J. Math. Res. & Exposition, 1988, No. 2

# Strong Consistency of Non-parametric Regression Estimates with Censored Data*

*Zheng Zukang* (郑祖康)
(Fudan University)

**Abstract**

Let $(X, Y)$ be an $R^d \times R$ valued random vector with $E|Y| < \infty$ and $(X_1, Y_1)$ $(X_2, Y_2)$, …, $(X_n, Y_n)$ be i.i.d. observations of $(X, Y)$. To estimate the regres sion function $m(x) = E(Y|X=x)$, Stone [1] suggested

$$m_n(x) = \sum_{i=1}^{n} W_{ni}(x) Y_i ,$$

where $W_{ni}(x) = W_{ni}(x, X_1, X_2, \cdots, X_n)$ $(i = 1, 2, \cdots, n)$ are weight functions. Dev- roye [2] and Chen Xiru [3] established the strong consistency of $m_n(x)$.

In this paper, we discuss the case that $\{Y_i\}$ are censored by $\{t_i\}$, where $\{t_i\}$ are i.i.d. random variables and also independent of $\{Y_i\}$. Under certain conditions we still obtain the strong consistency of $m_n(x)$.

Let $(X, Y)$ be an $R^d \times R$ valued random vector with $E|Y| < \infty$. Denote the reg ression function by $m(x) = E(Y|X=x)$ and let $(X_1, Y_1)$, $(X_2, Y_2)$, …, $(X_n, Y_n)$ be i.i.d. observations of $(X, Y)$. To estimate $m(x)$, Stone [1] suggested the following form

$$m_n(x) = \sum_{i=1}^{n} W_{ni}(x) Y_i \quad . \tag{1}$$

where $W_{ni}(x) = W_{ni}(x, X_1, X_2, \cdots, X_n)$ $(i = 1, 2, \cdots, n)$ are weight functions selected as following:

For a fixed $x \in R^d$, rerange the observations $(X_1, Y_1)$, …, $(X_n, Y_n)$ according to

$$\|X_{R_1} - x\| < \|X_{R_2} - x\| < \cdots < \|X_{R_n} - x\| , \tag{2}$$

and break ties by comparing indices, where $\|x\|$ can be taken, for example, as the usual Euclidean norm or $\|x\| = \max(|X^{(1)}|, |X^{(2)}|, \cdots, |X^{(d)}|)$ for $x = (X^{(1)}, X^{(2)}, \cdots, X^{(d)})$. Suppose that $\{V_{ni}, i \geq 1\}$ is a given series of weights, i.e. $V_{ni} \geq 0$, $\sum_{i=1}^{n} V_{ni} = 1$ for all $n$ and $i$. Then we take

* Received Oct. 4, 1985.

(3) $$W_{nR_i}(x) = V_{ni} \qquad i = 1, 2, \cdots, n.$$

Devroye[2] established the strong consistency of $m_n(x)$ under the following conditions:

(A1) Y is bounded.

(A2) There exists a sequence of positive integers $k = k_n$ such that

$$\frac{k}{n} \to 0, \ \frac{\log n}{k} \to 0 \ (\text{as } n \to \infty), \ \sup_n (k \max_{1<i<k} V_{ni}) < \infty, \ \sum_{i>k} V_{ni} = o(1) \text{ as } n \to \infty.$$

In 1985, Chen Xiru[3] improved the conditions by

(B1) $Y$ is bounded.

(B2) There exists a sequence of positive integers $k = k_n$ such that

$$\frac{k}{n} \to 0, \ \frac{\log n}{k} \to 0 \ (\text{as } n \to \infty), \ \sum_{i>k} V_{ni} = o(1) \text{ a.s.}, \sum_{i=1}^{k} V_{ni}^2 = o\left(\frac{1}{\log n}\right) \text{ a.s.}.$$

(B3) $\lim_{\varepsilon \to 0} C(\varepsilon) = 0$ a.s. where $C(\varepsilon) = \sup_n \{\max(\sum_i{}' V_{ni}$: the number of terms contained in $\Sigma'$ does not exceed $k\varepsilon)\}$.

In this paper we discuss the case that $\{Y_i\}$ are censored by random variables $\{t_i\}$. It means that we can not observe $Y_i$ and instead of $Z_i = \min(Y_i, t_i)$, $\delta_i = I_{(Y_i < t_i)}$. We always suppose that $t_i$ i.i.d. and independent of $\{Y_i\}$. Let $F_x$ be the distribution function of $Y$ for fixed $x$ and $G$ be the distribution function of $t$. Denote $\tau_{F_x} = \inf\{u: F_x(u) = 1\}$, $\tau_G = \inf\{u: G(u) = 1\}$. It is clear that if the censoring is too heavy we can not get the enough information of $Y_i$. As a basic assumption, it is reasonable that

(4) $$\sup_x \tau_{F_x} < \tau_G \leqslant \infty .$$

where $x$ over the range of $X$. We denote $H_x(t) = P_x(Z_i < t)$, $\tau_{H_x} = \inf\{u: H_x(u) = 1\}$. Thus $\tau_{H_x} = \tau_{F_x}$ for any $x$, furthermore we let

(5) $$y = G(\sup_X \tau_{F_x}) < 1 .$$

Now our problem here is how to fit the regression on the basis of only observing $(\delta_i, Z_i, X_i)$. A naive idea is that if $Y_i$ is censored we add something to it to make up for the censored part and if $Y_i$ is uncensored we also modify it appropriately to ensure unbiasedness in the sense that the modification $Y_i^*$ has the same expectation as $Y_i$. In view of this consideration, we always assume $G$ continuous and suggest using $Y_i^*$ of the form (for known $G$)

(6) $$Y_i^* = \delta_i \varphi_i(Z_i) + (1 - \delta_i)\varphi_2(Z_i)$$

where $\varphi_1$, $\varphi_2$ are continuous on $(-\infty, a]$ $(a < \tau_G)$

such that

(7) $$\begin{cases} \text{(i)} \ [1 - G(Y)]\varphi_1(Y) + \int_{-\infty}^{Y} \varphi_2(t) dG(t) = Y \\ \text{(ii)} \ \varphi_1, \varphi_2 \text{ are independent of distribution of } (X, Y) \text{ (but may depend on } G) \end{cases}$$

We used this technical in the censored data linear regression model[4], and we will show the success in non-parametric case. The class of all pairs $(\varphi_1, \varphi_2)$ of such functions will be denoted by $\widetilde{K}$. For simplicity we also use $\widetilde{K}$ to denote the class of all "estimator" $Y_i^* = \delta_i \varphi_i(Z_i) + (1-\delta_i)\varphi_2(Z_i)$ of $Y_i$ with $(\varphi_1, \varphi_2) \in \widetilde{K}$. Note that

$$
\begin{aligned}
E(Y_i^* | X_i) &= E_{X_i}[\delta_i \varphi_1(Z_i) + (1-\delta_i)\varphi_2(Z_i)] \\
&= \iint_{t>y} \varphi_1(y)\,\mathrm{d}G(t)\,\mathrm{d}F_{X_i}(y) + \iint_{t<y} \varphi_2(t)\,\mathrm{d}G(t)\,\mathrm{d}F_{X_i}(y) \\
&= \int_{-\infty}^{\infty} \varphi_1(y)\Big(\int_y^{\infty} \mathrm{d}G(t)\Big)\mathrm{d}F_{X_i}(y) + \int_{-\infty}^{\infty}\Big(\int_{-\infty}^{y} \varphi_2(t)\,\mathrm{d}G(t)\Big)\mathrm{d}F_{X_i}(y) \\
&= \int_{-\infty}^{\infty} [1-G(y)]\varphi_1(y)\,\mathrm{d}F_{X_i}(y) + \int_{-\infty}^{\infty}\Big(\int_{-\infty}^{y} \varphi_2(t)\,\mathrm{d}G(t)\Big)\mathrm{d}F_{X_i}(y) \\
&= \int_{-\infty}^{\infty} \Big\{[1-G(y)]\varphi_1(y) + \int_{-\infty}^{y} \varphi_2(t)\,\mathrm{d}G(t)\Big\}\mathrm{d}F_{X_i}(y) \\
&= \int_{-\infty}^{\infty} y\,\mathrm{d}F_{X_i}(y) = E(Y_i | X_i).
\end{aligned}
$$

We will give some examples below (omiting the substript $i$ for simplicity).

**Example 1** Suppose that we mant to keep $Y^* = Y$ when $Y$ is uncensored (i.e., $\delta = 1$). Then $\varphi_1(Z) = Z$. Assuming that $G$ has continuous positive density $g$, we have $\varphi_2(Z) = Z + G(Z)/g(Z)$ by (7). Therefore $Y^* = \delta Z + (1-\delta)(Z + G(Z)/g(Z))$.

**Example 2** In developing least squares estimates for the linear regression model with censored response, Koul, Susarla and Van Ryzin[5] proposed to replace the censored response by 0. This means that $\varphi_2(Z) = 0$. Then (7) leads the solution $\varphi_1(Z) = Z/(1-G(Z))$.

**Example 3** Suppose that we want to augment the censored and the uncensored data equally. This means that $\varphi_1(Z) = \varphi_2(Z)$. We have

$$
\varphi_1(Z) = \varphi_2(Z) = \int_{-\infty}^{Z} \frac{\mathrm{d}s}{1-G(s)} ,
$$

since
$$
\begin{aligned}
&[1-G(y)]\int_{-\infty}^{y} \frac{\mathrm{d}s}{1-G(s)} + \int_{-\infty}^{y}\Big(\int_{-\infty}^{y} \frac{\mathrm{d}s}{1-G(s)}\Big)\mathrm{d}G(t) \\
&= [1-G(y)]\int_{-\infty}^{y} \frac{\mathrm{d}s}{1-G(s)} + \int_{-\infty}^{y} \frac{G(y)-G(s)}{1-G(s)}\mathrm{d}s \\
&= [1-G(y)]\int_{-\infty}^{y} \frac{\mathrm{d}s}{1-G(s)} + \int_{-\infty}^{y}\mathrm{d}s - \int_{-\infty}^{y} \frac{1-G(y)}{1-G(s)}\mathrm{d}s = y.
\end{aligned}
$$

Now we turn to non-parametric regression estimates. We assume that $G(t)$ is known first and consider the estimator

(8) $$m_n^*(x) = \sum_{i=1}^{n} W_{ni}(x) Y_i^* .$$

**Theorem 1** If (B1), (B2), (B3) and (4) hold, $(\varphi_1,\varphi_2)\in\widetilde{K}$, then

$$\lim_{n\to\infty} m_n(x)=m(x) \quad \text{a.s..}$$

**Proof** $m_n^*(x)-m(x)=\sum_{i=1}^{n}W_{ni}(x)(Y_i^*-m(x))=\sum_{i=1}^{n}W_{ni}(x)(Y_i^*-Y_i)$

$$+\sum_{i=1}^{n}W_{ni}(x)(Y_i-m(X_i))+\sum_{i=1}^{n}W_{ni}(x)(m(X_i)-m(x))$$

$$=\sum_{i=1}^{n}W_{nR_i}(x)(Y_{R_i}-m(X_{R_i}))+\sum_{i=1}^{n}W_{nR_i}(x)(m(X_{R_i})-m(x))$$

$$+\sum_{i>k}W_{nR_i}(x)(Y_{R_i}-m(X_{R_i}))+\sum_{i=1}^{k}W_{nR_i}(x)(Y_{R_i}^*-Y_{R_i})+\sum_{i>k}W_{nR_i}(x)(Y_{R_i}^*-Y_{R_i})$$

$$\triangleq J_{1n}(x)+J_{2n}(x)+J_{3n}(x)+J_{4n}(x)+J_{5n}(x).$$

Chen Xiru proved that $J_{1n}(x)+J_{2n}(x)+J_{3n}(x)\to 0$ a.s. We only need to de deal with $J_{4n}(x)$ and $J_{5n}(x)$.

Since that $Y_i$ are bounded and $\gamma=G(\sup_X \tau_{F_X})<1$, there is a constant $A$ such that $-A<Y<A$ and $G(A)<1$. Therefore on $[-A,A]$ $\varphi_1$, $\varphi_2$ are continuous and there exists constant $B$ such that $|\varphi_1(Z)|\leqslant B$, $|\varphi_2(Z)|\leqslant B$ for $Z=\min(Y,t)\leqslant Y<A$. It leads that $|Y^*|=|\delta\varphi_1(Z)+(1-\delta)\varphi_2(Z)|<\max\{|\varphi_1(Z)|, |\varphi_2(Z)|\}\leqslant B$, i.e., $Y^*$ bounded. On the other hand $EY^*=E(E(Y^*|X))=E(E(Y|X))=EY$. Let $u_i=\frac{2(Y_i^*-Y_i)}{5(A+B)}$ be independent bounded ($|u_i|<2/5$) random variables with mean zero. For given $x, X_1,\cdots,X_n$ the conditional distribution of $\frac{2J_{4n}(x)}{5(A+B)}$ is the same as that of $\sum_{i=1}^{k}c_iu_i$, where $c_1,\cdots,c_k$ are constants which satisfy

$$d_n=\sum_{i=1}^{k}c_i^2=\sum_{i=1}^{k}W_{nR_i}^2(x)=o(\frac{1}{\log n}).$$

Now we can use the following inequality due to Tao Bo-Cheng Ping[6], just as Chen Xiru used for $J_{1n}$ part:

$$(9)\qquad E(\sum_{i=1}^{n}a_iu_i)^{2s}\leqslant 3^s(2s-1)!!\max_{1\leqslant i\leqslant n}Eu_i^{2s}\qquad (s=1,2,\cdots)$$

where $\{u_i\}$ are independent random variables with mean zero and $\{a_i\}$ satisfy

$$\sum a_i^2=1\qquad ((2s-1)!!=\frac{(2s)!}{2^s\cdot s!}).$$

Let $T_n=\sum_{i=1}^{k}c_iu_i/\sqrt{d_n}$, then

$$P(|\sum_{i=1}^{k}c_iu_i|\geqslant\varepsilon)=P(|T_n|\geqslant\varepsilon/\sqrt{d_n})\leqslant\exp(-\varepsilon^2/d_n)E(e^{T_n^2})$$

$$=\exp(-\varepsilon^2/d_n)\sum_{s=0}^{\infty}\frac{1}{s!}ET_n^{2s}\leqslant\exp(-\varepsilon^2/d_n)[1+\sum_{s=1}^{\infty}\frac{1}{s!}3^s(2s-1)!!(\frac{2}{5})^{2s}]$$

$\leqslant \exp(-\varepsilon^2/d_n)[1+\sum_{s=1}^{\infty}(\frac{3\cdot 2\cdot 2^2}{25})^s] \leqslant 25\exp(-\varepsilon^2/d_n)$ and $\sum_{n=1}^{\infty}P(|\sum_{i=1}^{k}c_iu_i|\geqslant\varepsilon)$

$\leqslant 25\sum_{n=1}^{\infty}\exp(-\varepsilon^2/d_n)<\infty$. Hence by Borel-Cantelli lemma we have proved hat for any fixed $X_i=x_i$ $\lim_{n\to\infty} J_{4n}(x,x_1,Y_1,\cdots,x_n,Y_n)=0$ a.s. This in turn proves that $\lim_{n\to\infty} J_{4n}(x)=0$ a.s..

On the other hand, by the boundedness of $Y_i^*-Y_i$, it is clear that $J_{5n}(x)\to$ $\to 0$ a.s. according to (B2).

If $G(t)$ is unknown, we can think that $t_i$ are censored by $Y_i$ and use Kaplan-Meier[7] estimator $\hat{G}_n(t)$ instead of $G(t)$, where

$$1-G_n(t)=\prod_{Z_i<t}(1-\frac{1}{n-i+1})^{(1-\delta_i)} \tag{10}$$

It is well known that

$$\sup_{-\infty<t<t_0}|\hat{G}_n(t)-G(t)|\to 0 \quad \text{a.s.} \quad (t_0<\sup_X \tau_{H_x}) \tag{11}$$

From now on, we use the notations $\varphi_1(Z_i,G)$, $\varphi_2(Z_i,G)$ in place of $\varphi_1(Z_i)$, $\varphi_2(Z_1)$ respectively to signify their dependence on $G$. For $G$ unknown case, it is natural to substitute it by an estimator $G_n^*$ and use $\varphi_1(Z_i,G_n^*)$, $\varphi_2(Z_i,G_n^*)$ instead

We will restrict $(\varphi_1,\varphi_2)$ to certain "nice" subsets of the class $\widetilde{K}$ to be defined below:

Let $\widetilde{K}^*$ be the class of all $(\varphi_1,\varphi_2)\in\widetilde{K}$ with the following boundedness property: For every $d$ with $1>d>0$ and every $s$, there exists $C$ such that

$$\max_{\substack{j=1,2\\ t\leqslant s}}|\varphi_j(t,G')|\leqslant C$$

for all distribution function $G'$ with $G'(s)\leqslant d$.

Let $\widetilde{K}_C^*$ be the class of all $(\varphi_1,\varphi_2)\in\widetilde{K}^*$ with the following continuity property at the censoring distribution $G$:

For every $\varepsilon>0$ and every $s$ with $G(s)<1$, there exists $\eta>0$ such that

$$\max_{\substack{j=1,2\\ t\leqslant s}}|\varphi_j(t,G')-\varphi_j(t,G)|\leqslant\varepsilon. \tag{12}$$

for all distribution function $G'$ with $\sup_{t\leqslant s}|G'(t)-G(t)|<\eta$.

We can verify that $(\varphi_1,\varphi_2)$ in example 2 and example 3 belong to $\widetilde{K}_C^*$. We suppose that $G$ and all the conditional distribution functions $F_X$ are continuous in the rest part of this paper.

**Theorem 2** Suppose that we know $\gamma=G(\sup_X \tau_{H_x})<1$. Define

$$\widetilde{G}_n(t)=\begin{cases}\hat{G}_n(t) & \text{(Kaplan-Meier e timator) if } \hat{G}_n(t)<\gamma,\\ \gamma & \text{if } t\leqslant \max Z_i;\ \text{and } \hat{G}_n(t)>\gamma.\end{cases}$$

and $\hat{Y}_i^*=\delta_i\varphi_1(Z_i,\widetilde{G}_n)+(1-\delta_i)\varphi_2(Z_i,\widetilde{G}_n)$. For fixed $x$,

$$\hat{m}_n^*(x)\triangleq\sum_{i=1}^n W_{ni}(x)\hat{Y}_i^*.$$

If $(\varphi_1,\varphi_2)\in\widetilde{\mathrm{K}}_c^*$ and (B1), (B2), (B3) hold, then $\lim_{n\to\infty}\hat{m}_n^*(x)=m(x)$ a.s.

**Proof**

$$\begin{aligned}\hat{m}_n^*(x)-m(x)=&\sum_{i=1}^k W_{nR_i}(x)(Y_{R_i}-m(X_{R_i}))+\sum_{i=1}^n W_{nR_i}(x)(m(X_{R_i})-m(x))\\ &+\sum_{i>k}W_{nR_i}(x)(Y_{R_i}-m(X_{R_i}))+\sum_{i=1}^k W_{nR_i}(x)(Y^*_{R_i}-Y_{R_i})\\ &+\sum_{i>k}W_{nR_i}(x)(Y^*_{R_i}-Y_{R_i})+\sum_{i=1}^n W_{ni}(x)(\hat{Y}_i^*-Y_i^*)\\ &\triangleq J_{1n}(x)+J_{2n}(x)+J_{3n}(x)+J_{4n}(x)+J_{5n}(x)+J_{6n}(x)\ .\end{aligned}$$

We have proved that $J_{1n}(x), J_{2n}(x), J_{3n}(x), J_{4n}(x), J_{5n}(x)$ converge to zero a.s. For $J_{6n}(x)$, let $\varepsilon>0$ arbitrary small and we can find $T^*<\infty$ such that $1>\inf_X H_X(T^*)>1-\varepsilon$. Thus

$$|J_{6n}(x)|<\left|\sum_{i=1}^n W_{ni}(x)(\hat{Y}_i^*-Y_i^*)I_{(Z_i<T^*)}\right|+\left|\sum_{i=1}^n W_{ni}(x)(\hat{Y}_i^*-Y_i^*)I_{(Z_i>T^*)}\right|\triangleq|I_{n1}|+|I_{n2}|;$$

$$\begin{aligned}|I_{n1}|&<\sup_i|\hat{Y}_i^*-Y_i^*|I_{(Z_i<T^*)}\\ &=\sup_i|(\delta_i\varphi_1(Z_i,\widetilde{G}_n)+(1-\delta_i)\varphi_2(Z_i,\widetilde{G}_n))-(\delta_i\varphi_1(Z_i,G)+(1-\delta_i)\varphi_2(Z_i,G))|I_{(Z_i<T^*)}\\ &\leqslant\sup_i\max_{j=1,2}[|\varphi_1(Z_i,\widetilde{G}_n)-\varphi_1(Z_i,G)|,\ |\varphi_2(Z_i,\widetilde{G}_n)-\varphi_2(Z_i,G)|]I_{(Z_i<T^*)}\ .\end{aligned}$$

By the definition of $\widetilde{\mathrm{K}}_c^*$ and the consistency of Kaplan-Meier estimator

$$|I_{n1}|\to 0 \text{ as } \sup_{u<T^*}|\widetilde{G}_n(u)-G(u)|\to 0.$$

$$\begin{aligned}|I_{n2}|&<\sum_{i=1}^n W_{ni}(x)|\hat{Y}_i^*-Y_i^*|I_{(Z_i>T^*)}\\ &\leqslant\sup_i[|\hat{Y}_i^*|+|Y_i^*|]\cdot\sum_{i=1}^n W_{ni}(x)I_{(Z_i>T^*)}\leqslant D\sum_{i=1}^n W_{ni}(x)I_{(Z_i>T^*)}\\ &=D\cdot\sum_{i=1}^n[W_{ni}(x)I_{(Z_i>T^*)}-E(W_{ni}(x)I_{(Z_i>T^*)}|X_1,\cdots X_n)]\\ &\quad+D\cdot\sum_{i=1}^n E(W_{ni}(x)I_{(Z_i>T^*)}|X_1,\cdots X_n)\\ &=D\cdot\sum_{i=1}^n W_{ni}(x)[I_{(Z_i>T^*)}-E(I_{(Z_i>T^*)}|X_i)]+D\cdot\sum_{i=1}^n W_{ni}(x)P(Z_i>T^*|X_i),\end{aligned}$$

where $D$ is a constant (by the definitions of $\widetilde{\mathrm{K}}^*$ and $\widetilde{G}_n$).

For given $X_i = x_i$, let $u_i' = I_{(Z_i > T^*)} - E(I_{(Z_i > T^*)}|X_i)$ and use (9) again, we get $\sum_{i=1}^{n} W_{ni}(x)[I_{(Z_i > T^*)} - E(I_{(Z_i > T^*)}|X_i)] \to 0$ a.s. On the other hand

$$\sum_{i=1}^{n} W_{ni}(x) E(I_{(Z_i > T^*)}|X_i) = \sum_{i=1}^{n} W_{ni}(x)(1 - H_{X_i}(T^*)) < \sum_{i=1}^{n} W_{ni}(x) \sup_X (1 - H_X(T^*)) < \varepsilon .$$

It completes the proof.

## References


[1] Stone, C. T., Consistent Nonparametric Regression, Ann. Stat., 1977, Vol 5, 595—645.

[2] Devroye, L. P., On the Almost Everywhere Convergence of Nonparametric Regression Function Estimates, Ann. Stat., 1981, Vol 9, 1310—1319.

[3] Chen Xiru (陈希孺), Almost Sure Convergence of Nonparametric Regression Estimates, Chin. Ann. of Math., 6B(1), 1985, 103—108.

[4] Zheng Zukang, A Class of Estimators for the Parameters in Linoar Regression with Censored Data, to appear.

[5] Koul, H., Susarla, V. and Van Ryzin, J., Regression Analysis with Randomly Right-Censored Data, Ann. Stat., 1981, Vol 9, 1276—1288.

[6] Tao Bo and Cheng Ping (陶波, 成平) 一个矩不等式, China 数学年刊, 1981, 451—461.

[7] Kaplan, E. L. and Meier, P., Nonparametric Estimation from Incomplete Observations, JASA, 1958, (53) 457—481.