TAIWANESE JOURNAL OF MATHEMATICS
Vol. 2, No. 3, pp. 313-319, September 1998

NEAREST NEIGHBOR MEDIAN ESTIMATION OF REGRESSION FUNCTION AND ITS DERIVATIVES*

Ying Yang

Abstract. Consider the fixed-design nonparametric median regression model $Y_i = g(x_i) + e_i$, $i \geq 1$. For estimating the regression function $g(x)$ and its derivative, the nearest neighbor median estimators $\tilde{g}_{n,h}(x)$ are employed, where h is the number of the nearest neighbors. Under mild regularity conditions, rates of convergence for the estimators are obtained.

1. Introduction

Consider the observations

$$Y_i = g(x_i) + e_i, i \geq 1, \tag{1}$$

where $g(\cdot)$ is an unknown regression function to be estimated and assumed to be continuous on [0, 1], $\{x_i, i \geq 1\}$ are nonrandom fixed-design points from the interval [0, 1], $\{e_i, i \geq 1\}$ are independent and identically distributed (iid) random errors with a unique median 0 and Y_i $(i \geq 1)$ are noisy observations of $g(x)$ at x_i $(i \geq 1)$. The practical importance of obtaining a nonparametric regression estimate of $g(x)$ has led to several estimators for $g(x)$, for examples, kernel estimates (Priestly and Chao, 1972; Gasser and Müller, 1979; Cheng and Lin, 1981 a,b; Georgiev, 1984; and the references therein) and weighted orthogonal series estimation (Yang, 1994). The presence of a small part of outliers may, however, cause a difficult explanation of the estimated regression function. Robust alternatives to the kernel methods have been proposed

Received November 5, 1996.
Communicated by I.-S. Chang.
1991 *Mathematics Subject Classification*: Primary 62G05.
Key words and phrases: Nonparametric regression, estimation of derivative, nearest neighbor median estimates, rate of convergence.
*Research Supported by NSFC.

(Härdle and Gasser, 1984). Robust spline smoothing and local median method in the random case were considered (Huber, 1979; Truong, 1989).

The estimation of derivatives from noisy observations is of importance in many areas of engineering and physics. Härdle and Gasser (1985) discussed the problem on robust kernel estimation of derivatives of regression functions. Holiday (1989) also discussed the estimation of derivatives of a nonparametric regression function when the data are correlated.

The results above-mentioned except Truong (1989) are considered in the nonparametric regression setup. However, the model (1) is not, generally speaking, a nonparametric regression, since the conditional expectation

$$E_{g(\cdot)}[Y_i] = g(x_i) + Ee_i$$

may not exist under Assumptions 1 and 2 below. The $g(x_i)$ has the meaning of conditional median.

The aim of this paper is to present a robust estimator for the nonparametric median function $g(x)$ and its derivative $g^{'}(x)$. In the nonparametric regression setup, similar problems have been studied by Holiday (1989). The rates of convergence for the estimators will be described in Section 2. Proofs of the conclusions are given in Section 3.

2. Assumptions and Main Results

Take a subseries $\{Y_i, 1 \le i \le n\}$ from the infinite series $\{Y_i, i \ge 1\}$. Without loss of generality, assume that $0 \equiv x_0 \le x_1 \le x_2 \cdots \le x_n \equiv 1$. For $x \in [0,1]$ and $n \ge 1$, let $\{D_{n1}(x), D_{n2}(x) \cdots, D_{nn}(x)\}$ be the order statistics of $\{x_1, x_2, \cdots, x_n\}$, ordered by the distances $\{|x - x_i|\}$, $1 \le i \le n$, with ties being broken by the chronological order. Let $Y_{ni}(x)$ and $e_{ni}(x)$ $(i = 1, 2, \cdots, n)$ denote the corresponding noisy observation and random error at $D_{ni}(x)$ $(i = 1, 2, \cdots, n)$, respectively. The following estimate

$$\begin{aligned} \tilde{g}_{n,h}(x) &= m(Y_{n1}(x), Y_{n2}(x), \cdots, Y_{nh}(x)) \\ &= \text{median of } Y_{n1}(x), Y_{n2}(x), \cdots, Y_{nh}(x) \end{aligned} \tag{2}$$

is called the *nearest neighbor median estimator* of $g(x)$, where the number of nearest neighbors h plays the role of the smoothing parameter. If h is even, then $\tilde{g}_{n,h}(x)$ is equal to the average of the two middle order statistics.

The estimate of $g^{'}(x)$, the derivative of $g(x)$, is defined by

$$D\tilde{g}_{n,h}(x) = \delta_{n,h}[\tilde{g}_{n,h}(x + \delta_{n,h}^{-1}) - \tilde{g}_{n,h}(x)], \tag{3}$$

where the integer sequence $\delta_{n,h} \to \infty$.

A set of mild regularity conditions will be useful throughout this paper.

Assumption 1. $e_1, e_2, \cdots, e_n, \cdots$ are iid random variables with median 0, i.e., $F(0) = \frac{1}{2}$, and common distribution function $F(x)$ and density function $f(x) = F(x)'$. $\exists \delta > 0$, $\exists c_1 > 0$ such that $f(x) \geq c_1$, $\forall x \in [-\delta, \delta]$;

Assumption 2. (i) $\exists L > 0$ such that $|g(x) - g(y)| \leq L|x - y|$ for $x, y \in [0, 1]$; (ii) $\exists M_0 > 0$ such that $|g^{''}(x)| \leq M$, $\forall x \in (0, 1)$;

Assumption 3. $\exists c_2 > 0$ such that $|x - D_{nh}(x)| \leq c_2 \frac{h}{n}$, where $h \sim n^{\frac{2}{3}}$.

Let $\alpha_n = n^{-\frac{1}{3}} \sqrt{\log n} \beta_n$, where β_n tends very slowly to infinity.

Theorem 2.1. *Suppose Assumptions* 1, 2(i), *and* 3 *hold and let* $h \sim n^{\frac{2}{3}}$. *Then*

$$\alpha_n^{-1} \sup_{0 \leq x \leq 1} |\tilde{g}_{n,h}(x) - g(x)| = o(1) \quad a.s..$$

Theorem 2.2. *Under the assumptions of Theorem* 2.1,

$$\sup_{0 \leq x \leq 1} E(\tilde{g}_{n,h}(x) - g(x))^2 = O\left(\left(\frac{h}{n}\right)^2 + \frac{1}{h}\right) + o(h^{-\frac{3}{2}}).$$

In particular, if $h \sim n^{\frac{2}{3}}$, *then*

$$\sup_{0 \leq x \leq 1} E(\tilde{g}_{n,h}(x) - g(x))^2 = O(n^{-\frac{2}{3}}).$$

Theorem 2.3. *Suppose Assumptions* 1, 2(ii), *and* 3 *hold and let* $h \sim n^{\frac{2}{3}}$ *and* $\delta_{n,h} = \frac{1}{\sqrt{\alpha_n}}$. *Then*

$$\sup_{1 \leq x \leq 1} |D\tilde{g}_{n,h}(x) - g^{'}(x)| = O(\sqrt{\alpha_n}) \quad \text{a.s..}$$

Remark 1. $x + (\delta_{n,h})^{-1} \in [0, 1]$ in (3) is required.

Remark 2. It should be noted that in the present work, the choice of h is left somewhat subjective. However, the smoothing parameters can be chosen using the cross-validation procedure. Zheng and Yang (1996) and Yang and Zheng (1996) suggested the following median cross-validation criterion to choose the number of nearest neighbors h:

$$\text{cv}(h) = \text{median of } |Y_1 - \tilde{g}_{n,h}^{-1}(x_1)|, |Y_2 - \tilde{g}_{n,h}^{-2}(x_2)|, \cdots, |Y_n - \tilde{g}_{n,h}^{-n}(x_n)|,$$

where $\tilde{g}^{-i}_{n,h}(x_i)$ is the deleted-one version of the estimator in (2), i.e., $\tilde{g}^{-i}_{n,h}(x_i) = m(Y_{n2}(x_i), Y_{n3}(x_i), \cdots, Y_{nh}(x_i))$. The interested reader can refer to Zheng and Yang (1996) and Yang and Zheng (1996) for details.

3. Proofs

Proof of Theorem 2.1. By Assumptions 2(i) and 3, for each $i, 1 \leq i \leq n$, and every $x \in [x_{i-1}, x_i]$, $1 \leq j \leq h$, $h \sim n^{\frac{2}{3}}$, we have

$$|g(D_{nj}(x)) - g(x)| \leq L|D_{nj}(x) - x| \leq Lc_2\frac{h}{n}. \tag{4}$$

Also, when $n \gg 1$, $\forall \varepsilon > 0$, $h \sim n^{\frac{2}{3}}$,

$$\epsilon\alpha_n - Lc_2\frac{h}{n} \sim \epsilon n^{-\frac{1}{3}}\sqrt{\log n}\beta_n - Lc_2 n^{-\frac{1}{3}} \geq \frac{1}{2}\epsilon\alpha_n, \tag{5}$$

and then by Assumption 1 and the mean value theorem

$$F\left(\epsilon\alpha_n - Lc_2\frac{h}{n}\right) - F(0) = \left(\epsilon\alpha_n - Lc_2\frac{h}{n}\right) f(\theta_{n,h}),$$

where $\theta_{n,h}$ lies between 0 and $\epsilon\alpha_n - Lc_2\frac{h}{n}$. If $h \sim n^{\frac{2}{3}}$, $n \gg 1$, then $0 < \epsilon\alpha_n - Lc_2\frac{h}{n} \leq \delta$. Thus

$$F\left(\epsilon\alpha_n - Lc_2\frac{h}{n}\right) - F(0) \geq c_1\left(\epsilon\alpha_n - Lc_2\frac{h}{n}\right) \geq \frac{1}{2}c_1\epsilon\alpha_n. \tag{6}$$

Therefore, for n sufficiently large, by (4) $\sim$ (6) we have

$$\begin{aligned}
&P\Big\{\sup_{0\leq x\leq 1}|\tilde{g}_{n,h}(x) - g(x)| \geq \epsilon\alpha_n\Big\} \\
&\leq P\Big\{\max_{1\leq i\leq n}\sup_{x_{i-1}\leq x\leq x_i}|\tilde{g}_{n,h}(x) - g(x)| \geq \epsilon\alpha_n\Big\} \\
&\leq n\max_{1\leq i\leq n}P\Big\{\sup_{x_{i-1}\leq x\leq x_i}|m(e_{n1}(x) + g(D_{n1}(x)), \cdots, e_{nh}(x) \\
&\qquad +g(D_{nh}(x))) - g(x)| \geq \epsilon\alpha_n\Big\} \\
&\leq n\max_{1\leq i\leq n}P\Big\{\sup_{x_{i-1}\leq x\leq x_i}|m(e_{n1}(x), \cdots, e_{nh}(x))| \\
&\qquad +\max_{1\leq j\leq h}|g(D_{nj}(x)) - g(x)| \geq \epsilon\alpha_n\Big\} \\
&\leq n\max_{1\leq i\leq n}P\Big\{\sup_{x_{i-1}\leq x\leq x_i}|m(e_{n1}(x), \cdots, e_{nh}(x))| + Lc_2\frac{h}{n} \geq \epsilon\alpha_n\Big\} \\
&\leq n\max_{1\leq i\leq n}P\Big\{\bigcup_{i=1}^{n}|m(e_{n1}(x_i), \cdots, e_{nh}(x_i))| \geq \epsilon\alpha_n - Lc_2\frac{h}{n}\Big\} \\
&\leq n^2\max_{1\leq i\leq n}P\Big\{|m(e_{n1}(x_i), \cdots, e_{nh}(x_i))| \geq \frac{1}{2}\epsilon\alpha_n\Big\}.
\end{aligned}$$

Since $m(e_{n1}(x_i), \cdots, e_{nh}(x_i))$ and $m(e_1, \cdots, e_h)$ have the same distribution,

$$
\begin{aligned}
(7)\qquad & P\Big\{\sup_{1\le x\le 1} |\tilde{g}_{n,h}(x) - g(x)| \ge \epsilon\alpha_n\Big\} \le n^2 P\Big\{|m(e_1, \cdots, e_h)| \ge \frac{1}{2}\epsilon\alpha_n\Big\} \\
& \le n^2 P\Big\{m(e_1, \cdots, e_h) \ge \frac{1}{2}\epsilon\alpha_n\Big\} + n^2 P\Big\{m(e_1, \cdots, e_h) \le -\frac{1}{2}\epsilon\alpha_n\Big\}.
\end{aligned}
$$

As to the second term on the right side in (7), by Hoeffding's inequality (Hoeffding, 1963)

$$
\begin{aligned}
& n^2 P\left\{m(e_1, \cdots, e_h) \ge \frac{1}{2}\epsilon\alpha_n\right\} = n^2 P\left\{\frac{1}{n}\sum_{i=1}^{h} I_{\{e_i \ge \frac{1}{2}\epsilon\alpha_n\}} \ge \frac{1}{2}\right\} \\
& = n^2 P\left\{\frac{1}{n}\sum_{i=1}^{h}\left[I_{\{e_i \ge \frac{1}{2}\epsilon\alpha_n\}} - P\left\{e_i \ge \frac{1}{2}\varepsilon\alpha_n\right\}\right] \ge \frac{1}{2} - P\left\{e_i \ge \frac{1}{2}\varepsilon\alpha_n\right\}\right\} \\
(8)\qquad & = n^2 P\left\{\frac{1}{n}\sum_{i=1}^{h}\left[I_{\{e_i \ge \frac{1}{2}\epsilon\alpha_n\}} - P\left\{e_i \ge \frac{1}{2}\epsilon\alpha_n\right\}\right] \ge F\left(\frac{1}{2}\epsilon\alpha_n\right) - F(0)\right\} \\
& \le n^2 P\left\{\frac{1}{n}\sum_{i=1}^{h}\left[I_{\{e_i \ge \frac{1}{2}\epsilon\alpha_n\}} - P\left\{e_i \ge \frac{1}{2}\epsilon\alpha_n\right\}\right] \ge \frac{1}{2}c_1\epsilon\alpha_n\right\} \\
& \le 2n^2 \exp\left\{-c_0\left(\frac{1}{2}c_1\epsilon\alpha_n\right)^2 h\right\} \\
& \sim 2n^2 \exp\left\{-\frac{1}{4}c_0 c_1^2 \epsilon^2 \beta_n^2 \log n\right\} \le \frac{1}{n^2}, \quad (n \gg 1).
\end{aligned}
$$

Similarly, when $n \gg 1$,

$$
(9)\qquad n^2 P\left\{m(e_1, \cdots, e_h) \le -\frac{1}{2}\epsilon\alpha_n\right\} \le \frac{1}{n^2}\,.
$$

By (7) $\sim$ (9),

$$
\sum_{n=1}^{\infty} P\Big\{\sup_{0\le x\le 1} |\tilde{g}_{n,h}(x) - g(x)| \ge \epsilon\alpha_n\Big\} < \infty,
$$

therefore, according to Borel-Cantelli lemma:

$$
\alpha_n^{-1} \sup_{1\le x\le 1} |\tilde{g}_{n,h}(x) - g(x)| = o(1) \quad \text{a.s..}
$$

This finishes the proof of the theorem. ■

Proof of Theorem 2.2. Note that

$$\begin{aligned}|\tilde{g}_{n,h}(x) - g(x))| &= |m(e_{n1}(x) + g(D_{n1}(x)), \cdots, e_{nh}(x) + g(D_{nh}(x))) - g(x)| \\ &\leq |m(e_{n1}(x), \cdots, e_{nh}(x))| + \max_{1 \leq j \leq h} |g(D_{nj}(x)) - g(x)| \\ &\leq |m(e_{n1}(x), \cdots, e_{nh}(x))| + Lc_2 \frac{h}{n}.\end{aligned}$$

Sine $m(e_{n1}(x), \cdots, e_{nh}(x))$ and $m(e_1, \cdots, e_h)$ have the same distribution, by Theorem C (Serfling, 1983, p. 101)

$$E[m(e_{n1}(x), \cdots, e_{nh}(x))]^2 = \frac{1}{4f^2(0)h} + o(h^{-\frac{3}{2}}).$$

So

$$\begin{aligned}E|\tilde{g}_{n,h}(x) - g(x))|^2 &\leq 2E[m(e_{n1}(x), \cdots, e_{nh}(x))]^2 + o(h^{-\frac{3}{2}}) + 2\left(Lc_2 \frac{h}{n}\right)^2 \\ &= O\left(\left(\frac{h}{n}\right)^2 + \frac{1}{h}\right) + o\left(h^{-\frac{3}{2}}\right).\end{aligned}$$

This finishes the proof of the theorem. ■

Proof of Theorem 2.3. Note that

$$\begin{aligned}&|D\tilde{g}_{n,h}(x) - g(x)| \\ =\ &|\delta_{n,h}[\tilde{g}_{n,h}(x + \delta_{n,h}^{-1}) - \tilde{g}_{n,h}(x)] - g^{'}(x)| \\ \leq\ &|\delta_{n,h}[\tilde{g}_{n,h}(x + \delta_{n,h}^{-1}) - g(x + \delta_{n,h}^{-1})]| \\ &\quad + |\delta_{n,h}[\tilde{g}_{n,h}(x) - g(x)]| + |\delta_{n,h}[g(x + \delta_{n,h}^{-1} - g(x)] - g^{'}(x)| \\ =\ &|\delta_{n,h}[g_{n,h}(x + \delta_{n,h}^{-1}) - g(x + \delta_{n,h}^{-1})]| \\ &\quad + |\delta_{n,h}[g_{n,h}(x) - g(x)]| + \left|\frac{1}{2} g^{''}(x) \delta_{n,h}^{-1} + o(|\delta_{n,h}^{-1}|)\right|,\end{aligned}$$

where $o(\cdot)$ is uniformly in $x \in [0, 1]$. Therefore for all n large enough, $h \sim n^{\frac{2}{3}}$, and $|\delta_{n,h}| \sim \alpha_n^{-\frac{1}{2}}$, we have

$$\begin{aligned}\sup_{x \in [0,1]} |D\tilde{g}_{n,h}(x) - g(x)| &\leq 2|\delta_{n,h}| \sup_{x \in [0,1]} |\tilde{g}_{n,h}(x) - g(x)| + M|\delta_{n,h}^{-1}| \\ &\leq 2c|\delta_{n,h}|\alpha_n + M|\delta_{n,h}^{-1}| \\ &= o\left(\frac{1}{\sqrt{\alpha_n}} \alpha_n\right) + O(\sqrt{\alpha_n}) = O(\sqrt{\alpha_n}) \quad \text{a.s.,}\end{aligned}$$

which implies the desired conclusion. ■

References

1. K. F. Cheng and P. E. Lin, Nonparametric estimation of regression function, *Z. Wahrsch. Verw. Gebiete.* **57** (1981), 223-233.
2. K. F. Cheng and P. E. Lin, Nonparametric estimation of regression function: limiting distribution, *Austral. J. Statist.* **23** (1981), 186–195.
3. T. Gasser and H. G. Müller, Kernel estimation of regression functions, in: Smoothing Techniques for Curve Estimation, T. Gasser and M. Rosenblatt, eds., Lecture Notes in Math., Vol. **757**, Springer-Verlag, Berlin, 1979, pp. 23-68.
4. A. A. Georgiev, Kernel estimates of functions and their derivatives with applications, *Statist. Probab. Lett.* **2** (1984), 45-50.
5. W. Härdle and T. Gasser, Robust nonparametric function fitting, *J. Roy. Statist. Soc. Ser. B* **46** (1984), 42-51.
6. W. Härdle and T. Gasser, On robust kernel estimation of derivatives of regression functions, *Scand. J. Statist.* **12** (1985), 233-240.
7. W. Hoeffding, Probability inequality for sums of bounded random variables, *J. Amer. Statist. Assoc.*, **58** (1963), 13-30.
8. D. B. Holiday, The estimation of derivatives of a nonparametric regression function when data are correlated, *Comm. Statist. Theory Methods* **18** (1989), 2107-2124.
9. P. J. Huber, Robustness in statistics, Proceedings of a Workshop 1978, Robert L. Launer and Graham N. Wilkinson, eds., Academic Press, New York, 1979.
10. M. B. Priestly and M. T. Chao, Nonparametric function fitting, *J. Roy. Statist. Soc. Ser. B* **34** (1972), 385-392.
11. R. J. Serfling, Approximation Theorems of Mathematical Statistics, John Wiley & Sons, New York, 1980.
12. Y. K. Truong, Asymptotic properties of kernel estimators based on local medians, *Ann. Statist.* **17** (1989), 606-617.
13. Y. Yang, Weighted orthogonal series estimation of regression curves, *J. Systems Sci. Math. Sci.* **14** (1994), 337-349. [In Chinese].
14. Y. Yang and Z. G. Zheng, Asymptotic properties for median cross-validated nearest neighbor median estimate in nonparametric regression, *Sci. China Ser. A*, to appear.
15. Z. G. Zheng and Y. Yang, Cross-validation and median criterion, *Statist. Sinica*, to appear.

Department of Applied Mathematics, Tsinghua University
Beijing 100084, China
E-mail: yyang@math.tsinghua.edu.cn