# On Survival Function Estimation in Dependent Partially Informative Random Censorship

**Abdushukurov Abdurahim Ahmedovich[1], Bozorov Suxrob Baxodirovich[2, ***]**

[1]Department of Applied Mathematics and Informatics, Tashkent Branch of Moscow State University Named After M. V. Lamanosov, Tashkent, Uzbekistan

[2]Faculty of Physics and Mathematics, Gulistan State University, Gulistan, Republic of Uzbekistan

**Email address:**
a_abdushukurov@rambler.ru (A. A. Ahmedovich), Suxrobbek_8912@mail.ru (B. S. Baxodirovich)
*Corresponding author

**Abstract:** In such areas as bio-medicine, engineering and insurance researchers are interested in positive variables, which are expressed as a time until a certain event. But observed data may be incomplete, because it is censored. Moreover, the random variables of interest (lifetimes) and censoring times can be influenced by other variable, often called prognostic factor or covariate. The basic problem is the estimation of survival function of lifetime. In this article we propose three asymptotical equivalent estimators of survival function in partially informative competing risks model. This paper deals with the estimation of a survival function with random right censoring and dependent censoring mechanism through covariate. We extend exponential – hazard, product - limit and relative - risk power estimators of survival functions in partially informative censoring model in which conditional on a covariate, the survival and censoring times are assumed to be independent. In this model, each observation is the minimum of one lifetime and two censoring times. The survival function of one of these censoring times is a power of the survival function of the lifetime. The distribution of the other censoring time has no relation with the distribution of the lifetime (non-informative censoring). For estimators we show their uniform strong consistency and convergence to same Gaussian process. Comparisons of estimators with the Jensen-Wiedmann's estimator are included.

**Keywords:** Random Censoring, Proportional Hazards Model, Exponential Hazard, Product-limit, Relative-risk, Survival Function

## 1. Introduction

In survival analysis it is often faced with censored lifetime data, i.e. with the only partially observable lifetimes. Let $\{X_i, i \geq 1\}$ be independent and identically distributed (i.i.d.) positive random variables (r.v.-s) (lifetimes) with common distribution function (d.f.) $F(t) = P(X_i \leq t)$, $t \in \square^+$. At first we consider the situation when $X_i^{'}s$ are censored from the right by i.i.d. and also by positive r.v.-s $\{Y_i, i \geq 1\}$, and the so-called censoring times, which have a common d.f. $G$, are independent from the $X_i^{'}s$. At the ninth stage of experiments, the observed data contains

$C^{(n)} = \{(Z_i, \delta_i), 1 \leq i \leq n\}$, where $Z_i = \min(X_i, Y_i)$ and $\delta_i = I(X_i \leq Y_i)$ with $I(\cdot)$ denoting the indicator function of the event in brackets. In this general right censoring model our focus is concentrated on estimation of d.f. $F$ with nuisance d.f. $G$. In considered model the most commonly used estimator of $F$ is the product-limit (PL) estimator of Kaplan and Meier. There is a tremendous literature on estimation and testing centered around the optimality properties of PL-estimator of $F$ if no additional assumptions are made in this model. This optimality is lost if informative censoring is present, i.e. if the d.f. $G$ of the censoring times carries additional information about d.f. $F$. A well-known model of informative censoring is called proportional hazards model (PHM) (or Koziol-Green model), where $1 - G$ is the

power of $1-F$:

$$1-G(t)=\left(1-F(t)\right)^{\beta}, \ \beta>0, t\in \Box^{1}.$$

The PHM of random censorship is an appealing and potentially useful special semi parametric model in which $p=P\left(X_i\leq Y_i\right)=\dfrac{1}{1+\beta}$, and the extreme case of $\beta=0$ or $p=1$ may be identified with the lack of censorship.

Abdushukurov (1984) and Cheng and Lin (1984) pointed out that in PHM the estimator $F_n^{ACL}(t)=1-\left(1-H_n(t)\right)^{P_n}$ of $F$, where

$$H_n(t)=\frac{1}{n}\sum_{i=1}^{n}I(Z_i\leq t) \text{ and } p_n=\frac{1}{n}\sum_{i=1}^{n}\delta_i,$$

is strictly better than PL – estimator in terms of asymptotic variance. S. Csörgő described a number of new results, giving a complete survey of the estimation theory based on Abdushukurov-Cheng-Lin estimator $F_n^{ACL}$ rather than the PL –estimator [8]. Namely, the estimation of survival, hazard, quantile, mean and percentile residual life, total time on test, Lorenz, density and hazard rate functions is considered, and boot strapping in this model is sketched [8]. Testing the model is also discussed (see, also [10, 12-14]). However, it has been pointed out by many authors that the assumptions of PHM are much restrictive. Therefore, it is desirable to develop generalization of the PHM that are more appropriate for practical situations. Abdushukurov [1, 2, 6], Abdushukurov, Makhmudova [4], Gather and Pawlitschko [9] have proposed some generalizations of PHM, where only a part of the censored observations is supposed to be informative.

In some cases it is not reasonable to assume independence between the lifetime and censoring variables. The dependence may be due to a covariate. For example, in competing risks situation, where some technical system fails due to one or more competing causes, only one observes the time to failure of the system and the corresponding failure cause. If a system with two failure causes $A_1$ and $A_2$ fails due to cause $A_1$, then the failure time of cause $A_2$ is randomly censored and vice versa. Since the failure times due to both causes are affected by the same stress and operating environment described by a covariate, it is likely the failure times that are positively correlated. In medical trials the survival and censoring times may be affected by a set of patients' covariates as age, blood pressure, cholesterol. Here in the considered model we assume that the survival times and the censoring times are conditionally independent in a given covariate. Take partially informative completing risks modelin the presence of covariate [4].

Let $\left\{(X_i,Y_{1i},Y_{2i}),i\geq 1\right\}$ be a sequence of independent and identically distributed triples of positive r.v.-s, where the components $X_i, Y_{1i}$ and $Y_{2i}$ are supposed to be conditionally independent given in a covariate $Z_i$. The $X_i$ 's

are lifetimes with a common continuous d.f. $F(t)$, $t\in \Box^{+}$. The $Y_{ki}$ 's, $k=1,2$, are censoring times with common continuous d.f.–s $G_k(t)$, $k=1,2$, $t\in \Box^{+}$, respectively. At the ninth stage of experiments the observed data contains a sample of triples $\left\{(\xi_i,\delta_i,Z_i),i=1,...n\right\}=C_{*}^{(n)}$ with $\xi_i=\min\left(X_i,Y_{1i},Y_{2i}\right)$ and

$$\delta_i=\begin{cases}1, & if \ \ X_i\leq \min\left(Y_{1i},Y_{2i}\right),\\ 0, & if \ \ Y_{1i}\leq \min\left(X_i,Y_{2i}\right),\\ -1, & if \ \ Y_{2i}\leq \min\left(X_i,Y_{1i}\right).\end{cases}$$

In $C_{*}^{(n)}$ the r.v.–s of interest $Y_{ki}$ 's are censored from the right by r.v.–s $\min\left(Y_{1i},Y_{1i}\right)$ and observable partially only in case of $\delta_i=1$. The estimation of d.f. $F$ and its functionals from sample $C_{*}^{(n)}$ is one of the main goals in survival analysis. Let's define the conditional d.f.–s of r.v.–s $Y_{ki}$, $k=1,2$, given in a covariate $Z_i=z$ as

$$F(t/z)=P\left(X_i\leq t/Z_i=z\right), \quad (t,z)\in \Box^{+}\times\Box,$$

$$G_k(t/z)=P\left(Y_{ki}\leq t/Z_i=z\right),k=1,2,(t,z)\in \Box^{+}\times\Box.$$

We also suppose that the censoring by r.v.–s $Y_{1i}$ for a given covariate is informative, i.e. the pairs $\left(X_i,Y_{1i}\right)$ follows the conditionally PHM in which the d.f. $G_1(t/z)$ is related to $F(t/z)$ as

$$1-G_1(t/z)=\left(1-F(t/z)\right)^{\beta}, \quad (t,z)\in \Box^{+}\times\Box. \tag{1}$$

Here $\beta$ is some fixed but unknown censoring parameter. This kind of partially informative random censoring model with nuisance parameter $(\beta,G_2)$ in lack of covariate $Z_i$ was considered by authors [1, 4, 9, 15].

Adapting some of ideas from [4] here we propose three asymptotical equivalent estimators of $F(t)$ through estimation of conditional d.f. $F(t/z)$ by exponential – hazard, product – limit and relative – risk power estimators using data from sample $C_{*}^{(n)}$.

## 2. Estimators of Survival Function

Let $H(t/z)$ is a conditional d.f. of $\xi_i$. Then by supposed independence of r.v.–s for a given covariate and from (1) we have

$$1-H(t/z)=\left(1-K(t/z)\right)\left(1-G_2(t/z)\right),$$

where

$$K(t/z) = P\big(\min(X_i, Y_{1i}) \le t / Z_i = z\big) =$$

$$= 1 - \big(1 - F(t/z)\big)\big(1 - G_1(t/z)\big) = 1 - \big(1 - F(t/z)\big)^{\beta+1},$$

$t \in \square^+, z \in \square$. For any d.f. $L(t)$, let

$$\tau_L = \sup\{t \in \square^+ : L(t) = 0\}, \quad T_L = \inf\{t \in \square^+ : L(t) = 1\},$$
$$L(t-) = \lim_{s \uparrow t} L(s), \quad \Delta L(t) = L(t) - L(t-).$$

Then by (1),

$$\tau_F = \tau_{G_1} = \tau_K, \quad T_F = T_{G_1} = T_K$$

and

$$\tilde{H}(t/z) = P\big(\xi_i \le t, \delta_i \ne -1 / Z_i = z\big) = P\big(\min(X_i, X_{1i}) \le \min(t, Y_{2i}) / Z_i = z\big) = \int_0^t \big(1 - G_2(u/z)\big) dK(u/z)$$

with $\tilde{H}(t/z) + \tilde{\tilde{H}}(t/z) = H(t/z)$ for all $t \in \square^+$, $z \in \square$. Let $\gamma = \dfrac{1}{\beta+1}$ and $p_m = P(\delta_i = m)$, $m = -1, 0, 1$. Then

$$P(\delta_i \ne -1) = \int \lim_{t \to \infty} \tilde{H}(t/z) dQ(z) = \int \int_0^\infty \big(1 - G_2(u/z)\big) dK(u/z) dQ(z)$$

$$= \int \left( \int_0^\infty \big(1 - G_2(u/z)\big) d\Big[1 - \big(1 - F(u/z)\big)^{\beta+1}\Big] \right) dQ(z) = \frac{1}{\gamma} \int \left( \int_0^\infty \big(1 - G_2(u/z)\big)\big(1 - F(u/z)\big)^\beta dF(u/z) \right) dQ(z). \quad (2)$$

and

$$P(\delta_i = 1) = \int P(\delta_i = 1 / Z_i = z) dQ(z) = \int P\big(X_i \le \min(Y_{1i}, Y_{2i}) / Z_i = z\big) dQ(z)$$

$$= \int \left( \int_0^\infty \big(1 - F(u/z)\big)^\beta \big(1 - G_2(u/z)\big) dF(u/z) \right) dQ(z). \quad (3)$$

From (2) and (3), we get $\gamma = P(\delta_i = 1) / P(\delta_i \ne -1) = P(\delta_i = 1 / \delta_i \ne -1)$. Hence the parameter $\gamma = \dfrac{1}{\beta+1}$ that can be consistently estimated by statistics

$$\gamma_n = \frac{\sum_{i=1}^n I(\delta_i = 1)}{\sum_{i=1}^n I(\delta_i \ne -1)} = \frac{p_{1n}}{p_{0n} + p_{1n}}, \quad (4)$$

where $p_{mn} = \dfrac{1}{n} \sum_{i=1}^n I(\delta_i = m)$ are estimators of probabilities $p_m$, $m = -1, 0, 1$. Let's define cumulative hazard functions (c.h.f.–s)

$$\tau_H = \max(\tau_F, \tau_{G_2}) \ge 0, \quad T_H = \min(T_F, T_{G_2}) \le \infty.$$

Let $Q(z)$ is d.f. of r.v.–s $Z_i$. Then

$$F(t) = \int F(t/z) dQ(z),$$

$$G_k(t) = \int G_k(t/z) dQ(z), \quad k = 1, 2,$$

$$K(t) = \int H(t/z) dQ(z) = P\big(\min(X_i, Y_{1i}) \le t\big),$$

$$H(t) = \int H(t/z) dQ(z) = P(\xi_i \le t).$$

In order to construct the estimators of $F(t)$, we need the following conditional sub distribution function:

$$\tilde{\Lambda}(t/z) = -\log\big(1 - K(t/z)\big) = -\frac{1}{\gamma} \cdot \log\big(1 - F(t/z)\big),$$

$$\tilde{\tilde{\Lambda}}(t/z) = -\log\big(1 - G_2(t/z)\big) \quad (5)$$

and

$$\Lambda(t/z) = -\log\big(1 - H(t/z)\big) = \tilde{\Lambda}(t/z) + \tilde{\tilde{\Lambda}}(t/z).$$

We suppose that d.f. $Q(z)$ have a density $q(z)$. Then c.h.f.–s (5) can be represented as

$$\tilde{\Lambda}(t/z) = \int_0^t \frac{q(z) d\tilde{H}(u/z)}{q(z)\big(1 - H(u/z)\big)} = \int_0^t \frac{d\tilde{A}(u;z)}{B(u;z)},$$

$$\tilde{\tilde{\Lambda}}(t/z) = \int_0^t \frac{q(z)d\tilde{\tilde{H}}(u/z)}{q(z)(1-H(u/z))} = \int_0^t \frac{d\tilde{\tilde{A}}(u;z)}{B(u;z)}, \qquad (6)$$

and

$$\Lambda(t/z) = \int_0^t \frac{q(z)dH(u/z)}{q(z)(1-H(u/z))} = \int_0^t \frac{dA(u;z)}{B(u;z)},$$

where

$$A(u;z) = \tilde{A}(u;z) + \tilde{\tilde{A}}(u;z),$$

$$\tilde{A}(u;z) = q(z)(1-G_2(u/z)),$$

$$\tilde{\tilde{A}}(u;z) = q(z)(1-K(u/z))$$

and

$$B(u/z) = q(z)(1-H(u/z)).$$

These functions can be estimated by statistics

$$A_n(u;z) = \tilde{A}_n(u;z) + \tilde{\tilde{A}}_n(u;z),$$

$$\tilde{A}_n(u;z) = \frac{1}{n}\sum_{i=1}^n I(\xi_i \le u, \delta_i \ne -1)\pi_a(z;Z_i),$$

$$\tilde{\tilde{A}}_n(u;z) = \frac{1}{n}\sum_{i=1}^n I(\xi_i \le u, \delta_i \ne -1)\pi_a(z;Z_i), \qquad (7)$$

and

$$B_n(u;z) = \frac{1}{n}\sum_{i=1}^n I(\xi_i \ge u)\pi_a(z;Z_i),$$

where $\pi_a(z;t) = \frac{1}{a_n}\pi\left(\frac{z-t}{a_n}\right)$ with kernel function $\pi(z)$ and bandwidth sequence $a = a_n \downarrow 0$ as $n \to \infty$. By substitution of estimators (7) into formulas (6) we obtain the estimators for c.h.f.–s

$$\tilde{\Lambda}_n(t/z) = \int_0^t \frac{d\tilde{A}_n(u;z)}{B_n(u;z)}, \quad \tilde{\tilde{\Lambda}}_n(t/z) = \int_0^t \frac{d\tilde{\tilde{A}}_n(u;z)}{B_n(u;z)},$$

and

$$\Lambda_n(t/z) = \tilde{\Lambda}_n(t/z) + \tilde{\tilde{\Lambda}}_n(t/z) = \int_0^t \frac{dA_n(u;z)}{B_n(u;z)}. \qquad (8)$$

In order to estimate the conditional d.f. $F(t/z)$ we use representation

$$1-F(t/z) = (1-K(t/z))^\gamma, \ (t,z)\in \Box^+ \times \Box, \qquad (9)$$

following from (1). For $1-K(t/z)$ we use the following exponential hazard type estimator of Altschuler – Breslow, PL - estimator of Kaplan – Meier and relative – risk power type estimator of Abdushukurov (see, [1–6]):

$$1-K_{1n}(t/z) = \exp\left(-\tilde{\Lambda}_n(t/z)\right),$$

$$1-K_{2n}(t/z) = \prod_{u\le t}\left(1-\Delta\tilde{\Lambda}_n(u/z)\right), \qquad (10)$$

$$1-K_{3n}(t/z) = \left[\prod_{u\le t}\left(1-\Delta\Lambda_n(u/z)\right)\right]^{R_n(t;z)},$$

where $R_n(t;z) = \tilde{\Lambda}_n(t/z)(\Lambda_n(t/z))^{-1}$ is estimator of

$$R(t;z) = \tilde{\Lambda}(t/z)(\Lambda(t/z))^{-1},$$

$$\Delta\tilde{\Lambda}_n(u/z) = \tilde{\Lambda}_n(u/z) - \tilde{\Lambda}_n(u-/z),$$

$$\Delta\Lambda_n(u/z) = \Lambda_n(u/z) - \Lambda_n(u-/z).$$

According to (9) using estimators (4) and (10) we get corresponding estimators of $1-F(t/z)$ as

$$1-F_{ln}(t/z) = (1-K_{ln}(t/z))^{\gamma n},$$
$$l = 1,2,3, \ (t,z)\in \Box^+ \times \Box. \qquad (11)$$

Finally, using statistics (11) we construct estimators of $1-F(t)$ by averaging as follows:

$$1-F_{ln}(t) = \int (1-F_{ln}(t/z))dQ_n(z),$$
$$l = 1,2,3, \ t\in \Box^+, \qquad (12)$$

where

$$Q_n(z) = \frac{1}{n}\sum_{i=1}^n I(Z_i \le z), \ z\in \Box,$$

is the empirical estimator of d.f. $Q(z)$.

Note that in lack of censoring by r.v.–s $Y_{1i}$ –s (i.e. $G_1(t) \equiv 0$) the estimator $K_{1n}(t/z)$ in (10) coincides with one considered in [11].

## 3. Asymptotic Properties of Estimators of Survival Function

In order to investigate the asymptotic properties of estimators (12) we need the following conditions.

Conditions I:

(I.1) The kernel function $\pi$ is bounded and Lipschitz condition of order 1 with respect to the Euclidean distance on $\Box$.

(I.2) $\int \pi(z)\, dz = 1, \int z\pi(z)\, dz = 0, \int z^2 |\pi(z)|\, dz < \infty.$

(I.3) The bandwidth sequence $\{a_n, n \geq 1\}$ satisfies: $a_n \to 0$ and $\dfrac{\log n}{na_n} \to 0$ as $n \to \infty$.

(I.4) The partial derivatives $\dfrac{\partial F(t/z)}{\partial t}$ and $\dfrac{\partial G_2(t/z)}{\partial t}$ exist and are continuous in $t$ for each $z$.

(I.5) The functions $q(z)$, $F(t/z)$ and $G_2(t/z)$ have bounded continuous first and second partial derivatives with respect to $z$.

(I.6) For any closed interval $[a,b] \subset \Box^+$, there exists a constants $p, \delta(\varepsilon) > 0$ such that

$$P(\xi_i > p / Z_i = z) \geq \delta(\varepsilon), \quad \forall z \in [a,b],$$

with $q(z) \geq \varepsilon$ and $\varepsilon > 0$ arbitrary small.

Note that in view of (1) the conditions (I.4) and (I.5) for d.f. $G_1(t/z)$ are hold too.

Moreover, from (I.6) we also have a chain of inequalities

$$1 - K(p) = P\left(\min(X_i, X_{1i}) > p\right) \geq P(\xi_i > p) = \int q(z) P(\xi_i > p / Z_i = z)\, dz \geq \varepsilon \cdot \int\limits_{\{z: q(z) \geq \varepsilon\}} P(\xi_i > p / Z_i = z)\, dz > 0.$$

The properties of estimators (12) are established from the corresponding properties of estimators (4),(7),(8) and (10). In accordance with results of Cheng [7] under Conditions *I* we obtain asymptotic unbiasedness of $\tilde{A}_n, \tilde{\tilde{A}}_n, A_n, B_n$ and hence the uniform strong consistency of $\tilde{\Lambda}_n(t/z)$ and $\tilde{\tilde{\Lambda}}_n(t/z)$ over a rectangle $[0,\tau] \times [a,b]$ with $\tau \in (0, T_H)$ and rate of convergence as follows:

$$\sup_{(t,z) \in [0,\tau] \times [a,b]} \left| \tilde{\Lambda}_n(t/z) - \tilde{\Lambda}(t/z) \right| \overset{a.s.}{=} O\left( \left( \frac{\log n}{na_n} \right)^{1/2} \right) + O\left( a_n^2 \right), \tag{13}$$

$$\sup_{(t,z) \in [0,\tau] \times [a,b]} \left| \tilde{\tilde{\Lambda}}_n(t/z) - \tilde{\tilde{\Lambda}}(t/z) \right| \overset{a.s.}{=} O\left( \left( \frac{\log n}{na_n} \right)^{1/2} \right) + O\left( a_n^2 \right). \tag{14}$$

Then

$$\sup_{(t,z) \in [0,\tau] \times [a,b]} \left| \Lambda_n(t/z) - \Lambda(t/z) \right| \overset{a.s.}{=} O\left( \left( \frac{\log n}{na_n} \right)^{1/2} \right) + O\left( a_n^2 \right), \tag{15}$$

$$\sup_{(t,z) \in [0,\tau] \times [a,b]} \left| K_n(t/z) - K(t/z) \right| \overset{a.s.}{=} O\left( \left( \frac{\log n}{na_n} \right)^{1/2} \right) + O\left( a_n^2 \right), \tag{16}$$

where (15) is consequence of (13), (14) and triangular inequality, (16) follows from (13) and inequality $|a - b| \leq |\log a - \log b|$, for $0 < a, b \leq 1$. It is easy to see that statistics (4) is strong consistent and asymptotically unbiased estimator of $\gamma$. From Consequence 3 in Abdushukurov and Nedzvedsky [3], for each $m = -1, 0, 1$ and $\varepsilon > 0$ we have

$$P\left( |p_{mn} - p_m| > \left( \frac{\varepsilon \log n}{n} \right)^{1/2} \right) \leq 2n^{-\varepsilon}. \tag{17}$$

Also if $\min(p_m, 1 - p_m) \geq 2\left( \dfrac{\varepsilon \log n}{n} \right)^{1/2}$, then

$$P\left( p_{mn}^{-1} > 2p_m^{-1} \right) \leq 2n^{-\varepsilon} \tag{18}$$

and

$$P\left( (1 - p_{mn})^{-1} > 2(1 - p_m)^{-1} \right) \leq 2n^{-\varepsilon} \tag{19}$$

Hence from (17)-(19) by Borel - Cantelly lemma for $m = -1, 0, 1$ we have

$$\left| p_{mn} - p_m \right| \overset{a.s}{=} O\left( \left( \frac{\log n}{n} \right)^{1/2} \right),$$

and with probability one

$$\frac{1}{p_{mn}} < \frac{2}{p_m}, \quad \frac{1}{1-p_{mn}} < \frac{2}{1-p_m}.$$

Adapting characterization of simple proportional hazard model under independent random censoring from the right (see, [1, 3-6, 8, 9]) we get following property of considered conditionally partially informative competing risks model. For a given covariate $Z_i = z$ partially observable (only if $\delta_i \neq -1$) r.v.-s $\min(X_i, Y_{1i})$ and indicators $I(X_i \leq Y_{1i})$ are independent if and only if the representation (1) is satisfied. Hence under occurrence of events $A_z^{(i)} = \{Z_i = z\} \cap \{\delta_i \neq -1\}$ the r.v.-s $\xi_i = \min(X_i, Y_{1i}, Y_{2i})$

and $\delta_i$ are conditionally independent only if the representation (1) is satisfied. This characterization property of considered model is very useful in investigating of asymptotical properties of estimators (12).

In the next theorem we show that all three statistics $F_{ln}(t), l = 1, 2, 3$, are a consistentestimators for the unconditional d.f. $F(t)$.

Theorem 1. *Under Conditions I statistics* $\{F_{ln}(t), \; l = 1, 2, 3\}$ *are a uniformly strongly consistent estimators for d.f. $F(t)$ on $[0, \tau]$:*

$$\sup_{0 \leq t \leq \tau} \left| F_{ln}(t) - F(t) \right| \underset{n \to \infty}{\to} 0, \; a.s$$

Moreover we also prove weak convergence of normed processes

$$\left\{ V_{ln}(t) = \sqrt{n} \left( F_{ln}(t) - F(t) \right), \; t \in [0, \tau], \; l = 1, 2, 3 \right\}$$

similar to Gaussian process.

In order to formulate these results we introduce Condition II.

Condition II:

(II.1) $E\left[ \left( 1 - F(t/Z) \right)^2 \int_0^t \frac{dF(u/Z)}{\left( 1 - F(y/z) \right)^2 \left( 1 - G_2(u/Z) \right)} \right] < \infty, \forall t \in [o, T].$

(II.2) $a_n^2 \sqrt{n} \to 0, \; \dfrac{\log n}{a_n \sqrt{n}} \to 0, \; n \to \infty.$

Theorem 2. *Under conditions I and II the processes converges weakly on Skorokhod space $D[0, \tau]$ to the same mean zero Gaussian process $\{V(t), t \geq 0\}$, with covariance function for $0 \leq s \leq t \leq \tau$:*

$$Cov(V(s), V(t)) = E\Big\{ \left( 1 - F(t/Z) \right) \left( 1 - F(s/Z) \right) \cdot$$

$$\left[ \gamma \int_0^s \frac{dF(u/Z)}{\left( 1 - F(u/Z) \right)^2 \left( 1 - G_2(t/Z) \right)} + 1 + \frac{\gamma(1-\gamma)}{p_0 + p_1} \log\left( 1 - K(t/Z) \right) \log\left( 1 - K(t/Z) \right) \right] \Big\} - \left( 1 - F(t) \right) \left( 1 - F(s) \right).$$

As Jensen and Wiedman [11], instead of estimators (12), we can consider following CIM-estimator (Conditional Independence Model) for $1 - F(t)$ without using informativeness of considered competing risks model:

$$1 - F_n^{JW}(t) = \int \exp\left( -\Lambda_n^*(u/z) \right) dQ_n(z), \tag{20}$$

where

$$\Lambda_n^*(t/z) = \int_0^t \frac{dA_n^*(u;z)}{B_n(u;z)},$$

$$A_n^*(u;z) = \frac{1}{n} \sum_{i=1}^n I(\xi_i \leq u, \delta_i = 1) \pi_a(z; Z_i).$$

Then, as proved in [11], under conditions of our theorem 2 the process $\left\{ \sqrt{n} \left( F_n^{JW}(t) - F(t) \right), t \geq 0, n \geq 1 \right\}$ converges weakly

on $D[0,\tau]$ to a mean zero Gaussian process $\{W(t), t \geq 0\}$ with covariance function for $0 \leq s \leq t \leq \tau$:

$$Cov\big(W(s), W(t)\big) = E\left\{\big(1-F(t/Z)\big)\big(1-F(s/Z)\big) \cdot \left[\int_0^s \frac{dF(u/Z)}{(1-F(u/Z))^2(1-G_2(u/Z))} + 1\right]\right\} - \big(1-F(t)\big)\big(1-F(s)\big).$$

Let

$$\sigma^2(t) = Var\{V(t)\} = As.\,var.\left\{\sqrt{n}F_{\ln}(t)\right\}, \ l = 1,2,3$$

and

$$\sigma^{*2}(t) = Var\{W(t)\} = As.\,var.\left\{\sqrt{n}F_n^{JW}(t)\right\}$$

are asymptotical variances of estimators (12) and (20) respectively. The next theorem shows that the estimators (12) are asymptotically effective than the estimator (20).

Theorem 3. *Under conditions of theorem 2, for all* $t \in [0,\tau]$, *we have*

$$\frac{\gamma}{p_0 + p_1} E\left[\big(1-F(t/Z)\big)^2 \cdot \log^2\big(1-K(t/Z)\big)\right] < \sigma^2(t) < \sigma^{*2}(t).$$

# 4. Discussion

We have introduced three new estimators (12) for the averaged unconditional d.f.

$$F(t) = \int F(t/z)dQ(z) \qquad (21)$$

in partially informative competing risks model in survival analysis. These estimators generalize the unconditional exponential hazard type estimator of Altschuler-Breslow, PL-estimator of Kaplan-Meier and relative-risk power type estimator of Abdushukurov. We have showed that all three statistics are strongly consistent estimators for the $F(t)$ and they converged to same mean zero Gaussian process with asymptotic variance $\sigma^2(t)$ which is less than the asymptotic variance of Jensen – Wiedman's CIM-estimator $F_n^{JW}(t)$ not taking into account the partially informativeness of considered competing risks model. It is also worthy to note that the new estimators are good competitors to the other known estimators. Our estimators are useful tools in deriving some functionals of interest in survival analysis. For instance, the conditional residual lifetime distribution defined via averaged unconditional d.f. (21) as

$$F_t(s) = \frac{F(t+s)-(t)}{1-F(t)} = \int \frac{P(t < X_i \leq t+s, Z_i = z)dQ(z)}{P(X_i > t, Z_i = z)} = \int \frac{F(t+s/z)-F(t/z)}{1-F(t/z)}dQ(z). \qquad (22)$$

One of main characteristics of d.f. (22) is its mean, i.e. mean conditional residual life function

$$\mu(t) = \big(1-F(t)\big)^{-1} \int_t^{+\infty}\big(1-F(s)\big)ds, \ t > 0. \qquad (23)$$

We can estimate functional $\mu(t)$ by plugging in estimators (12) instead of $F(t)$ in (23). But from section 3 we know that (Theorem 1) estimators (12) have consistent properties in some interval $[0,\tau]$. Therefore, we can consider the following truncated version of (23):

$$\mu^\tau(t) = \big(1-F(t)\big)^{-1} \int_t^\tau\big(1-F(s)\big)ds, \ 0 < t < \tau. \qquad (24)$$

Now we can estimate (24) by statistics

$$\mu_{\ln}^\tau(t) = \big(1-F_{\ln}(t)\big)^{-1} \int_t^\tau\big(1-F_{\ln}(s)\big)ds, \ 0 < t < \tau, \ l = 1,2,3.$$

# 5. Conclusion

In this paper, we have proposed class of semi parametric estimators of exponential-hazard, product-limit and relative - risk power types of survival function in partially informative censoring model. Considered model is competing risks model in which each observation is the minimum of one lifetime and two censoring times in the presence also of covariate. The survival function of one of these censoring times is a power of the survival function of the lifetime (this is the informative censoring) and the distribution of the other censoring time has no relation with the distribution of the lifetime (non-informative censoring). We have showed that proposed estimators are uniformly strong consistent and they converges to same Gaussian process. They are also asymptotical efficient than the Jensen-Wiedmanns CIM

(Conditional Independence Model) estimator.

# References

[1]   A.   A.   Abdushukurov   Nonparametric   estimation   of
      thedistributionfunction   based   on   relative   risk   function.
      Commun Statist: Theory & Meth. 1998. V. 27. N. 8. p. 1991-
      2012.

[2]   A. A. Abdushukurov Onnonparametricestimationof reliability
      indices by censoredsample. Theory Probab. Appl. 1999. V. 43.
      N. 1. p. 3-11.

[3]   A. A. Abdushukurov, D. T. Nedzvedsky Asymptotic properties
      of empirical processes on censored samples of ramdomsices.
      J. Math. Sciences. 2005. V. 127. N. 1. p. 931-939.

[4]   A.   A.   Abdushukurov,   D.   Makhmudova   Semiparametric
      estimation of distribution function in the informative model of
      competing risks. J. Math. Sciences. Springer. 2017. V. 227. N.
      2, p. 117-123.

[5]   A. A. Abdushukurov Nonparametric estimation based on
      incomplete observations. In: International Encyclopedia of
      Statistical Sciences. (Prof. Miodrag Lovric, Editor). Springer.
      2011. Pt. 14. p. 962-964.

[6]   A. A. AbdushukurovEstimates of unknown distributions from
      incomplete   observations   and   its   properties.   LAMBERT
      Academic Publishing. 2011. p. 301. (Russian).

[7]   P. E. ChengNonparametric estimation of survival curve under
      dependentcensorship. J. Statist. Plann. Infer. 1989 V. 23. p.
      181-191.

[8]   S. CsörgőEstimating in proportional hazards model of ran-
      dom censorship. // Statistics. 1988. v. 19. N. 3. p. 437-463.

[9]   U. Gather, J. Pawlitschko. Estimating the survival function
      under a generalized Koziol-Green model with partially
      informative censoring. Metrika. 1998. V. 48. p. 189-207.

[10]  T. Herbst Test of fit with the Koziol-Green model for random
      censership. Statist. Decisions. 1992. v. 10. p. 163-171.

[11]  U. Jensen, J. Wiedmann Estimation of a Survival Curve under
      Dependent Cenoring. Second Internat. Conf. on Math. Meth-s
      in Reliability. Bordeaux. France. July 4-7. 2000. V. 2. p. 571-
      574.

[12]  S. Kirmani, J. Y. Dauxois Testing the Koziol-Green model
      against monotone conditional odds for censoring. Statist.
      Probab. Lett. 2004. v. 63. N. 3. p. 327-334.

[13]  F. Siannis Applications of a parametric model for informative
      censoring. Biometrics. 2004. v. 60. N. 3. p. 704-714.

[14]  M. C. Wang, J. Qin, C. Chiang Analysing recurrent event data
      with informative censoring. J. Amer. Statist. Soc. 2001. v. 96.
      N. 455. p. 1057-1065.

[15]  H. Zhang, M. B. Rao On generalizedmaximumlikelihood
      estimation in the proportiona hazards model with partially
      informative censoring. Metrika. 2004. V. 59. p. 125-136.