ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
2022 Управление, вычислительная техника и информатика № 61
Tomsk State University Journal of Control and Computer Science

ОБРАБОТКА ИНФОРМАЦИИ

DATA PROCESSING

Original article
doi: 10.17223/19988605/61/3

Asymptotic properties of modified empirical Kac processes under general random censorship model

Abdurakhim A. Abdushukurov[1], Gulnoz S. Sayfulloyeva[2]

[1] Moscow State University named after M.V.Lomonosov, Tashkent Branch, Tashkent, Uzbekistan, a_abdushukurov@rambler.ru
[2] Navoi State Pedagogical Institute, Navoi, Uzbekistan, sayfullayevagulnoz@gmail.com

Abstract. In this paper, we consider a general random censorship model and prove asymptotic properties of modified empirical Kac processes. This model includes such important special cases as random censorship on the right and competing risks model. Our results use strong approximation method. Cumulative hazard processes also investigated in a similar manner in the general setting.

Keywords: censored data; empirical estimates; Kac estimate; strong approximation; Gaussian process

For citation: Abdushukurov, A.A., Sayfulloyeva, G.S. (2022) Asymptotic properties of modified empirical Kac processes under general random censorship model. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitelnaja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 26–36. doi: 10.17223/19988605/61/3

Научная статья
УДК 519.2
doi: 10.17223/19988605/61/3

Асимптотические свойства модифицированных эмпирических процессов Каца в общей модели случайного цензурирования

Абдурахим Ахмедович Абдушукуров[1], Гулноз Сайфуллоевна Сайфуллоева[2]

[1] Ташкентский филиал Московского государственного университета им. М.В. Ломоносова, Ташкент, Узбекистан, a_abdushukurov@rambler.ru
[2] Навоийский государственный педагогический институт, Навои, Узбекистан, sayfullayevagulnoz@gmail.com

Аннотация. Рассматривается общая модель случайного цензурирования, включающая в себя модели случайного цензурирования справа и конкурирующих рисков. В ней определены эмпирические процессы Каца и определены их модификации. Также исследованы кумулятивные процессы риска с доказательством свойств сильной аппроксимации последовательностями гауссовских процессов.

Ключевые слова: цензурированные данные; эмпирические оценки; оценка Каца; сильная аппроксимация; гауссовские процессы

***Для цитирования*:** Абдушукуров А.А., Сайфуллоева Г.С. Асимптотические свойства модифицированных эмпирических процессов Каца в общей модели случайного цензурирования // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 26–36. doi: 10.17223/19988605/61/3

The empirical distribution function has been widely used as an estimator for the distribution function of the elements of a random sample. It is not, however, appropriate when the observations are incomplete. Developing the corresponding theory of convergence of considered empirical and concentrated processes to a Gaussian process has been obtained by many scientists. A generalization of these results for the case of competing risks or when present various types of censorship considered by authors [see, for example, [1−3]]. These results have numerous statistical application in areas such as medical follow-up studies, life testing, actuarial sciences and demography (see, also, [4–6]). A general scheme of random censorship was considered by authors includes an competing risks model and random censoring from both sides.

1. Mathematical model

Let Z be a real random variable (r.v.) with distribution function (d.f.) $H(x) = P(Z \le x),\ \ x \in \mathbb{R}$. For a fixed integer $k \ge 1$ let $A^{(1)},...,A^{(k)}$ be pairwise disjoint random events, and define the subdistribution functions $H(x;i) = P(Z \le x, A^{(i)}),\ i \in \Im = \{1,...,k\}$. Suppose that when observing Z we are interested in the joint behaviour of the pairs $(Z, A^{(i)}),\ i \in \Im$. Let $\{(Z_j, A_j^{(1)},...,A_j^{(k)}),\ \ j \ge 1\}$ be a sequence of independent replicas of the $(Z, A^{(1)},...,A^{(k)})$ defined on some probability space $\{\Omega, A, P\}$. We assume throughout that the functions $H(x), H(x;1),...,H(x;k)$ are continuous. Let $H_n(x)$ denote the ordinary empirical d.f. of $Z_1,...,Z_n$ and introduce the empirical sub d.f. $H_n(x;i)$, $i \in \Im$

$$H_n(x;i) = \frac{1}{n}\sum_{j=1}^{n} \delta_j^{(i)} I(Z_j \le x),\ (x;i) \in \overline{\mathbb{R}} \times \Im,$$

where $\overline{R} = [-\infty;\infty]$, $\delta_j^{(i)} = I(A_j^{(i)})$ is an indicator of event $A_j^{(i)}$ and

$$H_n(x;1) + ... + H_n(x;k) = \frac{1}{n}\sum_{j=1}^{n} I(Z_j \le x) = H_n(x),\ x \in \overline{\mathbb{R}},$$

is the ordinary empirical d.f. Properties of many biometrical estimates depends on limit behaviours of proposed empirical statistics.

The following results are a straightforward consequences of exponential inequality of Dvoretzky-Kiefer-Wolfowits with exactly constant $D = 2$ from [7, 8]:

For all $n = 1,2,...$ and $\varepsilon > 0$:

$$P\left(\sup_{|x|<\infty} \left|H_n(x) - H(x)\right| > \left(\frac{(1+\varepsilon)}{2} \cdot \frac{\log n}{n}\right)^{1/2}\right) \le 2n^{-(1+\varepsilon)}, \tag{1}$$

and

$$P\left(\sup_{|x|<\infty} \left|H_n(x;i) - H(x;i)\right| > 2\left(\frac{(1+\varepsilon)}{2} \frac{\log n}{n}\right)^{1/2}\right) \le 4n^{-(1+\varepsilon)}. \tag{2}$$

A crucial role is played the vector-valued empirical process

$$\left\{a_n(t) = \left(a_n^{(0)}(t_0), a_n^{(1)}(t_1),...,a_n^{(k)}(t_k)\right),\ t = (t_0,...,t_k) \in \overline{\mathbb{R}}^{k+1}\right\};$$

where

$$a_n^{(0)}(x) = \sqrt{n}\left(H_n(x) - H(x)\right),\ a_n^{(i)}(x) = \sqrt{n}\left(H_n(x;i) - H(x;i)\right),\ i \in \Im.$$

The results of our approximation theorems presented here is, quite naturally, the approximation theorems of Komlós–Major–Tusnády's, for the ordinary empirical process with the approximation with the rate of order $n^{-1/2}\log n$. We will construct the approximation Gaussian processes in terms of Wiener sequences. The following theorem of Burke-Csörgő-Horváth [9, 10] is an extended analogue of Komlós–Major–Tusnády's result [11, 12].

Theorem A [9, 10]. If the underlying probability space $\{\Omega, A, P\}$ is rich enough, then one can define $k+1$ sequences of Gaussian processes $B_n^{(0)}(x), B_n^{(1)}(x), ..., B_n^{(k)}(x)$ such that for $a_n(t)$ and $B_n(t) = \left(B_n^{(0)}(x_0), B_n^{(1)}(x_1), ..., B_n^{(k)}(x_k)\right)$, $t = (t_0, ..., t_k)$, we have

$$P\left\{\sup_{t\in\overline{\mathbb{R}}^{k+1}} \left\|a_n(t) - B_n(t)\right\|^{(k+1)} > n^{-1/2}\left(M(\log n) + z\right)\right\} \le K\exp(-\lambda z), \tag{3}$$

for all real *z*, where $M = (2k+1)A_1$, $K = (2k+1)A_2$ and $\lambda = A_3/(2k+1)$ with A_1, A_2 and A_3 are absolute constants. Moreover, B_n itself is a $(k+1)$ dimensional vector-valued Gaussian process having the same covariance structure as the vector $a_n(t)$, namely $EB_n^{(i)}(x) = 0$, $(x,i) \in \overline{\mathbb{R}} \times \overline{\mathfrak{I}} = \mathfrak{I} \cup \{0\}$ and for any $i, j \in \mathfrak{I}$, $i \ne j$, $x, y \in \overline{\mathbb{R}}$:

$$\begin{gathered} EB_n^{(0)}(x)B_n^{(0)}(y) = \min\{H(x), H(y)\} - H(x)\cdot H(y), \\ EB_n^{(i)}(x)B_n^{(i)}(y) = \min\{H(x;i), H(y;i)\} - H(x;i)\cdot H(y;i), \\ EB_n^{(i)}(x)B_n^{(j)}(y) = -H(x;i)\cdot H(y;j), \\ EB_n^{(0)}(x)B_n^{(i)}(y) = \min\{H(x;i), H(y;j)\} - H(x)\cdot H(y;i). \end{gathered} \tag{4}$$

Note that in proving of theorem A (theorem 3.1 in [10]) authors constructed sequence of two – parametrical Gaussian processes $\mathbb{Q}^{(0)}(x;n), \mathbb{Q}^{(2)}(x;n), ..., \mathbb{Q}^{(k)}(x;n)$ such that for $a_n(t)$ and $\mathbb{Q}(t;n) = \left(\mathbb{Q}^{(0)}(x;n), ..., \mathbb{Q}^{(k)}(x;n)\right)$, $t \in \overline{\mathbb{R}}^{k+1}$ the following its Borel-Cantelly consequence of approximation have used

$$\sup_{t\in\overline{\mathbb{R}}^{k+1}} \left\|a_n(t) - n^{-1/2}\mathbb{Q}(t;n)\right\|^{(k+1)} \overset{a.s.}{=} O\left(n^{-1/2}\log^2 n\right),$$

where $\mathbb{Q}(t;n)$ is the $(k+1)$ dimensional vector-valued Gaussian process that $\mathbb{Q}(t;n) \overset{D}{=} n^{1/2}a_n(t)$. Hence

$$E\mathbb{Q}^{(i)}(x;n) = 0,\ (x,i) \in \overline{\mathbb{R}} \times \overline{\mathfrak{I}}$$

and for any $i, j \in \mathfrak{I}$, $i \ne j$, $x, y \in \overline{\mathbb{R}}$:

$$\begin{gathered} E\mathbb{Q}^{(0)}(x;n)\mathbb{Q}^{(0)}(y;m) = \min(n,m)\left\{\min\{H(x), H(y)\} - H(x)H(y)\right\}, \\ E\mathbb{Q}^{(0)}(x;n)\mathbb{Q}^{(0)}(y;m) = \min(n,m)\left\{\min\{H(x;i), H(y;j)\} - H(x)H(y;i)\right\}, \\ E\mathbb{Q}^{(i)}(x;n)\mathbb{Q}^{(i)}(y;m) = \min(n,m)\left\{\min\{H(x;i), H(y;i)\} - H(x;i)H(y;j)\right\}, \\ E\mathbb{Q}^{(i)}(x;n)\mathbb{Q}^{(j)}(y;m) = -\min(n,m)H(x;i)\cdot H(y;j). \end{gathered} \tag{5}$$

Observe that $\left\{\mathbb{Q}^{(i)}, i \in \overline{\mathfrak{I}}\right\}$ are Kiefer processes and they satisfying the distributional equality

$$\mathbb{Q}^{(i)}(x;n) \overset{D}{=} W^{(i)}\left(H(x;i);n\right) - H(x;i)W^{(i)}(1;n), \tag{6}$$

where $\left\{W^{(i)}(y;n),\ 0 \le y \le 1, n \ge 1, i \in \overline{\mathfrak{I}}\right\}$ itself are two-parametrical Wiener processes with $EW^{(i)}(y;n) = 0$ and

$$EW^{(i)}(y;n)W^{(i)}(u;m)=\min(n,m)\min(y,u), \quad i\in\Im.$$

It is important to note that though Kiefer processes $\{\mathbb{Q}^{(i)}, i\in\Im\}$ are dependent processes, but corresponding Wiener processes are independent. Indeed, from proof of theorem A are follows that

$$\mathbb{Q}^{(1)}(x;n)\overset{D}{=}K(H(x;1);n),$$

$$\mathbb{Q}^{(2)}(x;n)\overset{D}{=}K(H(x;2)-H(+\infty;1);n)-K(H(+\infty;1);n),$$

$$\dots$$

$$\mathbb{Q}^{(i)}(x;n)\overset{D}{=}K(H(x;i)+H(+\infty;1)+\dots+H(+\infty;i-1);n)-K(H(+\infty;1)+\dots+H(+\infty;i-1);n), \quad i\in\Im,$$

where $H(+\infty;i)=\lim_{x\uparrow+\infty}H(x;i), \quad H(+\infty;1)+\dots+H(+\infty;k)=1.$

The Kiefer processes $\{K(y;n),\ 0\le y\le 1,\ n\ge 1\}$ are represented through two-parametrical Wiener processes $\{W(y;n),\ 0\le y\le 1,\ n\ge 1\}$ by distributional equality

$$\{K(y;n),\ 0\le y\le 1,\ n\ge 1\}\overset{D}{=}\{W(y;n)-yW(1;n),\ 0\le y\le 1,\ n\ge 1\}. \tag{7}$$

Consequently, in view of (6) and (7) the Wiener process $\{W^{(i)}, i\in\Im\}$ also admits representations for all $(x;i)\in\overline{\mathbb{R}}\times\Im$:

$$W^{(1)}(H(x;1);n)\overset{D}{=}W(H(x;1);n),$$

$$W^{(2)}(H(x;2);n)\overset{D}{=}W(H(x;2)+H(+\infty;1);n)-W(H(+\infty;1);n),$$

$$\dots$$

$$W^{(i)}(H(x;i);n)\overset{D}{=}W(H(x;i)+H(+\infty;i-1);n)-W(H(+\infty;1)+\dots+H(+\infty;i-1);n).$$

Now by directly calculations of covariance of processes $\{W^{(i)}, i\in\Im\}$ it is easy to believing on its independency. This paper further structured as follows. In section 1 we introduce the classical Kac processes analogues and their modifications. For its we prove approximation results. Then in section 2 we propose corresponding estimators of hazard functions. For them we also prove approximation results.

2. Kac processes under general censoring

Authors [9] proved the general theorems to obtain approximation for the usual empirical and corresponding cumulative hazard estimates by Gaussian processes for the competing risk generalizations. We prove these results for a corresponding Kac-type processes.

Following of [12] we introduce the modified empirical d.f. of Kac by the following way. Along with sequence $\{Z_j, j\ge 1\}$ on a probability space $\{\Omega, A, P\}$ consider also a sequence $\{\nu_n, n\ge 1\}$ of r.v.-s having Poisson distribution with parameter $E\nu_n=n, \quad n=1,2,\dots$. Assume throughout that the two sequences $\{Z_j, j\ge 1\}$ and $\{\nu_n, n\ge 1\}$ are independent. Kac's empirical d.f. is

$$H_n^*(x)=\begin{cases}\dfrac{1}{n}\sum_{j=1}^{\nu_n}I(Z_j\le x), & if \quad \nu_n\ge 1 \ a.s.,\\ 0, \quad if \quad \nu_n=0 \ a.s.,\end{cases}$$

while the empirical sub-d.f. one is

$$H_n^*(x;i)=\begin{cases}\frac{1}{n}\sum_{j=1}^{\nu_n} I\left(Z_j\le x, A_j^{(i)}\right), i\in\Im, & if \quad \nu_n\ge 1 \ a.s.,\\ 0,\ i\in\Im \quad if \quad \nu_n=0\ a.s.,\end{cases}$$

with $H_n^*(x;1)+...+H_n^*(x;k)=H_n^*(x)$ for all $x\in\overline{R}$. Here we suppose that sequence $\{\nu_n, n\ge 1\}$ is independent of random vectors $\left\{\left(Z_j,\delta_j^{(1)},...,\delta_j^{(k)}\right), j\ge 1\right\}$, where $\delta_j^{(i)}=I\left(A_j^{(i)}\right)$. Note that statistics $H_n^*(x;i)$ (consequently also $H_n^*(x)$) are unbiased estimators of $H(x;i)$, $i\in\Im$ (consequently also of $H(x)$):

$$E\left(H_n^*(x;i)\right)=\frac{1}{n}E\left\{\sum_{m=1}^{\infty}E\left[\sum_{k=1}^{n}\delta_k^{(i)}\cdot I(Z_k\le x)\right],\nu_n=m\right\}=$$

$$=\frac{1}{n}E\left\{\sum_{m=1}^{\infty}E\left[\sum_{k=1}^{n}\delta_k^{(i)}\cdot I(Z_k\le x)/\nu_n=m\right]\cdot P(\nu_n=m)\right\}=$$

$$=\frac{1}{n}\sum_{m=1}^{\infty}H(x;i)mP(\nu_n=m)=\frac{1}{n}H(x;i)\sum_{m=1}^{\infty}m\cdot\frac{n^m e^{-n}}{m!}=$$

$$=H(x;i)e^{-n}\sum_{m=0}^{\infty}\frac{n^m}{m!}=H(x;i),\quad (x;i)\in\overline{\mathbb{R}}\times\Im.$$

Consequently,

$$E\left[H_n^*(x)\right]=\sum_{i=1}^{k}E\left[H_n^*(x;i)\right]=\sum_{i=1}^{k}H(x;i)=H(x),\ x\in\overline{\mathbb{R}}.$$

Let's define $a_n^{(i)*}(x)=\sqrt{n}\left(H_n^*(x;i)-H(x;i)\right)$, $i\in\Im$ and $a_n^{(0)*}(x)=\sqrt{n}\left(H_n^*(x)-H(x)\right)$ the empirical Kac processes.

Theorem 1. If the underlying probability space $\{\Omega, A, P\}$ is rich enough, then one can define $k+1$ sequences of Gaussian processes $W_n^{(0)}(x), W_n^{(1)}(x),...,W_n^{(k)}(x)$ such that for $a_n^*(t)=\left(a_n^{(0)*}(t_0),a_n^{(1)*}(t_1),...,a_n^{(k)*}(t_k)\right)$ and $W_n^*(t)=\left(W_n^{(0)}(t_0),W_n^{(1)}(t_1),...,W_n^{(k)}(t_k)\right)$, $t=(t_0,t_1,...,t_k)$, we have

$$P\left\{\sup_{t\in\overline{\mathbb{R}}^{k+1}}\left\|a_n^*(t)-W_n^*(t)\right\|^{(k+1)}>C^*n^{-1/2}\log n\right\}\le K^*n^{-r}, \tag{8}$$

where $r\ge 2$ is an arbitrary integer, $C^*=C^*(r)$-depends only on r and K^* is an absolute constant. Moreover, $W_n^*(t)$ itself is a $(k+1)$-dimensional vector-valued Gaussian process with expectation $EW_n^{(i)}(x)=0$, $(x,i)\in\overline{\mathbb{R}}\times\Im$ and for any $i,j\in\Im$, $i\ne j$, $x,y\in\overline{\mathbb{R}}$:

$$\begin{aligned}EW_n^{(0)}(x)W_n^{(0)}(y)&=\min\{H(x),H(y)\},\\ EW_n^{(i)}(x)W_n^{(j)}(y)&=\min\{H(x;i),H(y;j)\},\\ EW_n^{(i)}(x)W_n^{(0)}(y)&=\min\{H(x;i),H(y)\}.\end{aligned} \tag{9}$$

The basic relation between $a_n(t)$ and $a_n^*(t)$ is the following easily checked identity

$$a_n^*(x)=\sqrt{\frac{\nu_n}{n}}a_{\nu_n}^{(i)}(x)+H(x;i)\frac{(\nu_n-n)}{\sqrt{n}},\quad i\in\Im \tag{10}$$

Hence the approximating sequence have respectively the form

$$W_n^{(i)}(x)=B_{\nu_n}^{(i)}(x)+H(x;i)\frac{W^*(n)}{\sqrt{n}},\quad i\in\Im,$$

where $B_{\nu_n}^{(i)}(x)$ is a Poisson indexed Brownian bridge type process of Teorem A and $\{W^{(*)}(x),\ x\ge 0\}$ is a Wiener process. Easy to verify that $\{W_n^{(i)}(x),\ (x;i)\in\mathbb{R}\times\overline{\Im}\}\overset{D}{=}\{W^*(H(x;i)),(x,i)\in\overline{\mathbb{R}}\times\overline{\Im}\}$. The proof of Teorem 1 is coincides with the proof of theorem 1 of Stute in [13] hence it is omitted.

In so far as $\lim\limits_{x\uparrow+\infty} H_n^*(x)=H_n^*(+\infty)=\dfrac{\nu_n}{n}$, then by Stirlings formula

$$P(\nu_n=n)=P\left(H_n^*(+\infty)=1\right)=\frac{n^n e^{-n}}{n!}=\frac{1}{\sqrt{2\pi n}}\left(1+o(1)\right),\quad n\to\infty,$$

and

$$P\left(H_n^*(+\infty)>1\right)=P(\nu_n>n)=\sum_{k=n+1}^{\infty}\frac{n^k e^{-n}}{k!}=o(1),\quad n\to\infty.$$

Thus $H_n^*(x)$ with positive probability t_0 be greater than 1. In order to avoid these undesirable properties, we propose following modifications of Kac statistics:

$$\begin{gathered}H_n(x)=1-\left(1-H_n^*(x)\right)I\left(H_n^*(x)<1\right),\quad x\in\overline{\mathbb{R}},\\ H_n(x;i)=1-\left(1-H_n^*(x;i)\right)I\left(H_n^*(x;i)<1\right),\quad (x;i)\in\overline{\mathbb{R}}\times\Im.\end{gathered}\tag{11}$$

The following inequalities are useful in investigating of Kac processes.

Theorem 2. Let $\{\nu_n, n\ge 1\}$ be a sequence of Poisson r.v.-s with $E\nu_n=n.$ Then for any $\varepsilon>0$ such that

$$\frac{n}{\log n}\ge\frac{\varepsilon}{8\left(1+e/3\right)^2},\tag{12}$$

we have

$$P\left(\left|\nu_n-n\right|>\frac{1}{2}\left(\frac{\varepsilon}{2}n\log n\right)^{1/2}\right)\le 2n^{-\varepsilon w},\tag{13}$$

$$P\left(\sup_{|x|<\infty}\left|H_n^*(x;i)-H(x;i)\right|>2\left(\frac{\varepsilon\log n}{2n}\right)^{1/2}\right)\le 4n^{-4\varepsilon w},\ \ i\in\Im,\tag{14}$$

$$P\left(\sup_{|x|<\infty}\left|H_n(x;i)-H(x;i)\right|>2\left(\frac{\varepsilon\log n}{2n}\right)^{1/2}\right)\le 4n^{-4\varepsilon w},\ \ i\in\Im,\tag{15}$$

where $w=\left[16\left(1+e/3\right)\right]^{-1}$.

Proof. Let $\gamma_1,\gamma_2,...$ be a sequence of Poisson r.v.-s with $E\gamma_k=1$ for all $k=1,2,...$. Then $S_n=\nu_n-\nu=\sum\limits_{k=1}^{n}(\gamma_k-1)=\sum\limits_{k=1}^{n}\xi_k$ and $E\exp(t\xi_k)=e^{-t}\exp(t\gamma_1)=\exp(-(t+1))\sum\limits_{k=0}^{\infty}\dfrac{\left(e^t\right)^k}{k!}=\exp\left\{e^t-(t+1)\right\}.$

Using Taylor expansion for e^t, we get

$$E\exp(t\xi_k)=\exp\left\{1+t+\frac{t^2}{2}+\psi(t)-(t+1)\right\}=\exp\left\{\frac{t^2}{2}+\psi(t)\right\},$$

where $\psi(t)=\dfrac{t^3}{6}\exp(\theta t)$, $0<\theta<1$. Estimate $\psi(t)$ taking into account that $t^3\le t^2$ under $0\le t\le 1$: $\psi(t)\le\dfrac{t^3}{6}e\le e\dfrac{t^2}{6}$. Thus, $E\exp(t\xi_k)=\exp\left\{\dfrac{t^2}{2}\left(1+\dfrac{e}{3}\right)\right\},\quad 0\le t\le 1.$

The following result (theorem from [14]) is necessary for our further investigations.

Lemma 1 [14]. Let $\{\xi_n, n\ge 1\}$ be a sequence of independent r.v.-s with $E\xi_n=0,\quad n=1,2,...$. Suppose that $U,\lambda_1,...,\lambda_n$ positive real numbers such that

$$E\exp\left(t\xi_k\right)\le\exp\left(\frac{1}{2}\lambda_k t_k^2\right)\ \text{for}\ k=1,2,...,n\quad |t|\le U\,. \tag{16}$$

Let $\Lambda=\lambda_1+...+\lambda_n$. Then

$$P\left(\left|\xi_1+...+\xi_k\right|\ge z\right)\le\begin{cases}2\exp\left(-\dfrac{z^2}{2\Lambda}\right), & if\quad o\le z\le\Lambda U,\\[2mm] 2\exp\left(-\dfrac{Uz}{2}\right), & if\quad z>\Lambda U.\end{cases}$$

Let in lemma 1 $\lambda_k=1+e/3,\ U=1,\ z=\frac{1}{2}\left(\frac{\varepsilon}{2}n\log n\right)^{1/2}$, then we obtain (13). Here $0\le z=\frac{1}{2}\left(\frac{\varepsilon}{2}n\log n\right)^{1/2}\le$ $\le\left(1+e/3\right)n=\Lambda U$. Consider probability in (14). By total probability formula

$$P\left(\sup_{|x|<\infty}\left|H_n^*(x;i)-H(x;i)\right|>2\left(\frac{\varepsilon\log n}{2n}\right)^{1/2}\right)=$$

$$=P\left(\sup_{|x|<\infty}\left|H_n(x;i)-H(x;i)+\frac{1}{n}\sum_{k=n+1}^{\nu_n}\delta_k^{(i)}I\left(Z_k\le x\right)\right|>2\left(\frac{\varepsilon\log n}{2n}\right)^{1/2}\Big/\nu_n>n\right)\cdot P\left(\nu_n>n\right)+$$

$$+P\left(\sup_{|x|<\infty}\left|H(x;i)-H(x;i)-\frac{1}{n}\sum_{k=\nu_n+1}^{n}\delta_k^{(i)}I\left(Z_k\le x\right)\right|>2\left(\frac{\varepsilon\log n}{2n}\right)^{1/2}\Big/\nu_n\le n\right)\cdot P\left(\nu_n\le n\right)\le$$

$$\le P\left(\sup_{|x|<\infty}\left|H_n(x;i)-H(x;i)\right|>\left(\frac{\varepsilon\log n}{2n}\right)^{1/2}\right)+P\left(\sup_{|x|<\infty}\left|\frac{1}{n}\sum_{k=\min(n,\nu_n)+1}^{\max(n,\nu_n)}\delta_k^{(i)}I\left(Z_k\le x\right)\right|>\left(\frac{\varepsilon\log n}{2n}\right)^{1/2}\right)\le$$

$$\le 2n^{-4\varepsilon}+P\left(\left|\frac{\nu_n-n}{n}\right|>\left(\frac{\varepsilon\log n}{2n}\right)^{1/2}\right)\le 2n^{-4\varepsilon}+2n^{-4w\varepsilon}\le 4n^{-4w\varepsilon},\quad i\in\Im,$$

where we applied (2) and (13) that proves (14). Let's define $T_n^{(i)}=\inf\left\{x:\ H_n(x;i)=1\right\},i\in\Im$. If $x\ge T_n^{(i)}$ and $\nu_n>n,$ then $H_n(x;i)=1$ and $H_n^*(x;i)-H(x;i)\ge H_n^*(x;i)-H_n(x;i)\ge 0$. Then assuming $\nu_n>n$, we obtain

$$\begin{aligned}\sup_{|x|<\infty}\left|H_n(x;i)-H(x;i)\right|&=\left\{\max\left[\sup_{x<T_n^{(i)}}\left|H_n^*(x;i)-H(x;i)\right|,\sup_{x\ge T_n^{(i)}}\left|H_n(x;i)-H(x;i)\right|\right]\right\}\le\\ &\le\left\{\max\left[\sup_{x<T_n^{(i)}}\left|H_n^*(x;i)-H(x;i)\right|,\sup_{x\ge T_n^{(i)}}\left|H_n^*(x;i)-H(x;i)\right|\right]\right\}=\sup_{|x|<\infty}\left|H_n^*(x;i)-H(x;i)\right|,\ i\in\Im.\end{aligned} \tag{17}$$

Under $\nu_n\le n,$ it is obvious that $H_n(x;i)=H_n^*(x;i),$ for all $(x;i)\in\overline{\mathbb{R}}\times\Im$.

Now taking into account last two relations, total probability formula and (14) we obtain (15). Theorem 2 is proved.

Let $a_n(t)=\left(a_n^{(0)}(t_0),a_n^{(1)}(t_1),...,a_n^{(k)}(t_k)\right),$ where $a_n^{(0)}(x)=\sqrt{n}\left(H_n(x)-H(x)\right),$ $a_n^{(i)}(x)=\sqrt{n}\left(H_n(x;i)-\right.$ $\left.-H(x;i)\right),$ $(x;i)\in\overline{\mathbb{R}}\times\Im$. We shall prove an approximation theorem of the vector-valued modified empirical Kac process $a_n(t)$ by the appropriate Gaussian vector-valued process $W_n^*(t),t\in\overline{\mathbb{R}}^{k+1}$ from theorem 2.

Theorem 3. Let $\{T_n,n\ge 1\}$ be a numerical sequence satisfying, for each *n*, the condition $T_n<T_H=\inf\{x:H(x)=1\}\le\infty$ such that

$$\min_{i\in\Im}\left\{P\left(A^{(i)}\right)-H\left(T_n,i\right)\right\}\ge 1-H\left(T_n\right)\ge 2\left(\frac{r\log n}{2wn}\right)^{1/2}. \tag{18}$$

If for any $\varepsilon > 0$ condition (12) hold, then on a probability space of theorem 2 one can define $k+1$ sequences of mean zero Gaussian processes $W_n^{(0)}(x), W_n^{(1)}(x), ..., W_n^{(k)}(x)$ with the covariance structure (9) such that for $a_n(t)$ and $W_n^*(t) = \left(W_n^{(0)}(t_0), W_n^{(1)}(t_1), ..., W_n^{(k)}(t_k)\right)$ we have

$$P\left\{ \sup_{t \in (-\infty; T_n]^{k+1}} \left\| a_n(t) - W_n^*(t) \right\|^{(k+1)} > Cn^{-1/2} \log n \right\} \le Kn^{-\beta}, \tag{19}$$

where K is an absolute constant, $C = C(\varepsilon)$ and $\beta = \min(r, \varepsilon w)$ for any $\varepsilon > 0$.

Proof. It is easy to seen that probability in (19) can be estimated by sum

$$P\left\{ \sup_{x \le T_n} \left| a_n^{(0)}(x) - W_n^{(0)}(x) \right| > Cn^{-1/2} \log n \right\} + \sum_{i=1}^{k} P\left(\sup_{x \le T_n} \left| a_n^{(i)}(x) - W_n^{(i)}(x) \right| > Cn^{-1/2} \log n \right) = q_{1n} + q_{2n}. \tag{20}$$

Taking into account that for any $x \le T_n$, $H_n^*(x) \le H_n^*(T_n)$ and if $H_n^*(T_n) \le 1$, then $a_n^{(0)}(x) = a_n^{(0)*}(x)$ and by formula of total probability

$$\begin{aligned} q_{1n} &\le P\left(\sup_{x \le T_n} \left| a_n^{(0)}(x) - W_n^{(0)}(x) \right| > Cn^{-1/2} \log n \Big/ H_n^*(T_n) \le 1 \right) + P\left(H_n^*(T_n) > 1\right) \le \\ &\le P\left(\sup_{x \le T_n} \left| a_n^{(0)*}(x) - W_n^{(0)}(x) \right| > Cn^{-1/2} \log n \right) + P\left(H_n^*(T_n) > 1\right) \le \\ &\le Kn^{-r} + P\left(H_n^*(T_n) - H(T_n) > 1 - H(T_n)\right) \le \\ &\le Kn^{-r} + P\left(\sup_{|x| < \infty} \left| H_n^*(x) - H(x) \right| > \left(\frac{r \log n}{2wn} \right)^{1/2} \right) \le Ln^{-r}, \end{aligned} \tag{21}$$

where we have used theorem 1 and analogue of (14) for $H_n^* - H$, $L = K^* + 4$. Analogously,

$$\begin{aligned} q_{2n} &\le \sum_{i=1}^{k} P\left(\sup_{x \le T_n} \left| a_n^{(i)}(x) - W_n^{(i)}(x) \right| > Cn^{-1/2} \log n \right) + \sum_{i=1}^{k} P\left(H_n^*(T_n; i) > P\left(A^{(i)}\right) \right) \le \\ &\le \sum_{i=1}^{k} P\left(\sup_{x \le T_n} \left| a_n^{(i)*}(x) - W_n^{(i)}(x) \right| > Cn^{-1/2} \log n \right) + \sum_{i=1}^{k} P\left(\sup_{|x| < \infty} \left| a_n^{(i)*}(x) - W_n^{(i)}(x) \right| > Cn^{-1/2} \log n \right) + \\ &+ kP\left(\frac{|\nu_n - \nu|}{n} > \frac{1}{2} \left(\frac{4r \log n}{2wn} \right)^{1/2} \right) \le kLn^{-r} + 2kn^{-4r}, \end{aligned} \tag{22}$$

where we also have used inequalities (13), (15) and theorem 1. Now from (21) and (22) follows (19). Theorem 3 is proved.

3. Estimation of hazard function

In many practical situations, when we are interested in the joint behaviors of the pairs $\left\{ \left(Z, A^{(i)}\right), \ i \in \Im \right\}$, a crucial role is played by the so-called cumulative hazard functions $\left\{ S^{(i)}(x) = \exp\left(-\Lambda^{(i)}(x)\right), \ i \in \Im \right\}$, where $\Lambda^{(i)}(x)$ is the i-th hazard function $\left(\int_{-\infty}^{x} = \int_{(-\infty; x]} \right)$:

$$\Lambda^{(i)}(x) = \int_{-\infty}^{x} \frac{dH(u; i)}{1 - H(u)}, \quad i \in \Im,$$

with $\Lambda^{(1)}(x) + ... + \Lambda^{(k)}(x) = \Lambda(x) = \int_{-\infty}^{x} \frac{dH(u)}{1 - H(u)}$ is the corresponding hazard function of d.f. $H(x)$.

Consider two important special cases of considered generalized censorship model:

a) Let $\{X_1, X_2, ...\}$ be a sequence of independent r.v.-s with common continuous d.f. *F*. These are censored on the right by $\{Y_1, Y_2, ...\}$ a sequence of independent r.v.-s, independent of the *X* – sequence, with common continuous d.f. *G*. One can only observe the sequence of pairs $\{(Z_k, \delta_k), k = \overline{1,n}\}$, where $Z_j = \min(X_j, Y_j)$ and $\delta_j = \delta_j^{(1)}$ is the indicator of event $A_j = A_j^{(1)} = \{Z_j = X_j\}$. In this case $k = 2$, $1 - H(x) = (1 - F(x))(1 - G(x))$, $H(x;1) = \int_{-\infty}^{x} (1 - G(u))dF(u)$, thus $S^{(1)}(x) = S(x) = 1 - F(x)$. The useful special case when $1 - G(x) = (1 - F(x))^{\beta}$, $\beta > 0$, which corresponds to independence of r.v.-s Z_j and δ_j, $j \geq 1$.

b) For $k > 1$ consider independent sequences $\{Y_1^{(i)}, Y_2^{(i)}, ...\}$ $(i = 1, ..., k)$ of independent r.v.-s with common continuous d.f. *F* and let $Z_j = \min(Y_j^{(1)}, ..., Y_j^{(k)})$. One observes the sequences $\{(Z_j, \delta_j^{(i)}), i = \overline{1,k}\}_{j=1}^{n}$, where $\delta_j^{(i)}$ is the indicator of the event $A_j^{(i)} = \{Z_j = Y_j^{(i)}\}$. This is the competing risks model with $S^{(i)}(x) = 1 - F^{(i)}(x)$, $i \in \Im$.

Define the natural Kac-type estimator

$$\Lambda_n^{(i)}(x) = \int_{-\infty}^{x} \frac{dH(u;i)}{1 - H_n(u)}, \quad i \in \Im,$$

of $\Lambda^{(i)}(x)$, $i \in \Im$. Let $w_n^{(i)}(x) = \sqrt{n}\left(\Lambda_n^{(i)}(x) - \Lambda^{(i)}(x)\right)$, $i \in \Im$ is an Kac-type hazard process and $w_n(t) = \left(w_n^{(1)}(t_1), ..., w_n^{(k)}(t_k)\right)$, $t = (t_1, ..., t_k)$, $Y_n(t) = \left(Y_n^{(1)}(t_1), ..., Y_n^{(k)}(t_k)\right)$ corresponding vector process with

$$Y_n^{(i)}(x) = \int_{-\infty}^{x} \frac{W_n^{(0)}(u)dH(u;i)}{(1 - H(u))^2} + \frac{W_n^{(i)}(x)}{1 - H(x)} - \int_{-\infty}^{x} \frac{W_n^{(i)}(u)dH(u)}{(1 - H(u))^2}, \quad i \in \Im$$

and $\{W_n^{(0)}(x), W_n^{(1)}(x), ..., W_n^{(k)}(x)\}$ are Wiener processes with the covariance structure (9). Then for $i \in \Im$, $EY_n^{(i)}(x) = 0$ and

$$EY_n^{(i)}(x)Y_n^{(i)}(y) = C(x, y),$$

where $x, y \leq T_H = \inf\{x : H(x) = 1\} \leq \infty$.

Theorem 4. Let $\{T_n, n \geq 1\}$ be a numerical sequence satisfying for each *n*, the condition $T_n < T_H$ such that

$$\frac{n}{\log n} \geq \max\left\{32\varepsilon w^2, \frac{2rb_n^2}{w}, \frac{2\varepsilon b_n^2}{w}\right\}, \tag{23}$$

where $b_n = (1 - H(T_n))^{-1}$, $\varepsilon > 0$, $r \geq 2$. Then on a probability space of theorem 2

$$P\left(\sup_{t \in (-\infty; T_n]^{(k)}} \left\| w_n(t) - Y_n(t) \right\|^{(k)} > r(n)\right) \leq k\Phi_1 n^{-\beta}, \tag{24}$$

where $r(n) = \Phi_0 b_n^2 n^{-1/2} \log n$, $\Phi_0 = \Phi_0(\varepsilon, r)$, Φ_1 – (absolute) constants.

Proof. It is enough to prove that for each $i \in \Im$

$$P\left(\sup_{x \leq T_n} \left(w_n^{(i)}(x) - Y_n^{(i)}(x)\right) > r(n)\right) \leq \Phi_1 n^{-\beta}. \tag{25}$$

For difference we have representation for each $i \in \Im$:

$$w_n^{(i)}(x) - Y_n^{(i)}(x) = \int\limits_{-\infty}^{x} \frac{\left(a_n^{(0)}(u) - W_n^{(0)}(u)\right) dH(u;i)}{\left(1-H(u)\right)^2} + \frac{a_n^{(i)}(x) - W_n^{(i)}(x)}{1-H(x)} -$$

$$-\int\limits_{-\infty}^{x} \frac{\left(a_n^{(i)}(u) - W_n^{(i)}(u)\right) dH(u)}{\left(1-H(u)\right)^2} + n^{-1/2} \int\limits_{-\infty}^{x} \frac{\left(a_n^{(0)}(u)\right)^2 dH(u;i)}{\left(1-H(u)\right)^2 \left(1-H_n(u)\right)} +$$

$$+n^{-1/2} \int\limits_{-\infty}^{x} \frac{a_n^{(0)}(u)\, d\, a_n^{(i)}(u)}{\left(1-H(u)\right)\left(1-H_n(u)\right)} = \sum_{m=1}^{4} R_{mn}^{(i)}(x).$$

For sum $R_{1n}^{(i)}(x) + R_{2n}^{(i)}(x) + R_{3n}^{(i)}(x)$ using (15) and (19) we have

$$P\left(\sup_{x \le T_n} \left| \sum_{m=1}^{4} R_{mn}^{(i)}(x) \right| > 3Cn^{-1/2} \log n + \varepsilon n^{-1/2} b_n^3 \log n \right) \le 3Kn^{-\beta} + 2Ln^{-w\varepsilon} \le \left(3K + 2L\right) n^{-\beta},\ i \in \Im. \quad (26)$$

Rewrite $R_{4n}^{(i)}$ as

$$R_{4n}^{(i)}(x) = n^{-1/2} \int\limits_{-\infty}^{x} \frac{\left(a_n^{(0)}(u)\right)^2 d\left(H(u;i) - H(u;i)\right)}{\left(1-H(u)\right)^2 \left(1-H_n(u)\right)} + n^{-1/2} \int\limits_{-\infty}^{x} \frac{a_n^{(0)}(u)\, d\, a_n^{(i)}(u)}{\left(1-H(u)\right)^2} = \overline{R}_{4n}^{(i)}(x) + \overline{\overline{R}}_{4n}^{(i)}(x). \quad (27)$$

Then by (15) for $i \in \Im$

$$P\left(\sup_{x \le T_n} \left| \overline{R}_{4n}^{(i)}(x) \right| > 2\varepsilon n^{-1/2} b_n^3 \log n \right) \le 2Ln^{-w\varepsilon} \le 2Ln^{-\beta}. \quad (28)$$

There exists an absolute constant A such that

$$P\left(\sup_{x \le T_n} \left| \overline{R}_{4n}^{(i)}(x) \right| > 3An^{-1/2} b_n^2 \log n \right) \le P\left(H_n^*(T_n) > 1\right) +$$

$$+P\left(\sup_{x \le T_n} n^{-1/2} \left| \int\limits_{-\infty}^{x} \frac{a_n^{(0)*}(u)\, da_n^{(i)*}(u)}{\left(1-H(u)\right)^2} \right| > 3An^{-1/2} b_n^2 \log n \right) \le Ln^{-r} + p_n, \quad (29)$$

so that for any $x \le T_n, H_n^*(x) \le H_n^*(T_n)$ and if $H_n^*(T_n) \le 1,$ then $H_n^*(x;i) \le H_n^*(T_n)$ and hence $a_n^{(i)}(x) = a_n^{(i)*}(x)$ for $i \in \Im$. It is enough to estimate probability p_n. According to proof of theorem 1 in [13], supposing $a_{\nu_n}^{(0)}(x) = \sqrt{\nu_n}\left(H_{\nu_n}^*(x) - H(x)\right),\ a_{\nu_n}^{(i)}(x) = \sqrt{\nu_n}\left(H_{\nu_n}^*(x;i) - H(x;i)\right),\ i \in \Im$ and using representation (10), we have proved the theorem 4.

Conclusion

We consider Kac processes in a general censorship scheme, including competing risks model and random censoring from both sides. Our results uses strong approximation method. Cumulative hazard processes also investigated in a similar manner in the general setting. In paper we obtain corresponding approximation results for ordinary empirical processes, for a Kac processes and their modifications and for hazard processes. All results are new and have approximation rates of order $n^{-1/2} \log n.$

References

1. Abdushukurov, A.A. (1998) Nonparametric estimation of the distribution function based on relative risk function. *Communications in Statistics – Theory and Methods*. 27(8). pp. 1991–2012.

2. Abdushukurov, A.A. (1999) On nonparametric estimation of reliability indices by censored samples. *Theory of Probability and Its Applications*. 43(1). pp. 3–11.
3. Csörgő, S. & Horváth, L. (1982) On random censorship from the right. *Acta Scientiarum Mathematicarum*. 44. pp. 23–34.
4. Shulenin, V.P. (2010) Asymptotic and nonasymptotic properties of Hodges – Lehmann adaptiveestimators. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitelnaya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(11). pp. 96–112.
5. Shulenin, V.P. (2021) Asymptotic properties and robustness of trimmed versions estimates of standard deviation and mean absolute deviations. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitelnaya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 55. pp. 91–102. DOI: 10.17223/19988605/55/11
6. Zenkova, J.N. (2011) Using of information about Sf-equal-shoulder symmetry at the processing of censored data. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitelnaya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(15). pp. 32–45.
7. Dvoretzky, A., Kiefer, J. & Wolfowitz, J. (1956) Asymptotic minimax character of the sample distribution function and of the multinomial estimator. *The Annals of Mathematical Statistics.* 27. pp. 642–669.
8. Massart, P. (1990) The tight constant in the Dvoretzky-Kiefer-Wolfowits inequality. *The Annals of Probability.* 18(3). pp. 1269–1283.
9. Abdushukurov, A.A. (2009) *Statistics of Incomplete Observations*. Tashkent: University Press.
10. Burke, M.D., Csörgő, S. & Horváth, L. (1981) Strong approximations of some biometric estimates under random censorship. *Z. Wahrschein. Verw. Gebiete*. 56. pp. 87–112.
11. Mason, D.M. (2015) Classical empirical process theory and wighted approximation. *Comunicaciones del CIMAT*. N.I-15-03/09-11-2015/ (PE/CIMAT).
12. Mason, D.M. (2017) Selected definition and results from modern empirical theory. *Comunicaciones del CIMAT*. N.I-17-01, 16.03.2017. (PE).
13. Csörgő, S. (1980) Strong approximation of empirical Kac processes. *Carleton Math. Lect. Note.* 26. pp. 71–86.
14. Petrov, W.W. (1987) *Limit theorems for sum of random variables*. Moscow: Nauka.

Information about the authors:

Abdushukurov Abdurakhim A. (Doctor of Physical and Mathematical Sciences, Professor, Branch of Moscow State University in Tashkent, Tashkent, Uzbekistan). E-mail: a_abdushukurov@rambler.ru

Sayfulloyeva Gulnoz S. (Post-graduate Student of the Department of Probability Theory and Mathematical Statistics of the Navoi State Pedagogical Institute, Navoi, Uzbekistan). E-mail: sayfullayevagulnoz@gmail.com

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

***Информация об авторах*:**

Абдушукуров Абдурахим Ахмедович – профессор, доктор физико-математических наук, профессор Ташкентского филиала Московского государственного университета им. М.В. Ломоносова (Ташкент, Республика Узбекистан). E-mail: a_abdushukurov@rambler.ru

Сайфуллоева Гулноз Сайфуллоевна – докторант кафедры теории вероятностей и математической статистики Навоийского государственного педагогического института (Навои, Республика Узбекистан). E-mail: sayfullayevagulnoz@gmail.com

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 04.03.2022; принята к публикации 29.11.2022

Received 04.03.2022; accepted for publication 29.11.2022