УДК 519.2 Вестник СПбГУ. Математика. Механика. Астрономия. 2020. Т. 7 (65). Вып. 3
MSC 60F15

О комбинаторном усиленном законе больших чисел и ранговых статистиках*

А. Н. Фролов

Санкт-Петербургский государственный университет,
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: *Фролов А. Н.* О комбинаторном усиленном законе больших чисел и ранговых статистиках // Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия. 2020. Т. 7 (65). Вып. 3. С. 490–499.
https://doi.org/10.21638/spbu01.2020.311

Ранее автором был получен усиленный закон больших чисел (УЗБЧ) для комбинаторных сумм $\sum_i X_{ni\pi_n(i)}$, где $\|X_{nij}\|$ — матрица порядка n случайных величин с конечными четвертыми моментами, а $(\pi_n(1), \pi_n(2), \ldots, \pi_n(n))$ — случайная перестановка с равномерным распределением на множестве перестановок чисел $1, 2, \ldots, n$, не зависящая от случайных величин X_{nij}. Взаимная независимость элементов матрицы не предполагалась. В настоящей работе мы получим комбинаторный УЗБЧ при более общих предположениях, а также обсудим поведение ранговых статистик.

Ключевые слова: комбинаторные суммы, усиленный закон больших чисел, комбинаторный усиленный закон больших чисел, ранговые статистики, коэффициент ранговой корреляции Спирмена.

1. Введение. Пусть $\{\mathbf{X}_n = \|X_{nij}\|_{i,j=1}^n\}_{n=2}^\infty$ — последовательность матриц случайных величин и $\{\pi_n\}_{n=2}^\infty$ — последовательность случайных перестановок чисел $1, 2, \ldots, n$. Положим

$$S_n = \sum_{i=1}^n X_{ni\pi_n(i)}$$

для всех $n \geqslant 2$, где $\pi_n = (\pi_n(1), \pi_n(2), \ldots, \pi_n(n))$. Суммы S_n называются комбинаторными суммами.

Отметим, что комбинаторные суммы в предположении независимости компонент $\mathbf{X}_n$ и независимости с перестановкой π_n существенно отличаются от сумм независимых случайных величин. Последние суммы при увеличении числа слагаемых на единицу содержат предыдущие слагаемые. При таком же изменении номера комбинаторной суммы меняется перестановка, а это приводит к тому, что поменяться может значительное число слагаемых или даже все. При этом случай одинаково распределенных элементов матриц тривиален, так как все сводится к классическим суммам независимых одинаково распределенных (н. о. р.) случайных величин. Поэтому интерес представляют только матрицы с неодинаково распределенными элементами. При этом матрицы должны быть устроены так, чтобы распределения комбинаторных сумм не совпадали с распределениями сумм независимых случайных величин. Последний случай важен при оценке качества полученных результатов.

*Работа выполнена при финансовой поддержке РФФИ (грант № 18-01-00393).

Примем следующие терминологические соглашения. Если распределения центрированных и нормированных сумм S_n сходятся (в смысле слабой сходимости) к нормальному закону, то мы говорим о справедливости комбинаторной центральной предельной теоремы (ЦПТ). В случае подобной сходимости к вырожденному распределению, мы говорим о комбинаторном законе больших чисел (ЗБЧ). Заменяя слабую сходимость на сходимость почти наверное (п. н.), мы приходим к комбинаторному усиленному закону больших чисел (УЗБЧ). Отметим, что если строки каждой матрицы $\mathbf{X}_n$ независимы и состоят из одинаковых случайных величин, то S_n превращается в сумму независимых случайных величин. Таким образом, принимаемые нами соглашения аналогичны используемым в классической теории суммирования независимых случайных величин.

Без дополнительных предположений о характере зависимости элементов $\mathbf{X}_n$ и перестановки π_n и об их распределениях содержательной теории не построить. Поэтому мы далее следуем основной линии, предполагая их независимость и равномерную распределенность перестановки.

Дополнительно предположим сначала, что компоненты матриц $\mathbf{X}_n$ независимы, $\mathbf{E}X_{nij} = c_{nij}$ и

$$\sum_{j=1}^{n} c_{nij} = 0, \quad \sum_{i=1}^{n} c_{nij} = 0$$

для всех $1 \leqslant i, j \leqslant n$ и $n \geqslant 2$. Последнее условие на средние эквивалентно условию центрированности слагаемых в классической теории суммирования. Оно влечет $\mathbf{E}S_n = 0$, так как

$$\mathbf{E}X_{ni\pi_n(i)} = \frac{1}{n}\sum_{j=1}^{n} c_{nij} = 0.$$

Если дисперсии $\sigma_{nij}^2 = \mathbf{D}X_{nij}$ конечны для всех $1 \leqslant i, j \leqslant n$ и $n \geqslant 2$, то

$$B_n = \mathbf{D}S_n = \frac{1}{n-1}\sum_{i,j=1}^{n} c_{nij}^2 + \frac{1}{n}\sum_{i,j=1}^{n} \sigma_{nij}^2$$

для всех $n \geqslant 2$, а нормировкой в комбинаторной ЦПТ будет $\sqrt{B_n}$.

Неравенства Эссеена (для комбинаторных сумм) можно найти в работах Бара [1], Хо и Чена [2], Больтхаузена [3], Голдштейна [4], Неммани и Санторнчоста [5], Неммани и Раттановонга [6], Чена и Фанга [7] для случайных величин X_{nij} с конечными третьими моментами. Из этих результатов легко получаются условия, достаточные для комбинаторной ЦПТ, а ссылки на более ранние асимптотические результаты можно найти в библиографии этих работ. Фролов [8, 9] получил обобщения неравенств Эссеена для комбинаторных сумм на случай моментов порядка $2+\delta$, где $\delta \in (0,1]$, а также на случай бесконечных дисперсий. Результаты об умеренных и больших уклонениях комбинаторных сумм получены Фроловым в [10, 11]. Неравенства Эссеена для комбинаторных случайных сумм доказаны Фроловым в [12].

Наряду с ЦПТ, оценками точности остаточного члена в ней и большими уклонениями, в теории вероятностей и математической статистике играет важную роль УЗБЧ. По аналогии с классической теорией суммирования получение комбинаторного УЗБЧ также представляет интерес. Первый результат этого типа был получен Фроловым [13].

В работе [13] предполагалось, что элементы матриц $\mathbf{X}_n$ имеют четвертые моменты и не являются независимыми. Классическая техника, дающая результаты при минимальных моментных условиях, здесь не работает. Поэтому была проведена оценка четвертого момента S_n, а условия выражались в терминах максимумов сумм отклонений математических ожиданий произведений $\mathbf{E}X^{\alpha}_{nip}X^{\beta}_{njq}X^{\gamma}_{nkr}X^{\delta}_{nls}$ от произведений математических ожиданий $\mathbf{E}X^{\alpha}_{nip}\mathbf{E}X^{\beta}_{njq}\mathbf{E}X^{\gamma}_{nkr}\mathbf{E}X^{\delta}_{nls}$. Степени здесь неотрицательны и $\alpha+\beta+\gamma+\delta=4$. В настоящей работе мы сначала заменим максимальные условия на более общие — суммарные, а затем с помощью метода усечения распространим комбинаторный УЗБЧ на случай слагаемых без четвертых моментов. Кроме того, мы обсудим приложения к ранговым статистикам.

2. Комбинаторный УЗБЧ. Пусть $\{\mathbf{X}_n\}_{n=2}^{\infty}$ — последовательность матриц случайных величин такая, что $\mathbf{X}_n = \|X_{nij}\|_{i,j=1}^{n}$, $\mathbf{E}X_{nij} = c_{nij}$ для всех $1 \leqslant i, j \leqslant n$ и $n \geqslant 2$, и для любого $n \geqslant 2$ выполняется соотношение

$$c_{ni.} = \sum_{j=1}^{n} c_{nij} = 0, \quad c_{n.j} = \sum_{i=1}^{n} c_{nij} = 0 \quad \text{для всех} \quad 1 \leqslant i, j \leqslant n. \tag{1}$$

Пусть $\{\pi_n\}_{n=2}^{\infty}$ — последовательность случайных перестановок такая, что для всех $n \geqslant 2$ перестановка $\pi_n = (\pi_n(1), \pi_n(2), \ldots, \pi_n(n))$ имеет равномерное распределение на множестве всех перестановок чисел $1, 2, \ldots, n$. Предположим, что $\mathbf{X}_n$ и π_n независимы для любого $n \geqslant 2$.

Для всех $n \geqslant 2$ положим

$$S_n = \sum_{i=1}^{n} X_{ni\pi_n(i)}.$$

Отметим, что при нарушении условия (1) можно отцентрировать элементы $\mathbf{X}_n$, положив

$$X'_{nij} = X_{nij} - \frac{1}{n}c_{ni.} - \frac{1}{n}c_{n.j} + \frac{1}{n^2}c_{n..}, \quad \text{где} \quad c_{n..} = \sum_{i,j=1}^{n} c_{nij}.$$

Несложно проверить, что условие (1) выполнено с заменой c_{nij} на $\mathbf{E}X'_{nij}$.

В настоящей работе мы получим следующий вариант комбинаторного УЗБЧ.

Теорема 1. *Пусть $\mathbf{E}X^4_{nij} < \infty$ для всех $1 \leqslant i, j \leqslant n$ и $n \geqslant 2$. Для каждого n положим $C_n = \max\limits_{1\leqslant i,j\leqslant n} \mathbf{E}X^4_{nij}$ и $R_n = \max\limits_{1\leqslant i\leqslant 2}\{r_{ni}\}$, где*

$$r_{n1} = \frac{1}{(n)_4}\left|\sum_{i\neq j\neq k\neq l}\ \sum_{p\neq q\neq r\neq s}(\mathbf{E}X_{nip}X_{njq}X_{nkr}X_{nls} - c_{nip}c_{njq}c_{nkr}c_{nls})\right|,$$

$$r_{n2} = \frac{1}{(n)_3}\left|\sum_{i\neq j\neq k}\ \sum_{p\neq q\neq r}\left(\mathbf{E}X^2_{nip}X_{njq}X_{nkr} - \mathbf{E}X^2_{nip}c_{njq}c_{nkr}\right)\right|,$$

где $(n)_k = n(n-1)\cdot\ldots\cdot(n-k+1)$ для всех натуральных n и $k \leqslant n$ (здесь и далее мы считаем, что во всех суммах все индексы изменяются в пределах от 1 до n).

Пусть $\{b_n\}_{n=2}^{\infty}$ *— последовательность положительных постоянных. Предположим, что ряд* $\sum_n (C_n n^2 + R_n) b_n^{-4}$ *сходится.*
Тогда

$$\frac{S_n}{b_n} \to 0 \quad \text{п. н.} \tag{2}$$

Во всех предельных переходах мы считаем, что $n \to \infty$, если не оговорено противное.

Отметим, что если строки (или столбцы) матрицы $\mathbf{X}_n$ независимы, то $R_n = 0$. В частности, если все X_{nij} — вырожденные случайные величины для всех i, j и n, то $R_n = 0$. При этом в последнем случае сумма S_n, вообще говоря, невырождена.

Из этого результата вытекает теорема 1 работы [13]. Примеры и замечания последней работы сохраняют свою силу. В частности, если ряд из условия теоремы 1 расходится, то заключение теоремы 1 может не иметь места.

Доказательство. Положим $\xi_i = X_{ni\pi_n(i)}$ для всех $1 \leqslant i \leqslant n$. Мы имеем

$$\mathbf{E}S_n^4 = \sum_{i=1}^n \mathbf{E}\xi_i^4 + 4\sum_{i\neq j} \mathbf{E}\xi_i^3\xi_j + 3\sum_{i\neq j} \mathbf{E}\xi_i^2\xi_j^2 + 6\sum_{i\neq j\neq k} \mathbf{E}\xi_i^2\xi_j\xi_k + \sum_{i\neq j\neq k\neq l} \mathbf{E}\xi_i\xi_j\xi_k\xi_l. \tag{3}$$

В силу независимости $\mathbf{X}_n$ и π_n и равномерной распределенности π_n мы получим

$$\mathbf{E}\xi_i\xi_j\xi_k\xi_l = \frac{1}{(n)_4} \sum_{p\neq q\neq r\neq s} \mathbf{E}X_{nip}X_{njq}X_{nkr}X_{nls}.$$

Следовательно, выполняется неравенство

$$\left|\sum_{i\neq j\neq k\neq l} \mathbf{E}\xi_i\xi_j\xi_k\xi_l\right| \leqslant R_n + \frac{1}{(n)_4}|T(n)|, \quad \text{где} \quad T(n) = \sum_{i\neq j\neq k\neq l}\ \sum_{p\neq q\neq r\neq s} c_{nip}c_{njq}c_{nkr}c_{nls}.$$

В [13] на стр. 48–49 было показано, что

$$\left|\sum_{p\neq q\neq r\neq s} c_{nip}c_{njq}c_{nkr}c_{nls}\right| \leqslant 7n^2 C_n.$$

Отметим, что при доказательстве последнего неравенства существенно использовалось условие (1). Поэтому мы имеем

$$|T(n)| \leqslant 7n^2 C_n (n)_4.$$

Таким образом, мы получим

$$\left|\sum_{i\neq j\neq k\neq l} \mathbf{E}\xi_i\xi_j\xi_k\xi_l\right| \leqslant R_n + 7n^2 C_n. \tag{4}$$

Воспользуемся снова тем, что $\mathbf{X}_n$ и π_n независимы, а π_n равномерно распределена. Для всех $i \neq j \neq k$ мы имеем

$$\mathbf{E}\xi_i^2\xi_j\xi_k = \frac{1}{(n)_3} \sum_{p\neq q\neq r} \mathbf{E}X_{nip}^2 X_{njq}X_{nkr}.$$

Поэтому выполняется неравенство

$$\left|\sum_{i\neq j\neq k}\mathbf{E}\xi_i^2\xi_j\xi_k\right| \leqslant R_n + \frac{1}{(n)_3}|T'(n)|, \quad \text{где} \quad T'(n) = \sum_{i\neq j\neq k}\sum_{p\neq q\neq r}\mathbf{E}X_{nip}^2 c_{njq}c_{nkr}.$$

В [13] на стр. 49 с использованием условия (1) доказано, что

$$\left|\sum_{p\neq q\neq r}\mathbf{E}X_{nip}^2 c_{njq}c_{nkr}\right| \leqslant 2n^2 C_n.$$

Следовательно, имеет место неравенство

$$\left|\sum_{i\neq j\neq k}\mathbf{E}\xi_i^2\xi_j\xi_k\right| \leqslant 2n^2 C_n + R_n. \tag{5}$$

Используем снова тот факт, что $\mathbf{X}_n$ и π_n независимы, а перестановка π_n распределена равномерно. Мы имеем

$$\mathbf{E}\xi_i^2\xi_j^2 = \frac{1}{(n)_2}\sum_{p\neq q}\mathbf{E}X_{nip}^2 X_{njq}^2, \quad \mathbf{E}\xi_i^3\xi_j = \frac{1}{(n)_2}\sum_{p\neq q}\mathbf{E}X_{nip}^3 X_{njq}$$

для всех $i \neq j$ и

$$\mathbf{E}\xi_i^4 = \frac{1}{n}\sum_{p=1}^{n}\mathbf{E}X_{nip}^4 \leqslant C_n$$

для всех i. Используя неравенства Коши — Буняковского и Гёльдера, мы получим

$$\mathbf{E}X_{nip}^2 X_{njq}^2 \leqslant \sqrt{\mathbf{E}X_{nip}^4\mathbf{E}X_{njq}^4} \leqslant C_n$$

и

$$\left|\mathbf{E}X_{nip}^3 X_{njq}\right| \leqslant (\mathbf{E}X_{nip}^4)^{3/4}(\mathbf{E}X_{njq}^4)^{1/4} \leqslant C_n.$$

Следовательно, выполняются неравенства

$$\sum_{i=1}^{n}\mathbf{E}\xi_i^4 \leqslant nC_n, \quad \left|\sum_{i\neq j}\mathbf{E}\xi_i^3\xi_j\right| \leqslant n^2 C_n, \quad \sum_{i\neq j}\mathbf{E}\xi_i^2\xi_j^2 \leqslant n^2 C_n. \tag{6}$$

Из соотношений (3)–(6) вытекает неравенство

$$\mathbf{E}S_n^4 \leqslant 27n^2 C_n + 7R_n.$$

Отсюда следует, что

$$\sum_{n=1}^{\infty}\mathbf{P}\left(|S_n| \geqslant \varepsilon b_n\right) \leqslant \sum_{n=1}^{\infty}\frac{\mathbf{E}S_n^4}{\varepsilon b_n^4} \leqslant 27\sum_{n=1}^{\infty}\frac{n^2 C_n + R_n}{\varepsilon b_n^4} < \infty$$

для любого $\varepsilon > 0$. По лемме Бореля — Кантелли мы получим

$$\frac{S_n}{b_n} \to 0 \quad \text{п. н.}$$

Теорема доказана. □

С помощью метода усечения мы теперь получим обобщение теоремы 1 на случай $\mathbf{E}X_{nij}^4 = \infty$.

Пусть $\{b_n\}_{n=2}^{\infty}$ — последовательность положительных постоянных. Положим

$$Y_{nij} = X_{nij} - c_{nij}, \quad \bar{Y}_{nij} = Y_{nij}\mathbf{I}\left\{|Y_{nij}| < b_n\right\},$$

$$a_{nij} = \mathbf{E}Y_{nij}\mathbf{I}\left\{|Y_{nij}| < b_n\right\}, \quad a_{n.j} = \sum_{i=1}^{n} a_{nij}, \quad a_{ni.} = \sum_{j=1}^{n} a_{nij}, \quad a_{n..} = \sum_{i,j=1}^{n} a_{nij},$$

$$Z_{nij} = c_{nij} + \bar{Y}_{nij} - \frac{1}{n}a_{n.j} - \frac{1}{n}a_{ni.} + \frac{1}{n^2}a_{n..},$$

$$z_{nij} = c_{nij} + a_{nij} - \frac{1}{n}a_{n.j} - \frac{1}{n}a_{ni.} + \frac{1}{n^2}a_{n..},$$

$$P_n = \frac{1}{n}\sum_{i,j=1}^{n} \mathbf{P}(|Y_{nij}| \geqslant b_n),$$

где $\mathbf{I}\{\cdot\}$ — индикатор события в скобках.

Теорема 2. *Для каждого* $n \geqslant 2$ *положим* $C_n' = \max\limits_{1\leqslant i,j\leqslant n} \mathbf{E}Z_{nij}^4$ *и* $R_n' = \max\limits_{1\leqslant i\leqslant 2}\{r_{ni}'\}$, *где*

$$r_{n1}' = \frac{1}{(n)_4}\left|\sum_{i\neq j\neq k\neq l}\sum_{p\neq q\neq r\neq s}(\mathbf{E}Z_{nip}Z_{njq}Z_{nkr}Z_{nls} - z_{nip}z_{njq}z_{nkr}z_{nls})\right|,$$

$$r_{n2}' = \frac{1}{(n)_3}\left|\sum_{i\neq j\neq k}\sum_{p\neq q\neq r}\left(\mathbf{E}Z_{nip}^2Z_{njq}Z_{nkr} - \mathbf{E}Z_{nip}^2z_{njq}z_{nkr}\right)\right|.$$

Предположим, что ряды $\sum_n(C_n'n^2 + R_n')b_n^{-4}$ *и* $\sum_n P_n$ *сходятся и* $b_na_{n..}/n \to 0$. *Тогда*

$$\frac{S_n}{b_n} \to 0 \quad \textit{п. н.}$$

Так как случайные величины Z_{nij} — линейные комбинации усеченных центрированных X_{nij}, условия теоремы 2 не предполагают конечности четвертых моментов X_{nij}.

Доказательство. Положим

$$\bar{S}_n = \sum_{i=1}^{n} c_{ni\pi_n(i)} + \sum_{i=1}^{n} \bar{Y}_{ni\pi_n(i)}.$$

Тогда

$$e_n = \mathbf{E}\bar{S}_n = \frac{1}{n}a_{n..}.$$

Возьмем $\varepsilon > 0$. Если $|\bar{S}_n - e_n| < \varepsilon b_n$ и $|Y_{ni\pi_n(i)}| < b_n$ для всех i, то $|S_n - e_n| < \varepsilon b_n$. Поэтому

$$\{|S_n - e_n| \geqslant \varepsilon b_n\} \subset \left\{|\bar{S}_n - e_n| \geqslant \varepsilon b_n\right\} \cup \bigcup_{i=1}^{n} \left\{|Y_{ni\pi_n(i)}| \geqslant b_n\right\}.$$

Отсюда мы имеем

$$\mathbf{P}(|S_n - e_n| \geqslant \varepsilon b_n) \leqslant \mathbf{P}(|\bar{S}_n - e_n| \geqslant \varepsilon b_n) + \frac{1}{n} \sum_{i,j=1}^{n} \mathbf{P}(|Y_{nij}| \geqslant b_n).$$

Заметим, что

$$\bar{S}_n - e_n = \sum_{i=1}^{n} Z_{ni\pi_n(i)}.$$

В силу доказательства теоремы 1 ряд $\sum_n \mathbf{P}(|\bar{S}_n - e_n| \geqslant \varepsilon b_n)$ сходится. Ряд $\sum_n P_n$ сходится по условию. Следовательно, ряд $\sum_n \mathbf{P}(|S_n - e_n| \geqslant \varepsilon b_n)$ сходится. По лемме Бореля — Кантелли мы получаем требуемое. При этом следует учесть, что $e_n = o(b_n)$. Теорема доказана. $\square$

Проверка условий теоремы 2 упрощается, если распределения Y_{nij} симметричны. В этом случае $a_{nij} = 0$. Кроме того, если хвосты распределений Y_{nij} доминируются хвостом некоторой случайной величины Y, т. е. $\mathbf{P}(|Y_{nij}| \geqslant x) \leqslant \mathbf{P}(|Y| \geqslant x)$ для всех x, то $P_n \leqslant n\mathbf{P}(|Y| \geqslant b_n)$. Последнюю вероятность можно оценивать, если распределение Y известно. Другой вариант — предположить, что $D_n = \max_{i,j} \mathbf{E}|Y_{nij}|^q < \infty$ для некоторого $q > 0$. Тогда $P_n \leqslant nD_n b_n^{-q}$. Таким образом, могут быть легко построены примеры, в которых условия теоремы 2 применимы.

3. Ранговые статистики. Линейные ранговые статистики представляют собой комбинаторные суммы с $\mathbf{P}(X_{nij} = c_{nij}) = 1$ для всех i, j и n. Выбор различных матриц $\mathbf{X}_n$ дает большой набор статистик, используемых для проверки различных статистических гипотез. Вопросам, связанным с построением ранговых критериев, посвящена, например, книга [14].

Обсудим результат теоремы 1 на примере коэффициента ранговой корреляции Спирмена. Положим

$$c_{nij} = \frac{12}{n(n^2-1)} \left(i - \frac{n+1}{2}\right) \left(j - \frac{n+1}{2}\right), \quad 1 \leqslant i, j \leqslant n.$$

Ранговый коэффициент корреляции ρ Спирмена определяется соотношением

$$\rho = \sum_{i=1}^{n} c_{ni\pi_n(i)}.$$

Связь этой формулы с определением выборочного коэффициента корреляции между рангами и формулой для вычисления ρ, обычно приводимой в справочниках (см., например, [15]), можно найти в книге [16]. Мы имеем

$$|c_{nij}| \leqslant \frac{12}{n(n^2-1)} \frac{(n-1)^2}{4} = \frac{3(n-1)}{n(n+1)} \leqslant \frac{3}{n}$$

и $C_n = \max_{1\leqslant i,j\leqslant n} \mathbf{E}X^4_{nij} \leqslant 81n^{-4}$. Положим $b_n = n^{-1/4}\sqrt{\ln n}$. Тогда ряд $\sum_n n^2 C_n b_n^{-4}$ сходится. По теореме 1 мы имеем

$$\frac{n^{1/4}}{\sqrt{\ln n}}\,\rho \to 0 \quad \text{п. н.}$$

Использование четвертого момента вместо второго привело нас к содержательному результату. Если использовать второй момент, то по неравенству Чебышёва мы получим

$$\sum_{n=1}^{\infty} \mathbf{P}(|\rho| \geqslant \varepsilon b_n) \leqslant \sum_{n=1}^{\infty} \frac{\mathbf{D}\rho}{\varepsilon^2 b_n^2} = \sum_{n=1}^{\infty} \frac{1}{(n-1)\varepsilon^2 b_n^2}.$$

Это позволит получить УЗБЧ только с $b_n \to \infty$, т. е. на этом пути мы не можем даже доказать сходимость ρ к нулю. Теоретически нет проблем с использованием моментов более высокого порядка. Вместо формулы (4) мы получим формулу с большим числом слагаемых, состоящих из сумм смешанных моментов более высокого порядка. Экспоненциальные оценки из [11] позволяют улучшить результат для ρ, увеличив степень n до 0.5, но это уже относится к закону повторного логарифма, который мы обсудим в другой работе.

Коэффициент ρ применяется для проверки гипотезы независимости. При этом предполагается, что векторы рангов строятся по выборкам из непрерывных распределений. Это дает выборки без совпадений и однозначность в определении рангов. В реальности выборки могут содержать совпадения из-за округления или наличия атомов у распределений выборок. В книге [14, гл. 3, п. 8] изложены различные способы обработки совпадений. Например, можно для повторяющихся наблюдений назначать ранги с помощью дополнительного случайного эксперимента (рандомизация). Другой вариант — переопределить метки в c_{nij} так, чтобы на группах совпадающих наблюдений новые метки принимали бы одинаковые значения. Заметим, что число групп совпадений и число элементов в группах совпадений — случайные величины. Поэтому новые метки будут случайными величинами, что соответствует нашей модели.

Выше шла речь о ρ, но это лишь один пример из большого числа ранговых статистик. Другие примеры читатель найдет в книге [14]. Мы их приводить не будем, так как цель нашей статьи — обсуждение общих закономерностей поведения комбинаторных сумм.

Отметим, кроме того, что наш результат получен для, вообще говоря, зависимых случайных величин X_{nij}. Возвращаясь к ρ Спирмена, добавим, что теорема 1 дает дополнительное представление о характере зависимости, при которой статистика будет вести себя так же, как при независимости. Таким образом, мы приходим к лучшему пониманию ситуаций, в которых при использовании ранговых критериев может совершаться ошибка 2-го рода.

Ранговые статистики, упомянутые выше, не дают примеров с неограниченными случайными величинами X_{nij}. Поэтому мы приведем простую схему, упоминавшуюся в [7]. Из случайных величин $\xi_1, \ldots, \xi_n$ выберем наудачу $k = k(n) < n$ величин и просуммируем их. Такая сумма будет комбинаторной суммой, если положить $X_{nij} = \xi_j$ для $1 \leqslant i \leqslant k(n)$ и $1 \leqslant j \leqslant n$ и $X_{nij} = 0$ для всех остальных i и j.

Аналогичный подход приводит также к некоторым обобщениям ранговых статистик. Пусть, например, $\eta_1, \ldots, \eta_n$ — н. о. р. непрерывные центрированные случайные трехмерные векторы и $(r_1, \ldots, r_n)$ — вектор антирангов, построенный по их первым координатам. Тогда статистика (и комбинаторная сумма) $T_n = \sum_{i=1}^{k(n)} \eta_{r_i}^2 \eta_{r_i}^3$ будет оценкой ковариации 2-й и 3-й компонент вектора $\eta_1 = (\eta_1^1, \eta_1^2, \eta_1^3)$, когда мы отбрасываем $n - k(n)$ наблюдений, соответствующих большим значениям первой компоненты. В случае независимости компонент вектора η_1 сумма T_n представляется естественным обобщением некоторой ранговой статистики.

Автор благодарит рецензентов за ряд полезных замечаний, способствовавших улучшению текста статьи.

Литература

1. *von Bahr B.* Remainder term estimate in a combinatorial central limit theorem // Z. Wahrsch. verw. Geb. 1976. Vol. 35. P. 131–139.

2. *Ho S. T., Chen L. H. Y.* An L_p bounds for the remainder in a combinatorial central limit theorem // Ann. Probab. 1978. Vol. 6. P. 231–249.

3. *Bolthausen E.* An estimate of the remainder in a combinatorial central limit theorem // Z. Wahrsch. verw. Geb. 1984. Vol. 66. P. 379–386.

4. *Goldstein L.* Berry — Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing // J. Appl. Probab. 2005. Vol. 42. P. 661–683. https://doi.org/10.1239/jap/1127322019

5. *Neammanee K., Suntornchost J.* A uniform bound on a combinatorial central limit theorem // Stoch. Anal. Appl. 2005. Vol. 23. P. 559–578. https://doi.org/10.1081/SAP-200056686

6. *Neammanee K., Rattanawong P.* A constant on a uniform bound of a combinatorial central limit theorem // J. Math. Research. 2009. Vol. 1. P. 91–103. https://doi.org/10.5539/jmr.v1n2p91

7. *Chen L. H. Y., Fang X.* On the error bound in a combinatorial central limit theorem // Bernoulli. 2015. Vol. 21, N 1. P. 335–359. https://doi.org/10.3150/13-BEJ569

8. *Frolov A. N.* Esseen type bounds of the remainder in a combinatorial CLT // J. Statist. Planning and Inference. 2014. Vol. 149. P. 90–97. https://doi.org/10.1016/j.jspi.2014.01.004

9. *Frolov A. N.* Bounds of the remainder in a combinatorial central limit theorem // Statist. Probab. Letters. 2015. Vol. 105. P. 37–46. https://doi.org/10.1016/j.spl.2015.05.020

10. *Фролов А. Н.* О вероятностях умеренных уклонений комбинаторных сумм // Вестник С.-Петерб. ун-та. Сер. 1. Математика. Механика. Астрономия. 2015. Т. 2 (60). Вып. 1. С. 60–67.

11. *Frolov A. N.* On large deviations of combinatorial sums. 2019. ArXiv: 1901.04244.

12. *Frolov A. N.* On Esseen type inequalities for combinatorial random sums // Communications in Statistics — Theory and Methods. 2017. Vol. 46. Iss. 12. P. 5932–5940. https://doi.org/10.1080/03610926.2015.1115074

13. *Frolov A. N.* On a combinatorial strong law of large numbers // Istatistik: Journ. of Turkish Statist. Assoc. 2018. Vol. 11, no. 3. P. 46–52. Available at: http://jtsa.ieu.edu.tr/current/1.pdf (accessed: May 26, 2020).

14. *Гаек Я., Шидак З.* Теория ранговых критериев. М.: Наука, 1971.

15. *Большев Л. Н., Смирнов Н. В.* Таблицы математической статистики. М.: Наука, 1983.

16. *Фролов А. Н.* Краткий курс теории вероятностей и математической статистики. СПб.: Лань, 2017.

Статья поступила в редакцию 13 октября 2019 г.;
после доработки 15 февраля 2020 г.;
рекомендована в печать 19 марта 2020 г.

Контактная информация:

Фролов Андрей Николаевич — д-р физ.-мат. наук, проф.; a.frolov@spbu.ru

On combinatorial strong law of large numbers and rank statistics

A. N. Frolov

St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

For citation: Frolov A. N. On combinatorial strong law of large numbers and rank statistics. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy*, 2020, vol. 7 (65), issue 3, pp. 490–499. https://doi.org/10.21638/spbu01.2020.311 (In Russian)

The author had earlier obtained a strong law of large numbers for combinatorial sums $\sum_i X_{ni\pi_n(i)}$, where $\|X_{nij}\|$ is a matrix of order n from random variables with finite fourth moments and $(\pi_n(1), \pi_n(2), \ldots, \pi_n(n))$ is a random permutation having the uniform distribution on the set of all permutations of numbers $1, 2, \ldots, n$ and being independent from random variables X_{nij}. The mutual independence for entries of the matrix has not been assumed. In the present paper, we derive the combinatorial SLLN under more general assumptions and discuss the behaviour of rank statistics.

Keywords: combinatorial sums, strong law of large numbers, combinatorial strong law of large numbers, rank statistics, Spearman's coefficient of rank correlation.

References

1. von Bahr B., "Remainder term estimate in a combinatorial central limit theorem", *Z. Wahrsch. verw. Geb.* **35**, 131–139 (1976).

2. Ho S. T., Chen L. H. Y., "An L_p bounds for the remainder in a combinatorial central limit theorem", *Ann. Probab.* **6**, 231–249 (1978).

3. Bolthausen E., "An estimate of the remainder in a combinatorial central limit theorem", *Z. Wahrsch. verw. Geb.* **66**, 379–386 (1984).

4. Goldstein L., "Berry — Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing", *J. Appl. Probab.* **42**, 661–683 (2005). https://doi.org/10.1239/jap/1127322019

5. Neammanee K., Suntornchost J., "A uniform bound on a combinatorial central limit theorem", *Stoch. Anal. Appl.* **23**, 559–578 (2005). https://doi.org/10.1081/SAP-200056686

6. Neammanee K., Rattanawong P., "A constant on a uniform bound of a combinatorial central limit theorem", *J. Math. Research* **1**, 91–103 (2009). https://doi.org/10.5539/jmr.v1n2p91

7. Chen L. H. Y., Fang X., "On the error bound in a combinatorial central limit theorem", *Bernoulli*, **21**(1), 335–359 (2015). https://doi.org/10.3150/13-BEJ569

8. Frolov A. N., "Esseen type bounds of the remainder in a combinatorial CLT", *J. Statist. Planning and Inference* **149**, 90–97 (2014). https://doi.org/10.1016/j.jspi.2014.01.004

9. Frolov A. N., "Bounds of the remainder in a combinatorial central limit theorem", *Statist. Probab. Letters* **105**, 37–46 (2015). https://doi.org/10.1016/j.spl.2015.05.020

10. Frolov A. N., "On the probabilities of moderate deviations for combinatorial sums", *Vestnik St. Petersburg University. Mathematics* **48**, iss. 1, 23–28 (2015). https://doi.org/10.3103/S1063454115010045

11. Frolov A. N., "On large deviations of combinatorial sums" (2019). ArXiv: 1901.04244.

12. Frolov A. N., "On Esseen type inequalities for combinatorial random sums", *Communications in Statistics — Theory and Methods* **46**(12), 5932–5940 (2017). https://doi.org/10.1080/03610926.2015.1115074

13. Frolov A. N., "On a combinatorial strong law of large numbers", *Istatistik: Journ. of Turkish Statist. Assoc.* **11**(3), 46–52 (2018). Available at: http://jtsa.ieu.edu.tr/current/1.pdf (accessed: May 26, 2020).

14. Hajek J., Sidak Z., *Theory of rank tests* (Nauka Publ., Moscow, 1971). (In Russian)

15. Bol'shev L. N., Smirnov N. V., *Tables of mathematical statistics* (Nauka Publ., Moscow, 1983). (In Russian)

16. Frolov A. N., *Short course of probability theory and mathematical statistics* (Lan' Publ., St. Petersburg, 2017). (In Russian)

Received: October 13, 2019
Revised: February 15, 2020
Accepted: March 19, 2020

Author's information:

Andrei N. Frolov — a.frolov@spbu.ru