Semiparametric Estimator of Mean Conditional Residual Life Function under Informative Random Censoring from Both Sides

A. A. Abdushukurov, F. A. Abdikalikov

Department of Probability Theory and Mathematical Statistics, National University of Uzbekistan, Tashkent, Uzbekistan
Email: a_abdushukurov@rambler.ru

Received 14 January 2015; accepted 2 February 2015; published 5 February 2015

Abstract

In this paper we study estimator of mean residual life function in fixed design regression model when life times are subjected to informative random censoring from both sides. We prove an asymptotic normality of estimators.

Keywords

Informative Censoring, Power Estimator, Regression, Mean Residual Lifetime

1. Introduction

In survival data analysis, response random variable (r.v.) $Z$, the survival time of an individual (in medical study) or failure time of a machine (in industrial study) that usually can be influenced by r.v. $X$, is often called prognostic factor (or covariate). $X$ represents e.g. the dose of a drug for individual or some environmental conditions of a machine (temperature, pressure,…). Moreover, in such practical situations it often occurs that not all of survival times $Z_1, \ldots, Z_n$ of $n$ identical objects are complete observed, that they can be censored by other r.v.-s.

In this article we consider a regression model in which the response r.v.-s are subjected to random censoring from both sides.

We first introduce some notations. Let the support of covariate is the interval $[0,1]$ and we describe our regression results in the situation of fixed design points $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq 1$ at which we consider nonnegative independent responses $Z_{1}, \ldots, Z_{n}$. Suppose that these responses are censored from the left and right by
nonnegative r.v.-s $L_i, \ldots, L_n$ and $Y_i, \ldots, Y_n$ and the observed r.v.-s at design points $x_i$ are in fact
\[
\{\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}\}
\]
with $\xi_i = \max(L_i, \min(Z_i, Y_i))$, $\chi_i^{(0)} = I(\min(Z_i, Y_i) < L_i)$, $\chi_i^{(1)} = I(L_i \leq Z_i \leq Y_i)$ and $\chi_i^{(2)} = (L_i \leq Y_i < Z_i)$, where $I(A)$ denote the indicator of event $A$. Hence the observed data is consist of $n$

vectors:
\[
S^{(o)} = \left\{\{\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}\}, i = 1, \ldots, n\right\}.
\]

Assume that components of vectors $(Z_i, L_i, Y_i)$ are independent for a given covariate $X_i = x_i$. In sample $S^{(o)}$ the r.v.-s of interest $Z_i$’s are observable only when $\chi_i^{(1)} = 1$. Denote by $F_s$, $K_s$ and $G_s$ the conditional distribution functions (d.f.-s) of r.v.-s $Z_i$, $L_i$ and $Y_i$ respectively, given that $X = x$ and suppose that they are continuous.

Let $H_s$ and $N_s$ are conditional d.f.-s of $\xi_s$ and $\eta_s = \min(Z_s, Y_s)$ for $X = x$. Then it’s easy to see that
\[
H_s(t) = K_s(t)N_s(t) \quad \text{with} \quad N_s(t) = 1 - (1 - F_s(t))(1 - G_s(t)), \quad t \geq 0.
\]
In particular, if for all $x \in [0,1]$, $P(L \leq Y) = 1$, then we obtain the interval random censoring model.

The main problem in considered fixed design regression model is consist on estimation the conditional d.f. $F_s$ of lifetimes and its functionals from the samples $S^{(o)}$ under nuisance d.f.-s $K_s$ and $G_s$. The first product-limit type estimators for $F_s$ in the case of no censoring from the left (that is $P(L = \infty) = 1$ or $K_s(t) = 1$) proposed by Beran [1] and has been investigated by many authors (see, for example [2] [3]). In this article supposing that the random censoring from both sides is informative we use twice power type estimator of $F_s$ from [4] [5] for estimation the mean conditional residual life function. Suppose that d.f.-s $K_s$ and $G_s$ are expressed from $F_s$ by following parametric relationships for all $t \geq 0$:
\[
\left\{\begin{aligned}
1 - G_s(t) &= (1 - F_s(t))^{\theta_s}, \\
K_s(t) &= (N_s(t))^{\beta_s},
\end{aligned}\right.
\]
where $\theta_s$ and $\beta_s$ are positive unknown nuisance parameters, depending on the covariate value $x$. Informative model (1.1) include the well-known conditional proportional hazards model (PHM) of Koziol-Green, which follows under absence of left random censorship (that is $\beta_s = 0$). Estimation of $F_s$ in conditional PHM is considered in [6]. Model (1.1) one can considered as an extended two sided conditional PHM. In the case of no covariates, model (1.1) first is proposed in [7] [8].

It is not difficult to verify that from (1.1) one can obtain following expression of d.f. $F_s$:
\[
1 - F_s(t) = \left[1 - \left(H_s(t)\right)^{\gamma_s}\right]^{\lambda_s}, \quad t \geq 0,
\]
where $\lambda_s = \frac{1}{1 + \beta_s} = 1 - p_s^{(0)}$, $\gamma_s = \frac{1}{1 + \theta_s} = \frac{p_s^{(1)}}{1 - p_s^{(0)}}$ and $p_s^{(m)} = P\left(\chi_s^{(m)} = 1\right)$, $m = 0,1,2,$ with $p_s^{(0)} + p_s^{(1)} + p_s^{(2)} = 1$. Then estimator of $F_s$ one can constructed by natural plugging method as follows:
\[
1 - F_{sh}(t) = \left[1 - \left(H_{sh}(t)\right)^{\gamma_{sh}}\right]^{\lambda_{sh}}, \quad t \geq 0.
\]

Here $\gamma_{sh} = p_s^{(0)} \left(1 - p_s^{(0)}\right)^{-1}$, $\lambda_{sh} = 1 - p_s^{(0)}$, $p_s^{(m)} = \sum_{i=1}^{\#} \omega_m(x; h_s) \chi_s^{(m)}$, $m = 0,1,2$, and
\[
H_{sh}(t) = \sum_{i=1}^{\#} \omega_m(x; h_s) I(\xi_s \leq t),
\]
are smoothed estimators of $\lambda_s$, $\gamma_s$, $p_s^{(m)}$ and $H_s(t)$, used Gasser-Müllers weights $\{\omega_m(x; h_s)\}_{i=1}^{\#}$:
\[
\omega_m(x; h_s) = \int_{\xi_s}^{y} \frac{1}{h_s^n} \left(\frac{x - y}{h_s}\right) dy \left(\int_{0}^{\xi_s} \frac{1}{h_s^n} \left(\frac{x - y}{h_s}\right) dy\right)^{-1},
\]
$x_0 = 0$, $\pi(y)$ is a known probability density function (kernel), and $\{h_n\}$ is a sequence of positive constants tending to 0 as $n \to \infty$, called the bandwidth sequence. Note that in the case of no censoring from the left the estimator (1.3) coincides with estimator in conditional Koziol-Green model in [6]. Note also that a class of power type estimators for conditional d.f.s for several models authors have considered in book [9]. Estimator (1.3) was presented in [4] and its asymptotic properties have been investigated in [5]. Now we demonstrate some of these results.

2. Asymptotic Results for Estimator of Conditional Distribution Function

For asymptotic properties of estimator (1.3) we need some notations. For the design points $x_1, \cdots, x_n$ and kernel $\pi$ we denote

$$\Delta_n = \min_{i=0}^{n-1} (x_i - x_{i+1}), \quad \overline{\Delta}_n = \max_{i=0}^{n-1} (x_i - x_{i+1}),$$

$$\|\pi\|_v^2 = \int_{-\infty}^{\infty} \pi^2(y) dy, \quad m_v(\pi) = \int_{-\infty}^{\infty} y^v \pi(y) dy, \quad v = 1, 2.$$

Let $\tau_{F_s} = \sup \{ t \geq 0 : F(t) = 0 \}$ and $T_{F_s} = \inf \{ t \geq 0 : F(t) = 1 \}$ are lower and upper bounds of support of d.f. $F_s$. Then by (1.1):

$$\tau_{F_s} = \tau_{G_s} = \tau_{N_s} = \tau_{H_s} \quad \text{and} \quad T_{F_s} = T_{G_s} = T_{N_s} = T_{H_s}.$$

In [4] authors have proved the following property of two sided conditional PHM (1.1).

**Theorem 2.1 [5].** For a given covariate $x$, the model (1.1) holds if and only if r.v. $\xi$ and the vector $(\xi^{(0)}, \xi^{(1)}, \xi^{(2)})$ are independent.

This characterization of submodel (1.1) plays an important role for investigation the properties of estimator (1.3).

Let’s introduce some conditions:

(C1) As $n \to \infty$, $x_n \to 1$, $\Delta_n = O\left(\frac{1}{n}\right)$, $\overline{\Delta}_n - \Delta_n = o\left(\frac{1}{n}\right)$.

(C2) $\pi$ is a probability density function with compact support $[-M, M]$ for some $M > 0$, with $m_1(\pi) = 0$ and $|\pi(y) - \pi(y')| \leq C_y |y - y'|$, where $C_y$ is some constant.

(C3) $\dot{F}_s(t) = \frac{d}{dx} F_s(t)$ and $\ddot{F}_s(t) = \frac{d^2}{dx^2} F_s(t)$ exist and are continuous for $0 \leq x \leq 1$ and $\tau \leq t \leq T$, with $\tau_{F_s} < \tau < T < T_{F_s}$

(C4) $\dot{\theta}_s = \frac{d}{dx} \theta_s$ and $\ddot{\beta}_s = \frac{d^2}{dx^2} \beta_s$ exist and are continuous for $0 \leq x \leq 1$.

Let’s also denote: $r^{-1} = \sup_{t \in \tau \leq T} \left( (H_s(t))^{(0)} - H_s(t) \right)^{-1}$,

$$\dot{H}_s(t) = \frac{d}{dx} H_s(t), \quad \ddot{H}_s(t) = \frac{d^2}{dx^2} H_s(t), \quad \dot{\|\dot{H}\|} = \sup_{t \in \tau \leq T} \left| \dot{H}_s(t) \right|, \quad \ddot{\|\dot{H}\|} = \sup_{t \in \tau \leq T} \left| \ddot{H}_s(t) \right|.$$

$$\rho^{(m)}_s = \frac{d}{dx} \rho^{(m)}_s, \quad \dot{\rho}^{(m)}_s = \frac{d^2}{dx^2} \rho^{(m)}_s, \quad \dot{\|\rho^{(m)}_s\|} = \sup_{t \in \tau \leq T} \left| \rho^{(m)}_s \right|, \quad \ddot{\|\rho^{(m)}_s\|} = \sup_{t \in \tau \leq T} \left| \ddot{\rho}^{(m)}_s \right|, \quad m = 0, 1.$$

Note that existence of all these derivatives follows from conditions (C3) and (C4). Now we state some asymptotic results for estimator (1.3), which have proved in [5].

**Theorem 2.2 [5] (uniform strong consistency with rate).** Assume (C1)-(C4), $\tau_{F_s} < \tau < T < T_{F_s}$. If $h_n \to 0$, $\frac{nh_n^5}{\log n} = O(1)$, as $n \to \infty$, then

$$\sup_{\tau \leq t \leq T} \left| F_{ab}(t) - F_s(t) \right| = O\left( \left( \frac{\log n}{nh_n^5} \right)^{\frac{1}{2}} \right).$$
Theorem 2.3 [5] (almost sure asymptotic representation with weighted sums). Under the conditions of Theorem 2.2 with \( r > 0 \), we have for \( t \in (\tau_{(T)}, T) \):

\[
F_{ab}(t) - F_a(t) = \sum_{i=1}^{n} \omega_{n}(x_i) \Psi_{\omega}(\xi_i, \chi_{i}^{(0)}, \chi_{i}^{(1)}, \chi_{i}^{(2)}) + q_a(t, x),
\]

where

\[
\Psi_{\omega}(\xi_i, \chi_{i}^{(0)}, \chi_{i}^{(1)}, \chi_{i}^{(2)}) = (1 - F_a(t)) \left[ p_{x}^{(i)} \left( (H_x(t))^{\rho_{(0)}} - H_x(t) \right) \right]^{-1} \left( I(\xi_i \leq t) - H_x(t) \right)
\]

\[
\times \left[ \frac{p_{x}^{(i)}}{1 - p_{x}^{(0)}} \log \left[ 1 - (H_x(t))^{1 - \rho_{(0)}} \right] + \frac{p_{x}^{(i)}}{1 - p_{x}^{(0)}} H_x(t) \log H_x(t) \left( (H_x(t))^{\rho_{(0)}} - H_x(t) \right) \right]^{-1}
\]

\[
\times (\chi_{i}^{(0)} - p_{x}^{(0)}) - \frac{1}{1 - p_{x}^{(0)}} \log \left[ 1 - (H_x(t))^{1 - \rho_{(0)}} \right] \left( \chi_{i}^{(0)} - p_{x}^{(0)} \right).
\]

and as \( n \to \infty \),

\[
\sup_{x \leq t \leq } |q_a(t, x)|^{C_{\omega}} = O\left( \frac{\log n}{n} \right).
\]

Corollary. Under the conditions of Theorem 2.3, and as \( n \to \infty \), for \( \tau \leq t \leq T \):

\[
\left( n h_n \right)^{-1/2} (F_{ab}(t) - F_a(t)) \xrightarrow{\text{d}} \text{d} \left( a_{i}(t), \sigma_{\omega}(t) \right);
\]

where

\[
a_{i}(t) = \frac{1}{2} \left( 1 - F_{a}(t) \right) \left[ p_{x}^{(i)} \left( (H_x(t))^{\rho_{(0)}} - H_x(t) \right) \right]^{-1} \left( I(\xi_i \leq t) - H_x(t) \right)
\]

\[
- \left[ \frac{p_{x}^{(i)}}{1 - p_{x}^{(0)}} \log \left[ 1 - (H_x(t))^{1 - \rho_{(0)}} \right] + \frac{p_{x}^{(i)}}{1 - p_{x}^{(0)}} H_x(t) \log H_x(t) \left( (H_x(t))^{\rho_{(0)}} - H_x(t) \right) \right]^{-1} p_{x}^{(0)}
\]

\[
- \frac{1}{1 - p_{x}^{(0)}} \log \left[ 1 - (H_x(t))^{1 - \rho_{(0)}} \right] \left( \chi_{i}^{(0)} - p_{x}^{(0)} \right) m_{x}(x) C_{\omega}^{5/2},
\]

\[
\sigma_{\omega}(t) = \| \omega \| \left( 1 - F_{a}(t) \right)^{2} \gamma_{a}(t).
\]
with
\[
\gamma_s(t) = A_s(t) H_s(t) (1 - H_s(t)) + B_s(t) p_s^{(0)} (1 - p_s^{(0)}) + C_s(t) p_s^{(1)} (1 - p_s^{(1)}) - 2B_s(t) C_s(t) p_s^{(0)} p_s^{(1)},
\]
\[
A_s(t) = p_s^{(0)} \left[ (H_s(t))^{p_s^{(0)}} - H_s(t) \right]^{-1},
\]
\[
B_s(t) = - \left[ \frac{p_s^{(1)}}{1 - p_s^{(0)}} C_s(t) + \frac{A_s(t)}{1 - p_s^{(0)}} H_s(t) \log H_s(t) \right],
\]
\[
C_s(t) = - \frac{1}{(1 - p_s^{(0)})} \log \left[ 1 - (H_s(t))^{1 - p_s^{(0)}} \right].
\]

It is necessary to note that Theorems 2.1-2.4 are extended the corresponding theorems in conditional PHM of Koziol-Green from [6].

In the next Section 3 we use these theorems for investigation the properties of the estimator of mean conditional residual life function.

3. Asymptotic Normality of Estimator of Mean Conditional Residual Life Function

The conditional residual lifetime distribution defined as
\[
F_s(s|t) = P(Z_s - t \leq s|Z_s > t),
\]
i.e. the d.f. of residual lifetime, conditional on survival upon a given time \( t \) and at a given value of the covariate \( x \). Then for \( 0 < s < T_{F_x} \),
\[
F_s(s|t) = \frac{F_s(t+s) - F_s(t)}{1 - F_s(t)}. \tag{3.1}
\]

One of main characteristics of d.f. (3.1) is its mean, i.e. mean conditional residual life function
\[
\mu_s(t) = E(Z_s - t|Z_s > t) = \left(1 - F_s(t)\right)^{-1} \int_t^\infty (1 - F_s(s)) ds, \quad t > 0. \tag{3.2}
\]

We estimate functional \( \mu_s(t) \) by plugging in estimator (1.3) instead of \( F_s \) in (3.2). But from section 2 we know that estimator (1.3) have consistent properties in some interval \([\tau, T]\) with \( \tau < \tau T < T_{F_x} \). Therefore, we will consider the following truncated version of (3.2):
\[
\mu_s^T(t) = \left(1 - F_s(t)\right)^{-1} \int_t^T (1 - F_s(s)) ds, \quad \tau < t < T. \tag{3.3}
\]

Now we estimate (3.3) by statistics
\[
\mu_s^{Tn}(t) = \left(1 - F_{s|n}(t)\right)^{-1} \int_t^T (1 - F_{s|n}(s)) ds, \quad \tau < t < T. \tag{3.4}
\]

We have following asymptotic normality result.

**Theorem 3.1.** Assume (C1)-(C3) in \([\tau, T]\) with \( \tau < \tau T < T_{F_x} \).

(A) If \( nh_n^3 \to 0 \) and \( \log \left( \frac{nh_n^2}{T} \right) \to 0 \), as \( n \to \infty \), then
\[
(nh_n)^{1/2} \left( \mu_s^{Tn}(t) - \mu_s^T(t) \right) \to D N \left( 0, \beta_s^2(t) \right);
\]

(B) If \( h_n = Cn^{-1/5} \) for some \( C > 0 \), then as \( n \to \infty \),
\[
(nh_n)^{1/2} \left( \mu_s^{Tn}(t) - \mu_s^T(t) \right) \to D N \left( \alpha, \beta_s^2(t) \right).
Here
\[
\alpha_s(t) = \frac{1}{2} C^{1/2} m_s (\pi) \frac{1}{1 - F_s(t)} \left\{ \int_a \left( s - a_s(t) \right) ds - \int_a \left( 1 - F_s(s) \right) ds \right\},
\]
\[
\beta_s(t) = \left\| \mathbb{E} \left[ \frac{1}{(1 - F_s(t))^{1/2}} \int_a \left( 1 - F_s(s) \right) ds \right] \right\|^2 \gamma_s(t),
\]
and \( \gamma_s(t) \) from Theorem 2.4.

**Proof of theorem 3.1.** By standard manipulations and Theorem 2.3 we have that
\[
\mu_s^a(t) - \mu_s^b(t) = \int_a \left( \frac{1 - F_{ab}(s)}{1 - F_b(t)} - \frac{1 - F_b(s)}{1 - F_b(t)} \right) ds = M_{as} (t) + \sum_{i=1}^4 Q_{as}^{(i)} (t),
\]
where
\[
M_{as} (t) = \sum_{i=1}^4 \omega_a \left( x_i; h_a \right) \left[ -\frac{1}{1 - F_b(t)} \int_a \psi_a \left( \xi_i, Z_i^{(a)}, Z_i^{(b)} \right) + \frac{\psi_a \left( \xi_i, Z_i^{(a)}, Z_i^{(b)} \right)}{(1 - F_b(t))^2} \int_a \left( 1 - F_b(s) \right) ds \right],
\]
\[
Q_{as}^{(1)} (t) = -\frac{1}{1 - F_b(t)} \int_a q_a(s;x) ds,
\]
\[
Q_{as}^{(2)} (t) = \frac{q_a(t;x)}{(1 - F_b(t))} \int_a \left( 1 - F_b(s) \right) ds,
\]
\[
Q_{as}^{(3)} (t) = \frac{(F_{ab}(t) - F_b(t))^2}{(1 - F_b(t))^2} \int_a \left( F_{ab}(s) - F_b(s) \right) ds,
\]
\[
Q_{as}^{(4)} (t) = \frac{(F_{ab}(t) - F_b(t))^2}{(1 - F_b(t))^2} \int_a \left( 1 - F_{ab}(s) \right) ds.
\]

For \( Q_{as}^{(1)} \) and \( Q_{as}^{(2)} \) we use Theorem 2.3, for \( Q_{as}^{(3)} \) and \( Q_{as}^{(4)} \) Theorem 2.2. Then we see that all these remainder terms uniformly on \([\tau, T]\) almost surely have order \( O \left( \left( nh_a^{-1} \right) \log n \right) \).

Now statements (A) and (B) of theorem follows from corresponding statements of the theorem 2.4 by standard arguments.

Theorem 3.1 is proved.

**Acknowledgements**

This work is supported by Grant F4-01 of Fundamental Research Found of Uzbekistan.

**References**


[http://dx.doi.org/10.1023/A:1003166728321](http://dx.doi.org/10.1023/A:1003166728321)


Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.