Strong Consistency of BRPA Estimators for Maximizer of Nonparametric Regression Function

Zhu Chunhua[1],Gao Qibing[2]

(1. School of Mathematics and Statistics,Nanjing Audit University,Nanjing 211815,China)
(2. School of Mathematics,Nanjing Normal University,Nanjing 210023,China)

Abstract:The best-r-point-average(BRPA)estimator of the maximizer of regression function has certain merits in application. The strong consistency of the BRPA estimator is obtained under the certain conditions which extends the existing results. The results are illustrated by Monte-Carlo simulations.

Key words:BRPA estimator,nonparametric regression,order statistics,strong consistency
CLC number:O212.1 **Document code**:A **Article ID**:1001-4616(2015)04-0057-04

非参数回归函数最大值点的BRPA估计的强相合性

朱春华[1],高启兵[2]

(1.南京审计学院数学与统计学院,江苏 南京 211815)
(2.南京师范大学数学科学学院,江苏 南京 210023)

[摘要] 非参数函数的BRPA估计在实际中具有一定的应用价值.本文在一定条件下获得BRPA估计的强相合性,其推广现有文献的结果.本文结果通过蒙特卡洛方法验证.

[关键词] BRPA估计,非参数回归,次序统计量,强相合性

Consider a regression model:

$$Y = m(\boldsymbol{X}) + \varepsilon, \quad i = 1, \cdots, n \ , \tag{1}$$

where $m(\cdot)$ is a L-measurable function defined on a bounded convex closed set $D \subset \mathbf{R}^d$,with a unique maximum at $x_0 \in D$,that is,for any $x \in D$ and $x \neq x_0$,

$$m(x_0) \geqslant m(x). \tag{2}$$

Now,let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ be design sequences corresponding to the response variables $Y_1, \cdots, Y_n$, i.e.,

$$Y_i = m(\boldsymbol{X}_i) + \varepsilon_i, i = 1, 2, \ldots, n, \tag{3}$$

where $\varepsilon_1, \cdots, \varepsilon_n$, are i.i.d random errors. Suppose that $\{\boldsymbol{X}_i, i = 1, \cdots, n\}$ and $\{\varepsilon_i, i = 1, \cdots, n\}$ are independent if $\{\boldsymbol{X}_i, i = 1, \cdots, n\}$ are random.

The objective is to determine x_0 based on n observations $(\boldsymbol{X}_1, Y_1), \cdots, (\boldsymbol{X}_n, Y_n)$. The estimation of x_0 by traditional approach is done through the estimation of the regression function. That is,the function $m(\cdot)$ is estimated by some nonparametric method first and then the maximizer of the estimated function $\hat{m}(\cdot)$ is taken as the estimation of x_0 . The regression function can be estimated by various methods such as the kernel,the nearest neighbor,the

Received data:2015-01-18.

Foundation item:Supported by National Science Foundation of China(11271193),Humanities and Social Sciences Planning Foundation of Chinese Ministry of Education (11YJA910004) , Natural Science Foundation of the Jiangsu Higher Education Institutions of China (13KJD110004)and the Returned Overseas Fund of Nanjing Normal University(2014101XLH0194).

Corresponding author:Gao Qibing,associate professor,majored in Modern Regression analysis. E-mail:gaoqibing@njnu.edu.cn

orthogonal series or the smoothing splines, etc.. For details, we refer the reader to, among others, Ibragimov and Khas' minskii[1] and Muller[2].

In recent years another method called the best-r-point average (BRPA) is developed for the estimation of the maximizer of a regression function in a quite different spirit. The BRPA is coined by engineers in field studies. The method can be described as follows. Let $Y_{(1)} \leqslant \cdots \leqslant Y_{(n)}$ be the order statistics of response variables $\{Y_i, i = 1, \cdots, n\}$. Then, x_0 is estimated by $\hat{x}_0(r) = \frac{1}{r}\sum_{i=0}^{r-1}\boldsymbol{X}_{[n-i]} = \frac{1}{r}\sum_{i=0}^{r-1}\boldsymbol{X}_{l_{n-i}}$, with the average of those $\boldsymbol{X}_i's$ corresponding to the r largest order statistics $Y_{(i)}'s$, where $\boldsymbol{X}_{[i]} = X_{l_i}$ denotes the $\boldsymbol{X}_j$ corresponding to $Y_{(i)} = Y_{l_i}$. The estimator $\hat{x}_0(r)$ is called the BRPA estimator. The BRPA method was proposed by Chengchien (1990)[3], then the theoretical justifications of the BRPA method were first provided by Chen et al.[4] in which the weak consistency and certain rates of convergence of the BRPA estimator under some sufficient conditions were established. Since then, Bai et al. (1999[5], 2003[6]) and Wu et.al. (2000[7]) generalized the above results in different aspects. But strong consistency of BRPA is still not established.

In this paper, we will study the strong consistency under certain conditions, which extends the above existing results.

1　Main Results and Its Proof

Before giving main results, we introduce the following notations and conditions:

Let $A \subseteq \mathbf{R}^d$, $\boldsymbol{x} \subseteq \mathbf{R}^d$, c denotes a general constant which may change from one expression to another expression, $[a]$ denotes the largest integer which is less and equals to a and $|A|$ denotes Lebesgue measure of set A, $f(\boldsymbol{x}, A) = \sup\{d(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{y} \in A\}$, where $d(x, y)$ denotes the Euclidean distance between the points $\boldsymbol{x}$ and $\boldsymbol{y}$. For arbitrary $\delta > 0$, write $A(\delta) = \{\boldsymbol{x} : m(\boldsymbol{x}_0) - m(\boldsymbol{x}) \leqslant \delta\}$.

A1　$m(x)$ takes the largest value at the unique point $x_0 \in D \subset \mathbf{R}^d$.

A2　For arbitrary $\delta > 0$, $|A(\delta)| > 0$, and as $\delta \to 0$, $f(x_0, A(\delta)) \to 0$.

B1　When $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are fixed design sequences, they are designed by certain density function $g(x)$ defined on D.

B2　When $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are random sequences, $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are i.i.d. samples with the common density $g(x)$. And $\{\boldsymbol{X}_i, i = 1, \cdots, n\}$ and $\{\varepsilon_i, i = 1, \cdots, n\}$ are independent.

C　$\{\varepsilon_i, i = 1, \cdots, n\}$ are i.i.d. with the distribution function $F(x)$ and the density function $f(x)$ and $\omega(F) = \sup\{x : F(x) < 1\} < \infty$.

Theorem 1　Assume that the conditions A1, A2 and C are satisfied, $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are design sequences which satisfy the condition B1 or B2 and there exists some $c > 0$, for arbitrarily given $\boldsymbol{x} \in D, g(\boldsymbol{x}) > c$. Then, $\hat{x}_0(r) \to x_0$ almost surely.

Remark 1　A1 and A2 are used in Bai et.al. (1999[5], 2003[6]) and Wu et. al. (2000[7]). While B1 and B2 are ordinary assumptions which are used in Wu et. al. (2000[7]). As to C, we remark that practical errors are always finite.

For the proving Theorem 1, we first need the following lemma.

Lemma 1　Let $\varepsilon_i, i = 1, \cdots, n$ be i.i.d. random variables with the distribution $F(x)$ with $\omega(F) < \infty$ and the density $f(x)$, $\varepsilon_{(1)} \leqslant \cdots \leqslant \varepsilon_{(n)}$ are the order statistics, then, $\varepsilon_{(n)} - \varepsilon_{(n-k)}$ converges to 0 completely, i.e., for an arbitrary given $\delta > 0$, $\sum_{n=1}^{\infty} P(\varepsilon_{(n)} - \varepsilon_{(n-k)} > \delta) < \infty$, where $k = [n^{\beta}]$ and $2\beta < 1$.

Proof　Noting that

$$(\varepsilon_{(n)}, \varepsilon_{(n-k)}) \sim \frac{n!}{(n-k-1)!(k-1)!} F^{n-k-1}(x)(F(y) - F(x))^{k-1} f(x) f(y) \mathrm{d}x\mathrm{d}y.$$

It follows that

$$
\begin{aligned}
&P(\varepsilon_{(n)}-\varepsilon_{(n-k)}>\delta)=\int_{y-x>\delta}F^{n-k-1}(x)(F(y)-F(x))^{k-1}f(x)f(y)\mathrm{d}x\mathrm{d}y=\\
&\frac{n!}{(n-k-1)!(k-1)!}\int_{-\infty}^{\infty}[\int_{x+\delta}^{\infty}(F(y)-F(x))^{k-1}\mathrm{d}F(y)]F^{n-k-1}(x)\mathrm{d}F(x)=\\
&\frac{n!}{(n-k-1)!k!}\int_{-\infty}^{\infty}F^{n-k-1}(x)[(1-F(x))^{k}-(F(x+\delta)-F(x))^{k}]\mathrm{d}F(x)=\\
&\frac{n!}{(n-k-1)!k!}\int_{-\infty}^{\omega(F)-\delta}F^{n-k-1}(x)[(1-F(x))^{k}-(F(x+\delta)-F(x))^{k}]\mathrm{d}F(x)\leqslant\\
&\frac{n!}{(n-k-1)!k!}\int_{-\infty}^{\omega(F)-\delta}F^{n-k-1}(x)(1-F(x))^{k}\mathrm{d}F(x)=\\
&\frac{(k+1)(k+2)}{(n+1)(n+2)}\int_{-\infty}^{\omega(F)-\delta}\frac{(n+2)!}{(n-k-1)!(k+2)!}F^{n-k-1}(x)(1-F(x))^{k}\mathrm{d}F(x)\leqslant\\
&\frac{(k+1)(k+2)}{(n+1)(n+2)(1-F(\omega(F)-\delta))^{2}}\int_{0}^{1}\frac{(n+2)!}{(n-k-1)!(k+2)!}x^{n-k-1}(1-x)^{k+3-1}\mathrm{d}x=\\
&\frac{(k+1)(k+2)}{(n+1)(n+2)(1-F(\omega(F)-\delta))^{2}}\leqslant c\frac{1}{n^{(2-2\beta)}},
\end{aligned}
\tag{4}
$$

where the last inequality is satisfied with enough large n. Noting that $2-2\beta>1$, thus, for an arbitrary given $\delta>0$,

$$\sum_{n=1}^{\infty}P(\varepsilon_{(n)}-\varepsilon_{(n-k)}>\delta)<\infty. \tag{5}$$

This completes the proof of the Lemma 1.

Proof of Theorem 1

Write $k=[n^{\beta}]>r$ where $0\leqslant\beta<\frac{1}{2}$ and

$$L_{nk}=\{\varepsilon_{(n)}-\varepsilon_{(n-k)}>\frac{\delta}{2}\},\quad Q_{nk}=\{\#\{i\leqslant k,\boldsymbol{X}_{l_{n-i}}\notin A(\delta)\}<r\},$$

where $\#A$ denotes the cardinal number of the set A.

When the event $L_{nk}^{T}\bigcap Q_{nk}^{T}$ occurs, there exists at least r integer $i\leqslant k$ such that $X_{l_{n-i}}\in A(\delta)$, where $L_{nk}^{T}\bigcap Q_{nk}^{T}$ denotes the complementary event of event L_{nk} and Q_{nk} respectively.

Therefore, for all $X_j\notin A(2\delta)$, it follows that

$$Y_{l_{n-i}}=m(\boldsymbol{x}_{l_{n-i}})+\varepsilon_{(n-i)}\geqslant m(x_0)-\delta+\varepsilon_{(n)}-\frac{\delta}{2}\geqslant m(\boldsymbol{X}_j)+\varepsilon_j=Y_j\ ,$$

which implies

$$P(\hat{x}_0(r)\notin A(2\delta))\leqslant P(L_{nk})+P(Q_{nk}). \tag{7}$$

Hence,

$$\sum_{n=1}^{\infty}P(\hat{x}_0(r)\notin A(2\delta))\leqslant\sum_{n=1}^{\infty}P(L_{nk})+\sum_{n=1}^{\infty}P(Q_{nk}). \tag{8}$$

By Lemma 1, the first term is finite. Next, we move to estimate the second term.

Clearly,

$$P(Q_{nk})=P(B(k,|A(\delta)|)<r),$$

in which $B(n,p)$ denotes the binomial distribution with parameters n and p. So by Bernstein inequality,

$$\sum_{n=1}^{\infty}P(Q_{nk})\leqslant\sum\exp(-\frac{3}{8}[n^{\beta}]|A(\delta)|(|A(\delta)|-\frac{r}{[n^{\beta}]})^{2})\leqslant\sum_{n=1}^{\infty}\exp(-cn^{\beta})<\infty, \tag{9}$$

where $c>0$ which does not depend on n.

By the condition A1 and A2, which combined with Borel–Cantelli Lemma, Theorem 1 can be obtained easily. This completes the proof of Theorem 1.

2 Simulations

In this section, we perform some simulations to illustrate the asymptotic theory developed in Section 2 by Monte-Carlo methods. In our simulation study, we consider regression functions $f(x)=\exp(-(x-1)^{2})$ and error distribution $U[-1,1]$ and sample sizes n is taken to be 100 and 400, respectively. Each of the samples is generated as $y_i = f(x_i) + \varepsilon_i, i = 1,\ldots,n$, where the x_i's are random numbers generated from the standard normal distribution $N(0,1)$ and the x_i's are random numbers generated from the error distribution $U[-1,1]$ by using Matlab. N=1 000 repetitions are done for each sample size.

For each n, the mean and the averaged absolute deviation of each estimator with $r = 1,5$ are computed respectively as $\frac{1}{N}\sum_{k=1}^{N}\hat{x}_0^k(r)$ and $\frac{1}{N}\sum_{k=1}^{N}|\hat{x}_0^k(r) - x_0|$ in N=1 000 repetitions are given in Tables 1, which shows that the means of BRPA estimators are close to the true maximizer value x_0. As n became larger, the bias of parameter estimators and corresponding averaged absolute deviation becomes smaller.

Table 1 The mean and the averaged absolute deviation of BRPA estimator in N=1 000 repetitions. The values in parentheses are the averaged absolute deviation.

	$r = 1$	$r = 5$
$n = 100$	1.004 0(0.148 0)	1.005 0(0.120 0)
$n = 400$	0.999 0(0.095 0)	1.004 0(0.076 0)

Acknowledgement: We thank the editor, an associate editor and two referees for their valuable suggestions which have greatly improved this paper.

[参考文献]

[1] IBRAGIMOV I A, KHAS'MINSKII R Z. Estimates of the signal, its derivatives and point of maximum for Gaussian distributions. Theory probab[J]. Appl, 1980, 25: 703-720.

[2] MULLER H G. Adaptive nonparametric peak estimation[J]. Ann Statist, 1989, 17: 1 053-1 069.

[3] CHANGCHIEN G M. Optimization of blast furnace burden distribution[C]//CHAO M T, CHENG P E. Proceedings of the 1990 Taipei symposium in statistics, Taipei, 1990, 63-78.

[4] CHEN H, HUANG M N L, HUANG W J. Estimation of the location of the maximum of a regression function using extreme order statistics[J]. J Multivariate Anal, 1996, 57: 191-214.

[5] BAI Z D, HUANG M N L. On the consistency of the best-r-points-average estimator for the maximizer of a nonparametric regression function[J]. Sankhy_a, 1999, 61: 208-217.

[6] BAI Z D, CHENG Z H, WU Y H. Convergence rate of the best-r-point-average estimator for the maximizer of a nonparametric regression function[J]. J Multivariate Anal, 2003, 84; 319-334.

[7] WU Y H, WANG X M. The consistency of BRPA estimation of the maximizer of a multivariate regression function[J]. Chin J of Appl Probab and Stati, 2000, 16: 299-302.

[责任编辑:顾晓天]