X Congreso Galego de Estatística e Investigación de Operacións
Pontevedra, 3–4–5 de novembro de 2011

CLT of nonparametric regression estimator with truncated, censored and dependent data

Han-Ying Liang 1[1], Jacobo de Uña-Álvarez[2] and María del Carmen Iglesias-Pérez[2]

[1]Department of Mathematics, Tongji University, Shanghai, P. R. China
[2]Department of Statistics and OR, University of Vigo, Spain

ABSTRACT

Based on the idea of the Nadaraya-Watson (NW) kernel smoother and the technique of the local linear (LL) smoother, we construct for the first time the NW and LL estimators of conditional mean functions and their derivatives for a left truncated and right censored model. The target function includes the regression function, the conditional moment as well as the conditional distribution function as special cases. It is assumed that the lifetime observations form a stationary α-mixing sequence. Asymptotic normality of the estimators is established.

Palabras e frases crave: Asymptotic normality; Nadaraya-Watson and local linear smoothing; conditional mean function; truncated and censored data; α-mixing.

1. INTRODUCTIÓN

The study of the relationship between a variable of interest Y and a covariate X is one of the most important problems in Statistics. This can be done via estimating the function $m(x) = E(\phi(Y)|X = x)$, where the (known) transformation $\phi(\cdot)$ is introduced to include various targets of interest. For example, taking $\phi(y) = y^r$ we get the rth conditional moment, and if we take $\phi(y) = I(y < t)$ then $m(x)$ becomes the conditional distribution function of Y given $X = x$ at t. If we have n replications (X_i, Y_i) of (X, Y), several kernel smoothers consist in estimating $m(x)$ by calculating a weighted local average of the $\phi(Y_i)$ values. This can be written as

$$\sum_{i=1}^{n} w_i(x)\phi(Y_i), \tag{1}$$

where $w_i(x)$ is a given weight function which describing the degree of smoothing on the covariate. Special cases of (1) include the Nadaraya-Watson (NW) and the local linear (LL) estimators.
In this paper we consider dependent observations, then $\{(X_i, Y_i), i \geq 1\}$ denotes a

$\mathbb{R} \times \mathbb{R}$ valued stationary process from (X, Y). There is a large literature about the NW estimator of the regression function with complete dependent data. See, for example, Bosq (1998) and the references therein.

In some fields as Reliability, Survival Analysis or Economics, right-censored or/and left-truncated data are often encountered. The random variable Y can be regarded as the lifetime of a device, the survival time of a patient or the duration time of some economic variable. Recently, El Ghouch and Van Keilegom (2008) discussed for the first time the asymptotic normality for an estimator of $m(x)$ based on a transformed variable of Y, for the right-censored model with dependent data; Ould-Saïd and Lemdani (2006) constructed a new nonparametric estimator (of NW type) of $E(Y|X = x)$ for the left-truncation model and studied its asymptotic properties in independent and identically distributed (i.i.d.) setting, see also Liang et al. (2009) for dependent data. However, to the best of our knowledge, little is known about the estimator of $m(x)$ for left truncated and right censored (LTRC) model.
Based on the idea of the NW kernel smoother and the technique of the LL smoother, in this work, we construct for the first time the NW and LL estimators of $m(x)$ and its derivatives for the LTRC model, and establish the asymptotic normality for the NW and LL estimators of $m(x)$ under dependent observations. Firstly, we present some notation. In the next section we introduce estimators of $m(x)$ and the main results are formulated in Section 3.

Let (Y, T, W) be a random vector, where Y is the lifetime with distribution function (df) F, T is the random left truncation time with df L, and W denotes the random right censoring time with df G. In random LTRC model one observes (Z, T, δ) if $Z \geq T$, where $Z = \min(Y, W)$ and $\delta = I(Y \leq W)$, when $Z < T$ nothing is observed. Clearly, if Y is independent of W, then Z has df $H = 1 - (1 - F)(1 - G)$. Take $\theta = P(T \leq Z)$, then necessarily, we assume $\theta > 0$. Consider the presence of covariates and assume that X admits df $V(\cdot)$ and density $v(\cdot)$. Then denote by $(X_i, Z_i, T_i, \delta_i)$, $i = 1, 2, \cdots, n$ a stationary random sample from (X, Z, T, δ) which one observes $(T_i \leq Z_i, \forall i)$. Without loss of generality, we assume that Y, T and W are nonnegative random variables, as usual in survival analysis. Iglesias-Pérez and González-Manteiga (1999) defined a generalized product-limit estimator, GPLE, of the conditional survival function of Y given $X = x$ for the LTRC data, given by

$$1 - \hat{F}_n(y|x) = \prod_{i=1}^{n} \Big(1 - \frac{I(Z_i \leq y)\delta_i B_{ni}(x)}{\sum_{j=1}^{n} I(T_j \leq Z_i \leq Z_j) B_{nj}(x)}\Big), \tag{2}$$

where $B_{ni}(x) = K(\frac{x - X_i}{h_n}) / \sum_{j=1}^{n} K(\frac{x - X_j}{h_n})$, K denotes a kernel function, and $0 < h_n \to 0$ is a bandwidth parameter. Under the i.i.d. assumptions, Iglesias-Pérez and González-Manteiga (1999) obtained an almost sure representation and asymptotic normality of $\hat{F}_n(\cdot)$. Liang et al. (2010) investigated asymptotic properties of $\hat{F}_n(\cdot)$ with α-mixing data.

In the sequel, $\{(X_i, Z_i, T_i, \delta_i) =: \zeta_i, 1 \leq i \leq n\}$ is assumed to be a stationary α-mixing sequence of random vectors. Recall that the sequence $\{\zeta_k, k \geq 1\}$ is said to be α-mixing if the α-mixing coefficient

$$\alpha(n) \stackrel{\text{def}}{:=} \sup_{k\geq 1} \sup\{|P(AB) - P(A)P(B)| : A \in \mathcal{F}_{n+k}^{\infty}, B \in \mathcal{F}_1^k\}$$

converges to zero as $n \to \infty$, where $\mathcal{F}_l^m = \sigma\{\zeta_l, \zeta_{l+1}, \ldots, \zeta_m\}$ denotes the σ-algebra generated by $\zeta_l, \zeta_{l+1}, \ldots, \zeta_m$ with $l \leq m$. Among various mixing conditions used in the literature, α-mixing is reasonably weak, and has many practical applications. See Cai and Kim (2003) for motivation in the scope of Survival Analysis.

2. ESTIMATORS

In the sequel, for any df $Q(y) = P(\eta \leq y)$, its density function is denoted by $q(y)$, we denote the left and right support endpoints by $a_Q = \inf\{y : Q(y) > 0\}$ and $b_Q = \sup\{y : Q(y) < 1\}$, respectively.

For any given $x \in \mathbb{R}$, define $Q(y|x) = P(\eta \leq y|X = x)$ and $Q^*(y) = P(\eta \leq y|T \leq Z)$, their density functions are denoted by $q(y|x)$ and $q^*(y)$, respectively. Set $\theta(x) = P(T \leq Z|X = x)$, $C(y|x) = P(T \leq y \leq Z|X = x, T \leq Z)$ and $H_1^*(y|x) = P(Z \leq y, \delta = 1|X = x, T \leq Z)$.

We observe that

$$\begin{aligned} V^*(x) &= P(X \leq x|T \leq Z) = \theta^{-1}E\{I(X \leq x)E(I(T \leq Z)|X)\} \\ &= \theta^{-1}\int_{-\infty}^{x} P(T \leq Z|X = t)v(t)dt = \theta^{-1}\int_{-\infty}^{x} \theta(t)v(t)dt, \end{aligned}$$

which gives $v^*(x) = \theta^{-1}\theta(x)v(x)$. In addition, assume that Y, T and W are conditionally independent at $X = x$, and $F(y|x)$ and $G(y|x)$ are continuous with respect to y. Then it is easy to verify that

$$C(y|x) = \theta^{-1}(x)L(y|x)(1 - G(y|x))(1 - F(y|x)) = \theta^{-1}(x)L(y|x)(1 - H(y|x)),$$

$$H_1^*(y|x) = \theta^{-1}(x)\int_0^y L(t|x)(1 - G(t|x))f(t|x)dt,$$

which gives $h_1^*(y|x) = \theta^{-1}(x)L(y|x)(1 - G(y|x))f(y|x)$. The estimator of $C(y|x)$ is defined by

$$\hat{C}_n(y|x) = \sum_{i=1}^{n} I(T_i \leq y \leq Z_i)B_{ni}(x). \tag{3}$$

From now on we only consider classes of functions ϕ that satisfy the following conditions:

($\mathcal{H}$):($\mathcal{H}1$) ϕ vanishes outside the interval $[\tau_1, \tau_2]$ for some $a_{L(\cdot|x)} < \tau_1 \leq \tau_2 < b_{H(\cdot|x)}$;

($\mathcal{H}2$) ϕ is a bounded function on $[\tau_1, \tau_2]$.

Remark 1 *If the lifetime Y has no left truncation (i.e., $T = 0$), then one can choose $\tau_1 = 0$ in the condition ($\mathcal{H}$), in this case, ($\mathcal{H}$) reduces to assumption ($\mathcal{H}$) of El Ghouch and Van Keilegom (2008).*

Under condition $(\mathcal{H})$, if $\Lambda(\cdot)$ is a bounded function with bounded support, and the functions $v^*(t)$ and $f(y|t)$ are continuous at $t = x$ we find

$$
\begin{aligned}
& E\big\{\Lambda_{a_n}(X-x)[\delta\phi(Z)C^{-1}(Z|X)(1-F(Z|X)) - m(x)]\big\} \\
= \; & \frac{1}{a_n}\int_{\mathbb{R}}\int_{\mathbb{R}} \Lambda\Big(\frac{s-x}{a_n}\Big) v^*(s)\phi(y)C^{-1}(y|s)(1-F(y|s))h_1^*(y|s)dsdy - \frac{m(x)}{a_n}\int_{\mathbb{R}} \Lambda\Big(\frac{s-x}{a_n}\Big) v^*(s)ds \\
= \; & \int_{\mathbb{R}}\int_{\mathbb{R}} \Lambda(s)v^*(x+a_n s)\phi(y)f(y|x+a_n s)dsdy - \int_{\mathbb{R}} \Lambda(s)v^*(x+a_n s)ds\int_{\mathbb{R}}\phi(y)f(y|x)dy \\
= \; & \int_{\mathbb{R}} \Lambda(s)v^*(x+a_n s)\Big[\int_{\mathbb{R}}\phi(y)(f(y|x+a_n s)-f(y|x))dy\Big]ds \to 0, \qquad (4)
\end{aligned}
$$

where $\Lambda_{a_n}(\cdot) = \Lambda(\cdot/a_n)/a_n$ and $0 < a_n \to 0$.

Note that (4) suggests that the estimation of $m(x)$ can be viewed as a nonparametric regression of $Y_i^* = \delta_i\phi(Z_i)C^{-1}(Z_i|X_i)(1-F(Z_i|X_i))$ on X_i. Then, based on the idea of the NW kernel smoother, the estimator of $m(x)$ is defined by

$$
\hat{m}_{NW}(x) = \sum_{i=1}^n \hat{Y}_i^* w_i^{NW}(x) \quad \text{with} \quad w_i^{NW}(x) = \frac{\Lambda_{a_n}(X_i-x)}{\sum_{j=1}^n \Lambda_{a_n}(X_j-x)},
$$

where $\hat{Y}_i^* = \delta_i\phi(Z_i)\hat{C}_n^{-1}(Z_i|X_i)(1-\hat{F}_n(Z_i|X_i))$.

Next suppose that the 2th derivative of $m(z)$ at point x exists and is continuous. Then in a small neighbourhood of x, we can approximate $m(x)$ locally by a linear term

$$
m(z) \approx m(x) + m'(x)(z-x) := \beta_0 + \beta_1(z-x).
$$

Based on the idea of the LL smoother, the LL estimator $\hat{\mathbf{B}}_n(x) = (\hat{m}_{LL}(x), \hat{m}'_{LL}(x))^\tau$ of $\mathbf{B}(x) = (m(x), m'(x))^\tau$, is defined as $(\hat{\beta}_0, \hat{\beta}_1)^\tau$, which minimizes

$$
\sum_{i=1}^n \big(\hat{Y}_i^* - \beta_0 - \beta_1(X_i-x)\big)^2 \Lambda_{a_n}(X_i-x). \qquad (5)
$$

Let

$$
\mathbf{X} = \begin{pmatrix} 1 & (X_1-x) \\ \vdots & \vdots \\ 1 & (X_n-x) \end{pmatrix}, \quad \hat{\mathbf{Y}}^* = \begin{pmatrix} \hat{Y}_1^* \\ \vdots \\ \hat{Y}_n^* \end{pmatrix}, \quad \mathbf{W} = \mathrm{diag}(\Lambda_{a_n}(X_i-x)).
$$

Then, from (5) simple algebra shows that $\hat{\mathbf{B}}_n(x)$ can be expressed as $\hat{\mathbf{B}}_n(x) = (\mathbf{X}^\tau\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\tau\mathbf{W}\hat{\mathbf{Y}}^*$. Denote by $s_{n,j} = \frac{\theta}{n}\sum_{i=1}^n \big(\frac{X_i-x}{a_n}\big)^j \Lambda_{a_n}(X_i-x)$, $t_{n,j} = \frac{\theta}{n}\sum_{i=1}^n \big(\frac{X_i-x}{a_n}\big)^j \Lambda_{a_n}(X_i-x)\hat{Y}_i^*$,

$$
\mathbf{S}_n = \begin{pmatrix} s_{n,0} & s_{n,1} \\ s_{n,1} & s_{n,2} \end{pmatrix} \quad \text{and} \quad \mathbf{t}_n = \begin{pmatrix} t_{n,0} \\ t_{n,1} \end{pmatrix}.
$$

Then $\hat{\mathbf{B}}_n(x) = \mathrm{diag}(1, a_n^{-1})\mathbf{S}_n^{-1}\mathbf{t}_n$.

3. MAIN RESULTS

Throughout the paper, let $C, C_1, \cdots$ and $c_0, c, c_1, \cdots$ denote generic finite positive constants, whose values may change from line to line. $A_n = O(B_n)$ stands for $|A_n| \le C|B_n|$ and $f^{(i,j)}(x,y) := \partial^{i+j} f(x,y)/\partial x^i \partial y^j$; $U(x)$ represents a neighborhood of x. Let $I = [x_1, x_2]$ be an interval contained in the support of $m^*(\cdot)$, such that $\inf_{x\in I_\delta} m^*(x) \ge \delta_0 > 0$ for $I_\delta = [x_1 - \delta, x_2 + \delta]$ with small $\delta > 0$.

In order to formulate the main results, we need the following assumptions.

(A1) $K(\cdot)$ is a Lipschitz-continuous density function with bounded support and $\int_{\mathbb{R}} tK(t)dt = 0$.

(A2) (i) For $s \in I_\delta$, Y, T and W are conditionally independent at $X = s$;

(ii) τ_1 and τ_2 are two real numbers such that $a_{L(\cdot|x)} < \tau_1 \le \tau_2 < b_{H(\cdot|x)}$ and $a_{L(\cdot|x)} < a_{H(\cdot|x)}$ for $x \in I_\delta$.

(A3) The first 2 derivatives of functions $\theta(s)$ and $v(s)$ are bounded for $s \in I_\delta$, and the first 2 partial derivatives with respect to s of functions $L(y|s)$, $G(y|s)$, $F(y|s)$, $l(y|s)$, $g(y|s)$ and $f(y|s)$ are bounded for $(s,y) \in I_\delta \times \mathbb{R}$.

(A4) For all integers $j \ge 1$, the joint conditional density $v_j^*(\cdot,\cdot)$ of X_1 and X_{j+1} exists on $\mathbb{R}\times\mathbb{R}$ and satisfies $v_j^*(s_1,s_2) \le C_1$ for $(s_1,s_2) \in I_\delta \times I_\delta$.

(A5) $h_n^{-2}(nh_n/\ln(n))^{-(\lambda-3)/2} = O(1)$.

(A6) The kernel $\Lambda(\cdot)$ is a bounded function with bounded support, and satisfies that $\int_{\mathbb{R}} \Lambda(t)dt = 1$ and $\int_{\mathbb{R}} t\Lambda(t)dt = 0$.

(A7) Assume that $na_n \to \infty$, and that the sequence $\alpha(n)$ satisfies that there exist positive integers $q := q_n$ such that $q = o((na_n)^{1/2})$ and $\lim_{n\to\infty}(na_n^{-1})^{1/2}\alpha(q) = 0$.

(A8) Function $m(\cdot)$ has continuous second order derivatives at $x \in I$.

Set $\sigma^2(x) = \theta(x)\int_{\mathbb{R}} \frac{\phi^2(y)f(y|x)dy}{L(y|x)(1-G(y|x))} - m^2(x)$, $\mu_j = \int_{\mathbb{R}} t^j\Lambda(t)dt$, $\nu_j = \int_{\mathbb{R}} t^j\Lambda^2(t)dt$ and

$$\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix}.$$

Theorem 1 *Let $x \in I$ and $\alpha(k) = O(k^{-\lambda})$ for some $\lambda \ge 2.1$. Suppose that conditions $(\mathcal{H})$, (A1)-(A8) are satisfied, and that $\sigma^2(x) > 0$. If $na_n^3 \to \infty$, then*

$$(na_n)^{1/2}\big\{\hat{m}_{NW}(x) - m(x) - a_n^2\mu_2\big(2^{-1}m''(x) + m'(x)[v'(x)v^{-1}(x) + \theta'(x)\theta^{-1}(x)]\big) \\ +o_p(a_n^2) + O_p((\ln(n)/(nh_n))^{1/2} + h_n^2)\big\} \xrightarrow{\mathcal{D}} N(0, \theta\nu_0(\theta(x)v(x))^{-1}\sigma^2(x)).$$

Theorem 2 *Let $x \in I$ and $\alpha(k) = O(k^{-\lambda})$ for some $\lambda \geq 2.1$. Suppose that conditions $(\mathcal{H})$, (A1)-(A8) are satisfied, and that $\sigma^2(x) > 0$. Then*

$$(na_n)^{1/2}\big\{\mathrm{diag}(1, a_n)[\hat{\mathbf{B}}_n(x) - \mathbf{B}(x)] - 2^{-1}a_n^2 m''(x)\mathbf{S}^{-1}\mathbf{U} + o_p(a_n^2)$$
$$+O_p((\ln(n)/(nh_n))^{1/2} + h_n^2)\big\} \xrightarrow{\mathcal{D}} N\big(0, \theta\sigma^2(x)(\theta(x)v(x))^{-1}\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}\big),$$

i.e, for $0 \leq k \leq 1$ we have

$$(na_n^{2k+1})^{1/2}\big\{\hat{m}_{LL}^{(k)}(x) - m^{(k)}(x) - 2^{-1}a_n^{2-k}m''(x)B_k + o_p(a_n^{2-k})$$
$$+O_p(a_n^{-k}[(\ln(n)/(nh_n))^{1/2} + h_n^2])\big\} \xrightarrow{\mathcal{D}} N\big(0, \theta\sigma^2(x)(\theta(x)v(x))^{-1}V_k\big),$$

where B_k and V_k are, respectively, the k-th element of $\mathbf{S}^{-1}\mathbf{U}$ and the k-th diagonal element of $\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}$.

Corollary 1 *Under the assumptions of Theorem 2 we have*

$$(na_n)^{1/2}\big\{\hat{m}_{LL}(x) - m(x) - 2^{-1}a_n^2\mu_2 m''(x) + o_p(a_n^2) + O_p((\ln(n)/(nh_n))^{1/2} + h_n^2)\big\}$$
$$\xrightarrow{\mathcal{D}} N\big(0, \theta\nu_0(\theta(x)v(x))^{-1}\sigma^2(x)\big).$$

The above results are the key to construct confidence intervals of a certain asymptotic level for $m(x)$. A simulation study is in process to investigate the finite sample performance of the proposed estimators $\hat{m}_{NW}(x)$ and $\hat{m}_{LL}(x)$ in the case $\phi(y) = y$.

ACKNOWLEDGMENTS

Authors were supported by the Grants MTM2008-03129 of the Spanish Ministry of Science and Innovation, the Xunta de Galicia under the INBIOMED project (DX-PCTSUG, Ref. 2009/063) and Project 10PXIB300068PR of the Xunta de Galicia, Spain, and also by the National Natural Science Foundation of China (10871146).

REFERENCIAS

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* Springer, New York.

Cai, J. and Kim, J. (2003). Nonparametric quantile estimation with correlated failure time data. *Lifetime Data Analysis* **9**, 357-371.

El Ghouch, A. and Van Keilegom, I. (2008). Non-parametric regression with dependent censored data. *Scand. J. Statist.* **35(2)**, 228-247.

Iglesias-Pérez, C. and González-Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. *J. Nonparametr. Statist.* **10**, 213-244.

Liang, H. Y., Li, D. L. and Qi, Y. C. (2009). Strong convergence in nonparametric regression with truncated dependent data. *J. Multivariate Anal.* **100**, 162-174.

Liang, H. Y., de Uña-Álvarez, J. and Iglesias-Pérez, C. (2010). Asymptotic properties of conditional distribution estimator with truncated, censored and dependent data. Submitted.

Ould-Saïd, E. and Lemdani, M. (2006). Asymptotic properties of a nonparametric regression function estimator with randomly truncated data. *Ann. Inst. Statist. Math.* **58**, 357-378.