Siberian Mathematical Journal, Vol. 47, *No.* 6, *pp.* 975–979, 2006
Original Russian Text Copyright © 2006 *Baklanov E. A.*

THE STRONG LAW OF LARGE NUMBERS FOR *L*-STATISTICS WITH DEPENDENT DATA

E. A. Baklanov

UDC 519.214

Abstract: We prove the strong law of large numbers for the linear combinations of functions of order statistics (L-statistics) based on weakly dependent random variables. We also establish the Glivenko–Cantelli theorem for φ-mixing sequences of identically distributed random variables.

Keywords: L-statistics, stationary ergodic sequences, φ-mixing, Glivenko–Cantelli theorem, strong law of large numbers

1. Let $X_1, X_2, \dots$ be a sequence of random variables with common distribution function F. Let us consider the L-statistic

$$L_n = \frac{1}{n}\sum_{i=1}^{n} c_{ni} h(X_{n:i}), \tag{1}$$

where $X_{n:1} \le \dots \le X_{n:n}$ are the order statistics based on the sample $\{X_i, i \le n\}$, while h is a measurable function called a *kernel*, and c_{ni}, $i = 1, \dots, n$, are some constants called *weights*.

The aim of this article is to establish the strong law of large numbers (SLLN) for the L-statistics (1) based on sequences of weakly dependent random variables. The similar problems were considered in the papers [1] and [2], where the SLLN was proved for the aforementioned L-statistics based on stationary ergodic sequences. For example, in [2] the case of linear kernels ($h(x) = x$) and *asymptotically regular* weights was considered, i.e.

$$c_{ni} = n \int\limits_{(i-1)/n}^{i/n} J_n(t)\,dt, \tag{2}$$

with J_n an integrable function. In addition, the existence of a function J such that for all $t \in (0, 1)$

$$\int\limits_0^t J_n(s)\,ds \to \int\limits_0^t J(s)\,ds$$

was imposed there. The statistics (1) with linear kernels and *regular* weights, i.e. $J_n \equiv J$ in (2), were considered in [1]. In this paper we relax the regularity assumption on c_{ni} and, furthermore, consider the L-statistics (1) based on both stationary ergodic sequences and φ-mixing sequences. We do not require the *monotonicity* of the kernel in (1) either. Note that if h is a monotonic function then the L-statistic (1) can be represented as the statistic

$$\frac{1}{n}\sum_{i=1}^{n} c_{ni} Y_{n:i},$$

based on the sample $\{Y_i = h(X_i), i \le n\}$ (see [3] for more detail).

As an auxiliary result we establish the Glivenko–Cantelli theorem for φ-mixing sequences.

The author was financially supported by a grant of the President of the Russian Federation for Junior Scientists (Grant MK–2061.2005.1), the Russian Foundation for Basic Research (Grant 06–01–00738), and INTAS (Grant 03–51–5018).

Novosibirsk. Translated from *Sibirskiĭ Matematicheskiĭ Zhurnal*, Vol. 47, No. 6, pp. 1199–1204, November–December, 2006. Original article submitted March 23, 2006. Revision submitted May 31, 2006.

2. We start with introducing notations. Let $F^{-1}(t) = \inf\{x : F(x) \geq t\}$ be the quantile function corresponding to the distribution function F and let $U_1, U_2, \ldots$ be a sequence of random variables uniformly distributed on $[0,1]$. Since the joint distributions of the random vectors $(X_{n:1}, \ldots, X_{n:n})$ and $(F^{-1}(U_{n:1}), \ldots, F^{-1}(U_{n:n}))$ coincide, we have

$$L_n \stackrel{d}{=} \frac{1}{n} \sum_{i=1}^{n} c_{ni} H(U_{n:i}),$$

where $H(t) = h(F^{-1}(t))$, and $\stackrel{d}{=}$ stands for equality in distribution. Let us consider a sequence of functions $c_n(t) = c_{ni}$, $t \in ((i-1)/n, i/n]$, $i = 1, \ldots, n$, $c_n(0) = c_{n1}$. It is not difficult to see that in this case

$$L_n = \int_0^1 c_n(t) H(G_n^{-1}(t))\, dt,$$

where G_n^{-1} is the quantile function corresponding to the empirical distribution function G_n based on the sample $\{U_i, i \leq n\}$. We also introduce the following notation:

$$\mu_n = \int_0^1 c_n(t) H(t)\, dt; \qquad C_n(q) = \begin{cases} n^{-1} \sum_{i=1}^{n} |c_{ni}|^q & \text{if} \quad 1 \leq q < \infty, \\ \max_{i \leq n} |c_{ni}| & \text{if} \quad q = \infty. \end{cases}$$

We will use the following conditions on c_{ni} and H:

(i) the function H is continuous on $[0,1]$ and $\sup_{n \geq 1} C_n(1) < \infty$;

(ii) $\mathbf{E}|h(X_1)|^p < \infty$ and $\sup_{n \geq 1} C_n(q) < \infty$ $(1 \leq p < \infty,\ 1/p + 1/q = 1)$.

Assumptions (i) and (ii) guarantee the existence of μ_n. We note also that $C_n(\infty) = \|c_n\|_\infty = \sup_{0 \leq t \leq 1} |c_n(t)|$ and $C_n(q) = \|c_n\|_q^q = \int_0^1 |c_n(t)|^q\, dt$ for $1 \leq q < \infty$.

3. Let us formulate our main statement for stationary ergodic sequences.

Theorem 1. *Let $\{X_n, n \geq 1\}$ be a strictly stationary and ergodic sequence and let either* (i) *or* (ii) *hold. Then, as $n \to \infty$,*

$$L_n - \mu_n \to 0 \quad \text{a.s.} \tag{3}$$

Remark. Let us consider the case of regular weights, i.e. the coefficients c_{ni} are defined by (2) with $J_n = J$. Then

$$L_n = \sum_{i=1}^{n} H(U_{n:i}) \int_{(i-1)/n}^{i/n} J(t)\, dt = \int_0^1 J(t) H(G_n^{-1}(t))\, dt.$$

Hence, assuming $c_n(t) = J(t)$ in Theorem 1, we have

$$L_n \to \int_0^1 J(t) H(t)\, dt \quad \text{a.s.}$$

The convergence $\mu_n \to \mu$, $|\mu| < \infty$, implies that $L_n \to \mu$ a.s. In particular, if $c_n(t) \to c(t)$ uniformly in $t \in [0,1]$ then

$$\mu_n \to \int_0^1 c(t) H(t)\, dt.$$

Dropping the requirement that the coefficients c_{ni} are regular, it is easy to construct an example in which the assumptions of Theorem 1 are satisfied but the sequence $c_n(t)$ does not converge in any reasonable sense to a limit function. For simplicity, let $h(x) = x$ and let X_1 be uniformly distributed on [0, 1]. Put $c_{ni} = (i-1)\delta_n$ for $1 \le i \le k$ and $c_{ni} = (2k-i)\delta_n$ for $k+1 \le i \le 2k$, $k = k(n) = [n^{1/2}]$, $\delta_n = n^{-1/2}$. Thus, the function $c_n(t)$ is defined on the interval $[0, 2k/n]$. On the remaining part of $[0,1]$ we extend $c_n(t)$ periodically with period $2k/n$: $c_n(t) = c_n(t - 2k/n)$, $2k/n \le t \le 1$ (also see [3, p. 138]). Note that $0 \le c_n(t) \le 1$. One can show that in this case $\mu_n \to 1/4$. Thus, the assumptions of Theorem 1 are satisfied and so $L_n \to 1/4$ a. s.

4. We now formulate our main statement for mixing sequences. Let us define the mixing coefficients:

$$\varphi(n) = \sup_{k\ge 1} \sup\{|\mathbf{P}(B|A) - \mathbf{P}(B)| : A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+n}^{\infty}, \mathbf{P}(A) > 0\},$$

where $\mathcal{F}_1^k$ and $\mathcal{F}_{k+n}^{\infty}$ denote the σ-fields generated by $\{X_i, 1 \le i \le k\}$ and $\{X_i, i \ge k+n\}$ respectively. The sequence $\{X_i, i \ge 1\}$ is called *φ-mixing* (uniform mixing) if $\varphi(n) \to 0$ as $n \to \infty$.

Theorem 2. *Let $\{X_n, n \ge 1\}$ be a φ-mixing sequence of identically distributed random variables such that*

$$\sum_{n\ge 1} \varphi^{1/2}(2^n) < \infty, \tag{4}$$

and let one of the conditions (i) *or* (ii) *hold. Then* (3) *remains true.*

The proof of Theorem 2 essentially uses the result of the Lemma 1 below. The statement (a) of Lemma 1 is the SLLN for φ-mixing sequences. The statement (b), a Glivenko–Cantelli-type result for φ-mixing sequences, is of interest in its own right. We note that neither in Theorem 2 nor in Lemma 1 we assume the stationarity of $\{X_n\}$.

Lemma 1. *Let $\{X_n, n \ge 1\}$ be a φ-mixing sequence of identically distributed random variables such that the statement* (4) *holds. Then*

(a) *for every function f such that $\mathbf{E}|f(X_1)| < \infty$,*

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \mathbf{E}f(X_1) \quad \text{a. s.}; \tag{5}$$

(b)

$$\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0 \quad \text{a. s.}, \tag{6}$$

where F_n is the empirical distribution function based on the sample $\{X_i, i \le n\}$.

5. We now proceed to the proof of Theorem 1.

Lemma 2. *Let H be continuous on* $[0,1]$. *Then*

$$\sup_{0\le t\le 1} |H(G_n^{-1}(t)) - H(t)| \to 0 \quad \text{a. s.} \tag{7}$$

Proof of Lemma 2. Using the equality

$$\sup_{0\le t\le 1} |G_n^{-1}(t) - t| = \sup_{0\le t\le 1} |G_n(t) - t|$$

(for example, see [4, p. 95]) and the Glivenko–Cantelli theorem for stationary ergodic sequences, we deduce

$$\sup_{0\le t\le 1} |G_n^{-1}(t) - t| \to 0 \quad \text{a. s.},$$

i.e. $G_n^{-1}(t) \to t$ a. s. uniformly in $t \in [0,1]$ as $n \to \infty$. Since the function H is uniformly continuous on the compact set $[0,1]$, it follows that $H(G_n^{-1}(t)) \to H(t)$ a. s. uniformly in $t \in [0,1]$. This concludes the proof.

Let the condition (i) hold. Now, by Lemma 2,

$$|L_n - \mu_n| \le \int_0^1 |c_n(t)||H(G_n^{-1}(t)) - H(t)|\,dt \le C_n(1) \sup_{0\le t\le 1} |H(G_n^{-1}(t)) - H(t)| \to 0 \quad \text{a. s.}$$

Consequently, the proof of Theorem 1 in the first case is complete.

Lemma 3. *Let* $\mathbf{E}|h(X_1)|^p < \infty$. *Then*

$$\int\limits_0^1 |H(G_n^{-1}(t)) - H(t)|^p \, dt \to 0 \quad \text{a.s.} \tag{8}$$

Proof of Lemma 3. Note first that the set of all continuous functions on $[0,1]$ is everywhere dense in $L_p[0,1]$, $1 \le p < \infty$. Therefore, for all $\varepsilon > 0$ and all $f \in L_p[0,1]$ there exists a continuous function f_ε on $[0,1]$ such that $\int_0^1 |f(t) - f_\varepsilon(t)|^p \, dt < \varepsilon$. Since $\mathbf{E}|h(X_1)|^p = \int_0^1 |H(t)|^p \, dt < \infty$, this implies that there exists a continuous function H_ε on $[0,1]$ such that

$$\int\limits_0^1 |H(t) - H_\varepsilon(t)|^p dt < \varepsilon/2.$$

Further,

$$\int\limits_0^1 |H(G_n^{-1}(t)) - H(t)|^p \, dt \le 3^{p-1} \int\limits_0^1 |H(t) - H_\varepsilon(t)|^p \, dt$$

$$+3^{p-1} \int\limits_0^1 |H(G_n^{-1}(t)) - H_\varepsilon(G_n^{-1}(t))|^p \, dt + 3^{p-1} \int\limits_0^1 |H_\varepsilon(G_n^{-1}(t)) - H_\varepsilon(t)|^p \, dt. \tag{9}$$

From Lemma 2 it follows that $H_\varepsilon(G_n^{-1}(t)) \to H_\varepsilon(t)$ a.s. uniformly in t as $n \to \infty$. Hence, the last integral on the right-hand side of (9) converges to zero a.s. as $n \to \infty$. We now consider the second integral. By the ergodic theorem for stationary sequences,

$$\int\limits_0^1 |H(G_n^{-1}(t)) - H_\varepsilon(G_n^{-1}(t))|^p \, dt$$

$$= \frac{1}{n} \sum_{i=1}^n |H(U_i) - H_\varepsilon(U_i)|^p \to_{\text{a.s.}} \mathbf{E}|H(U_1) - H_\varepsilon(U_1)|^p$$

$$= \int\limits_0^1 |H(t) - H_\varepsilon(t)|^p \, dt < \varepsilon/2.$$

Consequently,

$$\limsup_{n\to\infty} \int\limits_0^1 |H(G_n^{-1}(t)) - H(t)| \, dt < 3^{p-1}\varepsilon \quad \text{a.s.}$$

Since ε is arbitrary, we obtain (8).

Assume that (ii) holds. By Hölder's inequality, we infer

$$|L_n - \mu_n| \le C_n^{1/q}(q) \biggl(\int\limits_0^1 |H(G_n^{-1}(t) - H(t))|^p \, dt \biggr)^{1/p} \quad \text{for } p > 1$$

and

$$|L_n - \mu_n| \le C_n(\infty) \int\limits_0^1 |H(G_n^{-1}(t) - H(t))| \, dt \quad \text{for } p = 1.$$

The statement (3) follows from Lemma 3. This completes the proof of Theorem 1.

6. We now prove Lemma 1. It is noted in [5, p. 353] that if the φ-mixing sequence has the mixing coefficient $\varphi(n)$, then for each measurable function f the sequence $\{f(X_n), n \geq 1\}$ is also φ-mixing and its mixing coefficient is less than or equal to $\varphi(n)$.

Hence, the condition (4) holds for mixing coefficients of $\{f(X_n), n \geq 1\}$. The statement (5) follows from the SLLN for φ-mixing sequences (see [6, p. 200]).

The statement (6) is an immediate corollary of (5) and the classical Glivenko–Cantelli theorem.

The proof of Theorem 2 is similar to the proof of Theorem 1. Indeed, (7) follows from the Glivenko–Cantelli theorem (6); using the SLLN (5), we arrive at (8). The proof of Theorem 2 is thus complete.

Acknowledgments. The author is grateful to the referee for a careful reading of the manuscript and for the helpful remarks that improve the original text.

References

1. *Aaronson J., Burton R., Dehling H., Gilat D., Hill T., and Weiss B.,* "Strong laws for L- and U-statistics," Trans. Amer. Math. Soc., **348**, No. 7, 2845–2866 (1996).
2. *Gilat D. and Helmers R.,* "On strong laws for generalized L-statistics with dependent data," Comment. Math. Univ. Carolinae, **38**, No. 1, 187–192 (1997).
3. *Baklanov E. A. and Borisov I. S.,* "Probability inequalities and limit theorems for generalized L-statistics," Lithuanian Math. J., **43**, No. 2, 125–140 (2003).
4. *Shorack G. R. and Wellner J. A.,* Empirical Processes with Applications to Statistics, John Wiley, New York (1986).
5. *Korolyuk V. S. and Borovskikh Yu. V.,* The Theory of U-Statistics [in Russian], Naukova Dumka, Kiev (1989).
6. *Lin Z. Y. and Lu C. R.,* Limit Theory for Mixing Dependent Random Variables, Kluwer, Beijing (1996).

Evgeny Baklanov
Novosibirsk State University, Novosibirsk, Russia
E-mail address: baklanov@mmf.nsu.ru