ANNALES
UNIVERSITATIS MARIAE CURIE - SKLODOWSKA
LUBLIN - POLONIA
VOL. LI.1, 13 SECTIO A 1997

PRZEMYSLAW MATULA (Lublin)

The Glivenko–Cantelli Lemma for a Class of Discrete Associated Random Variables

Dedicated to Professor Dominik Szynal on the occasion of his 60th birthday

ABSTRACT. We prove a Glivenko–Cantelli lemma for a class of discrete associated random variables. The obtained result applies in the case of lattice, in particular, integer-valued and binary random variables.

1. Introduction and the main result. Let $(X_n)_{n\in N}$ be a sequence of random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$. Here and in the sequel we assume that the random variables are associated, i. e., for every finite subcollection $X_{n_1}, X_{n_2}, ..., X_{n_k}$ and coordinatewise nondecreasing functions $f, g : \mathbb{R}^k \to \mathbb{R}$ the inequality

$$\mathrm{Cov}(f(X_{n_1}, X_{n_2}, ..., X_{n_k}), g(X_{n_1}, X_{n_2}, ..., X_{n_k})) \geq 0$$

holds, whenever this covariance is defined (cf. [7]).

1991 *Mathematics Subject Classification.* 60F15, 62G05.
Key words and phrases. Glivenko–Cantelli lemma, empirical distribution, associated random variables.

Associated processes are widely encountered in mathematical physics and statistics, in particular in reliability theory and in percolation theory (cf. [4], [7], [12], [13]). There is a number of limit theorems for associated sequences such as central limit theorem, strong law of large numbers, weak and strong invariance principle and the law of the iterated logarithm (cf. [2-4], [6], [8-13] and references therein). Asymptotic properties of empirical distribution and empirical survival function were considered in [1] and [8]. Hao Yu [8] studied the Glivenko–Cantelli lemma and weak convergence of empirical processes of associated sequences. He considered equidistributed random variables with continuous distribution function and pointed out that the Glivenko–Cantelli lemma remained open in the discrete case. Our goal is to fill this gap.

We shall consider associated random variables taking values in the set $S \subset \mathbb{R}$, such that for some $\delta > 0$, $\inf_{x,y \in S, x \neq y} |x - y| = \delta$. It is easy to see that S is at most countable, moreover any finite set and the set of integers satisfies the given condition. Associated processes of this kind are very important and were studied in [6] and [7].

Assume that $(X_n)_{n \in \mathbb{N}}$ is a sequence of r.v.'s with the same distribution function $F(x) = P(X_n \leq x)$. For each $n \geq 1$ put $S_n = \sum_{k=1}^{n} X_k$. The empirical distribution function of $X_1, \ldots, X_n$ is defined as

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} I[X_k \leq x], x \in \mathbb{R},$$

where $I[\cdot \leq x]$ is the indicator function.

Theorem. *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of associated random variables taking values in the set S and having the same distribution function $F(x)$. Assuming*

$$\sum_{n=2}^{\infty} \frac{1}{n^2} Cov(X_n, S_{n-1}) < \infty,$$

we have, as $n \to \infty$,

$$\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \to 0, \quad \textit{almost surely.}$$

Let us observe that the condition used in our Theorem is the same as in [8], therefore our result extends Theorem 2.1 of [8] on a larger class of associated sequences.

2. Proof of the main result and auxiliary lemmas. Let us put $g(x) = (-\frac{2}{\delta}|x| + 1)\, I_{(-\delta/2,\delta/2)}(x)$ and

$$G(x) = \sum_{x_k \le x, x_k \in S} g(x - x_k).$$

$G(x)$ is bounded and absolutely continuous with $|G(x)| \le 1$ and $|G'(x)| \le 2/\delta$, moreover $I[X_n \le x] = G(X_n)$, for $n \in \mathbb{N}$. Therefore, it follows from Lemma 1, that

$$\operatorname{Cov}(I[X_k \le x], I[X_m \le x]) = \operatorname{Cov}(G(X_k), G(X_m)) =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G'(x)G'(y)\,[P(X_k \le x, X_m \le y) - P(X_k \le x)P(X_m \le y)]\,dxdy \le$$

$$\le 4/\delta^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [P(X_k \le x, X_m \le y) - P(X_k \le x)P(X_m \le y)]\,dxdy =$$

$$= 4/\delta^2 \operatorname{Cov}(X_k, X_m), \text{for } k \ne m.$$

By Lemma 2, we get as $n \to \infty$

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} I[X_k \le x] \to F(x), \quad \text{almost surely.}$$

Similarly, taking $\widetilde{G}(x) = \sum_{x_k < x, x_k \in S} g(x - x_k)$ instead of $G(x)$, we prove that

$$F_n(x-0) = \frac{1}{n}\sum_{k=1}^{n} I[X_k < x] \to F(x-0), \text{ almost surely, as } n \to \infty.$$

Now, the proof may be completed as in the i.i.d. case (cf. Chung [5]).

For the sake of completeness we recall two results (cf. Theorem 2.3 of Hao Yu [8] and Theorem 2 of Birkel [2]).

Lemma 1. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be absolutely continuous functions in any finite interval. Then we have, for any random variables X and Y,*

$$\operatorname{Cov}(f(X), g(Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f'(x)g'(y)\Big[P(X \le x, Y \le y) - P(X \le x)P(Y \le y)\Big]\,dxdy.$$

Lemma 2. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of associated random variables with finite variance. Assume*

$$\sum_{n=1}^{\infty} \frac{1}{n^2} Cov(X_n, S_n) < \infty.$$

Then, as $n \to \infty$, we have $(S_n - ES_n)/n \to 0$ almost surely.

References

[1] Bagai, I., B.L.S. Prakasa Rao, *Estimation of the survival function for stationary associated processes*, Statist. Probab. Lett. **12** (1991), 385–391.

[2] Birkel, T., *A note on the strong law of large numbers for positively dependent random variables*, Statist. Probab. Lett. **7** (1988), 17–20.

[3] Birkel, T., *The invariance principle for associated processes*, Stochastic Process. Appl. **27** (1988), 57–71.

[4] Cox, J. T., G. Grimmett, *Central limit theorems for associated random variables and the percolation model*, Ann. Probab. **12** (1984), 514–528.

[5] K.L. Chung, K. L., *A course in probability theory*, Academic Press, New York, 1974.

[6] Dharmadhikari, A. D., M.M. Kuber, *The association in time of a binary semi-Markov process*, Statist. Probab. Lett. **14** (1992).

[7] Esary, J., F. Proschan, D. Walkup, *Association of random variables with applications*, Ann. Math. Statist. **38** (1967), 1466–1474.

[8] Hao Yu, *A Glivenko–Cantelli lemma and weak convergence for empirical processes of associated sequences*, Probab. Theory Related Fields **95** (1993), 357–370.

[9] Hao Yu, *A strong invariance principle for associated sequences*, Ann. Probab. **24** (1996), 2079–2097.

[10] Matuła, P., *A remark on the weak convergence of sums of associated random variables*, Ann. Univ. Mariae Curie-Skłodowska Sect. A **50** (1996), 115–123.

[11] Matuła, P., Z. Rychlik, *The invariance principle for nonstationary sequences of associated random variables*, Ann. Inst. H. Poincaré Probab. Statist. **26** (1990), 387–397.

[12] Newman, C. M., *Normal fluctuation and the FKG inequalities*, Comm. Math. Phys. **74** (1980), 119–128.

[13] Newman, C. M., *Asymptotic independence and limit theorems for positively and negatively dependent random variables*, in Y.L. Tong, ed., Inequalities in Statistics and Probability (IMS, Hayword CA) (1984), 127–140.

Instytut Matematyki UMCS
pl. Marii Curie–Skłodowskiej 1
20–031 Lublin, Poland
e-mail matula@golem.umcs.lublin.pl

received March 20, 1997