Journal of Siberian Federal University. Mathematics & Physics 2015, 8(2), 125–133

УДК 519.24

A Class of Special Empirical Processes of Independence

Abdurahim A. Abdushukurov*
Leyla R. Kakadjanova†
Dpt. Probability Theory and Mathematical Statistics
National University of Uzbekistan
VUZ Gorodok, Tashkent, 100174
Uzbekistan

Received 12.12.2014, received in revised form 13.02.2015, accepted 12.03.2015

In this paper we investigate the asymptotic properties of one class of empirical processes for certain classes of integrable functions.

Keywords: empirical processes, metric entropy, Glivenko-Cantelli theorem, Donsker's theorem.

Introduction

In this paper we investigate the limit properties of a class of empirical processes of independence indexed on a set of measurable functions. The necessity of considering such processes stems from practical situations where we are interested in joint properties of pairs consisting of random variables (r.v.-s) and events.

Let us consider the following sequence of experiments in which observed pairs are consisted of $\{(X_k, A_k), k \geqslant 1\}$, where X_k are random elements defined on a probability space $(\Omega, A, \mathbb{P})$ with values in a measurable space $(\mathfrak{X}, \mathfrak{B})$. Events A_k have a common probability $p \in (0,1)$. Let $\delta_k = I(A_k)$ be the indicator of the event A_k. At the $n-th$ step of experiment is observed the sample $\mathbb{S}^{(n)} = \{(X_k, \delta_k), 1 \leqslant k \leqslant n\}$. Each pair in the sample $\mathbb{S}^{(n)}$ induces a statistical model with the sample space $\mathfrak{X} \otimes \{0,1\}$, sigma-algebra of sets of the form $B \times D$ and induces distribution $\mathbb{Q}^*(B \times D) = \mathbb{P}(X_k \in B, \delta_k \in D)$, where $B \in \mathfrak{B}$, $D \subset \{0,1\}$. Let us define submeasures $\mathbb{Q}_1(B) = = \mathbb{Q}^*(B \times \{1\})$, $\mathbb{Q}_0(B) = \mathbb{Q}^*(B \times \{0\})$ and $\mathbb{Q}(B) = \mathbb{Q}^*(B \times \{0,1\}) = \mathbb{Q}_0(B) + \mathbb{Q}_1(B)$, $B \in \mathfrak{B}$. We also consider the hypothesis $\mathcal{H}$ of independence X_k and A_k for each $k \geqslant 1$. The validity of $\mathcal{H}$ can be tested by using the equations $\mathbb{Q}_1(B) = p\mathbb{Q}(B)$ or $\mathbb{Q}_0(B) = (1-p)\mathbb{Q}(B)$ for any $B \in \mathfrak{B}$. We define the measures $\Lambda(B) = \mathbb{Q}_1(B) - p\mathbb{Q}(B)$, $B \in \mathfrak{B}$. Thus, under the hypothesis $\mathcal{H} : \Lambda(B) = 0$, for any $B \in \mathfrak{B}$. Let us define the empirical measures for all $B \in \mathfrak{B}$:

$$\mathbb{Q}_{1n}(B) = \frac{1}{n}\sum_{k=1}^{n} \delta_k I(X_k \in B),$$

$$\mathbb{Q}_{0n}(B) = \frac{1}{n}\sum_{k=1}^{n} (1-\delta_k) I(X_k \in B),$$

$$\mathbb{Q}_n(B) = \frac{1}{n}\sum_{k=1}^{n} I(X_k \in B) = \mathbb{Q}_{0n}(B) + \mathbb{Q}_{1n}(B).$$

*a_abdushukurov@rambler.ru
†leyla_tvms@rambler.r

These measures are empirical estimates for $\mathbb{Q}_1$, $\mathbb{Q}_0$ and $\mathbb{Q}$ respectively. Since $p = \mathbb{Q}_1(\mathfrak{X})$ then estimate for p is $p_n = \mathbb{Q}_{1n}(\mathfrak{X}) = \frac{1}{n}\sum\limits_{k=1}^{n}\delta_k$. According to the strong law of large numbers (SLLN) for a fixed B when $n \to \infty$, $\mathbb{Q}_{jn}(B) \overset{\text{a.s.}}{\to} \mathbb{Q}_j(B)$, $j = 0, 1$ and consequently, $\mathbb{Q}_n(B) \overset{\text{a.s.}}{\to} \mathbb{Q}(B)$ and $p_n \overset{\text{a.s.}}{\to} p$. Thus, for each $B \in \mathfrak{B}$ at $n \to \infty$, $\Lambda_n(B) = \mathbb{Q}_{1n}(B) - p_n\mathbb{Q}_n(B) \overset{\text{a.s.}}{\to} \Lambda(B)$ and under validity of $\mathcal{H}$, $\Lambda_n(B) \overset{\text{a.s.}}{\to} 0$. Thus we are naturally led to the study of limit properties of processes of independence $\{\Lambda_n(B) - \Lambda(B)\}$ for a certain class $\mathcal{G}$ sets of B. In this paper we consider general classes of specially normalized empirical processes of independence indexed by a class of measurable functions.

1. Empirical processes of independence

Suppose that $\mathcal{F}$ be a set of measurable functions $f : \mathfrak{X} \to \mathbb{R}$. For the signed measure $\mathbb{G}$ and function $f \in \mathcal{F}$ we define the integral

$$\mathbb{G}f = \int_{\mathfrak{X}} f d\mathbb{G}.$$

Let us define $\mathcal{F}$ is indexed empirical process $\mathbb{G}_n : \mathcal{F} \in \mathbb{R}$ as:

$$f \mapsto \mathbb{G}_n f = \sqrt{n}(\mathbb{Q}_n - \mathbb{Q}) f = n^{-1/2}\sum_{k=1}^{n}(f(X_k) - \mathbb{Q}f),\ f \in \mathcal{F}.$$

Note that $\mathbb{G}_n f = \mathbb{G}_{0n} f + \mathbb{G}_{1n} f$, where $\{\mathbb{G}_{jn} f = \sqrt{n}(\mathbb{Q}_{jn} - \mathbb{Q}_j) f,\ j = 0, 1,\ f \in \mathcal{F}\}$ is subempirical processes. According to the SLLN and the central limit theorem (CLT) and under conditions $\mathbb{Q}|f| < \infty$, $\mathbb{Q}f^2 < \infty$ for the given function f we have

$$\mathbb{Q}_n f \overset{\text{a.s}}{\to} \mathbb{Q}f,\ \mathbb{G}_n f \Rightarrow N\left(0, \mathbb{Q}(f - \mathbb{Q}f)^2\right). \tag{1}$$

Uniformly variants for $f \in \mathcal{F}$ in statements (1) have well-developed theory. The generalized analogues of classical Glivenko-Cantelli theorem and Donsker's theorem for $\mathcal{F}$-indexed empirical processes can be found in [1–7]. One should mention the special case when $\mathcal{F}$ is the set of indicators of a class $\mathcal{G}$ of sets B:

$$\mathcal{F} = \{I(B) : B \in \mathcal{G}\}. \tag{2}$$

It is easy to see that in this case $\{\mathbb{G}_n f = \mathbb{G}_n(B) = \sqrt{n}(\mathbb{Q}_n(B) - \mathbb{Q}(B)),\ B \in \mathcal{G}\}$ and this process is called as $\mathcal{G}$-indexed. An example of such process is the classical empirical process obtained by $\mathfrak{X} = \mathbb{R}^m$, $G = \{(-\infty, x] : x \in \mathbb{R}^m\}$, $\mathbb{Q}((-\infty, x]) = H(x)$ and $\mathbb{Q}_n((-\infty, x]) = H_n(x)$ as $\{\mathbb{G}_n((-\infty, x]) = \sqrt{n}(H_n(x) - H(x)), x \in \mathbb{R}^m\}$.

Let us return to general $\mathcal{F}$-indexed processes $\{\mathbb{G}_n f, f \in \mathcal{F}\}$ and recall that there are various variants of the Glivenko-Cantelli theorem based on the theory of metric entropy under certain conditions on the set of measurable functions $\mathcal{F}$. These conditions ensure that $\|\mathbb{G}_n\|_{\mathcal{F}} = \sup\{|\mathbb{G}_n f| : f \in \mathcal{F}\}$ converges in probability to zero or it almost surely converges to zero. Such classes $\mathcal{F}$ are called the weak or strong Glivenko-Cantelli classes, respectively. Donsker-type theorems provide general conditions on $\mathcal{F}$ under which

$$\mathbb{G}_n f \Rightarrow \mathbb{G}f \text{ in } l^\infty(\mathcal{F}), \tag{3}$$

where $l^\infty(\mathcal{F})$ is the space of all bounded functions $f : \mathfrak{X} \to \mathbb{R}$ equipped with the supremum-norm $\|f\|_{\mathcal{F}}$ and $\Rightarrow$ means the weak convergence (see [6], p. 81).

Class $\mathcal{F}$ for which convergence (3) holds is called a Donsker class. Limiting field $\{\mathbb{G}f,\ f \in \mathcal{F}\}$ called $\mathbb{Q}$-Brownian bridge. It is a tight Borel measurable element of $l^{\infty}(\mathcal{F})$ and it is a Gaussian field with zero mean and covariance function

$$\mathbb{E}\mathbb{G}f_1\mathbb{G}f_2 = \mathbb{Q}\left(f_1 - \mathbb{Q}f_1\right)\left(f_2 - \mathbb{Q}f_2\right) = \mathbb{Q}f_1f_2 - \mathbb{Q}f_1\mathbb{Q}f_2. \tag{4}$$

$\mathbb{Q}$-Brownian bridge $\{\mathbb{G}f, f \in \mathcal{F}\}$ can be represented in terms of $\mathbb{Q}$-Brownian sheet $\{\mathbb{W}(f), f \in \mathcal{F}\}$ as

$$\mathbb{G}f \overset{d}{=} \mathbb{W}(f) - \mathbb{W}(1)\,\mathbb{Q}f,\ \ f \in \mathcal{F}, \tag{5}$$

where $\mathbb{E}\mathbb{W}(f) = 0,\ \ \mathbb{E}\mathbb{W}(f_1)\,\mathbb{E}\mathbb{W}(f_2) = \mathbb{Q}f_1f_2$ and $\mathbb{W}(1)$ is the value of $\mathbb{Q}$-Brownian sheet for $f \equiv 1$.

In connection with the problem of testing the hypothesis $\mathcal{H}$, we introduce $\mathcal{F}$-processes

$$\Lambda f = \mathbb{Q}_1 f - p\mathbb{Q}f,\ \ \Lambda_n f = \mathbb{Q}_{1n} f - p_n\mathbb{Q}_n f,\ \ f \in \mathcal{F}. \tag{6}$$

Let us note that for the given function f, when $n \to \infty$, $\mathbb{Q}_j\,|f| < \infty$, $j = 0, 1$, we have $\Lambda_n f \overset{\text{a.s.}}{\to} \Lambda f$ in accordance with SLLN and under validity of $\mathcal{H}$, $\Lambda f = 0$. It is easy to see that for the fixed f, variable $\sqrt{n}\,(\Lambda_n - \Lambda)\,f$ is a linear functional of subempirical processes provided that $\mathbb{Q}_j f^2 < \infty,\ j = 0, 1$, and it has the limit normal distribution with zero mean. In this paper we propose and study the following $\mathcal{F}$-indexed normalized process in order to test the hypothesis $\mathcal{H}$:

$$\Delta_n f = \int_{\mathfrak{X}} f d\Delta_n = \left(\frac{n}{p_n\left(1 - p_n\right)}\right)^{1/2} \left(\Lambda_n - \Lambda\right) f,\ \ f \in \mathcal{F}. \tag{7}$$

Process (7) has the important property: it converges to the same $\mathbb{Q}$-Brownian bridge $\{\mathbb{G}f\ ,\ f \in \mathcal{F}\}$ under validity of $\mathcal{H}$. Certain of the results presented in this paper can be found in reports [8–11].

2. Asymptotical results

Let $\mathcal{L}_q(\mathbb{Q})$ be the space of functions $f : \mathfrak{X} \to \mathbb{R}$ with the norm

$$\|f\|_{\mathbb{Q},q} = \left(\mathbb{Q}|f|^q\right)^{1/q} = \left\{\int_{\mathfrak{X}} |f|^q d\mathbb{Q}\right\}^{1/q}.$$

To prove the $\mathcal{F}$-uniform variants of Glivenko-Cantelli theorem and Donsker's theorem we define the complexity or entropy of class $\mathcal{F}$. To determine the entropy it is necessary to define the concept of ε-brackets. The ε-bracket in $\mathcal{L}_q(\mathbb{Q})$ is a pair of functions $\varphi, \psi \in \mathcal{L}_q(\mathbb{Q})$ such that $\mathbb{Q}\left(\varphi(X) \leqslant \psi(X)\right) = 1$ and $\|\psi - \varphi\|_{\mathbb{Q},q} \leqslant \varepsilon$, i.e. $\mathbb{Q}(\psi - \varphi)^q \leqslant \varepsilon^q$. Function $f \in \mathcal{F}$ is in (or covered by) bracket $[\varphi, \psi]$, if $\mathbb{Q}\left(\varphi(X) \leqslant f(X) \leqslant \psi(X)\right) = 1$. One should note that the functions φ and ψ may not belong to the class $\mathcal{F}$, but they must have finite norms. Bracketing (or covering) number $N_{[\,]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}\right)\right)$ is the minimum number of ε-brackets in $\mathcal{L}_q(\mathbb{Q})$ needed to cover $\mathcal{F}$ (see [1–7]):

$$N_{[\,]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}\right)\right) = \min \left\{ \begin{array}{l} k:\ \text{for some}\, f_1, ..., f_k \in \mathcal{L}_q\left(\mathbb{Q}\right), \\ \mathcal{F} \subset \bigcup\limits_{i,j} \left[f_i, f_j\right] : \|f_j - f_i\|_{\mathbb{Q},q} \leqslant \varepsilon. \end{array} \right.$$

Number $H_q\left(\varepsilon\right) = \log N_{[\,]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}\right)\right)$ is called the metric entropy with bracketing of the class $\mathcal{F}$ in $\mathcal{L}_q(\mathbb{Q})$. Number $H_{jq}\left(\varepsilon\right) = \log N_{[\,]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}_j\right)\right), j = 0, 1$ denotes the metric entropy of a class $\mathcal{F}$ in $\mathcal{L}_q(\mathbb{Q}_j)$, $j = 0,1$, respectively. To prove the weak convergence of $\mathcal{F}$-indexed empirical processes (7) we introduce the integral of the metric entropy with bracketing as

$$J_{j[\,]}^{(q)}(\delta) = J_{j[\,]}(\delta; \mathcal{F}; \mathcal{L}_q(\mathbb{Q}_j)) = \int_0^\delta (H_{jq}(\varepsilon))^{1/2} d\varepsilon, j = 0,1, \text{ for } 0 < \delta < 1.$$

Recall that numbers $N_{[\,]}(\cdot)$ converge to $+\infty$ at $\varepsilon \downarrow 0$. However, it is necessary for Donsker's theorem that they converge not very fast to $+\infty$. This speed is measured by the integrals $J_{j[\,]}^{(q)}(\delta)$ (see [6, 7]).

The following theorem shows validity of Glivenko-Cantelli type theorem for the process $\{\Delta_n f,\ f \in \mathcal{F}\}$. Here sign $*$ means a.s. convergence by outer probability.

Theorem 2.1. *Let the class $\mathcal{F}$ such that*

$$N_{[\,]}(\varepsilon, \mathcal{F}, \mathcal{L}_1(\mathbb{Q}_j)) < \infty, j = 0,1. \tag{8}$$

Then under validity of the hypothesis $\mathcal{H}$ and at $n \to \infty$

$$\left\| n^{-1/2} \Delta_n f \right\|_{\mathcal{F}}^* \overset{a.s.}{\to} 0. \tag{9}$$

Proof. According to SLLN when $n \to \infty$, $p_n \overset{\text{a.s.}}{\to} p \in (0,1)$. Therefore, convergence of (9) is equivalent to

$$\|\Lambda_n f\|_{\mathcal{F}}^* \overset{\text{a.s.}}{\to} 0,\ n \to \infty. \tag{10}$$

If hypothesis $\mathcal{H}$ is valid, then it is easy to verify that

$$\begin{aligned} \|\Lambda_n f\|_{\mathcal{F}} &\leqslant \|(\mathbb{Q}_{1n} - \mathbb{Q}_1) f\|_{\mathcal{F}} + p_n \|(\mathbb{Q}_n - \mathbb{Q}) f\|_{\mathcal{F}} + \|\mathbb{Q} f\|_{\mathcal{F}} \cdot |p_n - p| \leqslant \\ &\leqslant 2\|(\mathbb{Q}_{1n} - \mathbb{Q}_1) f\|_{\mathcal{F}} + \|(\mathbb{Q}_{0n} - \mathbb{Q}_0) f\|_{\mathcal{F}} + \|f\|_{\mathbb{Q},1} \cdot |p_n - p|, \end{aligned} \tag{11}$$

where

$$\|f\|_{\mathbb{Q},1} = \int_{\mathfrak{X}} |f| d\mathbb{Q} \leqslant \int_{\mathfrak{X}} |f|\, d\mathbb{Q}_1 + \int_{\mathfrak{X}} |f|\, d\mathbb{Q}_0 = \|f\|_{\mathbb{Q}_1,1} + \|f\|_{\mathbb{Q}_0,1} < \infty. \tag{12}$$

Under conditions (8) $\mathcal{F}$ is a Glivenko-Cantelli class with respect to measures $\mathbb{Q}_j$, $j = 0,1$. Hence, by Theorem 19.4 in [7] for each $\varepsilon > 0$:

$$\limsup_{n\to\infty} \left(\sup_{f \in \mathcal{F}} |(\mathbb{Q}_{jn} - \mathbb{Q}_j) f| \right)^* \leqslant \varepsilon. \tag{13}$$

Now relations (10) and (9) follow from (11)–(13). Theorem is proved. □

To prove the weak convergence of process (7) to a Gaussian process, we first investigate the limiting properties of two-dimensional empirical field $\{(\mathbb{A}_n f, \mathbb{A}_{1n} g),\ f, g \in \mathcal{F}\}$, where $\mathbb{A}_n f = n^{1/2}(\mathbb{Q}_n - \mathbb{Q}) f$ and $\mathbb{A}_{1n} g = n^{1/2}(\mathbb{Q}_{1n} - \mathbb{Q}_1) g$.

Theorem 2.2. *Let the class $\mathcal{F}$ such that*

$$\mathcal{F} \subset \mathcal{L}_2(\mathbb{Q}_j) \text{ and } J_{j[\,]}^{(2)}(1) < \infty,\ j = 0,1. \tag{14}$$

Then for $n \to \infty$ sequence $\{(\mathbb{A}_n f, \mathbb{A}_{1n} g),\ f, g \in \mathcal{F}\}$ of $\mathcal{F} \to \mathbb{R}^2$ maps weak converge in $l^\infty(\mathcal{F}) \times l^\infty(\mathcal{F})$ to the two-dimensional Gaussian field $\{(\mathbb{A} f, \mathbb{A}_1 g),\ f, g \in \mathcal{F}\}$ with zero mean and the following covariance structure for $f, g \in \mathcal{F}$:

$$\begin{aligned} \mathbb{E}(\mathbb{A} f \cdot \mathbb{A} g) &= \mathbb{Q} fg - \mathbb{Q} f \mathbb{Q} g, \\ \mathbb{E}(\mathbb{A}_1 f \cdot \mathbb{A}_1 g) &= \mathbb{Q}_1 fg - \mathbb{Q}_1 f \mathbb{Q}_1 g, \\ \mathbb{E}(\mathbb{A} f \cdot \mathbb{A}_1 g) &= \mathbb{Q}_1 fg - \mathbb{Q} f \mathbb{Q}_1 g. \end{aligned} \tag{15}$$

Proof. From the first condition in (14) it follows that for the fixed $f_i, g_i \in \mathcal{F} : \mathbb{Q}{f_i}^2 = \mathbb{Q}_0 f_i^2 + \mathbb{Q}_1 f_i^2 < \infty$ and $\mathbb{Q}_1 g_i^2 < \infty$, $i = \overline{1,m}$. Then according to multidimensional CLT finite dimensional distributions of vector $(\mathbb{A}_n f, \mathbb{A}_{1n} g)$ converge to multivariate Gaussian distribution with zero mean vector. Covariance matrix defined by structure (15) is the normalized sum of independent and identically distributed r.v.-s :

$$(\mathbb{A}_n f, \mathbb{A}_{1n} g) = n^{-1/2} \sum_{k=1}^{n} (f(X_k) - \mathbb{Q}f, \delta_k g(X_k) - \mathbb{Q}_1 g).$$

It remains to prove tightness of $(\mathbb{A}_n f,\ \mathbb{A}_{1n} g)$. Under conditions (14) and $n \to \infty$ we have following Donsker's theorems (see [6]):

$$\mathbb{A}_n f \Rightarrow \mathbb{A} f \text{ in } l^{\infty}(\mathcal{F}),\ \mathbb{A}_{1n} f \Rightarrow \mathbb{A}_1 f \text{ in } l^{\infty}(\mathcal{F}), \tag{16}$$

where limiting processes are respectively $\mathbb{Q}$- and $\mathbb{Q}_1$- Brownian bridges, i.e. tight Borel measurable elements of $l^{\infty}(\mathcal{F})$. Then the sequences of marginal distributions which induced by processes $\{\mathbb{A}_n f,\ \ f \in \mathcal{F}\}$ and $\{\mathbb{A}_{1n} f,\ \ f \in \mathcal{F}\}$ are tight (see, Lemma 1.3.8 in [6]). Process $\{(\mathbb{A}_n f, \mathbb{A}_{1n} g)\ ,\ f, g \in \mathcal{F}\}$ is element of space $l^{\infty}(\mathcal{F}) \times l^{\infty}(\mathcal{F})$ and by Lemma 1.4.3. in [6] also induces in this space the tight sequence of distributions. Theorem is proved. □

Remark. In formula (15) at $g \equiv 1$ we have $\mathbb{Q}_1 1 = p$ and

$$\mathbb{E}(\mathbb{A} f \cdot \mathbb{A}_1 1) = \mathbb{Q}_1 f - p\, \mathbb{Q} f,\ f \in \mathcal{F}. \tag{17}$$

Hence, when hypothesis $\mathcal{H}$ is valid then covariance (17) is equal to zero for all $f \in \mathcal{F}$. Thus under hypothesis $\mathcal{H}$ the Brownian bridge $\{\mathbb{A}\ f,\ \ f \in \mathcal{F}\}$ and r.v. $\mu_0 = \mathbb{A}_1 1$ with normal distribution $\mathcal{N}(0, p(1-p))$ are independent.

Let us introduce the empirical process $\left\{ n^{1/2}(\Lambda_n - \Lambda) f = \mathbb{G}_n^* f,\ f \in \mathcal{F} \right\}$. This process connected with process (7) by the following relation:

$$\mathbb{G}_n^* f = (p_n(1-p_n))^{1/2} \cdot \Delta_n f,\ f \in \mathcal{F}. \tag{18}$$

Process (18) plays a supporting role in study of basic process (7) which property of weak convergence to a $\mathbb{Q}$-Brownian bridge is contained in the following statement.

Theorem 2.3. *Under the conditions of Theorem 2.2 for $n \to \infty$*

$$\Delta_n f \Rightarrow \Delta f \text{ in } l^{\infty}(\mathcal{F}), \tag{19}$$

where $\{\Delta f,\ \ f \in \mathcal{F}\}$ is a Gaussian field with zero mean and under validity of the hypothesis $\mathcal{H}$ it coincides with $\mathbb{Q}$-Brownian bridge.

Proof. We consider process (18) and represent it in the form $\mathbb{G}_n^* f = \mathbb{A}_{1n} f - p_n \mathbb{A}_n f\ - \mu_n \mathbb{Q} f$, where $\mathbb{A}_n f = \mathbb{A}_{0n} f + \mathbb{A}_{1n} f$, $\mathbb{A}_{jn} f = n^{1/2}(\mathbb{Q}_{jn} - \mathbb{Q}_j) f$, $j = 0, 1$; $\mu_n = n^{1/2}(p_n - p) = \mathbb{A}_{1n} 1$. It is easy to see that $\mathbb{G}_n^* f$ is asymptotically equivalent (in terms of convergence to the same process) to the process $\mathbb{G}_n^0 f = \mathbb{A}_{1n} f - p\mathbb{A}_n f - \mu_n \mathbb{Q} f$. According to Theorem 2.2 for $n \to \infty$

$$\mathbb{G}_n^0 f \Rightarrow \mathbb{G}^0 f = \mathbb{A}_1 f - p\mathbb{A} f - \mu_0 \mathbb{Q} f \text{ in } l^{\infty}(\mathcal{F}). \tag{20}$$

Let us note that process $\left\{\mathbb{G}^0 f,\ f \in \mathcal{F}\right\}$ is a linear functional of Gaussian processes. It is also a Gaussian process with zero mean and covariance which calculated with the use of (15) and (17) for $f, g \in \mathcal{F}$ as

$$\mathbb{E}\mathbb{G}^0 f \mathbb{G}^0 g = \sum_{j=1}^{9} \mathbb{C}_j, \tag{21}$$

where

$$\mathbb{C}_1 = \mathbb{Q}_1 fg - \mathbb{Q}_1 f \mathbb{Q}_1 g; \qquad \mathbb{C}_2 = -p\left(\mathbb{Q}_1 fg - \mathbb{Q} f \mathbb{Q}_1 g\right); \qquad \mathbb{C}_3 = -\left(1-p\right) \mathbb{Q} f \mathbb{Q}_1 g;$$
$$\mathbb{C}_4 = -p\left(\mathbb{Q} fg - \mathbb{Q} g \mathbb{Q}_1 f\right); \qquad \mathbb{C}_5 = p^2\left(\mathbb{Q} fg - \mathbb{Q} f \mathbb{Q} g\right); \qquad \mathbb{C}_6 = -p\mathbb{Q} f\left(\mathbb{Q}_1 g - p\mathbb{Q} g\right);$$
$$\mathbb{C}_7 = -\left(1-p\right) \mathbb{Q} g \mathbb{Q}_1 f; \qquad \mathbb{C}_8 = -p\mathbb{Q} g\left(\mathbb{Q}_1 f - p\mathbb{Q} f\right); \qquad \mathbb{C}_9 = p\left(1-p\right) \mathbb{Q} f \mathbb{Q} g.$$

Under validity of the hypothesis $\mathcal{H}$ and taking into account the remark to Theorem 2.2 it is easy to verify that from (21) we have $\mathbb{E}\mathbb{G}^0 f\ \mathbb{G}^0 g = p\left(1-p\right)\left(\mathbb{Q} fg - \mathbb{Q} f \mathbb{Q} g\right)$. Then $\left[p\left(1-p\right)\right]^{-1/2}\mathbb{G}^0 f \overset{d}{=} \mathbb{G} f$. Thus we obtain a $\mathbb{Q}$-Brownian bridge with covariance (4). Therefore, according to (18) for $n \to \infty$

$$\Delta_n f \Rightarrow \left[p\left(1-p\right)\right]^{-1/2}\mathbb{G}^0 f \text{ in } l^\infty\left(\mathcal{F}\right)$$

and when hypothesis $\mathcal{H}$ is valid then

$$\Delta_n f \Rightarrow \mathbb{G} f \text{ in } l^\infty\left(\mathcal{F}\right).$$

□

Let us consider a generalization of Theorem 2.3 to the case of random sample size. Suppose that at n-th stage of observations a random number of observations from an infinite sequence of independent and identically distributed pairs $(X_1, \delta_1), (X_2, \delta_2), ...$ is available Here N_n is integer-valued nonnegative r.v. defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let the sequence N_n converges to infinity in the strong sense that there is a r.v. ν and at $n \to \infty$

$$\frac{N_n}{C_n} \xrightarrow{\text{p}} \nu, \tag{22}$$

Here $\mathbb{P}(\nu > 0) = 1$ and $C_n \to \infty$ is a deterministic sequence of numbers. Let $\{\Delta_{N_n} f,\ f \in \mathcal{F}\}$ be a sequence of normalized empirical processes of independence obtained from (7) by replacing index n to a random sequence N_n. The following theorem shows that this process has the same limiting distribution as $\{\Delta_n f, f \in \mathcal{F}\}$.

Theorem 2.4. *Under the conditions of Theorem 2.3 and (22) at $n \to \infty$*

$$\Delta_{N_n} f \Rightarrow \Delta f \ \textit{in}\ l^\infty(\mathcal{F}). \tag{23}$$

Consequently, from Theorem 2.3 and (23) under validity of hypothesis $\mathcal{H}$, distribution of Δf coincides with the distribution of $\mathbb{Q}$-Brownian bridge with covariance (4).

Proof is the consequence of Theorem 3.5.1 from [6] and Theorem 2.3 and hence details are omitted. □

Now suppose that $\{N_n,\ n \geqslant 1\}$ a sequence of Poisson r.v.-s with the mean n and independent identically distributed r.v.-s $(X_1, \delta_1), (X_2, \delta_2), ...$. Let us denote by $\{\Delta_n^* f,\ f \in \mathcal{F}\}$ a normalized empirical process of independence obtained from (7) by replacing the upper bounds n in all summations to N_n . Next theorem shows that the limiting process is the $\mathbb{Q}-$ Brownian sheet as defined in (5).

Theorem 2.5. *Under the conditions of Theorem 2.3 at $n \to \infty$*

$$\Delta_n^* f \Rightarrow \Delta^* f \ \textit{in}\ l^\infty(\mathcal{F}), \tag{24}$$

where by hypothesis $\mathcal{H}$, $\Delta^* f \overset{d}{=} \mathbb{W}(f)$, $f \in \mathcal{F}$.

Proof follows from Theorems 3.5.1, 3.5.3 from [6] and Theorem 3.4 if we take into consideration that $\frac{N_n}{n} \xrightarrow{\mathrm{p}} 1$, and processes $\mathbb{A}^*_{N_n} f = n^{1/2}(\sum_{k=1}^{N_n} f(X_k) - n\mathbb{Q}f)$ and $\mathbb{A}^*_{1N_n} f = n^{1/2}(\sum_{k=1}^{N_n} \delta_k f(X_k) - n\mathbb{Q}_1 f)$ have following standardized representations:

$$\mathbb{A}^*_{N_n} f = \sqrt{\frac{N_n}{n}} \mathbb{A}_{N_n} f + \sqrt{n}(\frac{N_n}{n} - 1)\mathbb{Q}f,$$

$$\mathbb{A}^*_{1N_n} f = \sqrt{\frac{N_n}{n}} \mathbb{A}_{1N_n} f + \sqrt{n}(\frac{N_n}{n} - 1)\mathbb{Q}_1 f.$$

The details are omitted. □

The results of Theorems 2.3–2.5 can be used to construct the statistics for testing the hypothesis $\mathcal{H}$. For example, from processes $\{\Delta_n f, f \in \mathcal{F}\}$, $\{\Delta_{N_n} f, f \in \mathcal{F}\}$ and $\{\Delta^*_n f, f \in \mathcal{F}\}$ one can construct the following Kolmogorov-type statistics $\mathbb{K}_n = \|\Delta_n f\|_{\mathcal{F}}$, $\mathbb{K}_{Nn} = \|\Delta_{Nn} f\|_{\mathcal{F}}$ and $\|\Delta^*_n f\|_{\mathcal{F}}$ which under validity of $\mathcal{H}$ have limiting distributions of r.v.-s $\mathbb{K}^0 = \|\mathbb{G}f\|_{\mathcal{F}}$ and $\mathbb{K}_n = \|\mathbb{W}(f)\|_{\mathcal{F}}$, respectively.

3. Application to random censoring

Let us consider a right random censoring model, where $X_i = \min\{T_i, C_i\}$ and $A_i = \{T_i \leqslant C_i\}$. Here r.v.-s T_i and C_i denote life times and censoring times. They are mutually independent with common continuous distribution functions F and G respectively ($F(0) = G(0) = 0$). Then considering data $\mathbb{S}^{(n)} = \{(X_i, \delta_i), 1 \leqslant i \leqslant n\}$ with $\delta_i = I(A_i)$, r.v.-s of interest T_i are observed when A_i occurs, i.e., $\delta_i = 1$. Take into account that X_i have common distribution function $H = 1 - (1-F)(1-G)$ and subdistributions defined as

$$\begin{aligned} \mathbb{Q}_0(B) &= \mathbb{P}(X_k \in B, \delta_k = 0) = \mathbb{P}(C_k \in B \cap [0, T_k)) \int_B (1 - F(t))G(dt), \\ \mathbb{Q}_1(B) &= \mathbb{P}(X_k \in B, \delta_k = 1) = \mathbb{P}(T_k \in B \cap [0, C_k]) \int_B (1 - G(t))F(dt). \end{aligned} \tag{25}$$

Now we consider simple proportional hazards model (PHM) or Koziol-Green model which is very useful in practical applications (see, for example, [12–16]). In PHM we assume the parametric relation

$$1 - G = (1 - F)^\beta \text{ for some } \beta > 0. \tag{26}$$

Taking into consideration (26), it is easy to see that $1 - F = (1 - H)^p$, where $p = \frac{1}{1+\beta} = \mathbb{P}(A_k)$. One of basic properties of PHM is that (26) holds when r.v.-s X_k and δ_k are independent. Such characteristic of PHM plays a basic role in constructing and studying estimators of many functionals of distribution F. The following sufficient maximum likelihood estimator of F was first introduced and studied [12–14]:

$$F_n(t) = 1 - (1 - H_n(t))^{p_n}, \tag{27}$$

where $H_n(t) = \frac{1}{n}\sum_{k=1}^{n} I(X_k \leqslant t)$ and $p_n = \frac{1}{n}\sum_{k=1}^{n} \delta_k$ are independent empirical estimators of $H(t)$ and p, respectively.

There are many papers devoted to statistical analysis of F_n. These papers are concerned with the superiority of methods for estimation and the testing in PHM and methods are based on F_n rather than on the product-limit estimator of Kaplan-Meier. Some references can be found in [16]. Hence the question arises as to when the advantages of the PHM can be used. In other words, there is now a need for testing of validity of PHM, i.e., for the composite hypothesis described by relation (26). But this relation is equivalent to hypothesis $\mathcal{H}$ on independence of r.v.-s $(X_1, ..., X_n)$ and $(\delta_1, ..., \delta_n)$.

Let us consider the following special empirical process (7):

$$\Delta_n(t) = \left(\frac{n}{p_n\,(1-p_n)}\right)^{1/2} \left(H_{1n}(t) - p_n H_n(t)\right), \; -\infty < t < \infty, \tag{28}$$

where $H_{1n}(t) = \frac{1}{n}\sum\limits_{k=1}^{n} I(X_k \leqslant t, \delta_k = 1)$. Then we have the consequence of Theorem 2.3: if $\mathcal{H}$ holds then as $n \to \infty$

$$\Delta_n\,(\cdot) \Rightarrow \mathbb{B}\,(H\,(\cdot))\,, \tag{29}$$

where $\{\mathbb{B}(y),\, 0 \leqslant y \leqslant 1\}$ is a Brownian bridge. Several statistics for testing $\mathcal{H}$ were considered [13–15]. Note that these statistics are based on relation (29) and corresponding tests are consistent. Moreover, by Theorems 2.3–2.5 one can consider more general classes of statistics using $\mathcal{F}$-indexed processes that are more flexible in applications than (28).

References

[1] K.S.Alexander, Probability inequalities for empirical processes and a law of the iterated logarithm, *Ann. Probab.*, **12**(1984), no. 4, 1041–1067.

[2] R.M.Dudley, Central limit theorems for empirical measures, *Ann. Probab.*, **6**(1978), 899–929.

[3] P.Gaensler, W.Stute, Empirical processes: a survey of results for independent and identically distributed random variables, *Ann. Probab.*, **7**(1979), no. 2, 193–243.

[4] E.Gine, J.Zinn, Some limit theorems for empirical processes, *Ann. Probab.* **12**(1984), no. 4, 929–989.

[5] G.R.Shorack, J.A.Wellner, Empirical processes with applications to statistics, *John Wiley & Sons*, 1986.

[6] A.W.Van der Vaart, J.A.Wellner, Weak convergence and empirical processes, Springer, 1996.

[7] A.W. Van der Vaart, Asymptotic Statistics, Cambridge University Press, 1998.

[8] A.A.Abdushukurov, L.R.Kakadjanova, About one class of empirical processes, Proceedings of the XVI International Conference on eventological mathematics and related issues, Krasnoyarsk, 2012, 31 (in Russian).

[9] A.A.Abdushukurov, L.R.Kakadjanova, Asymptotical Gaussian distribution of a class of empirical processes, Proc. of reports. International Conf. "Problems of modern topology and its applications", Tashkent, 2013, 100–102 (in Russian).

[10] A.A.Abdushukurov, L.R.Kakadjanova, An empirical processes of independence by class of measured functions, Materials of conf. "Statistics and its Applications", Tashkent, 2013, 24–29 (in Russian).

[11] A.A.Abdushukurov, L.R.Kakadjanova, An empirical processes of independence indexed by class of measured functions, *Acta National University of Uzbekistan*, **1**(2014), no. 2, 15–20 (in Russian).

[12] A.A.Abdushukurov, On some estimates of the distribution function under random censorship, Conference of young Scientists, Math. Inst. Acad. Sci. Uzbek SSR, VINITI, **8756**(1984) (in Russian).

[13] A.A.Abdushukurov, Nonparametric estimation in proportional hazards model, VINITI, **3448-87**(1987) (in Russian).

[14] A.A.Abdushukurov, Estimating of the probability density and intensity function of the Koziol-Green model of random censoring, *Sankhya: The Indian J. Statist.*, **48**(1987), 150–168.

[15] S.Csörgő, Estimation in the proportional hazards model of random censorship, *Statistics*, **19**(1988), no. 3, 437–463.

[16] S.Csörgő, J.J.Faraway, The paradoxical nature of the proportional hazards model of random censorship, *Statistics*, **31**(1998), 67–78.

Класс эмпирических процессов независимости

Абдурахим А. Абдушукуров
Лейла Р. Какаджанова

В данной статье мы исследуем асимптотические свойства одного класса эмпирических процессов для определенных классов интегрируемых функций.

Ключевые слова: эмпирические процессы, метрическая энтропия, теоремы Гливенко-Кантелли и Донскера.