample is.

In order to admit approximate solutions to the problem of minimizing $J_n(R)$, one can define MPD estimators as the ones minimizing $J_n(R)+\delta_n$, $\delta_n > 0$. Clearly, if δ_n tends to zero quickly enough, the asymptotic counterparts of the results presented in the paper are still valid.

2. The main result

THEOREM 2.1. *Let $\mathcal{R}^m(L,S) = \{R \in \mathcal{R}^m : \sup|R| \le L, \|R^{(m+1)}\|_1 \le S\}$. Then for each $i = 1,\ldots,m$*

$$\sup_{R\in\mathcal{R}^m(L,S)} E\|\hat{R}_n^{(i)} - R^{(i)}\|_1$$
$$\le \beta^{-i/(m+1)}\left\{M_1\beta^{1/2(m+1)}\left(\sum_{i=1}^n \sigma_{ii} + nL^2\right)^{1/2}/n\right.$$
$$\left. + M_2\beta\left[2\max\left(0,\sum_{i<j}\sigma_{ij}\right)\right]^{1/2}/n + M_3\beta\right\},$$

where M_1, M_2 and M_3 are constants given by (2.5), (2.6) *and* (2.7) *below.*

PROOF. Let k be a m-times differentiable function on $\mathbb{R}$ such that

(i) $\operatorname{supp} k = [-1,1]$,

(ii) $\int_{-1}^1 k(u)du = 1$,

(iii) $\int_{-1}^1 u^i k(u)du = 0$ for $i = 1,\ldots,m$.

Let us define

$$\tilde{R}_h(x) = h^{-1}\int k\left(\frac{x-t}{h}\right)R_n(t)dt$$

and for $i = 1,\ldots,m$,

$$\tilde{R}_h^{(i)}(x) = h^{-i-1}\int k^{(i)}\left(\frac{x-t}{h}\right)R_n(t)dt. \tag{2.1}$$

From Theorem 2.1 of Gajek (1989), we have that if the kernel estimator (2.1) has the bandwidth $h(n) = C_1\beta(n)^{1/(m+1)}$ with

$$C_1 = \left[i\|k^{(i)}\|_1(m-i)!/\int|v|^{m+1-i}|k(v)|dv\right]^{1/(m+1)}, \tag{2.2}$$

then for $i = 1, \ldots, m$

$$E\|\hat{R}_n^{(i)} - \tilde{R}_h^{(i)}\|_1 \leq C_2 \beta(n)^{-i/(m+1)} E J_n(\hat{R}_n),$$

where

$$C_2 = \frac{m+1}{m+1-i} \|k^{(i)}\|_1 C_1^{-i}. \tag{2.3}$$

Since $J_n(\hat{R}_n) \leq J_n(\tilde{R}_h)$, using the triangle inequality, we get

$$E\|\hat{R}_n^{(i)} - R^{(i)}\|_1 \leq C_2 \beta(n)^{-i/(m+1)} E J_n(\tilde{R}_h) + E\|\tilde{R}_h^{(i)} - R^{(i)}\|_1. \tag{2.4}$$

From Lemma A.2 of the Appendix, we have

$$\begin{aligned} E J_n(\tilde{R}_h) \leq\ & n^{-1} \beta^{1/2(m+1)} \left(\sum_{i=1}^n \sigma_{ii} + nL^2 \right)^{1/2} C_1^{1/2} \\ & \cdot \left\{ \int |v|^{1/2} |k(v)| dv + C_1^{-m-1} \int |v|^{1/2} |k^{(m+1)}(v)| dv \right\} \\ & + n^{-1} \beta^{1/(m+1)} \left[2 \max \left(0, \sum_{i<j} \sigma_{ij} \right) \right]^{1/2} C_1 \\ & \cdot \left[\int |v| |k(v)| dv + C_1^{-m-1} \int |v| |k^{(m+1)}(v)| dv \right] \\ & + \beta C_1^{m+1} \|R^{(m+1)}\|_1 \\ & \cdot \left[\int |v|^{m+1} |k(v)| dv / (m+1)! + C_1^{-m-1} \int |k(v)| dv \right]. \end{aligned}$$

Now, using Lemma A.1 of the Appendix and (2.4), we get

$$\begin{aligned} & E\|\hat{R}_n^{(i)} - R^{(i)}\|_1 \\ & \leq \beta^{-i/(m+1)} \left[M_1 \beta^{1/2(m+1)} \left(\sum_{i=1}^n \sigma_{ii} + nL^2 \right)^{1/2} / n \right. \\ & \qquad \left. + M_2 \beta^{1/(m+1)} \left[2 \max \left(0, \sum_{i<j} \sigma_{ij} \right) \right]^{1/2} / n + M_3 \beta \right], \end{aligned}$$

where

$$\begin{aligned} M_1 = C_1^{1/2} \Bigg\{ & C_2 \left[\int |v|^{1/2} |k(v)| dv + C_1^{-m-1} \int |v|^{1/2} |k^{(m+1)}(v)| dv \right] \\ & + C_1^{-i} \int |v|^{1/2} |k^{(i)}(v)| dv \Bigg\}, \end{aligned} \tag{2.5}$$

$$\begin{aligned} M_2 = C_1 \Bigg\{ & C_2 \left[\int |v| |k(v)| dv + C_1^{-m-1} \int |v| |k^{(m+1)}(v)| dv \right] \\ & + C_1^{-i} \int |v| |k^{(i)}(v)| dv \Bigg\} \end{aligned} \tag{2.6}$$

and

$$(2.7)\quad M_3 = SC_1^{m+1}\left\{C_2\left[\int |v|^{m+1}|k(v)|dv/(m+1)! + C_1^{-m-1}\int |k(v)|dv\right]\right.$$
$$\left. + C_1^{-i}\int |v|^{m+1-i}|k(v)|dv/(m+1-1)!\right\},$$

for C_1 and C_2 defined by (2.2) and (2.3), respectively. □

Assume that $\sum_{i<j}\sigma_{ij} \leq 0$. Then an optimal choice of β is given in the following corollary.

COROLLARY 2.1. *Let $\beta = M_4 n^{-(m+1)/(2m+1)}$ and $\sum_{i<j}\sigma_{ij} \leq 0$. Then for each $i = 1, \ldots, m$*

$$\sup_{R\in\mathcal{R}^m(L,S)} E\|\hat{R}_n^{(i)} - R^{(i)}\|_1$$
$$\leq n^{-(m+1)/(2m+1)}[M_1 M_4^{1/2(m+1)} + M_3 M_4]M_4^{-i/(m+1)}.$$

The optimal choice of M_4 in the above bound is

$$M_4 = \{2M_3(m+1-i)/[M_1(2i-1)]\}^{-2(m+1)/(2m+1)}.$$

Remark 2.1. As pointed out in Gajek (1989, 1990), the L_1-properties of the MPD estimator do not follow from the bounds for its L_2-risk via the Cauchy inequality.

Acknowledgement

The authors would like to thank the referee for detailed remarks which led to the improved version of the paper.

Appendix

LEMMA A.1. *Let $L = \sup_{0\leq t\leq 1}|r(t)|$. Then for each $i = 1, \ldots, m$*

$$E\|\tilde{R}_h^{(i)} - R^{(i)}\|_1 \leq h^{-i+1/2}n^{-1}\left(\sum_{i=1}^n \sigma_{ii} + nL^2\right)^{1/2}\int |v|^{1/2}|k^{(i)}(v)|dv$$
$$+ h^{-i+1}n^{-1}\left(2\max\left(0, \sum_{i<j}\sigma_{ij}\right)\right)^{1/2}\int |v||k^{(i)}(v)|dv$$
$$+ h^{m+1-i}\|R^{(m+1)}\|_1\frac{\int |v|^{m+1-i}|k(v)|dv}{(m+1-i)!}.$$

PROOF. Since

$$R^{(i)}(x) = h^{-i-1}\int k^{(i)}\left(\frac{x-t}{h}\right)R(t)dt + \iint_0^{hv}\frac{(z-hv)^m}{m!}R^{(m+1)}(x-z)k(v)dzdv,$$

therefore

$$\text{(A.1)}\qquad E|\tilde{R}_h^{(i)} - R^{(i)}| \le h^{-i}E\left|\int [R_n(x-hv) - R(x-hv)]k^{(i)}(v)dv\right| + \left|\iint_0^{hv}\frac{(z-hv)^m}{m!}R^{(m+1)}(x-z)k(v)dzdv\right|.$$

Observe that for $v>0$

$$\begin{aligned}\{E|R_n(x-hv) &- R_n(x) - [R(x-hv)-R(x)]|\}^2\\ &\le \operatorname{Var}[R_n(x-hv) - R_n(x)]\\ &= n^{-2}\operatorname{Var}\left[\sum_{i=1}^n Y_i\mathbf{1}_{(x-hv,x)}(T_i)\right]\\ &= n^{-2}\left[E\left\{\operatorname{Var}\left[\sum_{i=1}^n Y_i\mathbf{1}_{(x-hv,x)}(T_i)\mid T\right]\right\}\right.\\ &\qquad\left.+\operatorname{Var}\left\{E\left[\sum_{i=1}^n Y_i\mathbf{1}_{(x-hv,x)}(T_i)\mid T\right]\right\}\right],\end{aligned}$$

where $T=(T_1,\dots,T_n)$. Since

$$E\left\{\operatorname{Var}\left[\sum_{i=1}^n Y_i\mathbf{1}_{(x-hv,x)}(T_i)\mid T\right]\right\} = \min(x,hv)\sum_{i=1}^n\sigma_{ii} + 2\min{}^2(x,hv)\sum_{i<j}\sigma_{ij}$$

and

$$\begin{aligned}\operatorname{Var}\left\{E\left[\sum_{i=1}^n Y_i\mathbf{1}_{(x-hv,x)}(T_i)\mid T\right]\right\} &= n\operatorname{Var}[r(T_1)\mathbf{1}_{(x-hv,x)}(T_1)]\\ &\le nL^2\min(x,hv),\end{aligned}$$

therefore

$$\begin{aligned}\text{(A.2)}\quad E|R_n(x-hv) &- R_n(x) - R(x-hv) + R(x)|\\ &\le n^{-1}\left[h|v|\left(\sum_{i=1}^n\sigma_{ii}+nL^2\right)+2(hv)^2\max\left(0,\sum_{i<j}\sigma_{ij}\right)\right]^{1/2}\\ &\le n^{-1}\left[h|v|\left(\sum_{i=1}^n\sigma_{ii}+nL^2\right)\right]^{1/2}\\ &\quad + n^{-1}h|v|\left[2\max\left(0,\sum_{i<j}\sigma_{ij}\right)\right]^{1/2},\end{aligned}$$

where the last inequality follows from the fact that $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$, $a, b > 0$. It is easy to prove that (A.2) holds also for $v < 0$. Now, using (A.2) and the identity $\int k^{(i)}(v)dv = 0$ for $i = 1, \ldots, m$, we get

$$E\left|\int [R_n(x-hv) - R(x-hv)]k^{(i)}(v)dv\right|$$
$$\le n^{-1}h^{1/2}\left(\sum_{i=1}^n \sigma_{ii} + nL^2\right)^{1/2}\int |v|^{1/2}|k^{(i)}(v)|dv$$
$$+ n^{-1}h\left[2\max\left(0, \sum_{i<j}\sigma_{ij}\right)\right]^{1/2}\int |v||k^{(i)}(v)|dv.$$

Since

$$\left\|\int\left|\int_0^{hv}\frac{(z-hv)^{m-i}}{(m-i)!}R^{(m+1)}(\cdot+z)k(v)dz\right|dv\right\|_1$$
$$\le h^{m+1-i}\|R^{(m+1)}\|_1\frac{\int |v|^{m+1-i}|k(v)|dv}{(m+1-i)!},$$

therefore the assertion follows from (A.1). □

LEMMA A.2. *The following inequalities hold:*

(i)
$$E\|\tilde{R}_h^{(m+1)}\|_1 \le h^{-m-1/2}n^{-1}\left(\sum_{i=1}^n \sigma_{ii} + nL^2\right)^{1/2}\int |v|^{1/2}|k^{(m+1)}(v)|dv$$
$$+ h^{-m}n^{-1}\left[2\max\left(0, \sum_{i<j}\sigma_{ij}\right)\right]^{1/2}\int |v||k^{(m+1)}(v)|dv$$
$$+ \|R^{(m+1)}\|_1\int |k(v)|dv;$$

(ii)
$$E\|\tilde{R}_h - R_n\|_1 \le h^{1/2}n^{-1}\left(\sum_{i=1}^n \sigma_{ii} + nL^2\right)^{1/2}\int |v|^{1/2}|k(v)|dv$$
$$+ hn^{-1}\left[2\max\left(0, \sum_{i<j}\sigma_{ij}\right)\right]^{1/2}\int |v||k(v)|dv$$
$$+ h^{m+1}\|R^{(m+1)}\|_1\int |v|^{m+1}|k(v)|dv/(m+1)!.$$

PROOF. (i) Observe that

$$\tilde{R}_h^{(m+1)}(x) = h^{-m-1}\int [R_n(x-hv) - R_n(x) - R(x-hv) + R(x)]k^{(m+1)}(v)dv$$
$$+ h^{-m-1}\int [R(x-hv) - R(x)]k^{(m+1)}(v)dv$$

and

$$\int [R(x-hv)-R(x)]k^{(m+1)}(v)dv = h^{m+1}\int R^{(m+1)}(x-hv)k(v)dv.$$

The rest of the proof is similar to the proof of Lemma A.1.

(ii) Since

$$\int [R(x-hv)-R(x)]k(v)dv = -\iint_0^{hv} \frac{(z-hv)^m}{m!} R^{(m+1)}(x-z)k(v)dzdv,$$

therefore

$$|\tilde{R}_h(x)-R_n(x)| \le \left|\int [R_n(x-hv)-R_n(x)-R(x-hv)+R(x)]k(v)dv\right| + \left|\iint_0^{hv} \frac{(z-hv)^m}{m!} R^{(m+1)}(x-z)k(v)dzdv\right|.$$

Now, the assertion follows as in the previous cases. □

REFERENCES

Dodge, Y. (1987). *Statistical Data Analysis Based on the L_1-norm and Related Methods*, North-Holland, Tokyo.

Gajek, L. (1989). Estimating a density and its derivatives via the minimum distance method, *Probab. Theory Related Fields*, **80**, 601–617.

Gajek, L. (1990). The rate of convergence of the nonparametric minimum L_1-penalized distance estimator, *Probab. Math. Statist.* (to appear).

Jennen-Steinmetz, Ch. and Gasser, T. (1988). A unifying approach to nonparametric regression estimation, *J. Amer. Statist. Assoc.*, **83**, 1084–1089.

Pollard, D. (1990). Asymptotics for least absolute deviation regression estimators (preprint).

Silverman, D. (1984). Spline smoothing: the equivalent variable kernel method, *Ann. Statist.*, **12**, 898–916.

Silverman, D. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion), *Journal of the Royal Statistical Association, B*, **47**, 1–52.

van de Geer, S. (1990). Estimating a regression function, *Ann. Statist.*, **18**, 907–924.