Сибирский математический журнал
Ноябрь—декабрь, 2006. Том 47, № 6

УДК 519.214

УСИЛЕННЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ ДЛЯ L–СТАТИСТИК, ПОСТРОЕННЫХ ПО ЗАВИСИМЫМ НАБЛЮДЕНИЯМ

Е. А. Бакланов

Аннотация: Доказан усиленный закон больших чисел для линейных комбинаций функций от порядковых статистик (L-статистик), построенных по слабо зависимым случайным величинам. Также доказана теорема Гливенко — Кантелли для последовательностей одинаково распределенных случайных величин с логарифмически убывающими коэффициентами φ-перемешивания.

Ключевые слова: L-статистики, стационарные последовательности, φ-перемешивание, эргодичность, теорема Гливенко — Кантелли, усиленный закон больших чисел.

1. Пусть $X_1, X_2, \dots$ — стационарная (в узком смысле) последовательность случайных величин с общей функцией распределения F. Рассмотрим L-статистику

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

где $X_{n:1} \leq \ldots \leq X_{n:n}$ — порядковые статистики, построенные по выборке $\{X_i, i \leq n\}$, h — измеримая функция, называемая *ядром*, c_{ni}, $i = 1, \dots, n$, — некоторые постоянные, называемые *весами*.

В настоящей работе изучается усиленный закон больших чисел (УЗБЧ) для L-статистик вида (1), построенных по последовательностям слабо зависимых случайных величин. Аналогичные вопросы изучались в работах [1, 2], где УЗБЧ доказан для указанных L-статистик, построенных по стационарным эргодическим последовательностям. Так, например, в [2] изучались статистики (1) с линейным ядром ($h(x) = x$) и *асимптотически регулярными* весами, т. е.

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

где J_n — некоторые интегрируемые функции. Дополнительно в [2] предполагалось существование функции J такой, что для всех $t \in (0, 1)$

$$\int\limits_0^t J_n(s)\, ds \to \int\limits_0^t J(s)\, ds \quad \text{при } n \to \infty.$$

Работа выполнена при финансовой поддержке Совета по грантам президента РФ и государственной поддержке ведущих научных школ (код проекта МК–2061.2005.1), Российского фонда фундаментальных исследований (код проекта 06–01–00738) и INTAS (грант 03–51–5018).

В [1] также рассматривались статистики (1) с линейным ядром и *регулярными* весами, т. е. $J_n \equiv J$ в (2). В настоящей работе ослаблены ограничения на коэффициенты c_{ni} (не предполагается регулярности весов) и рассматриваются L-статистики (1), построенные как по стационарным эргодическим последовательностям, так и по последовательностям, удовлетворяющим условию φ-перемешивания. Кроме того, мы не требуем монотонности ядра в (1). В связи с этим отметим, что если h — монотонная функция, то L-статистика (1) представима в виде статистики

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

построенной по выборке $\{Y_i = h(X_i), i \le n\}$ (подробнее см. [3]).

В качестве вспомогательного результата получена теорема Гливенко — Кантелли для последовательностей с φ-перемешиванием.

2. Введем основные обозначения. Пусть $F^{-1}(t) = \inf\{x : F(x) \ge t\}$ — квантильное преобразование функции распределения F, и пусть $U_1, U_2, \dots$ — стационарная последовательность случайных величин с равномерным на отрезке $[0,1]$ распределением. Так как совместные распределения векторов $(X_{n:1}, \dots, X_{n:n})$ и $(F^{-1}(U_{n:1}), \dots, F^{-1}(U_{n:n}))$ совпадают, то

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

где $H(t) = h(F^{-1}(t))$, а символом $\stackrel{d}{=}$ обозначено равенство распределений. Рассмотрим последовательность функций $c_n(t) = c_{ni}$, $t \in ((i-1)/n, i/n]$, $i = 1, \dots, n$, $c_n(0) = c_{n1}$. Тогда, как нетрудно заметить, статистика L_n допускает следующее интегральное представление:

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

где G_n^{-1} — квантильное преобразование эмпирической функции распределения G_n, построенной по выборке $\{U_i, i \le n\}$.

Введем обозначения:

$$\mu_n = \int_0^1 c_n(t) H(t)\, dt, \quad C_n(q) = \begin{cases} n^{-1}\sum\limits_{i=1}^{n} |c_{ni}|^q & \text{при } 1 \le q < \infty, \\ \max\limits_{i \le n} |c_{ni}| & \text{при } q = \infty. \end{cases}$$

Далее мы будем использовать следующие условия на веса c_{ni} и функцию H:

(i) функция H непрерывна на $[0,1]$ и $\sup\limits_{n \ge 1} C_n(1) < \infty$.

(ii) $\mathbf{E}|h(X_1)|^p < \infty$ и $\sup\limits_{n \ge 1} C_n(q) < \infty$ $(1 \le p < \infty,\ 1/p + 1/q = 1)$.

Условия (i) и (ii) гарантируют существование величины μ_n. Отметим также, что $C_n(\infty) = \|c_n\|_\infty = \sup\limits_{0 \le t \le 1} |c_n(t)|$ и $C_n(q) = \|c_n\|_q^q = \int\limits_0^1 |c_n(t)|^q\, dt$ при $1 \le q < \infty$.

3. Сформулируем наше основное утверждение для стационарных эргодических последовательностей.

Теорема 1. *Пусть* $\{X_n, n \geq 1\}$ *— стационарная эргодическая последовательность, и пусть выполнено одно из условий* (i) *или* (ii). *Тогда*

$$L_n - \mu_n \to 0 \quad \text{п. н. при} \quad n \to \infty. \tag{3}$$

Доказательство теоремы 1 будет приведено в п. 5.

Замечание. Рассмотрим случай регулярных весов, т. е. случай, когда коэффициенты c_{ni} определяются соотношением (2) при $J_n = J$. Тогда

$$L_n = \sum_{i=1}^{n} H(U_{n:i}) \int\limits_{(i-1)/n}^{i/n} J(t)\,dt = \int\limits_0^1 J(t) H\big(G_n^{-1}(t)\big)\,dt.$$

Следовательно, положив $c_n(t) = J(t)$ в теореме 1, получим

$$L_n \to \int\limits_0^1 J(t) H(t)\,dt \quad \text{п. н. при } n \to \infty.$$

Также отметим, что если $\mu_n \to \mu$, $|\mu| < \infty$, то, очевидно, $L_n \to \mu$ п. н. В частности, если $c_n(t) \to c(t)$ равномерно по $t \in [0,1]$, то $\mu_n \to \int\limits_0^1 c(t) H(t)\,dt$.

Если отказаться от условия регулярности коэффициентов c_{ni}, то нетрудно построить пример последовательности $c_n(t)$, которая не сходится ни в каком разумном смысле к предельной функции, но удовлетворяет условиям теоремы 1. Пусть для простоты $h(x) = x$ и X_1 имеет равномерное на отрезке $[0,1]$ распределение. Положим $c_{ni} = (i-1)\delta_n$ при $1 \leq i \leq k$ и $c_{ni} = (2k-i)\delta_n$ при $k+1 \leq i \leq 2k$, $k = k(n) = [n^{1/2}]$, $\delta_n = n^{-1/2}$. Таким образом, функция $c_n(t)$ задана на отрезке $[0, 2k/n]$. На оставшуюся часть отрезка $[0,1]$ функцию $c_n(t)$ продолжаем так, чтобы она была периодической с периодом $2k/n$: $c_n(t) = c_n(t - 2k/n)$, $2k/n \leq t \leq 1$ (см. также [3, с. 165]). Отметим, что $0 \leq c_n(t) \leq 1$. Можно показать, что в данном случае $\mu_n \to 1/4$. Тем самым выполнены условия теоремы 1, и, следовательно, $L_n \to 1/4$ п. н.

4. Теперь приведем наше основное утверждение для последовательностей с перемешиванием. Введем коэффициенты перемешивания

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

где $\mathscr{F}_1^k$ и $\mathscr{F}_{k+n}^\infty$ — σ-алгебры, порожденные $\{X_i, 1 \leq i \leq k\}$ и $\{X_i, i \geq k+n\}$ соответственно. Последовательность $\{X_n, n \geq 1\}$ удовлетворяет условию *равномерно сильного перемешивания* (р.с.п.), если $\varphi(n) \to 0$ при $n \to \infty$. Последовательности, удовлетворяющие условию р.с.п., также будем назвать *последовательностями с* φ*-перемешиванием.*

Теорема 2. *Пусть* $\{X_n, n \geq 1\}$ *— последовательность одинаково распределенных случайных величин с* φ*-перемешиванием, для которой*

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

Пусть также выполнено одно из условий (i) *или* (ii). *Тогда имеет место утверждение* (3).

Доказательство теоремы 2 существенно использует результат, приведенный ниже в лемме 1. П. (a) леммы 1 содержит УЗБЧ для последовательностей с φ-перемешиванием. П. (b) леммы 1 — это вариант теоремы Гливенко — Кантелли для последовательностей с φ-перемешиванием, представляющий самостоятельный интерес. Отметим, что в теореме 2 и в лемме 1 мы не предполагаем стационарности последовательности $\{X_n\}$.

Лемма 1. *Пусть $\{X_n, n \geq 1\}$ — последовательность одинаково распределенных случайных величин с φ-перемешиванием, для которой выполнено условие* (4). *Тогда*

(a) *для любой функции f такой, что $\mathbf{E}|f(X_1)| < \infty$,*

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

(b) *имеет место сходимость*

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

где F_n — эмпирическая функция распределения, построенная по выборке $\{X_i, i \leq n\}$.

5. Приступим к доказательству теоремы 1.

Лемма 2. *Пусть функция H непрерывна на $[0, 1]$. Тогда*

$$\sup_{0\leq t\leq 1} \bigl|H\bigl(G_n^{-1}(t)\bigr) - H(t)\bigr| \to 0 \quad \text{п. н.} \tag{7}$$

Доказательство леммы 2. Так как

$$\sup_{0\leq t\leq 1} \bigl|G_n^{-1}(t) - t\bigr| = \sup_{0\leq t\leq 1} |G_n(t) - t|$$

(см., например, [4, с. 95]), в силу теоремы Гливенко — Кантелли для стационарных эргодических последовательностей имеем

$$\sup_{0\leq t\leq 1} \bigl|G_n^{-1}(t) - t\bigr| \to 0 \quad \text{п. н.},$$

т. е. $G_n^{-1}(t) \to t$ п. н. равномерно по $t \in [0, 1]$ при $n \to \infty$. Так как функция H равномерно непрерывна на компакте $[0, 1]$, то $H\bigl(G_n^{-1}(t)\bigr) \to H(t)$ равномерно по $t \in [0, 1]$, что и требовалось доказать.

Пусть выполнено условие (i). Тогда из леммы 2 получаем

$$|L_n - \mu_n| \leq \int_0^1 |c_n(t)|\bigl|H\bigl(G_n^{-1}(t)\bigr) - H(t)\bigr|\,dt \\ \leq C_n(1) \sup_{0\leq t\leq 1} \bigl|H\bigl(G_n^{-1}(t)\bigr) - H(t)\bigr| \to 0 \quad \text{п. н.}$$

Итак, утверждение теоремы 1 в случае (i) доказано.

Лемма 3. *Пусть $\mathbf{E}|h(X_1)|^p < \infty$. Тогда*

$$\int_0^1 \bigl|H\bigl(G_n^{-1}(t)\bigr) - H(t)\bigr|^p\,dt \to 0 \quad \text{п. н. при } n \to \infty. \tag{8}$$

Доказательство леммы 3. Так как множество непрерывных на отрезке $[0,1]$ функций всюду плотно в $L_p[0,1]$, $1 \le p < \infty$, для любого $\varepsilon > 0$ и для любой функции $f \in L_p[0,1]$ существует непрерывная на $[0,1]$ функция f_ε такая, что

$$\int_0^1 |f(t) - f_\varepsilon(t)|^p\,dt < \varepsilon.$$

Значит, в силу условия

$$\mathbf{E}|h(X_1)|^p = \int_0^1 |H(t)|^p\,dt < \infty$$

существует непрерывная на $[0,1]$ функция H_ε такая, что

$$\int_0^1 |H(t) - H_\varepsilon(t)|^p\,dt < \varepsilon/2.$$

Далее,

$$\int_0^1 \bigl|H\bigl(G_n^{-1}(t)\bigr) - H(t)\bigr|^p\,dt \le 3^{p-1}\int_0^1 |H(t) - H_\varepsilon(t)|^p\,dt$$
$$+\,3^{p-1}\int_0^1 \bigl|H\bigl(G_n^{-1}(t)\bigr) - H_\varepsilon\bigl(G_n^{-1}(t)\bigr)\bigr|^p\,dt + 3^{p-1}\int_0^1 \bigl|H_\varepsilon\bigl(G_n^{-1}(t)\bigr) - H_\varepsilon(t)\bigr|^p\,dt. \quad (9)$$

Отметим, что $H_\varepsilon\bigl(G_n^{-1}(t)\bigr) \to H_\varepsilon(t)$ п. н. равномерно по t при $n \to \infty$ в силу леммы 2. Следовательно, последний интеграл в правой части (9) сходится к нулю п. н. при $n \to \infty$. Рассмотрим второй интеграл. В силу эргодической теоремы для стационарных последовательностей имеем

$$\int_0^1 \bigl|H\bigl(G_n^{-1}(t)\bigr) - H_\varepsilon\bigl(G_n^{-1}(t)\bigr)\bigr|^p\,dt$$
$$= \frac{1}{n}\sum_{i=1}^n |H(U_i) - H_\varepsilon(U_i)|^p \to_{\text{п. н.}} \mathbf{E}|H(U_1) - H_\varepsilon(U_1)|^p = \int_0^1 |H(t) - H_\varepsilon(t)|^p\,dt < \varepsilon/2.$$

Таким образом,

$$\limsup_{n\to\infty}\int_0^1 \bigl|H\bigl(G_n^{-1}(t)\bigr) - H(t)\bigr|\,dt < 3^{p-1}\varepsilon \quad \text{п. н.}$$

Отсюда в силу произвольности ε получаем (8). Лемма доказана.

Вернемся к доказательству теоремы 1.

Пусть выполнено условие (ii). Применяя неравенство Гёльдера, имеем

$$|L_n - \mu_n| \le C_n^{1/q}(q)\left(\int_0^1 \bigl|H\bigl(G_n^{-1}(t) - H(t)\bigr)\bigr|^p\,dt\right)^{1/p} \qquad \text{при } p > 1$$

и

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

Утверждение (3) вытекает из леммы 3. Теорема 1 полностью доказана.

6. Докажем лемму 1. В [5, с. 353] отмечено, что если последовательность $\{X_n, n \ge 1\}$ удовлетворяет условию р.с.п. с коэффициентом перемешивания $\varphi(n)$, то для любой измеримой функции f последовательность $\{f(X_n), n \ge 1\}$ также удовлетворяет условию р.с.п. с коэффициентом перемешивания, не большим, чем $\varphi(n)$. Следовательно, для коэффициентов перемешивания последовательности $\{f(X_n), n \ge 1\}$ также выполняется условие (4). Утверждение (5) следует из УЗБЧ для последовательностей с φ-перемешиванием (см. [6, с. 200]).

Утверждение (6) есть непосредственное следствие (5) и классической теоремы Гливенко — Кантелли.

Доказательство теоремы 2 отличается от доказательства теоремы 1 лишь тем, что при выводе утверждения (7) используется теорема Гливенко — Кантелли (6), а при выводе (8) — УЗБЧ (5). Таким образом, теорема 2 доказана.

Автор выражает благодарность рецензенту за внимательное отношение к работе и ценные замечания, которые позволили существенно улучшить первоначальный текст.

ЛИТЕРАТУРА

1. *Aaronson J., Burton R., Dehling H., Gilat D., Hill T., Weiss B.* Strong laws for L- and U-statistics // Trans. Amer. Math. Soc. 1996. V. 348, N 7. P. 2845–2866.
2. *Gilat D., Helmers R.* On strong laws for generalized L-statistics with dependent data // Comment. Math. Univ. Carolinae. 1997. V. 38, N 1. P. 187–192.
3. *Бакланов Е. А., Борисов И. С.* Вероятностные неравенства и предельные теоремы для обобщенных L-статистик // Литов. мат. сб. 2003. Т. 43, № 2. С. 149–168.
4. *Shorack G. R., Wellner J. A.* Empirical processes with applications to statistics. New York: John Wiley, 1986.
5. *Королюк В. С., Боровских Ю. В.* Теория U-статистик. Киев: Наук. думка, 1989.
6. *Lin Z. Y., Lu C. R.* Limit theory for mixing dependent random variables. Beijing: Kluwer, 1996.

Статья поступила 23 марта 2006 г., окончательный вариант — 31 мая 2006 г.

Бакланов Евгений Анатольевич
Новосибирский гос. университет,
ул. Пирогова, 2, Новосибирск 630090
baklanov@mmf.nsu.ru