S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru

Том 11, стр. 857–862 (2014) УДК 519.234.7
MSC 62G30

О СУММЕ ДИСПЕРСИЙ ПОРЯДКОВЫХ СТАТИСТИК ДЛЯ ВЫБОРОК ЗАВИСИМЫХ НАБЛЮДЕНИЙ

И.С. БОРИСОВ

Abstract. We study asymptotic behavior of the sum of the variances of all order statistics based on samples of dependent observations when the sample size tends to infinity.

Keywords: Hoeffding's theorem, order statistics, dependent observations.

1. Введение

Пусть $X_1, X_2, \ldots$ — последовательность, вообще говоря, зависимых, но одинаково распределенных невырожденных случайных величин с общей функцией распределения $F(t)$ и конечным вторым моментом. Рассмотрим вариационный ряд $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$, построенный по выборке $X_1, X_2, \ldots, X_n$. В настоящей работе получена верхняя оценка скорости роста суммы $\sum_{j=1}^n \mathbf{D}(X_{(j)})$ при $n \to \infty$. В случае независимых наблюдений асимптотику $o(n)$ для указанной суммы можно извлечь из [1], где доказан следующий результат.

Теорема 1. *Пусть $X_1, X_2, \ldots, X_n$ – независимые одинаково распределенные случайный величины, $g(t)$ — непрерывная функция на R, а $h(t)$ — выпуклая неотрицательная функция, причем $|g(t)| \le h(t)$ при всех $t \in R$. Тогда если $\mathbf{E}h(X_1) < \infty$, то*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n g(\mathbf{E}X_{(j)}) = \mathbf{E}g(X_1). \tag{1}$$

Отметим, что, во-первых, моментные условия теоремы влекут за собой существование среднего рассматриваемых случайных величин, и во-вторых, в

Borisov, I.S., On the sum of the variances of the order statistics for samples of dependent observations.

Работа поддержана РФФИ (коды проектов: 13-01-12415 офи-м, 13-01-00511 и 14-01-00220).

Поступила 29 июля 2014 г., опубликована 21 ноября 2014 г.

силу перестановочности слагаемых справедливо следующее тождество по n:

$$\mathbf{E}g(X_1) = \frac{1}{n}\sum_{j=1}^{n}\mathbf{E}g(X_{(j)}).$$

Тогда при $h(t) = g(t) = t^2$ предельное соотношение (1) влечет за собой

Следствие 1. *Если в условиях предыдущей теоремы* $\mathbf{E}X_1^2 < \infty$, *то*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\mathbf{D}X_{(j)} = 0. \tag{2}$$

В [2] доказано, что предельное соотношение (1), а стало быть, и (2) останутся верными и в случае зависимых наблюдений, если только при $n \to \infty$

$$\frac{1}{n^2}\sum_{i,j=1}^{n}\mathbf{P}(\max\{X_i, X_j\} < t) \to F^2(t) \tag{3}$$

для каждого фиксированного $t \in R$.

В настоящей работе получено уточнение соотношения (2) при выполнении (3) и более сильном моментном условии

$$E\,|x_1|^{2+r} < \infty, \quad r > 0,$$

а также для ограниченных наблюдений $\{X_k\}$.

2. Основные результаты

Пусть $F_n(t) := \frac{1}{n}\sum_{j=1}^{n} I\{X_j < t\}$ — эмпирическая функция распределения, построенная по выборке $X_1, \ldots, X_n$, где $I(\cdot)$ – индикатор события. Обозначим

$$\delta_n(t) := \mathbf{E}|F_n(t) - F(t)|.$$

Справедлива следующая элементарная двусторонняя оценка:

$$\mathbf{E}\left(F_n(t) - F(t)\right)^2 \le \delta_n(t) \le \left(\mathbf{E}\left(F_n(t) - F(t)\right)^2\right)^{1/2}. \tag{4}$$

Отметим, что левое неравенство является следствием оценки $|F_n(t) - F(t)| \le 1$. Поскольку

$$\mathbf{E}\left(F_n(t) - F(t)\right)^2 = \frac{1}{n^2}\sum_{i,j=1}^{n}\mathbf{E}I\{X_i < t\}I\{X_j < t\} - F^2(t)$$

$$= \frac{1}{n^2}\sum_{i,j=1}^{n}\mathbf{P}(\max\{X_i, X_j) < t) - F^2(t), \tag{5}$$

то из (4) следует, что для того чтобы $\lim_{n\to\infty}\delta_n(t) = 0$ при всех $t \in R$ необходимо и достаточно выполнение условия (3).

Приведем несколько простых оценок величины $\delta_n(t)$, непосредственно вытекающих из (4) и (5), для той или иной формы зависимости наблюдений.

(i) Будем говорить, что последовательность $\{X_i\}$ состоит из *попарно отрицательно ассоциированных* случайных величин (что слабее *отрицательной ассоциированности*, см. [3]), если для любых невозрастающих функций f и g (или, что эквивалентно, любых неубывающих) случайные величины $f(X_i)$ и

$g(X_j)$ при всех $i \neq j$ и условии существования вторых моментов будут *неположительно коррелированными*. Беря в качестве f и g одну и ту же индикаторную функцию, мы в этом случае получаем из (4) и (5) оценку

$$\delta_n^2(t) \leq \frac{1}{n} F(t)(1 - F(t)). \tag{6}$$

Ясно, что попарно независимые наблюдения вкладываются в рассмотренную схему.

(ii) Пусть последовательность случайных величин $\{X_i\}$ удовлетворяет условию *сильного перемешивания* с коэффициентом $\alpha(k)$ (см. [4]), для которого при всех n справедлива оценка

$$\sum_{i=0}^{n} \alpha(k) \leq K n^{\varepsilon}, \tag{7}$$

где $\varepsilon \in [0, 1)$ и положительная постоянная K не зависят от n. Тогда из (4) и (5) следует оценка

$$\delta_n^2(t) \leq \frac{2K}{n^{1-\varepsilon}}. \tag{8}$$

(iii) Если последовательность случайных величин $\{X_i\}$ удовлетворяет условию *равномерно сильного перемешивания* с коэффициентом $\varphi(k)$ (см. [4]), для которого при всех n имеет место аналогичное (7) неравенство, то

$$\delta_n^2(t) \leq \frac{2K}{n^{1-\varepsilon}} \min\{F(t), 1 - F(t)\}. \tag{9}$$

При выводе оценки (9) мы учли, что ковариация между индикаторами двух событий совпадает с ковариацией между индикаторами их дополнений.

Отметим, что если ряды из α- или φ-коэффициентов в (7) сходятся (например, для m-зависимых последовательностей), то в оценках (8) и (9) нужно положить $\varepsilon = 0$. Кроме того, важно подчеркнуть, что в случаях (ii) и (iii) по аналогии со случаем (i) достаточно требовать лишь *попарное* "перемешивание" наблюдений.

Далее, для любой функции распределения $G(t)$, $t \in R$, символом $G^{-1}(s)$ будем обозначать ее квантильное преобразование:

$$G^{-1}(s) := \inf\{t \in R : G(t) \geq s\}, \ \ s \in (0, 1).$$

Отметим, что $G^{-1}(s)$ – монотонно неубывающая непрерывная слева функция (если рассматривается непрерывная слева версия функции распределения $G(t)$). Значения $G^{-1}(0)$ и $G^{-1}(1)$ определим как соответствующие односторонние пределы (конечные или бесконечные).

Сначала исследуем случай ограниченных наблюдений.

Теорема 2. *Если* $|X_1| \leq c$ *п. н., то*

$$\sum_{j=1}^{n} \mathbf{D} X_{(j)} \leq 2cn \int_{-c}^{c} \delta_n(t) dt. \tag{10}$$

Доказательство. Прежде всего отметим следующее ключевое представление порядковых статистик: $X_{(j)} = F_n^{-1}(s)$ при всех $s \in ((j-1)/n, j/n]$, $j = 1, \ldots, n$.

Тогда вновь использую отмеченную выше перестановочность слагаемых, получаем

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{D}X_{(j)} = \mathbf{E}X_1^2 - \frac{1}{n}\sum_{j=1}^{n}\left(\mathbf{E}X_{(j)}\right)^2 = \int_{-c}^{c} t^2 dF(t) - \frac{1}{n}\sum_{j=1}^{n}\left(\mathbf{E}F_n^{-1}(\frac{j}{n})\right)^2$$

$$= \int_0^1 \left(F^{-1}(s)\right)^2 ds - \int_0^1 \left(\mathbf{E}F_n^{-1}(s)\right)^2 ds$$

$$= \int_0^1 \left(F^{-1}(s) + \mathbf{E}F_n^{-1}(s)\right)\left(F^{-1}(s) - \mathbf{E}F_n^{-1}(s)\right) ds \le 2c\int_0^1 \left|F^{-1}(s) - \mathbf{E}F_n^{-1}(s)\right| ds$$

$$\le 2c\mathbf{E}\int_0^1 \left|F^{-1}(s) - F_n^{-1}(s)\right| ds = 2c\mathbf{E}\int_{-c}^{c} |F(t) - F_n(t)|\, dt = 2c\int_{-c}^{c} \delta_n(t)dt,$$

что и требовалось доказать. Отметим, что предпоследнее равенство, связывающее интегральные расстояния между функциями распределения и их квантильными преобразованиями, по-видимому, впервые было отмечено Ю. В. Прохоровым в [5]. □

Следствие 2. *В условиях теоремы* 2 *имеют место следующие оценки:*

1) *в случае* **(i)**

$$\sum_{j=1}^{n}\mathbf{D}X_{(j)} \le 2c\sqrt{n}\int_{-c}^{c}\sqrt{F(t)(1-F(t))}dt; \tag{11}$$

2) *в случае* **(ii)**

$$\sum_{j=1}^{n}\mathbf{D}X_{(j)} \le 4\sqrt{2K}c^2 n^{(1+\varepsilon)/2}; \tag{12}$$

3) *в случае* **(iii)**

$$\sum_{j=1}^{n}\mathbf{D}X_{(j)} \le 2c\sqrt{2K}n^{(1+\varepsilon)/2}\int_{-c}^{c}\sqrt{\min\{F(t), 1-F(t)\}}dt. \tag{13}$$

Теперь рассмотрим случай неограниченных случайных величин X_i.

Теорема 3. *Если* $E|X_1|^2 < \infty$, *то*

$$\sum_{j=1}^{n}\mathbf{D}X_{(j)} \le n\inf_{c>0}\left\{2c\int_{-c}^{c}\delta_n(t)dt + 4\mathbf{E}X_1^2 I\{|X_1| > c\}\right\}. \tag{14}$$

Доказательство. Введем в рассмотрение функцию

$$h_c(t) = \begin{cases} c, & \text{если} \quad t > c, \\ t, & \text{если} \quad -c \le t \le c\ , \\ -c, & \text{если} \quad t < -c, \end{cases}$$

где $c > 0$ выбрано произвольно. Теперь построим срезки исходных выборочных наблюдений по формуле $Y_j = h_c(X_j)$, $j = 1, \dots, n$. Очевидно, случайные величины Y_j одинаково распределены и $|Y_1| \le c$ п. н., т. е. выполнены условия теоремы 2.

Обозначим через $F^{(c)}(t)$ функцию распределения Y_1, а через $F_n^{(c)}(t)$ – эмпирическую функцию распределения, построенную по выборке $Y_1, \dots, Y_n$. Отметим следующие важные свойства введенных функций распределения:

1) $F^{(c)}(t) = F(t)$ для всех $t \in (-c, c)$;

2) $F_n^{(c)}(t) = F_n(t)$ для всех $t \in (-c, c)$.

Следовательно из теоремы 2 получаем

$$\sum_{j=1}^{n} \mathbf{D}Y_{(j)} \leq 2cn \int_{-c}^{c} \delta_n(t)dt. \tag{15}$$

Вернемся к оценке $\sum_{j=1}^{n} \mathbf{D}X_{(j)}$. Имеем

$$\sum_{j=1}^{n} \mathbf{D}X_{(j)} \leq \sum_{j=1}^{n} \mathbf{D}Y_{(j)} + \left| \sum_{j=1}^{n} \mathbf{D}X_{(j)} - \sum_{j=1}^{n} \mathbf{D}Y_{(j)} \right|$$

$$\leq \sum_{j=1}^{n} \mathbf{D}Y_{(j)} + n \left| \mathbf{E}X_1^2 - \mathbf{E}Y_1^2 \right| + \left| \sum_{j=1}^{n} \left(\mathbf{E}X_{(j)} \right)^2 - \sum_{j=1}^{n} \left(\mathbf{E}Y_{(j)} \right)^2 \right|, \tag{16}$$

Очевидно,

$$\left| \mathbf{E}X_1^2 - \mathbf{E}Y_1^2 \right| \leq \mathbf{E}X_1^2 I\{|X_1| > c\}. \tag{17}$$

Теперь оценим величину $\left| \sum_{j=1}^{n} \left(\mathbf{E}X_{(j)} \right)^2 - \sum_{j=1}^{n} \left(\mathbf{E}Y_{(j)} \right)^2 \right|$, используя по ходу неравенство Йенсена и перестановочность слагаемых вида $|X_{(j)}|^t I\{\left|X_{(j)}\right| > c\}$, $t = 1, 2$:

$$\left| \sum_{j=1}^{n} \left(\mathbf{E}X_{(j)} \right)^2 - \sum_{j=1}^{n} \left(\mathbf{E}Y_{(j)} \right)^2 \right| = \left| \sum_{j=1}^{n} \mathbf{E} \left(X_{(j)} + Y_{(j)} \right) \mathbf{E} \left(X_{(j)} - Y_{(j)} \right) \right|$$

$$\leq \sum_{j=1}^{n} \mathbf{E} \left| X_{(j)} \right| \mathbf{E} |X_{(j)}| I\{|X_{(j)}| > c\} + cn\mathbf{E}|X_1| I\{|X_1| > c\} \leq 2cn\mathbf{E}|X_1| I\{|X_1| > c\}$$

$$+ \sum_{j=1}^{n} \left(\mathbf{E}|X_{(j)}| I\{|X_{(j)}| > c\} \right)^2 \leq 3n\mathbf{E}X_1^2 I\{|X_1| > c\}. \tag{18}$$

Оценка (14) очевидным образом следует из (15)–(18). □

Легко видеть, что в условиях теоремы 3 инфимум в (14) стремится к нулю при $n \to \infty$, если только выполнено условие (3). Ниже мы приведем более детально вид правой части в (14) в случаях (i)–(iii) при существовании момента у X_1 порядка выше двух. При этом хвост второго момента в правой части (14) заменяем на очевидную оценку сверху $\mathbf{E}|X_1|^{2+r} c^{-r}$ и в каждом из трех вышеупомянутых случаев легко находим значение $c \equiv c(n)$, которое выравнивает порядки по n обоих слагаемых в правой части (14). В результате получаем

Следствие 3. *Если* $\mathbf{E} \left| X_1 \right|^{2+r} < \infty$ *для некоторого* $r > 0$, *то*

1) *в случае* **(i)**

$$\sum_{j=1}^{n} \mathbf{D}X_{(j)} \leq \left(2 \int_{-\infty}^{\infty} \sqrt{F(t) \left(1 - F(t)\right)} dt + 4\mathbf{E} \left| X_1 \right|^{2+r} \right) n^{\frac{r+2}{2r+2}}; \tag{19}$$

2) *в случае* **(ii)**

$$\sum_{j=1}^{n} \mathbf{D}X_{(j)} \leq 4 \left(\sqrt{2K} + \mathbf{E} \left| X_1 \right|^{2+r} \right) n^{\frac{(1+\varepsilon)r+4}{2r+4}}; \tag{20}$$

3) *в случае* **(iii)**

$$\sum_{j=1}^{n} \mathbf{D}X_{(j)} \leq \left(2\sqrt{2K}\int_{-\infty}^{\infty}\sqrt{\min\{F(t),1-F(t)\}}dt + 4\mathbf{E}\,|X_1|^{2+r}\right) n^{\frac{(1+\varepsilon)r+2}{2r+2}}. \quad (21)$$

Заметим, что интегралы в (19) и (21) конечны, что следует из моментного ограничения и неравенства Чебышева. Кроме того, при $r \to \infty$ порядок роста по n правых частей в (19)–(21) приближается к порядку роста соответствующих правых частей в (11)–(13).

Список литературы

[1] W. Hoeffding, *On the Distribution of the Expected Values of the Order Statistics*, Ann. Math. Statist. **24**:1 (1953), 93–100. MR0054197

[2] I. S. Borisov, *A note on a result by W. Hoeffding*, Statist. Probab. Lett. **87** (2014), 7–11. MR3168928

[3] А. В. Булинский, А. П. Шашкин, *Предельные теоремы для ассоциированных случайных полей и родственных систем*, М: Физматлит, 2008.

[4] И. А. Ибрагимов, Ю. В. Линник, *Независимые и стационарно связанные величины*, М: Наука, 1965. MR0202176

[5] Ю. В. Прохоров, *Сходимость случайных процессов и предельные теоремы теории вероятностей*, Теория вероятн. и ее примен. **1**:2 (1956), 177–238. Zbl 0075.29001

Игорь Семенович Борисов
Институт математики им. С. Л. Соболева СО РАН,
пр. академика Коптюга, 4,
630090, Новосибирск, Россия;
Новосибирский государственный университет,
ул. Пирогова, 2,
630090, Новосибирск, Россия
E-mail address: sibam@math.nsc.ru