Asymptotic theory of nonparametric regression estimates with censored data

SHI Peide (施沛德), WANG Haiyan (王海燕) & ZHANG Lihua (张利华)

Department of Probability and Statistics, Peking University, Beijing 100871, China
Correspondence should be addressed to Shi Peide

Received May 19, 1999

Abstract For regression analysis, some useful information may have been lost when the responses are right censored. To estimate nonparametric functions, several estimates based on censored data have been proposed and their consistency and convergence rates have been studied in literature, but the optimal rates of global convergence have not been obtained yet. Because of the possible information loss, one may think that it is impossible for an estimate based on censored data to achieve the optimal rates of global convergence for nonparametric regression, which were established by Stone based on complete data. This paper constructs a regression spline estimate of a general nonparametric regression function based on right-censored response data, and proves, under some regularity conditions, that this estimate achieves the optimal rates of global convergence for nonparametric regression. Since the parameters for the nonparametric regression estimate have to be chosen based on a data driven criterion, we also obtain the asymptotic optimality of AIC, AICC, GCV, C_p and FPE criteria in the process of selecting the parameters.

Keywords: nonparametric regression, censored data, regression spline, optimal rates of convergence.

Consider the following model:

$$Y = h(X) + \varepsilon, \tag{1}$$

where Y is a response, $h(\cdot)$ is a smooth function, X is a J-dimensional explanatory variable, and ε is a random error independent of X. Suppose that $\{(X_i, Y_i), 1 \leqslant i \leqslant n\}$ is an i.i.d. sample from model (1), where $\varepsilon_1, \cdots, \varepsilon_n$ are i.i.d. random errors with mean zero and finite variance. When the Y_i's are lifetime data with common distribution F, not all of the responses can be available. Specifically, one may observe (X_i, δ_i, z_i) with

$$z_i = \min\{Y_i, C_i\}, \qquad \delta_i = I(Y_i \leqslant C_i)$$

instead of (X_i, Y_i) itself, where the C_i's are independent and identically distributed as some undnown censoring distributions G, Y_i and $C \setminus - i$ are independent, and δ_i is an indicator of the status whether Y_i has been observed or not. This kind of data is widely used in practice. For example, the survival data in clinical trails or failure time data in reliability studies are often subject to such censoring[1].

For the case of $J = 1$, the pointwise consistency of the weighted estimate of $h(\cdot)$ has been studied by several authors[2,3]. The asymptotics of k-nearest neighbor estimate of h has been investigated by Qin[4]. The rate of convergence of the estimate for h has been obtained by ref. [5]. Unfortunately, all rates of the convergence of the estimates for h in literature are not optimal rates of global convergence according to Stone[6]. Since some information may have been lost when constructing the estimates of h based on the censored data, one may ask the following question: Is it possible for an estimate of h constructed with censored data to achieve the optimal rates of convergence?

In this paper, we project the unknown regression function h into a space of B-spline functions and obtain the estimate of h by minimizing the least squares criterion over the spline space based on right-censored response data. Although our estimate of h is constructed with the censored data, we prove, under some regularity conditions, that it achieves the optimal rates of global convergence for nonparametric regression established by ref. [6] based on complete data. In general, a small dimension of the spline space is enough to get a good approximation of h. Once the spline knots (the parameters of the spline space) are given, the estimate can be obtained easily based on the least square principle. The spline knots can be determined with a forward/backward algorithm which minimizes a model selection criterion (see refs. [7—9] for details).

As we have mentioned above, to obtain a nonparametric regression estimate, one has to determine the parameters for the estimate of h, such as the bandwidth for kernels, the penalty parameter for smoothing splines, the spline knots for regression splines, etc. The parameters can be chosen based on data via a criterion such as the generalized cross-validation (GCV), Akaike information criterion (AIC) and so on. Let $h_{\hat{\theta}}(\cdot)$ be the regression spline estimate of $h(\cdot)$ constructed based on censored data. Theoretically, the best criterion for choosing the parameters λ for $h_{\hat{\theta}}$ should be $U_n(\lambda) = n^{-1}\sum_{i=1}^{n}(h_{\hat{\theta}}(X_i) - h(X_i))^2$ since it measures the fidelity of the estimate constructed by applying censored data to the true function directly. Unfortunately, the true function is unknown. Therefore, we cannot use it to choose the spline knot sets in practice, but a variety of model selection criteria can be used to select the spline knot sets such as AIC, AICC, GCV, C_p and FPE criteria[10]. Let $\hat{\lambda}$ be the spline knot set obtained by minimizing AIC. It is of much interest to know the asymptotic optimality of AIC[11]; that is, $\hat{\lambda}$ satisfies

$$U_n(\hat{\lambda})/\inf_{\lambda} U_n(\lambda) - 1 = o_P(1). \tag{2}$$

Under some conditions, we show that (2) is true for AIC, AICC, GCV, C_p and FPE criteria.

1 Main results

Let

$$Z_{iG} = \frac{z_i\delta_i}{1 - G(z_i)}, \quad e_i = Z_{iG} - h(X_i), \quad Z_{i\hat{G}} = \frac{z_i\delta_i}{1 - \hat{G}(z_i)}$$

for $i = 1, \cdots, n$. Then, for given G, $e_1, \cdots, e_n$ are i.i.d. random variables

$$E(Z_{iG} \mid X_i) = E(Y_i \mid X_i) = h(X_i), \quad \text{and} \quad 0 < \mathrm{Var}(\varepsilon_1) \leqslant \sigma^2 < \infty,$$

where $\sigma^2 = \mathrm{Var}(e_1)$. Therefore,

$$Z_{iG} = h(X_i) + e_i, \quad 1 \leqslant i \leqslant n. \tag{3}$$

Based on (3), for given G, several estimates of $h(\cdot)$ can be obtained, such as smoothing splines, regression splines and kernels. For convenience, we consider regression spline estimate of $h(\cdot)$. To define our estimate formally, we shall introduce the spline basis functions as follows. We basically follow the method of Schumaker[8,12] by using B-spline function $h_\theta(x)$ to approximate the unknown smooth function $h(x)$ in (3). Without loss of generality, we assume that the regressor X is in $[0, 1]^J$. In general cases, the spline approximation can be applied to a rescaled variable.

Let $m \geqslant 0$, $k_{jn} > 0$, $j = 1, \cdots, J$, be integers. Given partitions $0 = t_{j0} < t_{j1} \cdots < t_{jk_{jn}} = 1$ of $[0, 1]$, we denote by $\pi_{ji}(x)(i = 1, 2, \cdots, k_{jn} + m)$ the normalized B-splines (of oreder $m + 1$) associated with an extended partition of $[0, 1]$ determined by $\{t_{ji}\}$, where $\max_k(t_{jk} - t_{jk-1})/\min_k(t_{jk} -$

$t_{jk-1}) \leqslant \alpha_0$ uniformly in j and n for some constant α_0. The details can be seen in ref. [13]. The spline knot t_{ji} is placed on the i/k_{jn}th quantile of the components of the observed regressors as in ref. [7], where a knot placement and deletion procedure was proposed.

Let $B_{i_1,\cdots,i_J} = B_{i_1,\cdots,i_J}(x_1,\cdots,x_J) = \prod_{j=1}^{J} \pi_{j,i_j}(x_j)$ and $\pi(x) = \pi(x_1,\cdots,x_J) = (B_{1,\cdots,1}\cdots, B_{p_1,1,\cdots,1},\cdots, B_{p_1,\cdots,p_{J-1},1},\cdots, B_{p_1,\cdots,p_{J-1},p_J})'$. Then $h_\theta(x) = \pi(x)^{\mathrm{T}}\theta$ and θ is an unknown projection parameter of $h(x)$.

For a J-tuple of nonnegative integers u, define $[u] = \sum_{j=1}^{J} u_j$ and let D^u denote the partially differential operator given by $D^u = \dfrac{\partial^{[u]}}{\partial x_1^{u_1}\cdots\partial x_J^{u_J}}$. Define $\mathscr{H}_r$ as the collection of all functions on $[0,1]^J$ such that the mth order partial derivative satisfies a Hölder condition of order γ with $r = m + \gamma > J/2$. That is, there exists a constant $W_0 \in (0,\infty)$ such that for each $h \in \mathscr{H}_r$

$$|D^u h(x) - D^u h(x^*)| \leqslant W_0 |x - x^*|^\gamma \quad \text{for all } x, x^* \in [0,1]^J \text{ and } [u] = m.$$

If the density function of X on $[0,1]^J$ is positive, any function of $\mathscr{H}_r$ can be approximated uniformly by its projection in the B-spline space in the order $O(k_n^{-r})$ so that the unknown regression function is well reexpressed by a spline function only if k_n increases with n, where $k_n = \min_j\{k_{jn}\}$. In the present paper, as in ref. [9] the Akaike information criterion may be adopted to select the number of knots for fitting the model $Z_{iG} = h_\theta(X_i) + e_i$ for $1 \leqslant i \leqslant n$ by minimizing

$$\sum_{i=1}^{n}(Z_{iG} - h_\theta(X_i))^2. \tag{4}$$

Usually, a small dimension of spline space is enough to get a good approximation.

In practice, however, G is generally unknown. Based on these randomly censored data, $(z_i, \delta_1),\cdots,(z_n,\delta_n)$, the nonparametric maximum likelihood estimator of G, is the product-limit estimator $\hat{G}$ (ref. [14]), defined as

$$\hat{G}_n(t) = \begin{cases} 1 - \prod\limits_{i:\delta_i=0 \text{ and } z_i\leqslant t}^{n}(1 - 1/N_n(z_i)) & \text{for } \delta_{(n)} = 0 \text{ and } t < z_{(n)}, \\ 1 & \text{for } \delta_{(n)} = 0 \text{ and } t \geqslant z_{(n)}, \\ \text{undefined} & \text{otherwise,} \end{cases}$$

where $z_{(1)},\cdots,z_{(n)}$ are the order statistics of $z_1,\cdots,z_n$, $\delta_{(1)},\cdots\delta_{(n)}$ are the indicators associated with $z_{(1)},\cdots,z_{(n)}$, respectively, and $N_n(t) = \sum_{i=1}^{n} I(z_i \geqslant t)$. With the estimate of G, the objective function in (4) can be modified as follows:

$$\sum_{i=1}^{n}(Z_{i\hat{G}} - h_\theta(X_i))^2. \tag{5}$$

Our estimate is defined by $h_{\hat{\theta}}(x) = \pi(x)'\hat{\theta}$, where $\hat{\theta}$ minimizes (5). The goal of this paper is to show that $h_{\hat{\theta}}$, the regression spline estimate of h, achieves the optimal rates of global convergence for nonparametric regression.

For given positive numbers a_n and b_n, $n \geqslant 1$, the notation $a_n \sim b_n$ means that a_n/b_n is bounded away from zero and infinity. Let Q be the distribution function of z_1 and $\tau = \inf\{t : Q(t) = 1\}$. Three assumptions are needed to introduce our results.

A1. The distribution of X is absolutely continuous with density function f. Furthermore, there exist two positive constants b_1 and b_2 such that $b_1 \leqslant f(x) \leqslant b_2$ for all $x \in [0,1]^J$.

A2. $h \in \mathscr{H}_r$.

A3. Both F and G are continuous distributions, and $\int_{-\infty}^{\tau}(1 - F(s-))^{-1} d\,G(s) < \infty$.

In addition, assume that $\sup_{x\in[0,1]^J} E\left[\frac{Y_1^2}{(1-G(Y_1))^3} \mid X_1 = x\right] < +\infty$ when $G(\tau)=1$, and that $E(Y_1^2) < \infty$ when $G(\tau) < 1$.

Theorem 1. Suppose that $k_{jn} \sim n^{1/(2r+J)}$ and Assumptions (A1)—(A3) are satisfied. Then the estimator $h_{\hat\theta}(x)$ minimizing (5) satisfies

$$n^{-1}\sum_{i=1}^{n}(h_{\hat\theta}(X_i) - h(X_i))^2 = O_P(n^{-2r/(2r+J)}).$$

Define $\| g \|_{\mathscr{S}}^2 = \int_{[0,1]^J} g^2(x) f(x) \mathrm{d}x$ whenever such integrals exist.

Theorem 2. Under the conditions of Theorem 1, we have

$$\| D^u h_{\hat\theta} - D^u h \|_{\mathscr{S}}^2 = O_p(n^{-(r-[u])/(2r+J)}), \quad 0 \leqslant [u] \leqslant m.$$

The rates of convergence for the estimate $h_{\hat\theta}$ based on censored data are the same as those of optimal global convergence rates established in ref. [6] based on complete data. The following theorem indicates that AIC criterion for selecting the regression parameters is asymptotically optimal. The presentation of Theorem 3 needs a useful result which is stated here for convenience.

Lemma 1[13]. For each $h \in \mathscr{H}_r$, there exists a constant $W_1 > 0$ depending only on m, J and W_1 such that

$$\begin{cases} h(x) = \pi(x)^{\tau}\theta^*(h) + R_{nx}, \\ \sup_{x\in[0,1]^J} | R_{nx} | \leqslant W_1 k_n^{-r}, \end{cases}$$

where $\theta^*(h)$ is a vector depending on h. Let $R_n = (R_{n1}, \cdots, R_{nn})'$ and $R_{ni} = R_{nX_i}$. Let Λ be a collection of possible spline knot sets, $N = N(\lambda)$ be the dimension of $\pi(x)$ for $\lambda \in \Lambda$, and $L_n(\lambda) = N(\lambda)\sigma^2/n + R'_n(I-H)R_n/n$. Then we have

Theorem 3. Suppose that Assumption (A1)—(A3) are satisfied, $\hat\lambda$ is the spline knot set obtained by minmizing AIC, $\inf_{\lambda\in\Lambda} nL_n(\lambda) \to \infty$ in probability, $\min_j k_{jn} \to \infty$ and $N/n \to 0$ as $n\to\infty$, and there exists some $s > 1$ such that $\sum_{\lambda\in\Lambda}\frac{1}{N^s(\lambda)} < \infty$ and $E(e_1^{4s}) < \infty$. Then (2) is true.

By definition, U_n and L_n are asymptotically equivalent (see the proof of Theorem 3). An assumption similar to $\inf_{\lambda\in\Lambda} nL_n(\lambda) \to \infty$ was assumed by Li[11], where $X_1, \cdots, X_n$ are fixed regressors. It is obvious that the argument used to prove (2) for AIC criterion is also applicable to AICC, GCV, C_p, and FPE criteria. Consequently, we have

Corollary. Under the conditions of Theorem 3, (2) remains true if the AIC criterion is replaced by AICC, GCV, C_p, or FPE.

2 Proof of Theorems 1 and 2

Theorem 2 follows from Theorem 1 and similar arguments in Lemmas 3.5 and 3.6 of ref. [13]. The details are therefore omitted here.

For simplicity, we assume uniform partitions $t_i = t_{ji} = i/k_n$, $i = 1, \cdots, k_n$ and $k_n = k_{1n} = \cdots, k_{Jn}$. Only nonessential modifications are needed to deal with the general quasi-uniform partitions described earlier.

Since $k_n \sim n^{1/(2r+J)}$ and $h(x) = \pi(x)'\theta^*(h) + R_{nx}$ with $\sup_x | R_{nx} | \leqslant k_n^{-r}$, to prove Theorem 1 we need only to show

$$n^{-1}\sum_{i=1}^{n}[\pi(X_i)'(\hat\theta - \theta^*(h))]^2 = O_p(n^{-2r/(2r+J)}). \quad (6)$$

By definition, $\sum_{i=1}^{n}\pi(X_i)(Z_{i\hat{G}} - \pi(X_i)'\hat{\theta}) = 0$ and

$$\begin{aligned}\hat{\theta} &= (\Pi'\Pi)^{-1}\Pi' Z_{\hat{G}} = (\Pi'\Pi)^{-1}(Z_{\hat{G}} - Z_G + Z_G)\\ &= (\Pi'\Pi)^{-1}\Pi'(Z_{\hat{G}} - Z_G) + \theta^*(h) + (\Pi'\Pi)^{-1}\Pi'(e + R_n),\end{aligned}$$

where $e = (e_1, \cdots, e_n)'$, $Z_{\hat{G}} = (Z_{1\hat{G}}, \cdots, Z_{n\hat{G}})'$, $Z_G = (Z_{1G}, \cdots, Z_{nG})'$, and $\Pi = (\pi(X_1), \cdots, \pi(X_n))'$. Thus we have

$$\begin{aligned}&\sum_{i=1}^{n}[\pi(X_i)'(\hat{\theta} - \theta^*(h))]^2 = (\hat{\theta} - \theta^*(h))'\Pi'\Pi(\hat{\theta} - \theta^*(h))\\ &= (Z_{\hat{G}} - Z_G)'\Pi(\Pi'\Pi)^{-1}\Pi'(Z_{\hat{G}} - Z_G) + (e + R_n)'\Pi(\Pi'\Pi)^{-1}(e + R_n)\\ &\quad + 2(e + R_n)'\Pi(\Pi'\Pi)^{-1}\Pi'(Z_{\hat{G}} - Z_G).\end{aligned}$$

Considering that $\Pi(\Pi'\Pi)^{-1}\Pi'$ is idempotent and with probability one, the smallest eigenvalue η_n of $\frac{k_n}{n}\Pi'\Pi$ satisfies

$$\eta_n \geqslant \eta > 0 \tag{7}$$

for some positive constant η when n is large enough (see Lemma 3.2 of ref. [13]). By Chebyshev inequality and taking conditioning expectation of $e'\Pi(\Pi'\Pi)^{-1}\Pi' e$ on $X_1, \cdots, X_n$, we can easily show

$$n^{-1}e'\Pi(\Pi'\Pi)^{-1}\Pi' e = O_P(k_n^J/n).$$

Note that $k_n \sim n^{1/(2r+J)}$. From the above results, to show (6) it suffices to verify the following lemma.

Lemma 2. Under the assumptions of Theorem 1,

$$n^{-1}(Z_{\hat{G}} - Z_G)'\Pi(\Pi'\Pi)^{-1}\Pi'(Z_{\hat{G}} - Z_G) = O_P(n^{-1}). \tag{8}$$

From refs. [15, 16], we conclude that $\sup_{t\leqslant \tau}\sqrt{n}\,|\hat{G}(t) - G(t\Lambda z_{(n)})|$ converges in distribution. Lemma 2 follows this result, (7), and Theorem 2.2 of ref. [17].

3 Proof of Theorem 3

For an n-vector B, let $\| B \|^2 = \sum_{i=1}^{n} B_i^2$. Let $ASR(\lambda) = \| Z_{\hat{G}} - \Pi\hat{\theta} \|^2/n$. Then

$$\mathrm{ASR}(\lambda) = U_n(\lambda) + \frac{1}{n}e'e - \frac{2}{n}e'He - \frac{2}{n}e'H(Z_{\hat{G}} - Z_G) - \frac{2}{n}e'(I - H)R_n,$$

where $H = \Pi(\Pi'\Pi)^{-1}\Pi'$. Note that AIC($\lambda$) has the following equivalent form:

$$\mathrm{AIC}(\lambda) = \mathrm{ASR}(\lambda)\exp(2N/n).$$

We have

$$\begin{aligned}\mathrm{AIC}(\lambda) &= U_n(\lambda) + e'e/n - 2e'He/n + 2tr(H)\sigma^2/n\\ &\quad - \frac{2}{n}e'(I - H)R_n + \frac{2N}{n}(\mathrm{ASR}(\lambda) - \sigma^2) - \frac{2}{n}e'H(Z_{\hat{G}} - Z_G) + V_n,\end{aligned}$$

where $V_n = \mathrm{ASR}(\lambda)\{\exp(2N/n) - 1 - 2N/n\}$. Since $e'e/n$ is independent of λ, Theorem 3 follows if we can show that in probability

$$\lim_{n\to\infty}\sup_{\lambda\in\Lambda} U_n(\lambda)/L_n(\lambda) = 1, \tag{9}$$

$$\lim_{n\to\infty}\sup_{\lambda\in\Lambda}\frac{2}{n}\,|e'He - \mathrm{tr}(H)\sigma^2|\,/L_n(\lambda) = 0, \tag{10}$$

$$\lim_{n\to\infty}\sup_{\lambda\in\Lambda}\frac{2}{n}\,|e'(I - H)R_n|\,/L_n(\lambda) = 0, \tag{11}$$

$$\lim_{n\to\infty}\sup_{\lambda\in\Lambda}\frac{2N}{n}\,|\mathrm{ASR}(\lambda) - \sigma^2|\,/L_n(\lambda) = 0, \tag{12}$$

$$\lim_{n\to\infty}\sup_{\lambda\in\Lambda}\frac{2}{n}\mid e'H(Z_{\hat{G}}-Z_G)\mid/L_n(\lambda)=0, \tag{13}$$

$$\lim_{n\to\infty}\sup_{\lambda\in\Lambda}\mid V_n\mid/L_n(\lambda)=0. \tag{14}$$

Since $nL_n(\lambda)\geqslant \mathrm{tr}(H)\sigma^2=N(\lambda\sigma^2)$ almost surely and $\sup_{\lambda\in\Lambda}\mathrm{ASR}(\lambda)=O_P(1)$, it is easy to show that (14) holds.

For (12),

$$\frac{N}{n}(\mathrm{ASR}(\lambda)-\sigma^2)=\frac{N}{n^2}\left(\frac{e'(I-2H)e}{n^2}-\sigma^2\right)+\frac{N}{n}U_n(\lambda)$$
$$-\frac{2N}{n}e'H(Z_{\hat{G}}-Z_G)-\frac{2N}{n}e'(I-H)R_n. \tag{15}$$

Then, a simple calculation gives

$$U_n(\lambda)=\frac{e'He}{n}+\frac{1}{n}(Z_{\hat{G}}-Z_G)'H(Z_{\hat{G}}-Z_G)+\frac{R'_n(I-H)R_n}{n}+\frac{2}{n}e'H(Z_{\hat{G}}-Z_G). \tag{16}$$

Therefore,

$$\mid U_n(\lambda)/L_n(\lambda)-1\mid=\mid U_n(\lambda)-L_n(\lambda)\mid/L_n(\lambda)=2\mid e'H(Z_{\hat{G}}-Z_G)\mid/[nL_n(\lambda)]$$
$$+\mid e'He-\sigma^2\mathrm{tr}(H)\mid/[nL_n(\lambda)]+[(Z_{\hat{G}}-Z_G)'H(Z_{\hat{G}}-Z_G)]/[nL_n(\lambda)].$$

Moreover, $\dfrac{(Z_{\hat{G}}-Z_G)'H(Z_{\hat{G}}-Z_G)}{nL_n(\lambda)}\leqslant\dfrac{(Z_{\hat{G}}-Z_G)'(Z_{\hat{G}}-Z_G)}{nL_n(\lambda)}=O_P\left(\dfrac{1}{nL_n(\lambda)}\right)$. Hence we have

$$\sup_{\lambda\in\Lambda}\frac{(Z_{\hat{G}}-Z_G)'H(Z_{\hat{G}}-Z_G)}{nL_n(\lambda)}\leqslant O_P\left(\frac{1}{\inf_{\lambda\in\Lambda}nL_n(\lambda)}\right)\to 0 \tag{17}$$

in probability as n tends to infinity. Combining (15)—(17), to show Theorem 3 we need only to show (10), (11) and (13).

First, for (10), given any $\delta>0$, from Chebychev's inequality and Theorem 2 of ref. [18] we obtain

$$\mathscr{P}\{\mid e'He-\mathrm{tr}(H)\sigma^2\mid>\delta\mid X_1,\cdots,X_n\}\leqslant D(s)E(e_1^{4s})(\mathrm{tr}(H^2))^s/\delta^{2s}\leqslant D_3(s)(nL_n(\lambda))^s,$$

where $D_3(s)=D(s)E(e_1^{4s})/[\delta^{2s}\sigma^{2s}]$. Therefore

$$\mathscr{P}\left\{\sup_{\lambda\in\Lambda}\frac{\mid e'He-\mathrm{tr}(H)\sigma^2\mid}{nL_n(\lambda)}>\delta\mid X_1,\cdots,X_n\right\}\leqslant\sum_{\lambda\in\Lambda}\frac{D_3(s)(nL_n(\lambda))^s}{n^{2s}L_n^{2s}(\lambda)}=\sum_{\lambda\in\Lambda}\frac{D_3(s)}{(nL_n(\lambda))^s}.$$

Inequality (7), together with the following two inequalities,

$$nL_n(\lambda)\geqslant\sigma^2\mathrm{tr}(H)\quad\text{and}\quad\sum_{\lambda\in\Lambda}\frac{1}{N^s(\lambda)}<\infty$$

implies that for any given $\zeta>0$ there is a subset of Λ denoted by Λ^* such that it contains only finite number of knot sets, and $\sum_{\lambda\in\Lambda\setminus\Lambda^*}\dfrac{D_3(s)}{N^s(\lambda)}<\zeta/2$. Consequently,

$$\sum_{\lambda\in\Lambda}\frac{D_3(s)}{(nL_n(\lambda))^s}\leqslant\sum_{\lambda\in\Lambda\setminus\Lambda^*}\frac{D_3(s)}{N^s(\lambda)}+D_3(s)\sum_{\lambda\in\Lambda^*}\frac{1}{(nL_n(\lambda))^s}\leqslant\zeta/2+D_3(s)\sum_{\lambda\in\Lambda^*}\frac{1}{(nL_n(\lambda))^s}$$

almost surely and the latter tends to 0 as n tends to infinity.

Then, for (11), with $E(e_1)^{4s}<\infty$ and Holder inequality, we can find a constant $C>0$ such that

$$n^{-2s}E(\mid e'(I-H)R_n\mid^{2s}\mid X_1,\cdots,X_n)\leqslant Cn^{-2s}(R'_n(I-H)R_n)^2\leqslant Cn^{-s}L_n(\lambda)^s.$$

From the last inequality and an argument similar to that used in the proof of (10), we have

$$\mathscr{P}\left\{\sup_{\lambda\in\Lambda}\left|\frac{e'(I-H)R_n}{nL_n(\lambda)}\right|>\delta\mid X_1,\cdots,X_n\right\}\leqslant\sum_{\lambda\in\Lambda}\frac{Cn^{-s}L_n(\lambda)^2n^{2s}}{[\delta nL_n(\lambda)]^{2s}}=\sum_{\lambda\in\Lambda}\frac{C}{\delta^{2s}[nL_n(\lambda)]^s},$$

which together with $nL_n(\lambda) \geqslant \sigma^2 N(\lambda)$ implies (11).

Finally, for (13), from (10) and Lemma 2, we can conclude

$$\left| \frac{1}{n} e' H(Z_{\hat{G}} - Z_G) \right| / L_n(\lambda) \leqslant \frac{O_P(1)(nL_n(\lambda))^{1/2}}{nL_n(\lambda)}.$$

Note that the convergence rate for $\frac{1}{n}\sum_{i=1}^{n}(Z_{i\hat{G}} - Z_{iG})^2$ given in Lemma 2 is independent of λ and so is the quantity $O_P(1)$. Thus, (13) follows from the last inequality and $\inf_{\lambda \in \Lambda} nL_n(\lambda) \to \infty$ in probability.

Acknowledgements This work was supported in part by the Postdoctoral Science Foundation and the National Natural Science Foundation of China.

References

1. Kalbfleisch, J. G., Prentice, R. L., The Statistical Analysis of Failure Time Data, New York: John Wiley, 1980.
2. Zheng, Z. K., Strong consistency of nonparametric regression estimates with censored data, in Proceedings of Sino-American Statsitical Conference, Beijing 1987, 627—630.
3. Wang, Q. H., Zheng, Z. G., Asymptotics of the estimates for semiparametric regression model with random censored data, Science in China (in Chinese), Series A, 1997, 27: 583.
4. Qin, G. S., The functional central limit theorem of k-nearest neighbor estimates, Acta Mathematicae Applicatae Sinica (in Chinese), 1995, 18: 559.
5. Qin, G. S., Cai, L., Estimation for the asymptotic variance of parametric estimates in partially linear model with censored data, Acta Mathematica Scientia, 1996, 16: 192.
6. Stone, C., Optimal global rates of convergence for nonparametric regression, Ann. Statist., 1982, 10: 1040.
7. Friedman, J. H., Silverman, B. W., Flexible parsimoneous smoothing and additive modeling (with discussion), Technometrics, 1989, 31: 3.
8. Shi, P. D., Li, G. Y., Global rates of convergence of B-spline M-estimates for nonparametric regression, Science in China, Ser. A, 1995, 38(3): 303.
9. He, X. M., Shi, P. D., Bivariate tensor-product splines in a partly linear model, Multivariate Analysis, 1996, 58: 162.
10. Shi, P. D., Tsai, C. L., A note on the unification of the Akaike information criterion, J. Royal Statist. Soc. (B), 1998, 60: 551.
11. Li, K. C., Asymptotic optimality for C_P, C_L, cross-validation and generalized cross-validation: discrete index set, Ann. Statist., 1987, 15, 958.
12. Schumaker, L. L., Spline Functions, New York: John Wiley, 1981.
13. Shi, P. D., Zheng, Z. G., Robust estimates in multivariate nonparametric regression via least absolute deviations, Acta Mathematica Scientia (Supp.), 1996, 16: 57.
14. Kaplan, E. L., Meier, R., Nonparametric estimation from incomplete observations, J. Amer. Statist. Assoc., 1958, 53: 457.
15. Gill, R. D., Large sample behaviour of the product-limit estimator on the whole line, Ann. Statist., 1983, 11: 49.
16. Gu, M. G., Lai, Z. L., Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation, Ann. Prob., 1990, 18: 160.
17. Zhou, M., Some properties of the Kaplan-Meier estimator for independent, nonidentically distributed random variables, Ann. Statist., 1991, 19: 2266.
18. Whittle, P., Bounds for the moments of linear and quadratic forms in independent variables, Theory Probab. Appl., 1960, 5: 302.