International Journal of Innovation in Science and Mathematics
Volume 2, Issue 3, ISSN (Online): 2347–9051

Estimation of Survival Functions from Dependent Random Right Censored Data

A. A. Abdushukurov, R. S. Muradov

Abstract – **In this article we consider the problem of estimating the survival function for dependent random censoring observations on the right.**

Keywords – **Random Censorship, Survival Function, Product-Limit Estimator, Relative-Risk Power Estimator, Copula-Graphic Estimator, Archimedean Copulas.**

I. Introduction

In the last years, we can see the interest in copula functions and their use in mathematical statistics. Copulas provide methods for describing the relationship between multidimensional and marginal distributions. The idea is that the joint distribution of the compounds formed with marginal distributions using dependent uniform distributions. At the beginning of copulas have been used in the development of theoretical-probabilistic aspects of metric spaces. Later, they became interested in the determination of non-parametric measures of dependence between random variables and thereby they gradually began to play a major role in probability theory and mathematical statistics.

Definition 1.[4,5]. A copula $C(u,v):[0,1]^2 \to [0,1]$ is a bivariate distribution function with uniform marginals.

Theorem 1. (Sclar, 1959). Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all x, y in R,

$$H(x,y) = C(F(x), G(y)). \qquad (1)$$

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on $Ran(F) \times Ran(G)$. Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (1) is a joint distribution function with margins F and G.

Furthermore, the representation (1) suggests that if the copula C were known, then substituting continuous marginal estimators for F and G would yield a plug-in estimate of their associated joint distribution function H. Moreover, in light of Sclar's result with arrive at the following functional definition of a copula.

Definition 2.[5]. Given a bivariate distribution function H with marginals F and G, the function defined as

$$C(u,v) = H\left(F^{-1}(u), G^{-1}(v)\right),$$

for $(u,v) \in [0,1]^2$, where $F^{-1}(u)$ and $G^{-1}(v)$ are the inverse functions of F and G respectively, is the copula corresponding to H.

In survival analysis our interest focuses on a nonnegative random variables (r.v.-s) denoting death times of biological organisms or failure times of mechanical systems. A difficulty in the analysis of survival data is the possibility that the survival times can be subjected to random censoring by other nonnegative r.v.-s and therefore we observe incomplete data. There are various types of censoring mechanisms. In this article we consider only right censoring model and problem of estimation of survival functions when the survival times and censoring times are dependent and propose new estimates of survival functions assuming that the dependence structure is described by a known copula function.

II. The Model and Estimation of Survival Function

On the probability space (Ω, A, P) we consider $\{(X_k, Y_k), k \ge 1\}$ - a sequence of independent and identically distributed pairs of nonnegative r.v.-s with common joint distribution function (d.f.) $H(x,y) = P(X_1 \le x, Y \le y)$, $(x,y) \in \overline{R}^{+2} = [0,\infty]^2$. We suppose that the marginal d.f.-s $F(x) = P(X_1 \le x) = H(x,\infty)$ and $G(y) = P(Y_1 \le y) = H(+\infty, y)$, $x, y \in \overline{R}^{+}$, are continuous and $F(0) = G(0) = 0$. Assume that the sequence $\{X_k, k \ge 1\}$ is right censored by the sequence $\{Y_k, k \ge 1\}$ and at n-th stage of the experiment the observation is available the sample $\mathrm{V}^{(n)} = \{(Z_k, \delta_k), 1 \le k \le n\}$, where $Z_k = \min(X_k, Y_k)$, $\delta_k = I(Z_k = X_k)$ and $I(A)$ is the indicator of the event A. Should be noted that it does not require independence of sequences $\{X_k\}$ and $\{Y_k\}$. The problem is consist in estimating of the survival function $S^X(x) = P(X_1 > x) = 1 - F(x)$, $x \in \overline{R}^{+}$, from the sample $\mathrm{V}^{(n)}$. Let $\overline{H}(x,y) = P(X_1 > x, Y_1 > y), (x,y) \in \overline{R}^{+2}$ - a joint survival function of the pairs (X_k, Y_k). According to Theorem of Sclar H and $\overline{H}$ can be submitted through the appropriate copula functions (see [4, 5]):

$$H(x,y) = C(F(x); G(y)), (x,y) \in \overline{R}^{+2},$$

$$\overline{H}(x,y) = C^*\left(S^X(x); S^Y(y)\right), (x,y) \in \overline{R}^{+2}, \qquad (2)$$

where copulas C and C^* are related as

$$C^*(u,v) = u + v - 1 + C(1-u; 1-v), (u,v) \in [0,1]^2. \qquad (3)$$

In the sequel in order to construct estimates for the survival function S^X, assume that C^* is Archimedean copula, i.e. $C^*(u,v)=\varphi^{[-1]}\left[\varphi(u)+\varphi(v)\right],(u,v)\in[0,1]^2$, where $\varphi:\ [0,1]\to\overline{R}^+$ is some generator function with the pseudo inverse $\varphi^{[-1]}$. Thus, by (2) and (3)

$$\overline{H}(x,y)=\varphi^{[-1]}\left[\varphi\left(S^X(x)\right)+\varphi\left(S^Y(y)\right)\right],\ (x,y)\in\overline{R}^{+2},$$

$$S^Z(x)=\varphi^{[-1]}\left[\varphi\left(S^X(x)\right)+\varphi\left(S^Y(x)\right)\right],\quad x\in\overline{R}^+. \quad (4)$$

We introduce a usual λ^X, λ^Z and "crude" $\lambda-$ hazard functions

$$\lambda^X(x)=\lim_{\Delta\downarrow 0}\frac{1}{\Delta}P\left(x<X_1\le x+\Delta/X_1>x\right),$$

$$\lambda^Z(x)=\lim_{\Delta\downarrow 0}\frac{1}{\Delta}P\left(x<Z_1\le x+\Delta/X_1>x,Y_1>x\right),$$

$$\lambda(x)=\lim_{\Delta\downarrow 0}\frac{1}{\Delta}P\left(x<X_1\le x+\Delta/X_1>x,Y_1>x\right).$$

In order to construct a copula estimates for S^X consider the following easily verifiable equality:

$$\lambda^X(x)S^X(x)\varphi'\left(S^X(x)\right)=\lambda(x)S^Z(x)\varphi'\left(S^Z(x)\right). \quad (5)$$

Integrating (5) over the interval $[0,x]$ and denoting by $\Lambda(x)=\int_0^x\lambda(t)dt$ and $\Lambda^X(x)=\int_0^x\lambda^X(t)dt$ corresponding cumulative hazard functions we obtain the integral equation

$$\int_0^x S^X(t)\varphi'\left(S^X(t)\right)d\Lambda^X(t)=\int_0^x S^Z(t)\varphi'\left(S^Z(t)\right)d\Lambda(t),\quad x\in\overline{R}^+. \quad (6)$$

Integral on the left side of (6) is equal to $-\varphi\left(S^X(x)\right)$ and then (6) takes the form

$$\varphi\left(S^X(x)\right)=-\int_0^x S^Z(t)\varphi'\left(S^Z(t)\right)d\Lambda(t),\quad x\in\overline{R}^+. \quad (7)$$

Hence we find the expression for the survival function S^X:

$$S^X(x)=\varphi^{[-1]}\left[-\int_0^x S^Z(t)\varphi'\left(S^Z(t)\right)d\Lambda(t)\right],\quad x\in\overline{R}^+. \quad (8)$$

Note that the survival function S^Z permit usual empirical estimation by the values Z_k observed in the sample $V^{(n)}$:

$$S_n^Z(x)=\frac{1}{n}\sum_{k=1}^n I\left(Z_k>x\right),\quad x\in\overline{R}^+. \quad (9)$$

Substituting (9) to the right of representation (8), we obtain a preliminary estimate of S^X as

$$\tilde{S}_n^X(x)=\varphi^{[-1]}\left[-\int_0^x I\left(S_n^Z(t-)>0\right)S_n^Z(t-)\varphi'\left(S_n^Z(t)\right)d\Lambda_n(t)\right],$$

(10)

where

$$\Lambda_n(t)=\frac{1}{n}\sum_{k=1}^n\frac{I\left(Z_k\le t,\delta_k=1\right)}{S_n^z\left(Z_k\right)-\frac{1}{n}}, \quad (11)$$

The corresponding estimate for

$$\Lambda(t)=\int_0^t\frac{dP\left(Z_1\le s,\delta_1=1\right)}{P\left(Z_1>s\right)}.$$

Estimate (10) plays a supporting role in the construction of the main estimates for S^X in the future. Let $N_k(t)==I\left(Z_k\le t,\delta_k=1\right)$. Define the counting processes $\overline{\mathrm{N}}_n(t)=\sum_{k=1}^n N_k(t)$ and $\mathrm{J}_n(t)=nS_n^Z(t-)=\sum_{k=1}^n I\left(Z_k\ge t\right)$. Then the estimates (10) and (11) can be represented as

$$\tilde{S}_n^X(x)=\varphi^{[-1]}\left[-\frac{1}{n}\int_0^x I\left(\mathrm{J}_n(t)>0\right)\varphi'\left(\frac{\mathrm{J}_n(t)}{n}\right)d\overline{\mathrm{N}}_n(t)\right], \quad (12)$$

$$\Lambda_n(t)=\int_0^t\frac{I\left(\mathrm{J}_n(s)>0\right)}{\mathrm{J}_n(s)}d\overline{\mathrm{N}}_n(s).$$

Given the analog left side of (6), i.e.

$$\varphi\left(S^Z(x)\right)=-\int_0^x S^Z(t)\varphi'\left(S^Z(t)\right)d\Lambda^Z(t), \quad (13)$$

where $\Lambda^Z(t)=\int_0^t\lambda^Z(s)ds$, together with (9) also obtain other estimate for S^Z as

$$\tilde{S}_n^Z(x)=\varphi^{[-1]}\left[-\frac{1}{n}\int_0^x I\left(\mathrm{J}_n(t)>0\right)\varphi'\left(\frac{\mathrm{J}_n(t)}{n}\right)d\overline{N}_n^Z(t)\right], \quad (14)$$

where $\Lambda_n^Z(t)=\int_0^t\frac{I\left(\mathrm{J}_n(s)>0\right)}{\mathrm{J}_n(s)}d\overline{N}_n^Z(s)$, is estimate for $\Lambda^Z(t)$ and $\overline{N}_n^Z(t)=n\left(1-S_n^Z(t)\right)=n-\mathrm{J}_n(t+)=$ $=\sum_{k=1}^n N_k^Z(t)=\sum_{k=1}^n I\left(Z_k\le t\right)$ - the counting process.

For S^X have the following obvious identity obtained from the representations (7) and (13):

$$S^X(x)=\varphi^{[-1]}\left[\varphi\left(S^Z(x)\right)\frac{\left(-\int_0^x S^Z(t)\varphi'\left(S^Z(t)\right)d\Lambda(t)\right)}{\left(-\int_0^x S^Z(t)\varphi'\left(S^Z(t)\right)d\Lambda^Z(t)\right)}\right].$$

(15)

Now substituting the empirical estimate of (9) under the first factor on the right of representation (15) and the corresponding estimates (12) and (14) instead of integrals we obtain the final estimate of S^X in the form

$$S_n^X(x)=\varphi^{[-1]}\left[\varphi\left(S_n^Z(x)\right)\frac{\left(-\int_0^x I\left(\mathrm{J}_n(t)>0\right)\varphi'\left(\frac{\mathrm{J}_n(t)}{n}\right)d\overline{N}_n(t)\right)}{\left(-\int_0^x I\left(\mathrm{J}_n(t)>0\right)\varphi'\left(\frac{\mathrm{J}_n(t)}{n}\right)d\overline{N}_n^Z(t)\right)}\right],$$

(16)

where

$$\varphi\left(S_n^Z(x)\right)=-\int_0^x I\left(\mathrm{J}_n(s)>0\right)\left[\varphi\left(\frac{\mathrm{J}_n(s)}{n}\right)-\varphi\left(\frac{\mathrm{J}_n(s)}{n}-\frac{1}{n}\right)\right]d\overline{\mathrm{N}}_n^Z(s),$$

is estimator of $\varphi\left(S^Z(x)\right)$.

IJISM Source of Knowledge
International Journal of Innovation in Science and Mathematics
Volume 2, Issue 3, ISSN (Online): 2347–9051

III. Main Results

In fact, we suppose that in (15) the generator function φ is strong (that is $\varphi(0)=\infty$) and hence $\varphi^{[-1]}=\varphi^{-1}$ is usual inverse function.

Denote $Z^{(n)}=\sup\{x\ge 0: \mathrm{J}_n(x)>0\}$, $T_Z=\sup\{x\ge 0: S^Z(x)>0\}$, $\Psi(x)=-x\varphi'(x)$. Introduce the regularity conditions with respect to S^X, S^Z and the copula generator φ. By Λ^* in conditions below denote both of Λ and Λ^Z: (C1) The strong generator function $\varphi(\cdot)$ is strictly decreasing on (0,1] and is sufficiently smooth in the sense that the first two derivatives of the functions $\varphi(x)$ and $\Psi(x)$ are bounded for $x\in[\varepsilon,1]$, where $\varepsilon>0$ is arbitrary. Moreover, the first derivative φ' is bounded away from zero on $[0,1]$;

(C2) $0<\int_0^{T_Z}\left[\Psi\left(S^Z(x)\right)\right]^m d\Lambda^*(x)<\infty$ for $m=1,2$;

(C3) $\int_0^{T_Z}\left|\Psi'\left(S^Z(x)\right)\right| d\Lambda^*(x)<\infty$;

(C4) $\lim\sup_{x\to T_Z}\int_x^{T_Z}\frac{\Psi\left(S^Z(t)\right)}{S^Z(t)}d\Lambda^*(t)=0$;

(C5) $S^X(\cdot)-$is continuous on $[0,T_Z]$ if $T_Z<\infty$. Otherwise, $S^X(\infty)=\lim_{x\to\infty}S^X(x)$.

Note that the condition (C1) - (C5) satisfies, for example, copula generators Clayton-Frank.

At first we state the strong consistency of estimator (16) on the interval $[0,T]$ where $T=T_Z$ if $T_Z<\infty$ and $T=Z^{(n)}$, if $T_Z=\infty$. In fact, these results are also valid throughout half $[0,\infty)$ and $T_Z=\infty$, because $S_n^X(x)=0$ for $x>Z^{(n)}$, $Z^{(n)}\xrightarrow{p}T_Z=\infty$ and $S^X\left(Z^{(n)}\right)\xrightarrow{p}S^X(T_Z)= =S^X(\infty)=0$ for $n\to\infty$.

Theorem 2. Let conditions (C1)-(C3) are hold. Then for $n\to\infty$

(A) $\sup_{0\le x\le T}\left|\Lambda_n^Z(x)-\Lambda^Z(x)\right|\xrightarrow{P}0$; (B) $\sup_{0\le x\le T}\left|\Lambda_n(x)-\Lambda(x)\right|\xrightarrow{P}0$;

(C) $\sup_{0\le x\le T}\left|\varphi\left(\tilde{S}_n^Z(x)\right)-\varphi\left(S^Z(x)\right)\right|\xrightarrow{P}0$;

(D) $\sup_{0\le x\le T}\left|\varphi\left(S_n^Z(x)\right)-\varphi\left(\tilde{S}_n^Z(x)\right)\right|\overset{a.s.}{=}O\left(\frac{1}{n}\right)$;

(E) $\sup_{0\le x\le T}\left|\varphi\left(S_n^Z(x)\right)-\varphi\left(S^Z(x)\right)\right|\xrightarrow{P}0$;

(F) $\sup_{0\le x\le T}\left|\varphi\left(\tilde{S}_n^X(x)\right)-\varphi\left(S^X(x)\right)\right|\xrightarrow{P}0$;

(G) $\sup_{0\le x<\infty}\left|S_n^X(x)-S^X(x)\right|\xrightarrow{P}0$.

The proof of the theorem 2. (A) For all $x\in[0,T]$ by using the identity $1\equiv I\left(\mathrm{J}_n(x)\ge 0\right)=I\left(\mathrm{J}_n(x)>0\right)+ +I\left(\mathrm{J}_n(x)=0\right)$ we have

$$\Lambda_n^Z(x)-\Lambda^Z(x)=\int_0^x\frac{I\left(\mathrm{J}_n(t)>0\right)}{\mathrm{J}_n(t)}d\overline{N}_n^Z(t)-\int_0^x I\left(\mathrm{J}_n(t)\ge 0\right)d\Lambda^Z(t)=\int_0^x\frac{I\left(\mathrm{J}_n(t)>0\right)}{\mathrm{J}_n(t)}d\mathrm{M}_n^Z(t)-\int_0^x I\left(\mathrm{J}_n(t)=0\right)d\Lambda^Z(t)=\mathrm{A}_{1n}(x)+\square_{1n}(x). \quad (17)$$

Let $\tau\le T$ so that $S^Z(\tau)>0$. Then for $x\in[0,\tau]$ (using (C2) when $m=0$) we have, $\left|\square_{1n}(x)\right|\le$ $\le I\left(\mathrm{J}_n(x)=0\right)\cdot\int_0^x d\Lambda^Z(t)<I\left(\mathrm{J}_n(\tau)=0\right)\Lambda^Z(\tau)$, where $\Lambda^Z(\tau)<\infty$ and in accordance with SLLN under $n\to\infty$ have $\frac{\mathrm{J}_n(\tau)}{n}=S_n^Z(\tau)\xrightarrow{a.s.}S^Z(\tau)>0.$ Consequently, $\mathrm{J}_n(\tau)\xrightarrow{a.s.}\infty$ and from here $I\left(\mathrm{J}_n(\tau)=0\right)\xrightarrow{a.s.}0$. Thus, when $n\to\infty$

$$\sup_{0\le x\le T}\left|\square_{1n}(x)\right|\xrightarrow{a.s.}0. \quad (18)$$

Integrand in $\mathrm{A}_{1n}(x)$ is bounded predictable random process (since it is adapted process on $[0,x]$ and continuous from the left) and, therefore, $\mathrm{A}_{1n}(x)$ is a locally square-integrable martingale $\left(\mathrm{A}_{1n}(x)\in\mathrm{M}_{\ell oc}^2\left(\mathrm{F}_p^{(n)}\right)\right)$ with quadratic characteristics

$$<\mathrm{A}_{1n},\mathrm{A}_{1n}>(x)=\int_0^x\frac{I\left(\mathrm{J}_n(t)>0\right)}{\mathrm{J}_n(t)}d\Lambda^Z(t).$$

Then $\mathrm{A}_{1n}^2(x)-<\mathrm{A}_{1n},\mathrm{A}_{1n}>(x)$ is also a martingale with respect to filtration $\mathrm{F}_p^{(n)}$ and by the Lenglart's inequality $\forall\varepsilon$ and $\eta>0$ [1]:

$$P\left(\sup_{0\le x\le\tau}\left|\mathrm{A}_{1n}(x)\right|>\varepsilon\right)\le P\left(\sup_{0\le x\le\tau}\mathrm{A}_{1n}^2(x)>\varepsilon^2\right)\le\frac{\eta}{\varepsilon^2}+ +P\left(\int_0^\tau\frac{I\left(\mathrm{J}_n(t)>0\right)}{\mathrm{J}_n(t)}d\Lambda^Z(t)>\eta\right)< <\frac{\eta}{\varepsilon^2}+P\left(\frac{\Lambda^Z(\tau)}{\mathrm{J}_n(\tau)}>\eta\right)=\hat{\imath}(1),\quad n\to\infty,$$

Because $\varepsilon,\ \eta-$arbitrary and $\mathrm{J}_n(\tau)\xrightarrow{a.s.}\infty$. Thus, when $n\to\infty$

$$\sup_{0\le x\le\tau}\left|A_{1n}(x)\right|\xrightarrow{P}0. \qquad (19)$$

Now consider the interval $(\tau,T]$. If $T_Z=\infty$, then the proof is obvious. Let $T_Z<\infty$. Then we choose $\varepsilon>0$ a sufficiently small and for $\tau=T-\varepsilon$ have

$$\sup_{0\le x\le T}\left|\Lambda_n^Z(x)-\Lambda^Z(x)\right|\le\sup_{0\le x\le\tau}\left|A_{1n}(x)\right|+\sup_{\tau<x\le T}\left|A_{1n}(x)\right|+$$
$$+\sup_{0\le x\le T}\left|\square_{1n}(x)\right|\le\sup_{0\le x\le\tau}\left|A_{1n}(x)\right|+\sup_{0\le x\le T}\left|\square_{1n}(x)\right|+$$
$$+\left|A_{1n}(T_Z)-A_{1n}(T_Z-\varepsilon)\right|+\left|A_{1n}(T_Z-\varepsilon)\right|. \qquad (20)$$

Now, using (25), (26) and tending $\varepsilon>0$ to zero from (27) we obtain the assertion (A) of the theorem. (B) repeats the proof of (A) , we need only replace $\overline{N}_n^Z$ and Λ^Z and respectively on $\overline{N}_n$ and Λ. Let us prove (C). It is easy to verify (see (14)), the following representations using to the indicator identity from proof of (A):

$$\varphi\left(\tilde S_n^Z(x)\right)-\varphi\left(S^Z(x)\right)=-\frac{1}{n}\int_0^x I\left(J_n(t)>0\right)\varphi'\left(\frac{J_n(t)}{n}\right)\cdot dM_n^Z(t)+$$
$$+\int_0^x I\left(J_n(t)>0\right)\left[\Psi\left(\frac{J_n(t)}{n}\right)-\Psi\left(S^Z(t)\right)\right]\cdot d\Lambda^Z(t)-$$
$$-\int_0^x I\left(J_n(t)=0\right)\Psi\left(S^Z(t)\right)d\Lambda^Z(t)=A_{2n}(x)+A_{3n}(x)+A_{4n}(x). \qquad (21)$$

For $x\in[0,\tau]$ at $n\to\infty$

$$\left|A_{4n}(x)\right|\le I\left(J_n(x)=0\right)\int_0^\tau\Psi\left(S^Z(t)\right)d\Lambda^Z(t)<$$
$$<I\left(J_n(\tau)=0\right)\int_0^T\Psi\left(S^Z(t)\right)d\Lambda^Z(t)\xrightarrow{p}0, \qquad (22)$$

where we use condition (C2), under m = 1 and arguments of the proof of (25). Obviously, $A_{2n}(x)\in M_{loc}^2\left(F_p^{(n)}\right)$ and quadratic characteristic of this martingale is

$$<A_{2n},A_{2n}>(x)=\int_0^x I\left(J_n(t)>0\right)\left[\varphi'\left(\frac{J_n(t)}{n}\right)\right]^2\frac{J_n(t)}{n^2}\,dM_n^Z(t)=$$
$$=\int_0^x\frac{I\left(J_n(t)>0\right)}{J_n(t)}\left[\Psi\left(\frac{J_n(t)}{n}\right)\right]^2 d\Lambda^Z(t).$$

Therefore, $A_{2n}^2(x)-<A_{2n},A_{2n}>(x)$ is a also martingale and by Lenglart's inequality, $\forall\varepsilon,\eta>0$ we have

$$P\left(\sup_{0\le x\le\tau}\left|A_{2n}(x)\right|>\varepsilon\right)\le$$
$$\le\frac{\eta}{\varepsilon^2}+P\left(\int_0^\tau\frac{I\left(J_n(t)>0\right)}{J_n(t)}\left[\Psi\left(\frac{J_n(t)}{n}\right)\right]^2 d\Lambda^Z(t)>\eta\right)<$$
$$<\frac{\eta}{\varepsilon^2}+P\left(\frac{1}{J_n(\tau)}\int_0^\tau\left[\Psi\left(\frac{J_n(t)}{n}\right)\right]^2 d\Lambda^Z(t)>\eta\right).$$

According to Glivenko-Cantelli theorem for $n\to\infty$

$$\sup_{0\le x\le\infty}\left|\frac{J_n(x)}{n}-S^Z(x)\right|\xrightarrow{a.s.}0. \qquad (23)$$

Moreover, due to the boundedness of Ψ and Ψ' on $\left[S^Z(\tau),1\right]$ (condition (C1)), for $n\to\infty$ we have

$$\sup_{0\le x\le\tau}\left|\Psi^2\left(\frac{J_n(x)}{n}\right)-\Psi^2\left(S^Z(x)\right)\right|\xrightarrow{a.s.}0$$

Then by condition (C2) with $n\to\infty$

$$\int_0^\tau\left[\Psi\left(\frac{J_n(t)}{n}\right)\right]^2 d\Lambda^Z(t)\xrightarrow{a.s.}\int_0^\tau\Psi^2\left(S^Z(t)\right)d\Lambda^Z(t)<\infty,$$

and consequently, taking into account $J_n(\tau)\xrightarrow{a.s.}\infty$ we have convergence to zero of probability in the right side of (30), i.e.

$$\sup_{0\le x\le\tau}\left|A_{2n}(x)\right|\xrightarrow{p}0. \qquad (24)$$

By mean value theorem

$$\sup_{0\le x\le\tau}\left|A_{3n}(x)\right|\le$$
$$\le\sup_{0\le x\le\tau}\int_0^x I\left(J_n(t)>0\right)\cdot\left|\Psi'\left(\theta_n(t)\right)\right|\cdot\left|\frac{J_n(t)}{n}-S^Z(t)\right|d\Lambda^Z(t)\le$$
$$\le\sup_{0\le x\le\infty}\left|\frac{J_n(x)}{n}-S^Z(x)\right|\cdot\int_0^\tau\left|\Psi'\left(\theta_n(t)\right)\right|d\Lambda^Z(t),$$

where $\theta_n(t)\in\left[\min\left\{\frac{J_n(t)}{n},S^Z(t)\right\};\ \max\left\{\frac{J_n(t)}{n},S^Z(t)\right\}\right]$.

Now, by using (31) and condition (C3) at $n\to\infty$ have

$$\sup_{0\le x\le\tau}\left|A_{3n}(x)\right|\xrightarrow{p}0. \qquad (25)$$

From (29), (32) and (33) at $n\to\infty$

$$\sup_{0\le x\le\tau}\left|\varphi\left(\tilde S_n^Z(x)\right)-\varphi\left(S^Z(x)\right)\right|\xrightarrow{p}0, \qquad (26)$$

where the number $\tau<T$ so that $S^Z(\tau)>0$. In order to proof (C) as in the proof of (A) set $\tau=T-\varepsilon$. Given the monotonity of S^Z and φ we have

$$\sup_{0\le x\le T}\left|\varphi\left(\tilde S_n^Z(x)\right)-\varphi\left(S^Z(x)\right)\right|\le$$
$$\le\sup_{0\le x\le T-\varepsilon}\left|\varphi\left(\tilde S_n^Z(x)\right)-\varphi\left(S^Z(x)\right)\right|+$$
$$+\sup_{T-\varepsilon\le x\le T}\left|\varphi\left(\tilde S_n^Z(x)\right)-\varphi\left(S^Z(x)\right)\right|\le$$
$$\le\sup_{0\le x\le T-\varepsilon}\left|\varphi\left(\tilde S_n^Z(x)\right)-\varphi\left(S^Z(x)\right)\right|+$$
$$+\left|\varphi\left(\tilde S_n^Z(T)\right)-\varphi\left(\tilde S_n^Z(T-\varepsilon)\right)\right|+$$
$$+\left|\varphi\left(\tilde S_n^Z(T-\varepsilon)\right)-\varphi\left(S^Z(T-\varepsilon)\right)\right|+$$

International Journal of Innovation in Science and Mathematics
Volume 2, Issue 3, ISSN (Online): 2347–9051

$$+\left|\varphi\left(S^Z\left(T-\varepsilon\right)\right)-\varphi\left(S^Z\left(T\right)\right)\right|. \qquad (27)$$

Now, using (34) and letting ε tend to zero, we obtain (C). Let us prove (D). According to Taylor's expansion, condition (C1) and (31) under $n\to\infty$ we get

$$\sup_{0\le x\le T}\left|\varphi\left(S_n^Z\left(x\right)\right)-\varphi\left(\tilde{S}_n^Z\left(x\right)\right)\right|\le$$

$$\le\int_0^T\left|\varphi\left(\frac{\mathrm{J}_n\left(t\right)}{n}\right)-\varphi\left(\frac{\mathrm{J}_n\left(t\right)}{n}-\frac{1}{n}\right)-\right.$$

$$\left.-\frac{1}{n}\varphi'\left(\frac{\mathrm{J}_n\left(t\right)}{n}\right)\right|d\overline{\square}_n^Z\left(t\right)\le\frac{1}{2n^2}\int_0^T\left|\varphi''\left(\theta_n\left(t\right)\right)\right|d\overline{\square}_n^Z\left(t\right)\le$$

$$\le\frac{1}{2n}\sup_{0\le x\le T}\left|\varphi''\left(\theta_n\left(\delta\right)\right)\right|\overset{a.s.}{=}\mathrm{O}\left(\frac{1}{n}\right),$$

i.e. (D) is true. Clearly (E) is a consequence of (C) and (D). The proof of (F) is identical with that of (C). Now turn to the proof of the main statement (G) of uniform consistency of S_n^X. Consider the representation (23), where according to (D), (F) and (31) at $n\to\infty$

$$\sup_{0\le x\le T}\left|\mathrm{A}_{1n}^*\left(x\right)\right|\overset{p}{\to}0,\quad\sup_{0\le x\le T}\left|\mathrm{A}_{2n}^*\left(x\right)\right|\overset{a.s.}{=}\mathrm{O}\left(\frac{1}{n}\right),$$

$$\sup_{0\le x\le T}\left|\mathrm{A}_{3n}^*\left(x\right)\right|=o_p\left(\frac{1}{n}\right), \qquad (27)$$

Hence, by (23),

$$\sup_{0\le x\le T}\left|\varphi\left(S_n^X\left(x\right)\right)-\varphi\left(S^X\left(x\right)\right)\right|\overset{p}{\to}0. \qquad (28)$$

Now by the mean value theorem and condition (C1) from (28) we obtain (G). The theorem 2 is proved.

Now we demonstrate result on asymptotic normality of estimator (16). Introduce the stopped processes

$$q_n\left(x\right)=n^{1/2}\left(S_n^X\left(x\wedge Z^{(n)}\right)-S^X\left(x\wedge Z^{(n)}\right)\right),$$

where $a\wedge b=\min\left(a,b\right)$. Let $q\left(x\right)=e\left(x\right)\left[\varphi'\left(S^X\left(x\right)\right)\right]^{-1}+$ $+\xi\left[\varphi'\left(S^X\left(T_Z\right)\right)\right]^{-1}$, where $e\left(x\right)$ is mean zero Gaussian process with covariance function

$$A\left(x_1,x_2\right)=\int_0^{x_1\wedge x_2}S^Z\left(t\right)\left[\varphi'\left(S^Z\left(t\right)\right)\right]^2d\Lambda\left(t\right)+$$

$$+2\int_0^{x_1\wedge x_2}\int_0^t S^Z\left(t\right)\left(1-S^Z\left(s\right)\right)\Psi'\left(S^Z\left(t\right)\right)\Psi'\left(S^Z\left(s\right)\right)d\Lambda\left(s\right)d\Lambda\left(t\right)+$$

$$+\int_{x_1\wedge x_2}^{x_1\vee x_2}S^Z\left(t\right)\Psi'\left(S^Z\left(t\right)\right)d\Lambda\left(t\right)\int_0^{x_1\wedge x_2}\left[\left(1-S^Z\left(s\right)\right)\Psi'\left(S^Z\left(s\right)\right)+\right.$$

$$\left.+\varphi'\left(S^Z\left(s\right)\right)\right]d\Lambda\left(s\right),$$

$x_1\vee x_2=\max\left(x_1,x_2\right)$, $\xi\overset{D}{=}\square\left(0,\sigma_0^2\right)$ and $\sigma_0{}^2\left(x\right)=$ $=\lim_{x\to T_Z}A(x,x)$. Let $C(x)=\lim_{x\to T_Z}A(t,x)$.

Theorem 3. Let conditions (C1)-(C5) are hold, $\sigma_0^2<\infty$, and for every $x\in\left[0,T_Z\right)$: $C\left(x\right)<\infty$. Then for $n\to\infty$:

$$q_n\left(x\right)\overset{D}{\Rightarrow}q\left(x\right)\text{ in }D\left[0,T_Z\right]. \qquad (29)$$

The proof of the theorem 3. First, examine the process

$$\mathrm{D}_n\left(x\right)=n^{1/2}\left(\varphi\left(S_n^X\left(x\wedge Z^{(n)}\right)\right)-\varphi\left(S^X\left(x\wedge Z^{(n)}\right)\right)\right)$$

and show that when $n\to\infty$

$$\mathrm{D}_n\left(x\right)\overset{D}{\Rightarrow}\mathrm{D}\left(x\right)\quad\text{in }D\left[0,T_Z\right]. \qquad (30)$$

According to the representation (21)

$$\mathrm{D}_n\left(x\right)=\sum_{m=1}^3 n^{1/2}\mathrm{A}_{mn}^*\left(x\wedge Z^{(n)}\right). \qquad (31)$$

Since $Z^{(n)}\overset{\text{a.s.}}{\to}T_Z$ when $n\to\infty$, according to (27)

$$\sup_{0\le x\le T_Z}n^{1/2}\left|\mathrm{A}_{mn}^*\left(x\wedge z^{(n)}\right)\right|=\mathrm{O}_p\left(n^{-1/2}\right),\quad m=2,3 \qquad (32)$$

Therefore, to establish (30), taking into account (31) and (32), it suffices to prove

$$n^{1/2}\mathrm{A}_{1n}^*\left(x\wedge Z^{(n)}\right)\overset{D}{\Rightarrow}\mathrm{D}\left(x\right)\text{ in }D\left[0,T_Z\right]. \qquad (33)$$

Using formulas (7), (10) and (11), we have

$$n^{1/2}\mathrm{A}_{1n}^*\left(x\wedge Z^{(n)}\right)=n^{1/2}\left(-\frac{1}{n}\int_0^{x\wedge Z^{(n)}}I\left(S_n^Z\left(t\right)>0\right)S_n^Z\left(t-\right)\cdot\right.$$

$$\cdot\varphi'\left(S_n^Z\left(t\right)\right)d\Lambda_n\left(t\right)+$$

$$\left.+\int_0^{x\wedge Z^{(n)}}I\left(S_n^Z\left(t\right)>0\right)S^Z\left(t\right)\varphi'\left(S^Z\left(t\right)\right)d\Lambda\left(t\right)\right)=$$

$$=n^{1/2}\left(-\frac{1}{n}\int_0^{x\wedge Z^{(n)}}I\left(S_n^Z\left(t\right)>0\right)\cdot\varphi'\left(S_n^Z\left(t\right)\right)d\mathrm{M}_n\left(t\right)+\right.$$

$$\left.+\int_0^{x\wedge Z^{(n)}}I\left(S_n^Z\left(t\right)>0\right)\cdot\left[\Psi\left(S_n^Z\left(t\right)\right)-\Psi\left(S^Z\left(t\right)\right)\right]d\Lambda\left(t\right)\right)+\mathrm{O}\left(\frac{1}{n}\right),$$

where have used (2.2.5) and the equation $S_n^Z\left(t-\right)=S_n^Z\left(t\right)+\dfrac{1}{n}$. *By (1.2.4) subject to the conditions (C1),*

$$\varphi'\left(S_n^Z\left(t\right)\right)=\varphi'\left(S^Z\left(t\right)\right)+\mathrm{O}_p\left(\left(n^{-1}\ln n\right)^{1/2}\right),$$

$$\Psi\left(S_n^Z\left(t\right)\right)-\Psi\left(S^Z\left(t\right)\right)=$$

$$=\Psi'\left(S^Z\left(t\right)\right)\left(S_n^Z\left(t\right)-S^Z\left(t\right)\right)+\mathrm{O}_p\left(n^{-1}\ln n\right). \qquad (34)$$

we obtain from (33), (34) we have

$$n^{1/2}\mathrm{A}_{1n}^*\left(x\wedge Z^{(n)}\right)=B_{1n}\left(x\right)+B_{2n}\left(x\right)+o_p\left(1\right), \qquad (35)$$

where

$$B_{1n}\left(x\right)=-n^{1/2}\int_0^{x\wedge Z^{(n)}}\varphi'\left(S^Z\left(t\right)\right)d\mathrm{M}_n\left(t\right),$$

$$B_{2n}\left(x\right)=\int_0^{x\wedge Z^{(n)}}\Psi'\left(S^Z\left(t\right)\right)\varepsilon_n\left(t\right)d\Lambda\left(t\right).$$

$$\varepsilon_n\left(t\right)=n^{1/2}\left(S^Z\left(t\right)-S^Z\left(t\right)\right).$$

IJISM Source of Knowledge
International Journal of Innovation in Science and Mathematics
Volume 2, Issue 3, ISSN (Online): 2347–9051

According to (35), the convergence (33) follows from the convergence

$$B_n(x) \overset{D}{\Rightarrow} \mathrm{D}(x) \text{ in } D[0,T_Z], \quad (36)$$

where $B_n(x)=B_{1n}(x)+B_{2n}(x)$. In the paper [6] (see equation (16)) states that for any $\tilde{o}_0$, such that $S^Z(x_0)>0$, $B_n(x)$ converges weakly to a $\mathrm{D}(x)$ in $D[0,x_0]$. Therefore, to prove (36) according to the criterion of weak convergence the density is $B_n(x)$, for $\hat{e}=1,2$ and for any $\varepsilon>0$:

$$\lim_{y\to T_Z}\limsup_{n\to\infty} P\left(\sup_{x\in(y,T_Z)}\left|B_{kn}(x)-B_{kn}(y)\right|>\varepsilon\right)=0. \quad (37)$$

For $k=1$

$$B_{1n}(x)-B_{1n}(y)=-n^{-1/2}\int_{y\wedge Z^{(n)}}^{x\wedge Z^{(n)}}\varphi'(S^Z(t))d\mathrm{M}_n(t). \quad (38)$$

Note that (38) is a martingale integral form with the stopping time, and then by the inequality Lenglart for $\forall\varepsilon,\eta>0$ we have

$$P\left(\sup_{x\in[y,T_Z]}\left|B_{1n}(x)-B_{1n}(y)\right|>\varepsilon\right)\le\frac{\eta}{\varepsilon^2}+$$

$$+P\left(\frac{1}{n}\int_{y\wedge Z^{(n)}}^{T_Z}\left(\varphi'(S^Z(t))\right)^2\mathrm{J}_n(t)d\Lambda(t)>\eta\right)\le\frac{\eta}{\varepsilon^2}+$$

$$+P\left(\int_y^{T_Z}\left(\varphi'(S^Z(t))^2 S_n^Z(t)\right)d\Lambda(t)>\eta\right). \quad (39)$$

According to theorem Glivenko-Cantelli for $n\to\infty$

$$\int_y^{T_Z}\left(\varphi'(S^Z(t))\right)^2 S_n^Z(t)d\Lambda(t)\overset{P}{\to}\int_y^{T_Z}\left(\frac{\Psi(S^Z(t))}{S^Z(t)}\right)^2 d\Lambda(t).$$

Consequently, for $y\to T_Z$ in view of the condition (C4) and the arbitrariness $\eta>0$ converge to zero, i.e. (37) for $k=1$ rightly. Since the empirical process $\varepsilon_n(t)$ converges weakly in $D[0,T_Z]$ to a Brownian bridge $B(1-S^z(t))$ by the theorem of Doob-Donsker and in view of condition (C3) and presentation

$$B_{2n}(x)-B_{2n}(y)=\int_{y\wedge Z^{(n)}}^{x\wedge Z^{(n)}}\Psi'(S^Z(t))\varepsilon_n(t)d\Lambda(t)$$

verify the validity of (37) and in the case for $k=2$. Now the density $B_n(x)$ follows from (37) by the triangle inequality. Thus, (30) holds. To prove (29) it suffices to note that under condition (C1)

$$q_n(x)=\mathrm{D}_n(x)\left[\varphi'\left(S^X\left(x\wedge Z^{(n)}\right)\right)\right]^{-1}+o_p(1)= \quad (40)$$

$$=\mathrm{D}_n(x)\left\{\left[\varphi'(S^X(x))\right]^{-1}I\left[0,Z^{(n)}\right)+\left[\varphi'\left(S^X\left(Z^{(n)}\right)\right)\right]^{-1}I\left\{Z^{(n)}\right\}\right\}+o_p(1).$$

Now (29) follows from (40) for $n\to\infty$. The theorem 3 is proved.□

Remark. Consider independent censoring model (i.e. $\{X_k\}$ and $\{Y_k\}$ are mutually independent). In this case in (2) $C(u;v)=uv=C^*(u;v)$, $u,v\in[0,1]$ and hence $\varphi(u)=-\log u$, $u\in[0,1]$ and $\varphi^{[-1]}(t)=\varphi^{-1}(t)=$ $=\exp(-t)$, so that

$$S^Z(x)=S^X(x)S^Y(x),\ x\in\overline{R}^+. \quad (41)$$

It is easy to verify that from (12) and (16) respectively we obtain the exponential-hazard estimator

$$\tilde{S}_n^X(x)=\exp\left\{-\int_0^x\frac{I(\mathrm{J}_n(t)>0)}{\mathrm{J}_n(t)}d\overline{\mathrm{N}}_n(t)\right\}, \quad (42)$$

and relative-risk power estimator of Abdushukurov (1998) (see[1]):

$$S_n^X(x)=\left[S_n^Z(x)\right]^{R_n(x)},\ R_n(x)=\frac{\Lambda_n(x)}{\Lambda_n^Z(x)}. \quad (43)$$

Note that the estimator (12) is investigated in [3]. Moreover the Zeng-Klein's (1994) copula-graphic estimator is (see [3,6]):

$$\hat{S}_n^X(x)=\varphi^{[-1]}\left[\int_0^x I(\mathrm{J}_n(t)>0)\left(\varphi\left(\frac{\mathrm{J}_n(t)-1}{n}\right)-\varphi\left(\frac{\mathrm{J}_n(t)}{n}\right)\right)\right]d\overline{\mathrm{N}}_n(t),$$

(44)

which in independence model (41) is reduced to well - known Kaplan- Meier product - limit estimator (see [9])

$$\hat{S}_n^X(x)=\prod_{t\le x}\left\{1-\frac{d\overline{\mathrm{N}}_n(t)}{\mathrm{J}_n(t)}\right\}. \quad (45)$$

Let $\tilde{S}_n^Y$, S_n^Y and $\hat{S}_n^Y$ are respectively estimators of S^Y of exponential-hazard, relative-risk power and product-limit structures obtained from formulas (42), (43) and (45) by using events $\delta_k=0$ instead of $\delta_k=1$. Then we have:

(a) $\tilde{S}_n^X(x)\tilde{S}_n^Y(x)=\exp\{-\Lambda_n^Z(x)\}\ne S_n^Z(x)$ and for $x\ge Z_{(n)}=\max\{Z_k,1\le k\le n\}$, $\max\{\tilde{S}_n^X(x),\tilde{S}_n^Y(x)\}<1$;

(b) $S_n^X(x)S_n^Y(x)=S_n^Z(x)$ for all $x\in\overline{R}^+$ and $S_n^X(x)=S_n^Y(x)=0$, for $x\ge Z_{(n)}$;

(c) $\hat{S}_n^X(x)\hat{S}_n^Y(x)\ne S_n^Z(x)$ and for $x\ge Z_{(n)}$ the estimators $\hat{S}_n^X$ and $\hat{S}_n^Y$ are undefined. Moreover the estimators $\hat{S}_n^X$ and $\hat{S}_n^Y$ require also the condition $P(X_k=Y_k)=0$, $k=1,2,\dots,$ which in many practical situations is not hold. Thus only the relative-risk power estimators have identifiability properties with independence censoring model satisfying empirical analogue of equality (41). Analogously a new estimator (16) is more suitable estimator for $\hat{S}_n^X$ than the estimators (12) and (43).

International Journal of Innovation in Science and Mathematics
Volume 2, Issue 3, ISSN (Online): 2347–9051

IV. Conclusions

In figures 1 and 2 below we demonstrate plots of estimators (12), (16) and (44) of $\hat{S}_n^X$ using well-known Channing House data of size n=97 (see [1], [7], [8]). Here, thin-solid line stands for $\tilde{S}_n^X$, medium-one for $\hat{S}_n^X$ and thick-solid line stands for a new estimator S_n^X. Note that estimate S_n^X is defined in whole line.

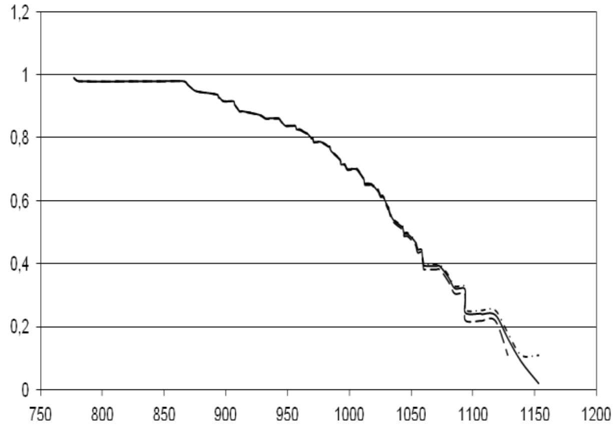

Fig.1. Plots of estimates $\tilde{S}_n^X$ (thin-solid), $\hat{S}_n^X$ (medium one) and S_n^X (thick-solid) for copula generator $\varphi(u) = -\ln u, u \in [0,1]$.

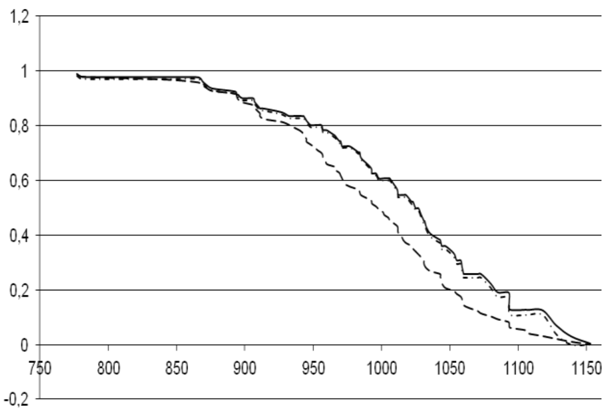

Fig.2. Plots of estimates $\tilde{S}_n^X$ (thin-solid), $\hat{S}_n^X$ (medium one) and S_n^X (thick-solid) for copula generator $\varphi(u) = (-\ln u)^2, u \in [0,1]$.

Acknowledgements

This work was supported by the grant F4-01 Republic of Uzbekistan.

References

[1] A.A. Abdushukurov Estimation of unknown distributions by incomplete observations and its properties, LAMBERT Academic Publishing, 2011.

[2] A.A. Abdushukurov, N.T. Dushatov, R.S. Muradov Estimating of functionals of a multidimensional distribution by censored observations with using copula functions. In: Statistical Methods of Estimation and Hyphoteses Testing. Perm State University. Issue 23, 2011. 36-47.

[3] Y. Li, R.C. Tiwari, S. Guha Mixture cure survival models with dependent censoring. J. Royal Statist. Soc. B. v. 69. Part 3. 2007 285-306.

[4] R.S. Muradov, A.A. Abdushukurov Estimation of multivariate distributions and its mixtures by incomplete data. LAMBERT Academic Publishing, 2011.

[5] R.B. Nelsen An introduction to copulas. - Springer, New York. 2006.

[6] L.-P. Rivest, M.T. Wells A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. J. Multivar.Anal.v.79. 2001. 138-155.

[7] S. Csörgő Estimating in the proportional hazards model of random censorship. Statistics. v.19. №3.1988. p.437-463.

[8] B. Efron Censored data and the bootstrap. J.A.S.A. v.76. №374. 1981. p.312-319.

[9] E.L. Kaplan, P. Meier Nonparametric estimation from incomplete observations. J.A.S.A. v.53. 1958. p.457-481.

Author's Profile

Abdushukurov Abdurahim Ahmedovich **(born 1959)**.

Head of Department of Probability Theory and Mathematical Statistics, faculty of Mechanics and Mathematics, National University of Uzbekistan. Prof. A.A. Abdushukurov's scientific interests include problems of nonparametric statistical estimation by incomplete observations. He proposed and studied few new functionals from unknown sub-distributions and using them built nonparametric estimates in generalized models of incomplete observations. These models include models of random censoring and competing risks. Estimates are studied as in fixed as well as in random sample sizes. While studying estimates properties, along with analytical methods of probability theory and mathematical statistics, he used asymptotical theory of empirical processes, method of martingales, strong and weak approximations and U-statistics. Prof. A.A. Abdushukurov's main results are followings:
He proposed basis functionals of exponential, product and power structures and studied their properties. Using these basis functionals he built nonparametric estimates in generalized models of incomplete observations.
Set properties of uniform strong consistency with speed of convergence – laws of iterated logarithm type, both in homogenous and in variable random censoring.
Proved generalized Gaussian approximations in weak and strong forms with non-improvable speeds of approximation. He proved that three classes of estimates converge to the same Gaussian process.
He found and studied similar properties of nonparametric estimates with random Poisson size of sample and in Bayesian estimating using a priori Dirichlet distribution. Estimates on their own are new. He is co-author of celebrated ACL (Abdushukurov-Cheng-Lin)-estimate of survival function.
He proposed and systematically studied nonparametric estimates of multivariate survival functions, when incomplete data are non-homogenous, dependent, censored random vectors. In Cox's regression model with random censoring from both sides he built and studied three classes of parametric-nonparametric estimates for basis functionals of survival function. Using method of strong approximation in several models of random censoring he obtained presentations for loglikelihood ratio statistics and its truncated analogy through stochastic integrals from two-parametrical Wiener processes. Majority of these results are essential investments to the theory of nonparametric statistics of incomplete observations. They can be applied to the analysis of survival data in insurance, demography and astronomy.
Prof. A.A. Abdushukurov reads lectures and conducts practical lessons for Bachelor's students in the following courses: "Probability theory and applied statistics", "Mathematical statistics", "Multivariate statistical analysis", "Statistics of incomplete observations" and special courses in "Non-classic models of statistics", "Regression analysis" and "Statistical hypothesis testing" for Master's students using latest advances in the field of mathematical statistics.

Under his supervision there were total of 35 Master's dissertations through the department of "Probability Theory and Mathematical Statistics" and currently under Prof. A.A. Abdushukurov supervision defended his dissertation 5 post-graduate students.

Prof. A.A. Abdushukurov is the author about of 230 scientific and methodical works.

Muradov Rustamjon Sobitkhonovich (born 1985)

Ph.D. of Department of Probability Theory and Mathematical Statistics, Institute of Mathematics, National University of Uzbekistan. R.S. Muradov's scientific interests include problems of parametric and nonparametric statistical estimation by the model dependent random censoring. He is the author about of 30 scientific and methodical works.