New Trends in Mathematical Sciences
http://www.ntmsci.com

On estimation of mean conditional residual life function under random censored data

A. A. Abdushukurov[1] *and F. A. Abdikalikov*[1]

[1] Department of Probability Theory and Mathematical Statistics, National University of Uzbekistan 100174, Tashkent, Uzbekistan.

Received: 26 July 2014, Revised: 4 December 2014, Accepted: 4 December 2014
Published online: 8 January 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: Mean residual life function, Fixed design regression model,Asymptotic normality.

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, often called prognostic factor (or covariate). X represent e.g. the dose of a drug for individual or some environmental conditions of a machine (temperature, pressure,). Moreover, in such practical situations often occurs that not all of survival times $Z_1,...,Z_n$ 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 $0 \le x_1 \le x_2 \le ... \le x_n \le 1$ at which we consider nonnegative independent responses $Z_1,...,Z_n$. Suppose that these responses are censored from the left and right by nonnegative v.-s $L_1,...,L_n$ and $Y_1,...,Y_n$ and the observed r.v.-s at design points x_i are in fact $\left\{\xi_i,\chi_i^{(0)},\chi_i^{(1)},\chi_i^{(2)}\right\}$ with $\xi_i = \max\left(L_i,\min(Z_i,Y_i)\right)$, $\chi_i^{(0)} = I\left(\min(Z_i,Y_i<L_i)\right)$, $\chi_i^{(1)} = I\left(L_i \le Z_i \le Y_i\right)$ and $\chi_i^{(2)} = (L_i \le Y_i < Z_i)$, where $I(A)$ denote the indicator of event A. Hence the observed data is consist of n vectors:

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

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_is 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 easy to see that $H_x(t) = K_x(t)N_x(t)$ with $N_x(t) = 1-(1-F_x(t))(1-G_x(t)), t \ge 0$. In particular, if for all $x \in [0,1]$, $P(L_x \le Y_x) = 1$, 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 functional 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 $K_x(t) = 1$) proposed by Beran [6] and has been investigated by many

* Corresponding author e-mail: a_abdushukurov@rambler.ru

authors (see, for example [7,8]). In this article supposing that the random censoring from both sides is informative we use twice power type estimator of F_x from [3,4] 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 $t \geq 0$:

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

whereθ_x and $\beta_x$are positive unknown nuisance parameters, depending on the co-variate value x. Informative model (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 [9]. Model (1) one can considered as an extended two sided conditional PHM. In the case of no covariates, model (1) first is proposed in [1,2]. It is not difficult to verify that from (1) one can obtain following expression of d.f.F_x:

$$1 - F_x(t) = \left[1 - (H_x(t))^{\lambda_x}\right]^{\gamma_x}, \quad t \geq 0, \tag{2}$$

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 $p_x^{(m)} = P\left(\chi_x^{(m)} = 1\right)$, $m = 0,1,2$, 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:

$$1 - F_{xh}(t) = \left\{1 - [H_{xh}(t)]^{\lambda_{xh}}\right\}^{\gamma_{xh}}, \; t \geq 0. \tag{3}$$

Here $\gamma_{xh} = p_{xh}^{(1)}\left(1 - p_{xh}^{(0)}\right)^{-1}$, $\lambda_{xh} = 1 - p_{xh}^{(0)}$, $p_{xh}^{(m)} = \sum\limits_{i=1}^{n} \omega_{ni}(x;h_n)\chi_i^{(m)}$, $m = 0,1,2$, and $H_{xh}(t) = \sum\limits_{i=1}^{n} \omega_{ni}(x;h_n) I(\xi_i \leq t)$, are smoothed estimators of $\lambda_x, \gamma_x, p_x^{(m)}$ and $H_x(t)$ used Gausser-Müllers wights $\{\omega_{ni}(x;h_n)\}_{i=1}^{n}$:

$$\omega_{ni}(x;h_n) = \int\limits_{x_{i-1}}^{x_i} \frac{1}{h_n}\pi\left(\frac{x-y}{h_n}\right)dy \left(\int\limits_{0}^{x_n} \frac{1}{h_n}\pi\left(\frac{x-y}{h_n}\right)dy\right)^{-1},$$

$x_0 = 0$, $\pi(y)$ is a known probability density function(kernel), and $\{h_n\}$ is a sequence of positive constants tending to 0 as $n \to \infty$, called the bandwidth sequence. Note that in the case of no censoring from the left the estimator (3) is coincides with estimator in conditional Koziol-Green model in [9]. Note also that a class of power type estimators for conditional d.f.-s for several models authors have considered in book [5]. Estimator (3) was presented in [3] and its asymptotic properties has been investigated in [4]. Now we demonstrate some of these results.

2 Asymptotic results for estimator of conditional distribution function

For asymptotic properties of estimator (3) we need some notations. For the design points $x_1,\ldots,x_n$ and kernel πwe denote $\underline{\Delta}_n = \min\limits_{1\leq i\leq n}(x_i - x_{i-1})$, $\overline{\Delta}_n = \max\limits_{1\leq i\leq n}(x_i - x_{i-1})$, $\|\pi\|_2^2 = \int\limits_{-\infty}^{\infty}\pi^2(y)\,dy$, $m_\nu(\pi) = \int\limits_{-\infty}^{\infty} y^\nu \pi(y)\,dy$, $\nu = 1,2$. 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): $\tau_{F_x} = \tau_{G_x} = \tau_{N_x} = \tau_{K_x} = \tau_{H_x}$ and $T_{F_x} = T_{G_x} = T_{N_x} = T_{K_x} = T_{H_x}$. In [4] authors have proved the following property of two sided conditional PHM (1).

Theorem 2.1 [4]. For a given covariate x, the model (1) holds if and only if r.v.ξ_x and the vector $\left(\chi_x^{(0)}, \chi_x^{(1)}, \chi_x^{(2)}\right)$ are independent. This characterization of submodel (1) plays an important role for investigation the properties of estimator

(3). Lets introduce some conditions:
(C1) As $n \to \infty$, $x_n \to 1$, $\overline{\Delta}_n = O\left(\frac{1}{n}\right)$, $\overline{\Delta}_n - \underline{\Delta}_n = o\left(\frac{1}{n}\right)$.
(C2) π is a probability density function with compact support [-M,M] for some $M > 0$, with $m_1(\pi) = 0$ and $|\pi(y) - \pi(y')| \le C_\pi |y - y'|$, 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 $0 \le x \le 1$ and $\tau \le t \le T$, with $\tau_{F_x} < \tau < T < T_{F_x}$.
(C4) $\dot{\theta}_x = \frac{\partial}{\partial x}\theta_x$ and $\dot{\beta}_x = \frac{\partial}{\partial x}\beta_x$ exist and are continuous for $0 \le x \le 1$.
Lets also denote: $r^{-1} = \sup\limits_{\tau \le t \le T} \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1}$,

$$\dot{H}_x(t) = \frac{\partial}{\partial x} H_x(t), \;\; \ddot{H}_x(t) = \frac{\partial^2}{\partial x^2} H_x(t),$$

$$\left\| \dot{H} \right\| = \sup_{(x;t) \in [0,1] \times [\tau,T]} \left| \dot{H}_x(t) \right|,$$

$$\left\| \ddot{H} \right\| = \sup_{(x;t) \in [0,1] \times [\tau,T]} \left| \ddot{H}_x(t) \right|,$$

$$\dot{p}_x^{(m)} = \frac{d}{dx} p_x^{(m)}, \ddot{p}_x^{(m)} = \frac{d^2}{dx^2} p_x^{(m)},$$

$$\left\| \dot{p}_x^{(m)} \right\| = \sup_{0 \le x \le 1} \left| \dot{p}_x^{(m)} \right|, \;\; \left\| \ddot{p}_x^{(m)} \right\| = \sup_{0 \le x \le 1} \left| \ddot{p}_x^{(m)} \right|, \; m = 0, 1.$$

Note that existence of all these derivatives follows from conditions (C3) and (C4). Now we state some asymptotic results for estimator (3), which have proved in [4].

Theorem 2.2 [4] (uniform strong consistency with rate). Assume (C1)-(C4), $\tau_{F_x} < \tau < T < T_{F_x}$. If $h_n \to 0$, $\frac{nh_n^5}{\log n} = O(1)$, as $n \to \infty$, then

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

Theorem 2.3 [4] (almost sure asymptotic representation with weighted sums). Under the conditions of theorem 2.2 with $r > 0$, we have for $t \in (\tau_{F_x}, T_{F_x})$:

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

where

$$\Psi_{tx}\left(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)} \right) = (1 - F_x(t)) \left\{ p_x^{(1)} \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1} \cdot \right.$$

$$\cdot \left(I(\xi_i \le t) - H_x(t) \right) - \left[\frac{p_x^{(1)}}{\left(1 - p_x^{(0)}\right)^2} \log \left[1 - (H_x(t))^{1 - p_x^{(0)}} \right] + \right.$$

$$\left. + \frac{p_x^{(1)}}{1 - p_x^{(0)}} H_x(t) \log H_x(t) \left[(H_x(t))^{p_x^{(0)}} - H_x(t) \right]^{-1} \right] \left(\chi_i^{(0)} - p_x^{(0)} \right) -$$

$$\left. - \frac{1}{1 - p_x^{(0)}} \log \left[1 - (H_x(t))^{1 - p_x^{(0)}} \right] \left(\chi_i^{(1)} - p_x^{(1)} \right) \right\}$$

and as $n \to \infty$,

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

Corollary. Under the conditions of theorem 2.3, and as $n \to \infty$, for $\tau \le t \le T$:

$$(nh_n)^{1/2}(F_{xh}(t) - F_x(t)) \overset{\text{a.s.}}{=} (nh_n)^{1/2} \sum_{i=1}^{n} \omega_{ni}(x;h_n)\Psi_{tx}(\xi_i, \chi_i^{(0)}, \chi_i^{(1)}, \chi_i^{(2)}) +$$

$$+O\left(\frac{\log n}{(nh_n)^{1/2}}\right).$$

Theorem 2.4 [4] (asymptotic normality). Assume (C1)-(C4). $\tau_{F_x} < \tau < T < T_{F_x}$. (A) If $nh_n^5 \to 0$ and $(nh_n)^{1/2}\log n \to 0$, then for $\tau \le t \le T$, as $n \to \infty$,

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

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

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

where

$$a_x(t) = \frac{1}{2}(1 - F_x(t))\left\{ p_x^{(1)}\left[(H_x(t))^{p_x^{(0)}} - H_x(t)\right]^{-1} \ddot{H}_x(t) - \right.$$

$$-\left[\frac{p_x^{(1)}}{\left(1 - p_x^{(0)}\right)^2}\log\left[1 - (H_x(t))^{1-p_x^{(0)}}\right] + \right.$$

$$\left. + \frac{p_x^{(1)}}{\left(1 - p_x^{(0)}\right)} H_x(t)\log H_x(t)\left[(H_x(t))^{p_x^{(0)}} - H_x(t)\right]^{-1}\right]\ddot{p}_x^{(0)} -$$

$$\left. \frac{1}{1 - p_x^{(0)}} \text{lo g}\left[1 - (H_x(t))^{1-p_x^{(0)}}\right]\ddot{p}_x^{(1)}\right\} m_2(\pi)C^{5/2},$$

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

with

$$\gamma_x(t) = A_x^2(t)H_x(t)(1 - H_x(t)) + B_x^2(t)p_x^{(0)}(1 - p_x^{(0)}) +$$

$$+C_x^2(t)p_x^{(1)}(1 - p_x^{(1)}) - 2B_x(t)C_x(t)p_x^{(0)}p_x^{(1)},$$

$$A_x(t) = p_x^{(1)}\left[(H_x(t))^{p_x^{(0)}} - H_x(t)\right]^{-1},$$

$$B_x(t) = -\left[\frac{p_x^{(1)}}{(1 - p_x^{(0)})}C_x(t) + \frac{A_x(t)}{(1 - p_x^{(0)})}H_x(t)\log H_x(t)\right],$$

$$C_x(t) = -\frac{1}{(1 - p_x^{(0)})}\log\left[1 - (H_x(t))^{1-p_x^{(0)}}\right].$$

It is necessary to note that theorems 2.1-2.4 are extended the corresponding theorems in conditional PHM of Koziol-Green from [9]. 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 $F_x(s/t) = P(Z_x - t \le s/Z_x > t)$, 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 $0 < s < T_{F_x}$,

$$F_x(s/t) = \frac{F_x(t+s) - F_x(t)}{1 - F_x(t)}. \tag{4}$$

One of main characteristics of d.f. (4) is its mean, i.e. mean conditional residual life function

$$\mu_x(t) = E(Z_x - t/Z_x > t) = (1 - F_x(t))^{-1} \int_t^{\infty} (1 - F_x(s))\, ds, \ \ t > 0. \tag{5}$$

We estimate functional $\mu_x(t)$ by plugging in estimator (3) instead of F_x in (5). But from section 2 we know that estimator (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 (5):

$$\mu_x^T(t) = (1 - F_x(t))^{-1} \int_t^{T} (1 - F_x(s))\, ds, \quad \tau < t < T. \tag{6}$$

Now we estimate (6) by statistics

$$\mu_{xh}^T(t) = (1 - F_{xh}(t))^{-1} \int_t^{T} (1 - F_{xh}(s))\, ds, \quad \tau < t < T. \tag{7}$$

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

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

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

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

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

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

$$\left.+\frac{\Psi_{sx}\left(\xi_i,\chi_i^{(0)},\chi_i^{(1)},\chi_i^{(2)}\right)}{(1-F_x(t))^2}\int_t^T(1-F_x(s))\,ds\right],$$

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

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

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

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

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 $O\left((nh_n)^{-1}\log n\right)$. Now statements (A) and (B) of theorem follows from corresponding statements of the theorem 2.4 by standard arguments. Theorem 3.1 proved.

References

[1] Abdushukurov A.A., *Model of random censoring from both sides and independence criterian for it.*, Doclady Acad. Sci. Uzbekistan,(In Russian)., (1994).

[2] Abdushukurov A.A., *Nonparametric estimation of distribution function based on relative risk function.*, Commun Statist. : Theory and Methods,(1998).

[3] Abdukalikov F. A., Abdushukurov A.A., *Informative regression model under random censorship from both sides and estimating of survival function*, Book of Abstracts of XXX Internet. Seminar on Stability Problems for stochastic Models. Svetlogorsk. Russia, (2012).

[4] Abdukalikov F. A., Abdushukurov A.A., *Semiparametrical estimation of conditional survival function in informative regression model of random censorship from both sides*, In Statistical Methods of Estimation and Hypothisis Testing. Perm State Univ. Perm Russia, (In Russian),(2012).

[5] Abdukalikov F. A., Abdushukurov A.A., *An Investigating of power-type estimators of Lifetime functions in Regression Models*, LAMBERT Academic Publishing. LAP. Germany, (In Russian), (2012).

[6] Beran R , *Nonparametric regression with randomly censored survival data*, Technical Report. Univ. California. Berkeley, (1981).

[7] Van Keilegom., Veraverbeke N., *Uniform strong convergence for the conditional Kaplan-Meir estimator and its quantiles*, Commun. Statist.:Theory and Methods, (1996).

[8] Van Keilegom., Veraverbeke N., *Estimation and bootstrap with censored data in fixed design nonparametric regression*, Ann. Iust. Statist. Math,, (1997).

[9] Veraverbeke N., Cadarso-Suares C., *Estimation of conditional distribution in a conditional Koziol-Green model*, Test, (2000).