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 \(\{(X_k, A_k), k \geq 1\}\), where \(X_k\) are random elements defined on a probability space \((\Omega, A, P)\) with values in a measurable space \((X, \mathcal{B})\). 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 \(S(n) = \{(X_k, \delta_k), 1 \leq k \leq n\}\). Each pair in the sample \(S(n)\) induces a statistical model with the sample space \(X \otimes \{0, 1\}\), sigma-algebra of sets of the form \(B \times D\) and induces distribution \(Q^*(B \times D) = P(X \in B, \delta \in D)\), where \(B \in \mathcal{B}, D \subset \{0, 1\}\). Let us define submeasures \(Q_1(B) = Q^*(B \times \{1\}), Q_0(B) = Q^*(B \times \{0\})\) and \(Q(B) = Q^*(B \times \{0, 1\}) = Q_0(B) + Q_1(B)\), \(B \in \mathcal{B}\). We also consider the hypothesis \(\mathcal{H}\) of independence \(X_k\) and \(A_k\) for each \(k \geq 1\). The validity of \(\mathcal{H}\) can be tested by using the equations \(Q_1(B) = pQ(B)\) or \(Q_0(B) = (1 - p)Q(B)\) for any \(B \in \mathcal{B}\). We define the measures \(\Lambda(B) = Q_1(B) - pQ(B), B \in \mathcal{B}\). Thus, under the hypothesis \(\mathcal{H} : \Lambda(B) = 0\), for any \(B \in \mathcal{B}\). Let us define the empirical measures for all \(B \in \mathcal{B}\):

\[
Q_{1n}(B) = \frac{1}{n} \sum_{k=1}^{n} \delta_k I(X_k \in B),
\]

\[
Q_{0n}(B) = \frac{1}{n} \sum_{k=1}^{n} (1 - \delta_k) I(X_k \in B),
\]

\[
Q_n(B) = \frac{1}{n} \sum_{k=1}^{n} I(X_k \in B) = Q_{0n}(B) + Q_{1n}(B).
\]
These measures are empirical estimates for $Q_1$, $Q_0$ and $Q$ respectively. Since $p = Q_1(\mathcal{X})$ then estimate for $p$ is $p_n = Q_{1n}(\mathcal{X}) = \frac{1}{n} \sum_{k=1}^{n} \delta_k$. According to the strong law of large numbers (SLLN) for a fixed $B$ when $n \to \infty$, $Q_{jn}(B) \overset{a.s.}{\to} Q_j(B)$, $j = 0, 1$ and consequently, $Q_n(B) \overset{a.s.}{\to} Q(B)$ and $p_n \overset{a.s.}{\to} p$. Thus, for each $B \in \mathcal{B}$ at $n \to \infty$, $\Lambda_n(B) = Q_{1n}(B) - p_n Q_n(B) \overset{a.s.}{\to} \Lambda(B)$ and under validity of $\mathcal{H}$, $\Lambda_n(B) \overset{a.s.}{\to} 0$. 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 $f : \mathcal{X} \to \mathbb{R}$. For the signed measure $\mathcal{G}$ and function $f \in \mathcal{F}$ we define the integral

$$Gf = \int \mathcal{X} f \, d\mathcal{G}.$$ 

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

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

Note that $G_n f = G_0 n f + G_{1n} f$, where $\{G_{jn} f = \sqrt{n} (Q_{jn} - Q_j) f, \quad j = 0, 1, f \in \mathcal{F}\}$ is subempirical processes. According to the SLLN and the central limit theorem (CLT) and under conditions $Q |f| < \infty$, $Q f^2 < \infty$ for the given function $f$ we have

$$Q_n f \overset{a.s.}{\to} Q f, \quad G_n f \Rightarrow N \left(0, Q(f - Q f)^2 \right). \quad (1)$$

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 $B$:

$$\mathcal{F} = \{1(B) : B \in \mathcal{G}\}. \quad (2)$$

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

Let us return to general $\mathcal{F}$-indexed processes $\{G_n f, f \in \mathcal{F}\}$ 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 $\|G_n\|_{\mathcal{F}} = \sup \{\|G_n f\| : f \in \mathcal{F}\}$ 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

$$G_n f \Rightarrow G f \text{ in } l^\infty(\mathcal{F}), \quad (3)$$

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

Class $\mathcal{F}$ for which convergence (3) holds is called a Donsker class. Limiting field $\{Gf, f \in \mathcal{F}\}$ called $Q$-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

$$
\mathbb{E}Gf_1Gf_2 = Q(f_1 - Qf_1)(f_2 - Qf_2) = Qf_1f_2 - Qf_1Qf_2.
$$

(4)

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

$$
\mathbb{E}W(f) = Qf, \quad f \in \mathcal{F},
$$

(5)

where $\mathbb{E}W(f) = 0$. $\mathbb{E}W(f_1)\mathbb{E}W(f_2) = Qf_1f_2$ and $\mathbb{E}W(1)$ is the value of $Q$-Brownian sheet for $f = 1$.

Let us note that for the given function $f$, the concept of $Q$ may not belong to the class $\mathcal{F}$ may not belong to the class $\mathcal{F}$. In connection with the problem of testing the hypothesis $H$ we introduce $\mathcal{F}$-processes

$$
\Lambda f = Q_1f - pQf, \quad \Lambda_n f = Q_1nf - p_nQ_nf, \quad f \in \mathcal{F}.
$$

(6)

Let us note that for the given function $f$, when $n \to \infty$, $Q_j|f| < \infty$, $j = 0, 1$, we have $\Lambda_n f \overset{a.s.}{\to} \Lambda f$ in accordance with SLLN and under validity of $H$, $\Lambda f = 0$. 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 $Q_j f^2 < \infty$, $j = 0, 1$, 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 $H$:

$$
\Delta_n f = \int_X fd\Delta_n = \left(\frac{n}{p_n(1 - p_n)}\right)^{1/2}(\Lambda_n - \Lambda) f, \quad f \in \mathcal{F}.
$$

(7)

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

2. Asymptotical results

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

$$
\|f\|_{Q,q} = (Q|f|^q)_1^{1/q} = \left\{\int_X |f|^q dQ\right\}^{1/q}.
$$

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 $\varepsilon$-brackets. The $\varepsilon$-bracket in $\mathcal{L}_q(Q)$ is a pair of functions $\varphi, \psi \in \mathcal{L}_q(Q)$ such that $Q(\varphi(X) \leq \psi(X)) = 1$ and $\|\psi - \varphi\|_{Q,q} \leq \varepsilon$, i.e. $Q(\psi - \varphi)^q \leq \varepsilon^q$. Function $f \in \mathcal{F}$ is in (or covered by) bracket $[\varphi, \psi]$, if $Q(\varphi(X) \leq f(X) \leq \psi(X)) = 1$. One should note that the functions $\varphi$ and $\psi$ may not belong to the class $\mathcal{F}$, but they must have finite norms. Bracketing (or covering) number $N(\varepsilon, \mathcal{F}, \mathcal{L}_q(Q))$ is the minimum number of $\varepsilon$-brackets in $\mathcal{L}_q(Q)$ needed to cover $\mathcal{F}$ (see [1–7]):

$$
N(\varepsilon, \mathcal{F}, \mathcal{L}_q(Q)) = \min \left\{ k : \text{for some } f_1, \ldots, f_k \in \mathcal{L}_q(Q), \mathcal{F} \subset \bigcup_{i,j} [f_i, f_j] : \|f_j - f_i\|_{Q,q} \leq \varepsilon. \right\}
$$

Number $H_q(\varepsilon) = \log N(\varepsilon, \mathcal{F}, \mathcal{L}_q(Q))$ is called the metric entropy with bracketing of the class $\mathcal{F}$ in $\mathcal{L}_q(Q)$. Number $H_q(\varepsilon) = \log N(\varepsilon, \mathcal{F}, \mathcal{L}_q(Q_j))$; $j = 0, 1$ denotes the metric entropy.
of a class $\mathcal{F}$ in $\mathcal{L}_q(Q_j)$, $j = 0, 1$, respectively. To prove the weak convergence of $\mathcal{F}$-indexed empirical processes (7) we introduce the integral of the metric entropy with bracketing as

$$J^{(q)}_{\delta} (\delta) = J_{\delta} (\delta; \mathcal{F}; \mathcal{L}_q(Q_j)) = \int_0^\delta (H_{\delta q}(\varepsilon))^{1/2} d\varepsilon, \quad j = 0, 1, \text{ for } 0 < \delta < 1.$$ 

Recall that numbers $N_{\delta} (\cdot)$ 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 $J^{(q)}_{\delta}$ (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

$$N_{\delta} (\varepsilon, \mathcal{F}, \mathcal{L}_q(Q_j)) < \infty, \quad j = 0, 1. \tag{8}$$

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

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

**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

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

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

$$\|\Lambda_n f\|_{\mathcal{F}} \leq \|Q_{n1} - Q_1\|f\|_{\mathcal{F}} + p_n \|Q_n - Q\|f\|_{\mathcal{F}} + \|Q_0f\|_{\mathcal{F}} \cdot |p_n - p| \leq$$

$$\leq 2\|Q_{n1} - Q_1\|f\|_{\mathcal{F}} + \|Q_{n0} - Q_0\|f\|_{\mathcal{F}} + \|Q_{1,1}\|f\|_{\mathcal{F}} \cdot |p_n - p|, \tag{11}$$

where

$$\|f\|_{Q_1} = \int_X |f| dQ \leq \int_X |f| dQ_1 + \int_X |f| dQ_0 = \|f\|_{Q_{1,1}} + \|f\|_{Q_{0,1}} < \infty. \tag{12}$$

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

$$\limsup_{n \to \infty} \left( \sup_{f \in \mathcal{F}} \left| (Q_{jn} - Q_j) f \right| \right)^* \overset{\text{a.s.}}{\to} \varepsilon. \tag{13}$$

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

To prove the weak convergence of process (7) to a Gaussian process, we first investigate the limiting properties of two-dimensional empirical field $\{(A_n f, A_{1n} g), f, g \in \mathcal{F}\}$, where $A_n f = n^{1/2} (Q_n - Q) f$ and $A_{1n} g = n^{1/2} (Q_{1n} - Q_1) g$.

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

$$\mathcal{F} \subset \mathcal{L}_q(Q_j) \text{ and } J^{(q)}_{\delta} (\delta) < \infty, \quad j = 0, 1. \tag{14}$$

Then for $n \to \infty$ sequence $\{(A_n f, A_{1n} g), f, g \in \mathcal{F}\}$ 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 $\{(A f, A_1 g), f, g \in \mathcal{F}\}$ with zero mean and the following covariance structure for $f, g \in \mathcal{F}$:

$$E(A f \cdot A g) = Q f g - Q f Q g, \tag{15}$$

$$E(A_{11} f \cdot A_{12} g) = Q_{11} f g - Q_{11} f Q_{12} g,$$

$$E(A_{12} f \cdot A_{11} g) = Q_{12} f g - Q_{12} f Q_{11} g.$$
Proof. From the first condition in (14) it follows that for the fixed \( f_i, q_i \in \mathcal{F} : \mathbf{Q} f_i^2 = Q_0 f_i^2 + Q_1 f_i^2 < \infty \) and \( Q_1 g_i^2 < \infty, \ i = 1, \ldots, m. \) Then according to multidimensional CLT finite dimensional distributions of vector \((A_n f, A_1 n g)\) 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:

\[
(A_n f, A_1 n g) = n^{-1/2} \sum_{k=1}^{n} (f(X_k) - \mathbf{Q} f, \delta_k g(X_k) - \mathbf{Q} g).
\]

It remains to prove tightness of \((A_n f, A_1 n g)\). Under conditions (14) and \( n \to \infty \) we have following Donsker’s theorems (see [6]):

\[
A_n f \Rightarrow A f \text{ in } l^\infty (\mathcal{F}), \ A_1 n f \Rightarrow A_1 f \text{ in } l^\infty (\mathcal{F}),
\]

where limiting processes are respectively \( \mathbf{Q} - \) and \( \mathbf{Q} _1 - \) Brownian bridges, i.e. tight Borel measurable elements of \( l^\infty (\mathcal{F}) \). Then the sequences of marginal distributions which induced by processes \( \{ A_n f, \ f \in \mathcal{F} \} \) and \( \{ A_1 n f, \ f \in \mathcal{F} \} \) are tight (see, Lemma 1.3.8 in [6]). Process \( \{(A_n f, A_1 n g), \ f, g \in \mathcal{F}\} \) 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. \( \square \)

Remark. In formula (15) at \( g \equiv 1 \) we have \( Q_1 1 = \rho \) and

\[
\mathbb{E} (\mathbf{A} f \cdot \mathbf{A} 1) = Q_1 1 - \rho Q f, \ f \in \mathcal{F}.
\]

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 \( \{ \mathbf{A} f, \ f \in \mathcal{F} \} \) and r.v. \( \mu_0 = \mathbf{A} 1 \) with normal distribution \( \mathcal{N}(0, P(1 - P)) \) are independent.

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

\[
G_n^* f = (p_n (1 - p_n))^{1/2} \cdot \Delta_n f, \ f \in \mathcal{F}.
\]

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

Theorem 2.3. Under the conditions of Theorem 2.2 for \( n \to \infty \)

\[
\Delta_n f \Rightarrow \Delta f \text{ in } l^\infty (\mathcal{F}),
\]

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 \( \mathbf{Q} - \) Brownian bridge.

Proof. We consider process (18) and represent it in the form \( G_n^0 f = \mathbf{A} n 1 f - p_n \mathbf{A} n f - \mu_n \mathbf{Q} f, \)

where \( \mathbf{A} n f = \mathbf{A} 0 n f + \mathbf{A} 1 n f, \ A_j n f = n^{1/2} (Q_j n - \mathbf{Q} j) f, \ j = 0, 1; \mu_n = n^{1/2} (p_n - P) = A_1 n 1. \) It is easy to see that \( G_n^0 f \) is asymptotically equivalent (in terms of convergence to the same process) to the process \( G_n^0 f = \mathbf{A} n 1 f - p \mathbf{A} n f - \mu \mathbf{Q} f. \) According to Theorem 2.2 for \( n \to \infty \)

\[
G_n^0 f \Rightarrow G^0 f = \mathbf{A} 1 f - p \mathbf{A} f - \mu \mathbf{Q} f \text{ in } l^\infty (\mathcal{F}).
\]

Let us note that process \( \{ G^0 f, \ f \in \mathcal{F} \} \) 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

\[
\mathbb{E} G^0 f G^0 g = \sum_{j=1}^{9} C_j,
\]
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 $\mathbb{E} G f G g = p(1-p) (Q f g - Q f Q g)$. Then $[p(1-p)]^{-1/2} G^0 f \overset{p}{\Rightarrow} G f$. Thus we obtain a $Q$-Brownian bridge with covariance (4). Therefore, according to (18) for $n \to \infty$

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

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

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

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 $(X_1, \delta_1), (X_2, \delta_2), \ldots$ 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. $\nu$ and at $n \to \infty$

$$\frac{N_n}{C_n} \overset{p}{\to} \nu, \quad (22)$$

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 $n \to \infty$

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

Consequently, from Theorem 2.3 and (23) under validity of hypothesis $\mathcal{H}$, distribution of $\Delta f$ coincides with the distribution of $Q$-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 $\{N_n, n \geq 1\}$ a sequence of Poisson r.v.-s with the mean $n$ and independent identically distributed r.v.-s $(X_1, \delta_1), (X_2, \delta_2), \ldots$ . 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 $Q$-Brownian sheet as defined in (5).

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

$$\Delta^*_n f \Rightarrow \Delta^* f \text{ in } l^\infty (\mathcal{F}), \quad (24)$$

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 \( \frac{N_n}{n} \xrightarrow{P} 1 \), and processes \( \mathcal{A}_{N_n}^* f = n^{1/2} (\sum_{k=1}^{N_n} f(X_k) - nQf) \) and \( \mathcal{A}_{1N_n}^* f = n^{1/2} (\sum_{k=1}^{N_n} \delta_k f(X_k) - nQ_1f) \) have following standardized representations:

\[
\mathcal{A}_{N_n}^* f = \sqrt{\frac{N_n}{n}} \mathcal{A}_{N_n} f + \sqrt{n} (\frac{N_n}{n} - 1)Qf,
\]

\[
\mathcal{A}_{1N_n}^* f = \sqrt{\frac{N_n}{n}} \mathcal{A}_{1N_n} f + \sqrt{n} (\frac{N_n}{n} - 1)Q_1f.
\]

The details are omitted. \( \square \)

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 \( \mathcal{K}_n = ||\Delta_n f||_\mathcal{F} \), \( \mathcal{K}_{N_n} = ||\Delta_{N_n} f||_\mathcal{F} \) and \( ||\Delta_n^* f||_\mathcal{F} \) which under validity of \( \mathcal{H} \) have limiting distributions of r.v.-s \( \mathcal{K}^0 = ||Gf||_\mathcal{F} \) and \( \mathcal{K}_n = ||W(f)||_\mathcal{F} \), respectively.

### 3. Application to random censoring

Let us consider a right random censoring model, where \( X_i = \min\{T_i, C_i\} \) and \( A_i = \{T_i \leq C_i\} \). 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 \( S^{(n)} = \{(X_i, \delta_i), 1 \leq i \leq n\} \) 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

\[
Q_0(B) = \mathbb{P}(X_k \in B, \delta_k = 0) = \mathbb{P}(C_k \in B \cap [0, T_k]) \int_B (1 - F(t))G(dt),
\]

\[
Q_1(B) = \mathbb{P}(X_k \in B, \delta_k = 1) = \mathbb{P}(T_k \in B \cap [0, C_k]) \int_B (1 - G(t))F(dt).
\]

Now we consider simple proportional hazards model (PHM) or Kozioł-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 \( \delta_k \) 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 \( H_n(t) = \frac{1}{n} \sum_{k=1}^{n} I(X_k \leq t) \) and \( p_n = \frac{1}{n} \sum_{k=1}^{n} \delta_k \) are independent empirical estimators of \( H(t) \) and \( p \), respectively.

---

Abduraim A. Abdushukurov, Leyla R. Kakadjanova A Class of Special Empirical of Independence
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 $H$ on independence of r.v.-s $(X_1, ..., X_n)$ and $(\delta_1, ..., \delta_n)$.

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

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

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

$$\Delta_n(\cdot) \Rightarrow B(H(\cdot)), \quad (29)$$

where $\{B(y), 0 \leq y \leq 1\}$ is a Brownian bridge. Several statistics for testing $H$ 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 $F$-indexed processes that are more flexible in applications than (28).

References


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 Глишенко-Кантелли и Донскера.