[Bulletin of National University of Uzbekistan: Mathematics and Natural Sciences](https://bulletin.nuu.uz/journal)

Volume 4 | Issue 3 Article 2

9-15-2023

Asymptotic results for empirical processes in informative model of random censorship from both sides

Abdurakhim Abdushukurov
Moscow State University Tashkent branch, a_abdushukurov@rambler.ru

Dilshod Mansurov
Navoi State Pedagogical Institute, mathematicianmd@gmail.com

Recommended Citation

Abdushukurov, Abdurakhim and Mansurov, Dilshod (2023) "Asymptotic results for empirical processes in informative model of random censorship from both sides," *Bulletin of National University of Uzbekistan: Mathematics and Natural Sciences*: Vol. 4: Iss. 3, Article 2.
DOI: https://doi.org/10.56017/2181-1318.1195

ASYMPTOTIC RESULTS FOR EMPIRICAL PROCESSES IN INFORMATIVE MODEL OF RANDOM CENSORSHIP FROM BOTH SIDES

Abdushukurov A.A.[1], Mansurov D.R.[2]
[1] *Moscow State University Tashkent branch, Tashkent, Uzbekistan*
[2] *Navoi State Pedagogical Institute, Navoi, Uzbekistan*
e-mail: a_abdushukurov@rambler.ru, mathematicianmd@gmail

Abstract

In the paper, the empirical process in informative model of random censorship from both sides is investigated. For it, the limit Gaussian process with mean zero is founded. Under investigating of empirical process, the characterization properties of the considered informative model is used. The properties of the semiparametric estimator by using methods of numerical modeling are discussed.

Keywords: *Random censorship, informative model, proportional hazards, empirical process, Gaussian process.*

Mathematics Subject Classification (2010): *62N01, 62N05.*

Introduction

In biomedical studies of individuals for survival, in engineering tests of technical devices for reliability, there may be cases when the test objects fall under observation after a certain random period after the start of testing. This phenomenon is called delayed entry or left random censoring since in this case, the random variable (r.v.) X that characterizes the lifetime of the tested object becomes to observation under the condition $X \geq L$. This L is the moment when the object was placed under surveillance. In addition to this, r.v. X may also be censored from the right by some other r.v. Y. When the operating time, we are interested in before failure X will be subject to random censorship from both sides by random vector $(L,\ Y)$. Let r.v.-s $L,\ X$ and Y are mutually independent with continuous distribution functions (d.f.) K, F and G respectively.

1 Informative model from both sides

In the informative model of random censorship the survival distribution of the censoring variables is some power of the survival distribution of the lifetimes. It often called also as proportional hazards model (PHM) so far as it equivalent to proportionality of corresponding hazard functions. The PHM is an appealing and potentially useful special semiparametric model of random censorship. One of the great advantages of this model that it makes possible easy and heuristically clear interpretation of results and conditions formulated in the general censorship model. PHM was considered in the framework of right random censorship by authors [16], and beginning with [9] it is often referred to as Koziol-Green model. Csörgő [10] gave a complete survey of the estimation theory for many reliability functions based on power-type estimator often called as ACL-(Abdushukurov-Cheng-Lin) estimator (for more information see, [2, 3, 4, 5, 6, 10, 11, 12, 13, 15, 20, 21])

In this paper we present some asymptotic properties of power-type semiparametric estimator of survival function of the lifetime in informative model of random censorship from both sides. We establish weak convergence result for empirical process and present some numerical modeling comparative results.

Let $\{(X_k,\, L_k,\, Y_k)\,,\ k \geq 1\}$ be a sequence of independent realizations of the triple $(X,\, L,\, Y)$ and

$$S^{(n)} = \{(\mathrm{Z}_i,\, \Delta_i)\,,\, i = 1,\, \dots, n\}$$

-the observed sample, where $\mathrm{Z}_i = \max\{L_i,\, \min\{X_i,\, Y_i\}\}\,,\ \Delta_i = \left(\delta_i^{(0)},\, \delta_i^{(1)},\, \delta_i^{(2)}\right)$, with $\delta_i^{(0)} = I\left(\min\left(X_i,\, Y_i\right) < L_i\right),\ \delta_i^{(1)} = I\left(L_i \leq X_i < Y_i\right),\ \delta_i^{(2)} = I\left(L_i \leq Y_i < X_i\right)$ and $I\left(A\right)$ standing for indicator of the event A. Note that in sample $S^{(n)}$ the number of observed r.v.-s X_i is equal to $\delta_1^{(1)} + \dots + \delta_n^{(1)}$. The statistical task is consist in estimating of d.f. F from a sample $S^{(n)}$. However, such a general statement of the estimating problem, d.f.-s K and G are considered as a nuisance. In this paper, we will investigate the evaluation of d.f. F in the case of informative censoring from two sources, when d.f.-s K and G functionally depend on F. To describe such a model by H and N we denote d.f.-s of r.v.-s Z_i and $V_i = \min\left(X_i,\, Y_i\right)$. Then it is easy to see, that

$$H\left(x\right) = K\left(x\right) N\left(x\right),\quad N\left(x\right) = 1 - \left(1 - F\left(x\right)\right)\left(1 - G\left(x\right)\right),\quad x \in \mathrm{R}^1. \tag{1}$$

Assume that there are positive unknown parameters $\theta,\, \beta$ such that following representations are valid for all $x \in \mathrm{R}^1$:

$$\begin{cases} 1 - G\left(x\right) = \left(1 - F\left(x\right)\right)^{\theta}, \\ K\left(x\right) = \left(N\left(x\right)\right)^{\beta}, \end{cases} \tag{2}$$

where the parameter β is responsible for the power of censoring from the left, and θ-from the right. The proximity of the values of these parameters to zero determines the weakness of the corresponding censorship. Such a special model of random censorship on both sides was introduced in [3] and generalized for the case of competing risks in [4]. In this model, the censoring parameters θ and β, which determine the deepness of censorship from both sides. From formulas (1) and (2), it is not difficult to derive the following representation for d.f. F:

$$1 - F\left(x\right) = \left[1 - \left(H\left(x\right)\right)^{\lambda}\right]^{\gamma},\quad x \in \mathrm{R}^1, \tag{3}$$

where $\lambda = \frac{1}{1+\beta},\ \ \gamma = \frac{1}{1+\theta}$ and therefore, the closeness of the parameters λ and γ to 1, denotes the weakness of the censoring. Using representation (3), we can construct a semiparametric estimate for F over a sample $S^{(n)}$ by estimating a triple $\left(H\left(x\right),\, \lambda,\, \gamma\right)$.

For this purpose, we also use the characterization properties of the model under consideration. In this regard, we define the sub-distributions $\left\{T^{(m)}(x)=P\left(\mathrm{Z}_i\le x,\delta_i^{(m)}=1\right),m=0,1,2\right\}$ by the following formulas:

$$T^{(0)}(x)=\int_{-\infty}^{x}N(s)\,dK(s)\,,\ \ T^{(1)}(x)=\int_{-\infty}^{x}K(s)\,(1-G(s))\,dF(s)\,,$$

$$T^{(2)}(x)=\int_{-\infty}^{x}K(s)\,(1-F(s))\,dG(s)\,.$$

It's easy to see that $T^{(0)}(x)+T^{(1)}(x)+T^{(2)}(x)=H(x)\,,\ \ x\in\mathrm{R}^1$. It should be noted that this model under consideration generalizes the well-known proportional hazards model of Koziol-Green, which studied by the authors [2, 3, 4, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 20, 21] which obtained from (2) under $\beta=0$, i.e. in the absence of random left side censoring ($K(x)\equiv 1$), and is a special case of the model from [5, 6] in the absence of a covariate. The following theorem from [3] characterizes the model (2).

Theorem 1. *[3] Equalities* (2) *hold if and only if the r.v.-s* Z_i *and* Δ_i *are independent.*

Taking into account theorems 1 and (2), it is easy to establish that

$$T^{(0)}(x)=(1-\lambda)\,H(x)\,,T^{(1)}(x)=\gamma\lambda H(x)\,,T^{(2)}(x)=(1-\gamma)\,\lambda H(x)\,. \tag{4}$$

Passing to the limit in relations (4) for $t\to+\infty$, we have

$$P\left(\delta_i^{(0)}=1\right)=1-\lambda,P\left(\delta_i^{(1)}=1\right)=\gamma\lambda,P\left(\delta_i^{(2)}=1\right)=(1-\gamma)\,\lambda. \tag{5}$$

Estimating the probabilities $p^{(m)}=P\left(\delta_i^{(m)}=1\right),\ \ m=0,\,1,\,2$ by the corresponding frequencies $p_n^{(m)}=\frac{1}{n}\sum_{i=1}^{n}\delta_i^{(m)},\ \ m=0,\,1,\,2$ from formulas (5), we find the following estimates of the parameters λ and γ: $\lambda_n=1-p_n^{(0)}$, $\gamma_n=p_n^{(1)}\left(1-p_n^{(0)}\right)^{-1}$. For d.f. $H(x)$ we use an empirical formula

$$H_n(x)=\frac{1}{n}\sum_{i=1}^{n}I(\mathrm{Z}_i\le x)\,,\ \ x\in\mathrm{R}^1.$$

Now, because of (3), we construct the corresponding estimate for d.f. $F(x)$ by the plug in method:

$$F_n(x)=1-\left[1-(H_n(x))^{\lambda_n}\right]^{\gamma_n},\ \ x\in\mathrm{R}^1. \tag{6}$$

It is easy to see that because of representations (3) in the model (2):

$$\begin{aligned}T_F=T_G=T_N=T_K=T_H=\inf\left\{x\in\mathrm{R}^1:\ H(x)=1\right\},\\ \tau_F=\tau_G=\tau_N=\tau_K=\tau_H=\sup\left\{x\in\mathrm{R}^1:\ H(x)=0\right\}.\end{aligned}$$

We introduce a normalized process constructed by estimator (6):

$$\left\{Q_n(x)=\sqrt{n}\,(F_n(x)-F(x))\,,\ \ x\in R^1\right\}, \tag{7}$$

Select the interval $D = [\tau, T] \subset \mathrm{R}^1$ so that $\tau_H < \tau \leq T < T_H$ and $q = \sup\limits_{x \in D} \left[(H(x))^{p^{(0)}} - H(x) \right]^{-1} > 0.$

The statment about the limiting Gaussian property of the sequence (7) is the consent of the next theorem.

Theorem 2. *Suppose that $q > 0$. Then the sequence of random processes $\{Q_n(x), \ x \in D\}$ converges weakly to a central Gaussian process $\{A(x), x \in D\}$ with a covariance structure for $x_1, x_2 \in D$:*

$$\begin{aligned}
&cov\{A(x_1), A(x_2)\} = (1 - F(x_1))(1 - F(x_2)) \times \\
&\times \{a(x_1) a(x_2) [H(\min(x_1, x_2)) - H(x_1) H(x_2)] + b(x_1) b(x_2) \times \\
&\times p^{(0)} \left(1 - p^{(0)}\right) + c(x_1) c(x_2) p^{(1)} \left(1 - p^{(1)}\right) - 2b(\min(x_1, x_2)) c(\min(x_1, x_2)) p^{(0)} p^{(1)}\},
\end{aligned}$$

where

$$\begin{aligned}
a(x) &= p^{(1)} \left[(H(x))^{p^{(0)}} - H(x) \right]^{-1}, \\
b(x) &= -\left[\tfrac{p^{(1)}}{\left(1 - p^{(0)}\right)} c(x) + \tfrac{a(x)}{\left(1 - p^{(0)}\right)} H(x) \log H(x) \right], \\
c(x) &= -\tfrac{1}{\left(1 - p^{(0)}\right)} \log \left[1 - (H(x))^{1 - p^{(0)}} \right].
\end{aligned}$$

Proof. At first we establish the following asymptotic representation, which decomposes $F_n(x) - F(x)$ into normalized sum of independent and identically distributed random functions and a remainder term:

$$F_n(x) - F(x) = \frac{1}{n} \sum_{i=1}^{n} \Psi_x(Z_i, \Delta_i) + r_n(x), \tag{8}$$

where

$$\begin{aligned}
&\Psi_x(Z_i, \Delta_i) = (1 - F(x)) \left\{ p^{(1)} \left[(H(x))^{p^{(0)}} - H(x) \right]^{-1} \right\} (I(Z_i \leq x) - H(x)) - \\
&- \left[\tfrac{p^{(1)}}{(1 - p^{(0)})^2} \log \left[1 - (H(x))^{1 - p^{(0)}} \right] + \tfrac{p^{(1)}}{1 - p^{(0)}} H(x) \log H(x) \left[(H(x))^{p^{(0)}} - H(x) \right]^{-1} \right] \times \\
&\times \left(\delta_i^{(0)} - p^{(0)} \right) - \tfrac{1}{1 - p^{(0)}} \log \left[1 - (H(x))^{1 - p^{(0)}} \right] \left(\delta_i^{(1)} - p^{(1)} \right),
\end{aligned}$$

and

$$\sup_{x \in D} |r_n(x)| \overset{a.s.}{=} \mathrm{O}\left(\frac{\log n}{n} \right). \tag{9}$$

Consider a function

$$f(u, y, z) = \left(1 - u^{1-y}\right)^{\frac{z}{1-y}}, \ \ u, y, z \in (0, 1).$$

Then by (3) and (6),

$$1 - F(x) = f\left(H(x), p^{(0)}, p^{(1)}\right) = f_0, \ \ 1 - F_n(x) = f\left(H_n(x), p_n^{(0)}, p_n^{(1)}\right) = f_n.$$

also $F_n(x) - F(x) = -(f_n - f_0)$. By second order Taylor expansion

$$-(f_n - f_0) = \frac{1}{n}\sum_{i=1}^{n} \Psi_x(Z_i,\, \Delta_i) + q_n(x). \tag{10}$$

Here under $(u,\, y,\, z) = \left(H_n(x),\, p_n^{(0)},\, p_n^{(1)}\right)$ and $(u_0,\, y_0,\, z_0) = \left(H(x),\, p^{(0)},\, p^{(1)}\right)$ we have

$$\begin{aligned} q_n(x) = \tfrac{1}{2}\left[f''_{uu}\cdot(u-u_0)^2 + f''_{yy}\cdot(y-y_0)^2 + f''_{zz}\cdot(z-z_0)^2\right] + \\ + f''_{uy}\cdot(u-u_0)(y-y_0) + f''_{uz}\cdot(u-u_0)(z-z_0) + f''_{yz}\cdot(y-y_0)(z-z_0), \end{aligned}$$

where $f_{uy} = f_{yu},\ f_{uz} = f_{zu},\ f_{yz} = f_{zy}$ and second order derivatives calculated at intermediate point $(u_*,\, y_*,\, z_*) = \left(H_*(t),\, p_*^{(0)},\, p_*^{(1)}\right)$. By using elementary inequality for $\alpha,\, \gamma > 0$

$$0 \le \vartheta^{\gamma}\left|\log\vartheta\right|^{\alpha} \le \left(\frac{\alpha}{\gamma}\right) e^{-\alpha},\ 0 < \vartheta \le 1,$$

from (10) we obtain

$$\sup_{x\in D}\left|r_n(x)\right| \overset{a.s.}{=} \sum_{j=1}^{5} r_{jn}, \tag{11}$$

where

$$\begin{aligned} r_{1n} &= \mathrm{O}\left(\left(\sup_{x\in D}\left|H_n(x) - H(x)\right|\right)^2\right), \\ r_{2n} &= \mathrm{O}\left(\left(p_n^{(0)} - p^{(0)}\right)^2\right) + \mathrm{O}\left(\left(p_n^{(1)} - p^{(1)}\right)^2\right), \\ r_{3n} &= \mathrm{O}\left(\left|p_n^{(0)} - p^{(0)}\right| \cdot \sup_{x\in D}\left|H_n(x) - H(x)\right|\right), \\ r_{4n} &= \mathrm{O}\left(\left|p_n^{(1)} - p^{(1)}\right| \cdot \sup_{x\in D}\left|H_n(x) - H(x)\right|\right), \\ r_{5n} &= \mathrm{O}\left(\left|p_n^{(0)} - p^{(0)}\right| \cdot \left|p_n^{(1)} - p^{(1)}\right|\right). \end{aligned}$$

By Dvoretzky-Kiefer-Wolfowitz exponential inequality for empirical estimators [14, 19], we have

$$\begin{aligned} &\sup_{x\in D}\left|H_n(x) - H(x)\right| \overset{a.s.}{=} \mathrm{O}\left(\left(\tfrac{\log n}{n}\right)^{1/2}\right), \\ &\left|p_n^{(m)} - p^{(m)}\right| \le \sup_{x\in D}\left|T_n^{(m)}(x) - T^{(m)}(x)\right| \overset{a.s.}{=} \mathrm{O}\left(\left(\tfrac{\log n}{n}\right)^{1/2}\right), \end{aligned} \tag{12}$$

where for $m = 0,\, 1$, $T_n^{(m)}(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(Z_i \le x,\, \delta_i^{(m)} = 1\right)$. Now (8) and (9) follows from formulas (10)-(12). From (8) follows that the empirical process $\sqrt{n}\left(F_n(x) - F(x)\right)$ is strong approximated by normalized sum $A_n(x) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\Psi_x(Z_i,\, \Delta_i)$ with the rate $\mathrm{O}\left(n^{-\frac{1}{2}}\log n\right)$ a.s.

Hence, it sufficient to prove the weak convergence of the process $A_n(x)$ to $A(x)$, $x \in D$ (see, Theorem 4.1 in [7]). We first show convergence of the finite dimensional distributions, i.e. for any $N = 1, 2, ...$ and $\tau \le x_1 \le ... \le x_N \le T$: $(A_n(x_1), ..., A_n(x_N)) \xrightarrow{d} N\left(0; (\gamma_{x_i x_j})\right)$. Since $A_n(x_i) = \sum_{k=1}^{N} A_{nki}$ where $A_{nki} = n^{-\frac{1}{2}} \Psi_{x_i}(Z_k, \Delta_k)$, it suffices to check that (see, [1]):

$$\lim_{n\to\infty} \sum_{k=1}^{n} EA_{nki}A_{nkj} = \gamma_{x_i x_j}, \ (1 \le i, \ j \le N) \tag{13}$$

$$\lim_{n\to\infty} \sum_{k=1}^{n} \int_{\{|A_{nk}|>\varepsilon\}} |A_{nk}|^2 \, dP = 0, \tag{14}$$

for every $\varepsilon > 0$, where $|A_{nk}|^2 = \sum_{i=1}^{N} A_{nki}^2$. Now it is easy to see that

$$\sum_{k=1}^{n} EA_{nki}A_{nkj} = \frac{1}{n} \sum_{k=1}^{n} Cov\left(\Psi_{x_i}(Z_k, \delta_k), \Psi_{xj}(Z_k, \delta_k)\right) = \gamma_{x_i x_j},$$

i.e. (13) is hold.

Also, since the functions $\Psi_{x_i}(Z_k, \delta_k)$ are uniformly bounded, we have that $\max_{1\le k\le n} |A_{nk}| \overset{a.s.}{=} \mathrm{O}\left(n^{-\frac{1}{2}}\right)$ and hence,

$$\sum_{k=1}^{n} \int_{\{|A_{nk}|>\varepsilon\}} |A_{nk}|^2 \, dP \le \int_{\left\{\max\limits_{1\le k\le n} |A_{nk}|>\varepsilon\right\}} \sum_{k=1}^{n} |A_{nk}|^2 \, dP \le$$
$$\le \mathrm{O}(1) \, P\left(\max_{1\le k\le n} |A_{nk}| > \varepsilon\right) = o(1), \tag{15}$$

i.e. (14) also hold. Now from (13) and (14) follows convergence of finite dimensional distributions. Finally, we verify the moment conditions for tightness for all $\tau \le r \le s \le t \le T$:

$$E\left[(A_n(r) - A_n(s))^2 (A_n(s) - A_n(t))^2\right] \le (\gamma(r) - \gamma(s))(\gamma(s) - \gamma(t)) + c_n, \tag{16}$$

For some continuous and non-decreasing function $\gamma(s)$ and some sequence $c_n = o(1)$.

Let's denote $\Psi_{x_i} = \Psi_x (Z_i,\, \Delta_i)$. Then, the left hand side of (16) equals

$$n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} E\left[(\Psi_{ri}-\Psi_{si})(\Psi_{rj}-\Psi_{sj})(\Psi_{sk}-\Psi_{tk})(\Psi_{sl}-\Psi_{tl})\right] =$$

$$= n^{-2}\{\sum_{i=1}^{n} E\left[(\Psi_{ri}-\Psi_{si})^2\right]\cdot \sum_{j=1,j\neq i}^{n} E\left[(\Psi_{sj}-\Psi_{tj})^2\right] +$$

$$+2\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} E\left[(\Psi_{ri}-\Psi_{si})(\Psi_{si}-\Psi_{ti})\right] E\left[(\Psi_{rj}-\Psi_{sj})(\Psi_{sj}-\Psi_{tj})\right] +$$

$$+\sum_{i=1}^{n} E\left[(\Psi_{ri}-\Psi_{si})^2(\Psi_{si}-\Psi_{ti})^2\right]\} \leq$$

$$\leq 3n^{-2}\sum_{i=1}^{n} E\left[(\Psi_{ri}-\Psi_{si})^2\right]\sum_{i=1}^{n} E\left[(\Psi_{si}-\Psi_{ti})^2\right] + \mathrm{O}\left(n^{-3}\right),$$

using the inequalities of Cauchy-Bunyakovsky and Hölder and the fact that the Ψ functions are uniformly bounded. Now, we have

$$n^{-2}\sum_{i=1}^{n} E\left[(\Psi_{ri}-\Psi_{si})^2\right] = n^{-2}\sum_{i=1}^{n}\left[D\Psi_{ri}+D\Psi_{si}-2Cov(\Psi_{ri},\,\Psi_{si})\right].$$

From here follows (16) with correspondingly $\gamma(s)$, the details are omitted. Note that the covariation structure of limiting Gaussian process $\{A(x),\, x\in D\}$ is calculated by using of independence of r.v.-s Z_i and $\Delta_i = \left(\delta_i^{(0)},\, \delta_i^{(1)},\, \delta_i^{(2)}\right)$ according to theorem 1. Theorem 2 is proved.

Theorem 2 was announced in [18].

2 Studies of the dependence of the estimator on unknown parameters of censorship

Numerical studies were conducted using the python programming language to clarify how much the estimate $F_n(x)$ depends on the parameters, demonstrating its proximity to the estimated function $F(x)$ (see, figures 1-8). As $F(x)$ we use the standard normal and exponential d.f.-s. Simulated volume data was used for this purpose is $n = 5000$. From figures 1-8, we can see that when $0<\beta,\,\theta\leq 1$ the estimator $F_n(x)$ is well approximated to $F(x)$ (figures 1-4) in other cases, with an increase in censoring on the right at $\theta = 4$ (figures 5, 6, 9, 10) and an increase in censoring on the left at $\beta = 4$ (figures 7, 8), the discrepancy between the estimate and d.f. is noticeable, right and left respectively.

It is known that it $F(x)$ - is a continuous function, $F_n^{\ni}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$- corresponding to an empirical function, constructed from a complete sample $X^{(n)} =$

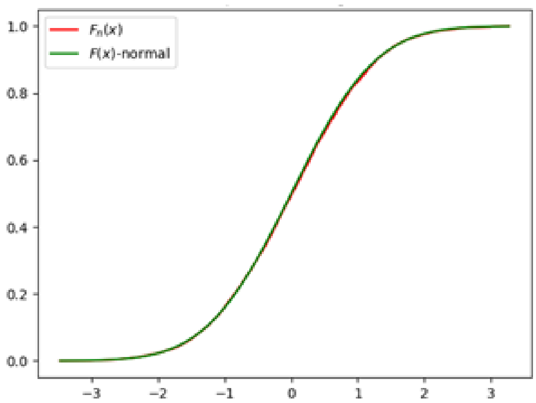

Figure 1: $\beta = 0.3, \quad \theta = 0.2, \quad n = 5000$

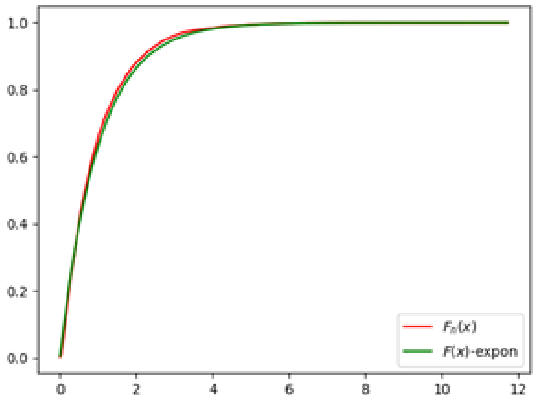

Figure 2: $\beta = 0.3, \quad \theta = 0.2, \quad n = 5000$

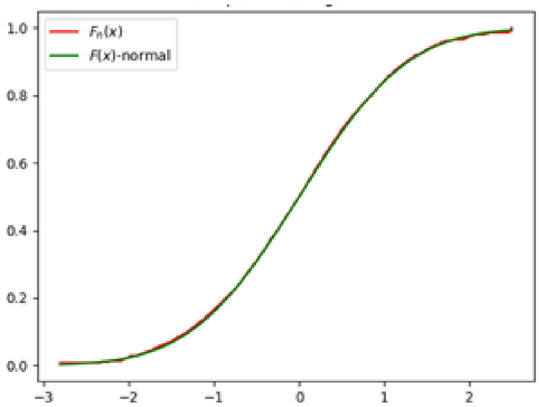

Figure 3: $\beta = 1, \quad \theta = 1, \quad n = 5000$

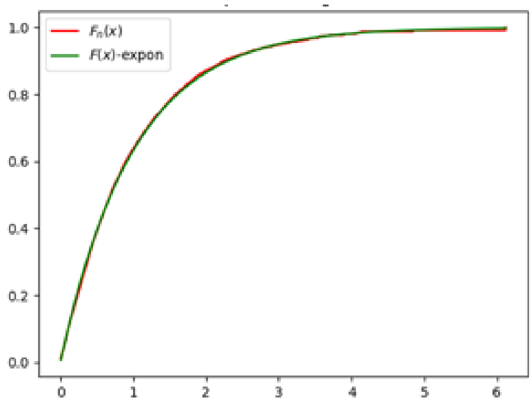

Figure 4: $\beta = 1, \quad \theta = 1, \quad n = 5000$

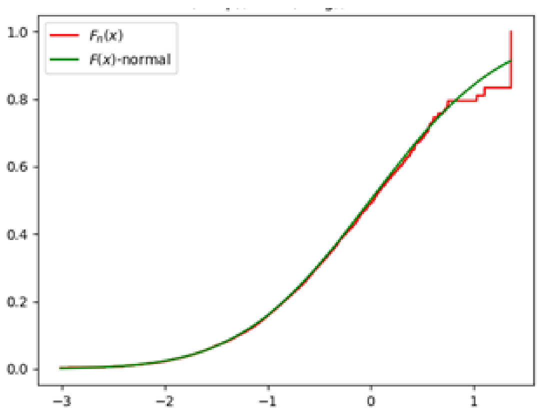

Figure 5: $\beta = 1, \quad \theta = 4, \quad n = 5000$

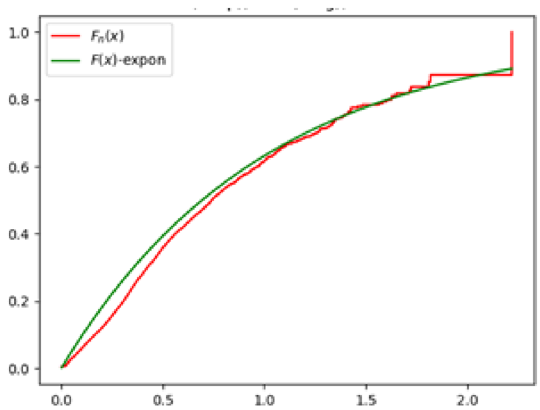

Figure 6: $\beta = 1, \quad \theta = 4, \quad n = 5000$

$(X_1, \ldots, X_n)$ in the absence of censoring, then by Donsker's theorem the random process $\{\sqrt{n} \cdot (F_n^{\ni}(x) - F(x)), \, x \in R^1\}$ weakly converges to a Brownian bridge $\mathrm{B}(F(x))$

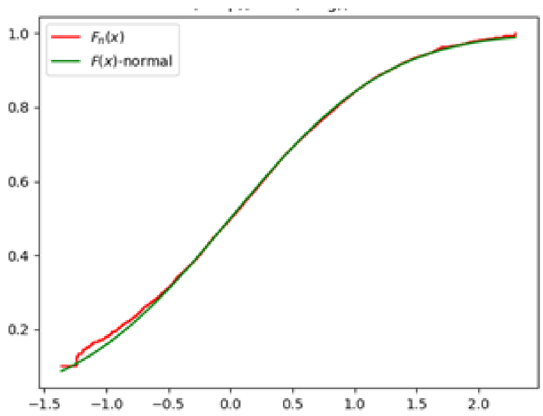

Figure 7: $\beta = 4, \quad \theta = 1, \quad n = 5000$

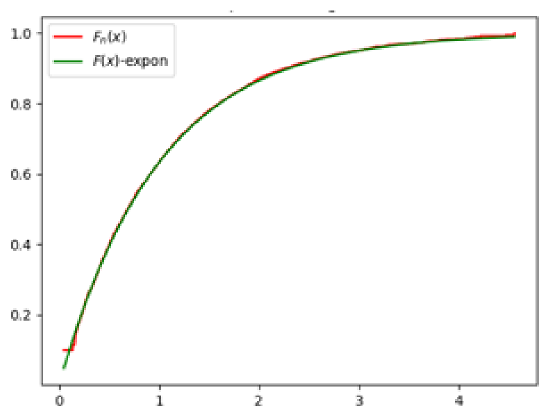

Figure 8: $\beta = 4, \quad \theta = 1, \quad n = 5000$

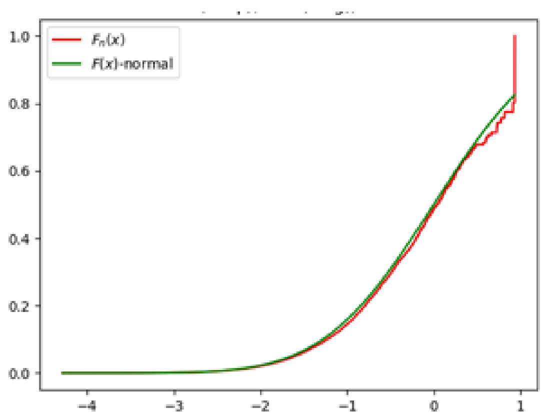

Figure 9: $\beta = 0.3, \quad \theta = 4, \quad n = 5000$

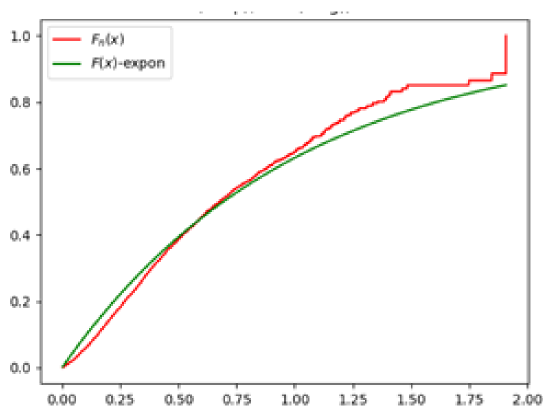

Figure 10: $\beta = 0.3, \quad \theta = 4, \quad n = 5000$

with zero mean and covariance $F\left(\min\left(x_1,\, x_2\right)\right) - F\left(x_1\right) F\left(x_2\right)$ and

$$\lim_{n\to\infty} P\left\{\sqrt{n}\cdot\sup_x \left|F_n^{\ni}\left(x\right) - F\left(x\right)\right| \le t\right\} = K\left(t\right),$$

where $K\left(t\right)$-is the Kolmogorov distribution. These results are very useful in constructing a confidence interval for the d.f. $F\left(x\right)$ and also when testing the hypothesis $H_0:\ \ F = F_0$,where F_0 is the given d.f. Now let us consider the estimation (6) and the empirical process (7) for $s \equiv 1$. It follows from theorem 2 that the limiting Gaussian process $\left\{A\left(x\right),\, x \in R^1\right\}$ with its complex covariance structure is far from a Brownian bridge. However, it is easy to verify that in the absence of random censoring on both sides (i.e., at $\beta = \theta \equiv 0$), $\lambda = \gamma \equiv 1$, $p^{(0)} = p^{(2)} \equiv 0$, $p^{(1)} \equiv 1$, $H\left(x\right) \equiv F\left(x\right)$ and this covariance coincides with the Brownian bridge covariance. Therefore, it is interesting to compare the histograms of statistical $\sqrt{n}\cdot\sup\limits_x\left|F_n\left(x\right) - F\left(x\right)\right|$ data with the distribution density of Kolmogorov statistics for different levels of censorship.

Analyzing the above histograms, we can see that for the values of the corresponding parameters $\beta,\, \theta\, \in\, [0,\, 1]$, the marginal distribution practically does not differ

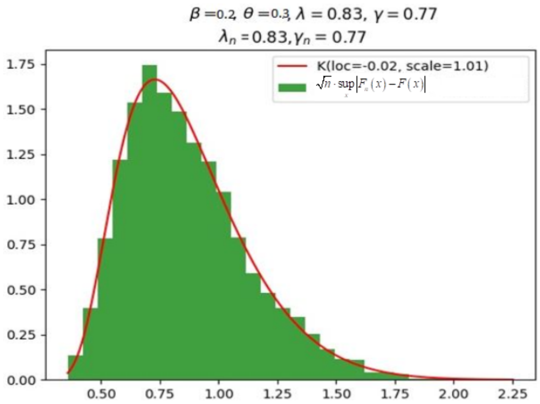

Figure 11: $\beta = 0.2, \quad \theta = 0.3, \quad n = 3500$

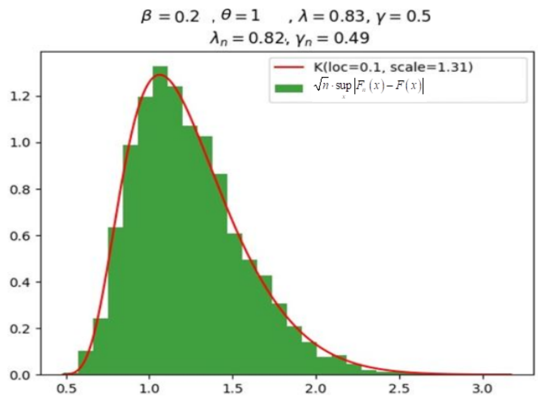

Figure 12: $\beta = 0.2, \quad \theta = 1, \quad n = 3500$

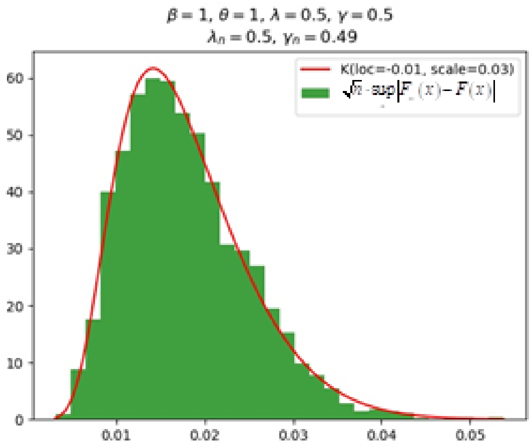

Figure 13: $\beta = 1, \quad \theta = 1, \quad n = 3500$

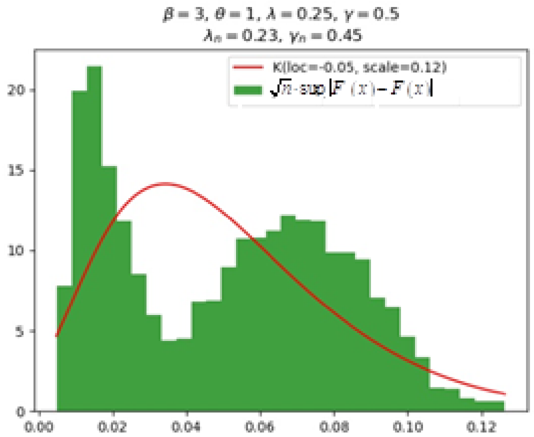

Figure 14: $\beta = 3, \quad \theta = 1, \quad n = 3500$

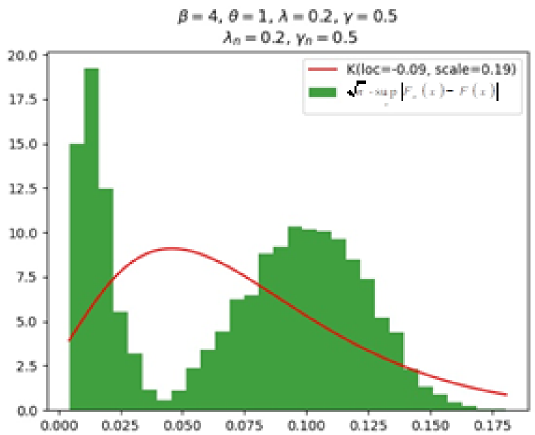

Figure 15: $\beta = 4, \quad \theta = 1, \quad n = 3500$

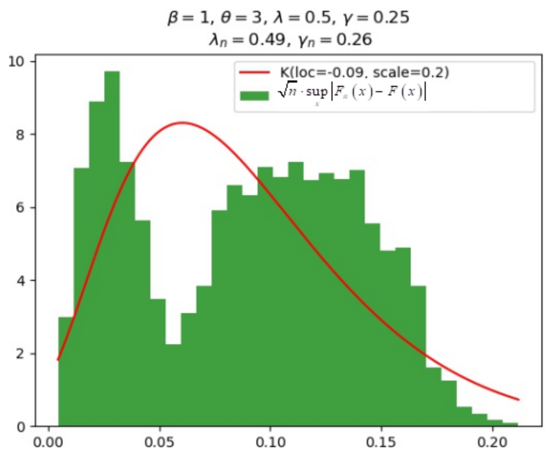

Figure 16: $\beta = 1, \quad \theta = 3, \quad n = 3500$

from the Kolmogorov distribution (see Fig. 11-13). In other cases, when the depth of censoring increases, the discrepancy between distributions is significant

References

[1] Araujo A., Gine E. The central limit theorem for real and Banach valued random variables. Willey. New York (1980).

[2] Abdushukurov A.A. Estimation of probability density and intensity function of the Koziol-Green model of random cencoring. Sankhya: Indian J. Statistics, Ser.A. v. 48. p. 150-168 (1987)

[3] Abdushukurov A.A. Random cencorship model from both sides and independence test for it. Report of Acad. Sci. Rep. Uz. Issue 11. p. 8-9 (1994) (In Russian).

[4] Abdushukurov A.A. Nonparametric estimation of the distribution function based on relative risk function. Commun. Statist: Th. & Meth. v. 27. N. 8. p. 1991-2012 (1998).

[5] Abdushukurov A.A., Abdikalikov F.A. Semiparametric estimator of mean conditional residual life function under informative random cencoring from both sides. Applied Mathematics. v. 6. p. 319-325 (2015).

[6] Abdikalikov F.A., Abdushukurov A.A.Semiparametric estimation of conditional survival function in informative regression model of random censorship from both sides. Statisticheskie Metody Otsenivaniya i Proverki Gipotes. Perm. Russia. Perm State Univ. Press. Issue 23. p. 145-162 (2012) (In Russian).

[7] Billingsley P. Convergence of probability measures. Willey. New York (1968).

[8] Chen P.E., Lin G. D. Maximum likehood estimation of survival function under the Koziol-Green proportional hazards model. Statist. Probab. Letters. v. 5. p. 75-80 (1987).

[9] Csörgő S., Horväth L. On the Koziol-Green model of random censorship. Biometrika. v. 68. p. 391-401 (1981).

[10] Csörgő S. Estimating in proportional hazards model of random censorship. Statistics. v. 19. p. 437-463 (1988).

[11] Csörgő S. Testing for the prorortional hazard model of random censorship. Proc. fourth Prague Symp. Asymp. Statist. Carles Univ. Press. Prague. p. 41-53 (1989)

[12] Csörgő S., Mielniczuk J. Density estimation in the simple proportional hazards model. Statist. Probab. Letters. v. 6. p. 419-426 (1988)

[13] Csörgő S., Faraway J.J. The paradoxical nature of the proportional hazards model of random censorship. Statistics. v. 31. p. 67-78 (1998).

[14] Dvoretzky A., Kiefer J., Wolfowitz J. Asymptotic minimax character of the sample distribution function and of the multinomil estimator. Ann. Math. Statist. v. 27. p. 642-669 (1956).

[15] Ghorai J. The asymptotic distribution of the suprema of the standardized empirical processes under the Koziol-Green model. Statist. Probab. Letters. v. 41. p. 303-313 (1999).

[16] Koziol J.A., Green S.B. A Cramer-von Mises statistic for randomly censored data. Biometrika. v. 63. N. 3. p. 465-476 (1976).

[17] Hollander M., Pena E. Families of confidence bands for the survival function under the general right censorship model and the Koziol-Green model. Canadian J. Statist. v. 17. N. 1. p. 59-74 (1989).

[18] Mansurov D.R. Sequential empirical processes in informative models of incomplete observations. Materials of international conf. "Teoriya funcsiy odnogo i mnogich compleksnych peremennich" November 26-28. Nukus. p. 165-168 (2020)

[19] Massart P. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. Ann. Probab. v. 18. N. 3. p. 1269-1283 (1990)

[20] Pawlitschko J. A Comparison of survival function estimators in the Koziol-Green model. Statistics. v. 32. p. 277-291 (1999).

[21] de Una-Álvares J. Kernel distribution function estimation under the Koziol-Green model. J. Stat. Plan. Infer. v. 87. p. 199-219 (2000).