

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

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

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 a 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, \dots, 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 $0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq 1$ at which we consider nonnegative independent responses Z_1, \dots, Z_n . Suppose that these responses are censored from the left and right by

nonnegative r.v.-s L_1, \dots, L_n and Y_1, \dots, 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 $\chi_i^{(2)} = I(L_i \leq Y_i < Z_i)$, where $I(A)$ denote the indicator of event A . Hence the observed data is consist of n vectors:

$$S^{(n)} = \left\{ \left(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}, X_i \right), i = 1, \dots, n \right\}.$$

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

Let H_x and N_x are conditional d.f.-s of ξ_x and $\eta_x = \min(Z_x, Y_x)$ for $X = x$. Then it's easy to see that $H_x(t) = K_x(t)N_x(t)$ with $N_x(t) = 1 - (1 - F_x(t))(1 - G_x(t))$, $t \geq 0$. In particular, if for all $x \in [0, 1]$, $P(L_x \leq Y_x) = 1$, 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_x of lifetimes and its functionals from the samples $S^{(n)}$ under nuisance d.f.-s K_x and G_x . The first product-limit type estimators for F_x in the case of no censoring from the left (that is $P(L_x = -\infty) = 1$ or $K_x(t) \equiv 1$) 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_x from [4] [5] for estimation the mean conditional residual life function. Suppose that d.f.-s K_x and G_x are expressed from F_x by following parametric relationships for all $t \geq 0$:

$$\begin{cases} 1 - G_x(t) = (1 - F_x(t))^{\theta_x}, \\ K_x(t) = (N_x(t))^{\beta_x}, \end{cases} \tag{1.1}$$

where θ_x and β_x 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 $\beta_x \equiv 0$). Estimation of F_x 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_x :

$$1 - F_x(t) = \left[1 - (H_x(t))^{\lambda_x} \right]^{\gamma_x}, \quad t \geq 0, \tag{1.2}$$

where $\lambda_x = \frac{1}{1 + \beta_x} = 1 - p_x^{(0)}$, $\gamma_x = \frac{1}{1 + \theta_x} = \frac{p_x^{(1)}}{1 - p_x^{(0)}}$ and $p_x^{(m)} = P(\chi_x^{(m)} = 1)$, $m = 0, 1, 2$, with

$p_x^{(0)} + p_x^{(1)} + p_x^{(2)} = 1$. Then estimator of F_x one can constructed by natural plugging method as follows:

$$1 - F_{xh}(t) = \left\{ 1 - [H_{xh}(t)]^{\lambda_{xh}} \right\}^{\gamma_{xh}}, \quad t \geq 0. \tag{1.3}$$

Here $\gamma_{xh} = p_{xh}^{(1)}(1 - p_{xh}^{(0)})^{-1}$, $\lambda_{xh} = 1 - p_{xh}^{(0)}$,

$$p_{xh}^{(m)} = \sum_{i=1}^n \omega_{ni}(x; h_n) \chi_i^{(m)}, \quad m = 0, 1, 2,$$

and

$$H_{xh}(t) = \sum_{i=1}^n \omega_{ni}(x; h_n) I(\xi_i \leq t),$$

are smoothed estimators of $\lambda_x, \gamma_x, p_x^{(m)}$ and $H_x(t)$, used Gasser-Müllers weights $\{\omega_{ni}(x; h_n)\}_{i=1}^n$:

$$\omega_{ni}(x; h_n) = \int_{x_{i-1}}^{x_i} \frac{1}{h_n} \pi\left(\frac{x-y}{h_n}\right) dy \left(\int_0^{x_i} \frac{1}{h_n} \pi\left(\frac{x-y}{h_n}\right) dy \right)^{-1},$$

$x_0 = 0$, $\pi(y)$ is a known probability density function (kernel), and $\{h_n\}$ is a sequence of positive constants tending to 0 as $n \rightarrow \infty$, called the bandwidth sequence. Note that in the case of no censoring from the left the estimator (1.3) is 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, \dots, x_n and kernel π we denote

$$\underline{\Delta}_n = \min_{1 \leq i \leq n} (x_i - x_{i-1}), \quad \bar{\Delta}_n = \max_{1 \leq i \leq n} (x_i - x_{i-1}),$$

$$\|\pi\|_2^2 = \int_{-\infty}^{\infty} \pi^2(y) dy, \quad m_\nu(\pi) = \int_{-\infty}^{\infty} y^\nu \pi(y) dy, \quad \nu = 1, 2.$$

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

$$\tau_{F_x} = \tau_{G_x} = \tau_{N_x} = \tau_{K_x} = \tau_{H_x} \quad \text{and} \quad T_{F_x} = T_{G_x} = T_{N_x} = T_{K_x} = T_{H_x}.$$

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. ξ_x and the vector $(\chi_x^{(0)}, \chi_x^{(1)}, \chi_x^{(2)})$ 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 $n \rightarrow \infty$, $x_n \rightarrow 1$, $\bar{\Delta}_n = O\left(\frac{1}{n}\right)$, $\bar{\Delta}_n - \underline{\Delta}_n = o\left(\frac{1}{n}\right)$.

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

(C3) $\dot{F}_x(t) = \frac{\partial}{\partial x} F_x(t)$ and $\ddot{F}_x(t) = \frac{\partial^2}{\partial x^2} F_x(t)$ exist and are continuous for $0 \leq x \leq 1$ and $\tau \leq t \leq T$, with $\tau_{F_x} < \tau < T < T_{F_x}$.

(C4) $\dot{\theta}_x = \frac{d}{dx} \theta_x$ and $\dot{\beta}_x = \frac{d}{dx} \beta_x$ exist and are continuous for $0 \leq x \leq 1$.

Let's also denote: $r^{-1} = \sup_{\tau \leq t \leq T} \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1}$,

$$\dot{H}_x(t) = \frac{\partial}{\partial x} H_x(t), \quad \ddot{H}_x(t) = \frac{\partial^2}{\partial x^2} H_x(t), \quad \|\dot{H}\| = \sup_{(x;t) \in [0,1] \times [\tau, T]} |\dot{H}_x(t)|, \quad \|\ddot{H}\| = \sup_{(x;t) \in [0,1] \times [\tau, T]} |\ddot{H}_x(t)|,$$

$$\dot{p}_x^{(m)} = \frac{d}{dx} p_x^{(m)}, \quad \ddot{p}_x^{(m)} = \frac{d^2}{dx^2} p_x^{(m)}, \quad \|\dot{p}_x^{(m)}\| = \sup_{0 \leq x \leq 1} |\dot{p}_x^{(m)}|, \quad \|\ddot{p}_x^{(m)}\| = \sup_{0 \leq x \leq 1} |\ddot{p}_x^{(m)}|, \quad m = 0, 1.$$

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), $\tau_{F_x} < \tau < T < T_{F_x}$. If $h_n \rightarrow 0$, $\frac{nh_n^5}{\log n} = O(1)$, as $n \rightarrow \infty$, then

$$\sup_{\tau \leq t \leq T} |F_{xh}(t) - F_x(t)| = O\left(\left(\frac{\log n}{nh_n} \right)^{1/2} \right).$$

Theorem 2.3 [5] (almost sure asymptotic representation with weighted sums). Under the conditions of Theorem 2.2 with $r > 0$, we have for $t \in (\tau_{F_x}, T_{F_x})$:

$$F_{xh}(t) - F_x(t) = \sum_{i=1}^n \omega_{ni}(x; h_n) \Psi_{ix}(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}) + q_n(t, x),$$

where

$$\begin{aligned} \Psi_{ix}(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}) = & (1 - F_x(t)) \left\{ p_x^{(1)} \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1} (I(\xi_i \leq t) - H_x(t)) \right. \\ & - \left[\frac{p_x^{(1)}}{(1 - p_x^{(0)})^2} \log \left[1 - (H_x(t))^{1 - p_x^{(0)}} \right] + \frac{p_x^{(1)}}{1 - p_x^{(0)}} H_x(t) \log H_x(t) \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1} \right] \\ & \left. \times \left(\chi_i^{(0)} - p_x^{(0)} \right) - \frac{1}{1 - p_x^{(0)}} \log \left[1 - (H_x(t))^{1 - p_x^{(0)}} \right] \left(\chi_i^{(1)} - p_x^{(1)} \right) \right\}. \end{aligned}$$

and as $n \rightarrow \infty$,

$$\sup_{\tau \leq t \leq T} |q_n(t, x)| \stackrel{\text{a.s.}}{=} O\left(\frac{\log n}{nh_n}\right).$$

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

$$(nh_n)^{1/2} (F_{xh}(t) - F_x(t)) \stackrel{\text{a.s.}}{=} (nh_n)^{1/2} \sum_{i=1}^n \omega_{ni}(x; h_n) \Psi_{ix}(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}) + O\left(\frac{\log n}{(nh_n)^{1/2}}\right).$$

Theorem 2.4 [5] (asymptotic normality). Assume (C1)-(C4). $\tau_{F_x} < \tau < T < T_{F_x}$.

(A) If $nh_n^5 \rightarrow 0$ and $(nh_n)^{1/2} \log n \rightarrow 0$, then for $\tau \leq t \leq T$, as $n \rightarrow \infty$,

$$(nh_n)^{1/2} (F_{xh}(t) - F_x(t)) \xrightarrow{d} N(0, \sigma_x^2(t));$$

(B) If $h_n = Cn^{-1/5}$ for some $C > 0$, then for $\tau \leq t \leq T$, as $n \rightarrow \infty$,

$$(nh_n)^{1/2} (F_{xh}(t) - F_x(t)) \xrightarrow{d} N(a_x(t), \sigma_x^2(t)),$$

where

$$\begin{aligned} a_x(t) = & \frac{1}{2} (1 - F_x(t)) \left\{ p_x^{(1)} \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1} \ddot{H}_x(t) \right. \\ & - \left[\frac{p_x^{(1)}}{(1 - p_x^{(0)})^2} \log \left[1 - (H_x(t))^{1 - p_x^{(0)}} \right] + \frac{p_x^{(1)}}{(1 - p_x^{(0)})} H_x(t) \log H_x(t) \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1} \right] \ddot{p}_x^{(0)} \\ & \left. - \frac{1}{1 - p_x^{(0)}} \log \left[1 - (H_x(t))^{1 - p_x^{(0)}} \right] \ddot{p}_x^{(1)} \right\} m_2(\pi) C^{5/2}, \end{aligned}$$

$$\sigma_x^2(t) = \|\pi\|_2^2 (1 - F_x(t))^2 \gamma_x(t),$$

with

$$\begin{aligned} \gamma_x(t) &= A_x^2(t)H_x(t)(1-H_x(t)) + B_x^2(t)p_x^{(0)}(1-p_x^{(0)}) + C_x^2(t)p_x^{(1)}(1-p_x^{(1)}) - 2B_x(t)C_x(t)p_x^{(0)}p_x^{(1)}, \\ A_x(t) &= p_x^{(1)} \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1}, \\ B_x(t) &= - \left[\frac{p_x^{(1)}}{(1-p_x^{(0)})} C_x(t) + \frac{A_x(t)}{(1-p_x^{(0)})} H_x(t) \log H_x(t) \right], \\ C_x(t) &= - \frac{1}{(1-p_x^{(0)})} \log \left[1 - (H_x(t))^{1-p_x^{(0)}} \right]. \end{aligned}$$

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

$$F_x(s/t) = P(Z_x - t \leq s/Z_x > t),$$

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 $0 < s < T_{F_x}$,

$$F_x(s/t) = \frac{F_x(t+s) - F_x(t)}{1 - F_x(t)}. \tag{3.1}$$

One of main characteristics of d.f. (3.1) is its mean, i.e. mean conditional residual life function

$$\mu_x(t) = E(Z_x - t/Z_x > t) = (1 - F_x(t))^{-1} \int_t^\infty (1 - F_x(s)) ds, \quad t > 0. \tag{3.2}$$

We estimate functional $\mu_x(t)$ by plugging in estimator (1.3) instead of F_x in (3.2). But from section 2 we know that estimator (1.3) have consistent properties in some interval $[\tau, T]$ with $\tau_{F_x} < \tau < T < T_{F_x}$. Therefore, we will consider the following truncated version of (3.2):

$$\mu_x^T(t) = (1 - F_x(t))^{-1} \int_t^T (1 - F_x(s)) ds, \quad \tau < t < T. \tag{3.3}$$

Now we estimate (3.3) by statistics

$$\mu_{xh}^T(t) = (1 - F_{xh}(t))^{-1} \int_t^T (1 - F_{xh}(s)) ds, \quad \tau < t < T. \tag{3.4}$$

We have following asymptotic normality result.

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

(A) If $nh_n^5 \rightarrow 0$ and $\frac{\log n}{(nh_n)^{1/2}} \rightarrow 0$, as $n \rightarrow \infty$, then

$$(nh_n)^{1/2} (\mu_{xh}^T(t) - \mu_x^T(t)) \xrightarrow{d} N(0, \beta_x^2(t));$$

(B) If $h_n = Cn^{-1/5}$ for some $C > 0$, then as $n \rightarrow \infty$,

$$(nh_n)^{1/2} (\mu_{xh}^T(t) - \mu_x^T(t)) \xrightarrow{d} N(\alpha_x(t), \beta_x^2(t)).$$

Here

$$\alpha_x(t) = \frac{1}{2} C^{5/2} m_2(\pi) \frac{1}{1-F_x(t)} \left\{ \int_t^T a_x(s) ds - a_x(t) \int_t^T (1-F_x(s)) ds \right\},$$

$$\beta_x^2(t) = \|\pi\|_2^2 \frac{1}{(1-F_x(t))^2} \int_t^T \left(\int_t^T (1-F_x(s)) ds \right)^2 d\gamma_x(s),$$

and $\gamma_x(t)$ from Theorem 2.4.

Proof of theorem 3.1. By standard manipulations and Theorem 2.3 we have that

$$\mu_{xh}^T(t) - \mu_x^T(t) = \int_t^T \left[\frac{1-F_{xh}(s)}{1-F_{xh}(t)} - \frac{1-F_x(s)}{1-F_x(t)} \right] ds = M_{nx}(t) + \sum_{k=1}^4 Q_{nx}^{(k)}(t),$$

where

$$M_{nx}(t) = \sum_{i=1}^n \omega_{ni}(x; h_n) \left[-\frac{1}{1-F_x(t)} \int_t^T \Psi_{sx}(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}) + \frac{\Psi_{tx}(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)})}{(1-F_x(t))^2} \int_t^T (1-F_x(s)) ds \right],$$

$$Q_{nx}^{(1)}(t) = -\frac{1}{1-F_x(t)} \int_t^T q_n(s; x) ds,$$

$$Q_{nx}^{(2)}(t) = \frac{q_n(t; x)}{(1-F_x(t))^2} \int_t^T (1-F_x(s)) ds,$$

$$Q_{nx}^{(3)}(t) = \frac{(F_{xh}(t) - F_x(t))}{(1-F_x(t))^2} \int_t^T (F_{xh}(s) - F_x(s)) ds,$$

$$Q_{nx}^{(4)}(t) = \frac{(F_{xh}(t) - F_x(t))^2}{(1-F_x(t))^2} \int_t^T (1-F_{xh}(s)) ds.$$

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

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

- [1] Beran, R (1981) Nonparametric Regression with Randomly Censored Survival Data. Technical Report, University of California, Berkeley, 19.
- [2] Van Keilegom, I. and Veraverbeke, N. (1996) Uniform Strong Convergence for the Conditional Kaplan-Meier Estimator and Its Quantiles. *Communications in Statistics—Theory and Methods*, **25**, 2251-2265.
- [3] Van Keilegom, I. and Veraverbeke, N. (1997) Estimation and Bootstrap with Censored Data in Fixed Design Nonparametric Regression. *Annals of the Institute of Statistical Mathematics*, **49**, 467-491. <http://dx.doi.org/10.1023/A:1003166728321>
- [4] Abdikalikov, F.A. and Abdushukurov, A.A. (2012) Informative Regression Model under Random Censorship from Both Sides and Estimating of Survival Function. Book of Abstracts, *XXX Internet. Seminar on Stability Problems for stochastic Models*, Svetlogorsk, 24-28 September 2012, 3-4.

- [5] Abdikalikov, F.A. and Abdushukurov, A.A. (2012) Semiparametrical Estimation of Conditional Survival Function in Informative Regression Model of Random Censorship from Both Sides. In: *Statistical Methods of Estimation and Hypothesis Testing*, Perm State University, Perm, Issue 24, 145-162. (In Russian)
- [6] Veraverbeke, N. and Cadarso-Suárez, C. (2000) Estimation of the Conditional Distribution in a Conditional Koziol-Green Model. *Test*, **9**, 97-122.
- [7] Abdushukurov, A.A. (1994) Model of Random Censoring from Both Sides and Independence Criterion for It. *Doclady Acad. Sci. Uzbekistan*, **11**, 8-9. (In Russian)
- [8] Abdushukurov, A.A. (1998) Nonparametric Estimation of Distribution Function Based on Relative Risk Function. *Communications in Statistics—Theory and Methods*, **27**, 1991-2012. <http://dx.doi.org/10.1080/03610929808832205>
- [9] Abdikalikov, F.A. and Abdushukurov, A.A. (2012) An Investigating of Power-Type Estimators of Lifetime Functions in Regression Models. LAMBERT Academic Publishing (LAP), Germany, 89. (In Russian)

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](#).

