İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION
Vol. 11, No. 3, December 2018, pp. 46–52
ISSN 1300-4077 | 18 | 3 | 46 | 52

İSTATİSTİK

ON A COMBINATORIAL STRONG LAW OF LARGE NUMBERS

Andrei N. Frolov *
Dept. of Mathematics and Mechanics,
St. Petersburg State University,
Universitetskii prosp. 28, Stary Peterhof, St. Petersburg, Russia

Abstract: We derive strong laws of large numbers for combinatorial sums $\sum_i X_{ni\pi_n(i)}$, where $\|X_{nij}\|$ are $n \times n$ matrices of random variables with finite fourth moments and $(\pi_n(1), \ldots, \pi_n(n))$ are uniformly distributed random permutations of $1, \ldots, n$ independent with X's. We do not assume the independence of X's, but this case is included as well. Examples are discussed.

Key words: Combinatorial central limit theorem; combinatorial sums; strong law of large numbers
History: Submitted: 30 August 2018; Revised: 9 October 2018; Accepted: 15 November 2018

1. Introduction

Let $\{\|X_{nij}\|_{i,j=1}^n\}_{n=2}^\infty$ be a sequence of matrices of random variables and $\{\pi_n\}_{n=2}^\infty$ be a sequence of random permutations of $1, 2, \ldots, n$. Put

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

for all $n \geqslant 2$, where $\pi_n = (\pi_n(1), \pi_n(2), \ldots, \pi_n(n))$. Sums S_n are called the combinatorial sums.

If distributions of centered and normalized combinatorial sums converge weakly to the normal law, then one says that a combinatorial central limit theorem (CLT) holds true. If centered and normalized combinatorial sums converge almost surely (a.s.) to a constant, then one says that a combinatorial strong law of large numbers (SLLN) holds. Replacing strong convergence by convergence in probability, one arrives at a combinatorial weak law of large numbers (WLLN).

One cannot construct an interesting theory without additional assumptions on type of dependence of X's and π_n and their distributions. We follow a general line in which X's and π_n are independent and π_n has the uniform distribution.

Assume that for every n, components of $\|X_{nij}\|$ are independent, matrix $\|X_{nij}\|_{i,j=1}^n$ and permutation π_n are independent and π_n has the uniform distribution on the set of permutations of $1, 2, \ldots, n$. Moreover, we also assume that $EX_{nij} = c_{nij}$ and

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

for all $1 \leqslant i, j \leqslant n$ and n. In results for combinatorial sums, the last condition provides that combinatorial sums S_n are centered at zero. Indeed,

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

E-mail address: Andrei.Frolov@pobox.spbu.ru

* This investigation was partially supported by RFBR, research project N. 18–01–00393

If $\sigma^2_{nij} = DX_{nij} = EX^2_{nij} - (EX_{nij})^2$ for all $1 \leqslant i,j \leqslant n$ and $n \geqslant 2$, then we have

$$B_n = DS_n = \frac{1}{n-1}\sum_{i,j=1}^{n} c^2_{nij} + \frac{1}{n}\sum_{i,j=1}^{n} \sigma^2_{nij},$$

for all n. Hence, the norming sequence in combinatorial CLT is $\sqrt{B_n}$.

One can easy derive sufficient conditions for the combinatorial CLT from Esseen inequalities which give bounds for the accuracy of the normal approximation of distributions of $S_n/\sqrt{B_n}$. One can find such inequalities in von Bahr [1], Ho and Chen [2], Botlthausen [3], Goldstein [4], Neammanee and Suntornchost [5], Neammanee and Rattanawong [6], Chen and Fang [7] for X's with finite third moments. Earlier asymptotic results on combinatorial CLT may be found in references therein. Frolov [8,9] derived generalizations of Esseen bounds for combinatorial sums to the cases of finite moments of order $2+\delta$, $\delta \in (0,1]$ and infinite variations. Moderate deviations for combinatorial sums have been investigated in Frolov [10]. Esseen bounds for combinatorial random sums may be found in Frolov [11].

Together with CLT and large deviations, SLLN plays an important role in probability and statistics. In this paper, we derive the combinatorial SLLN. Note that properties of combinatorial sums S_n for independent X's are quite different from those of sums of independent random variables. First, combinatorial sums are sums of dependent random variables. Second, many summands of S_n and S_{n+1} can be different even when $\|X_{nij}\|$ is a sub-matrix of $\|X_{n+1,i,j}\|$ with $X_{nij} = X_{n+1,i,j}$ for all $1 \leqslant i,j \leqslant n$. This is the result of randomness of permutations π_n. It follows that we have no monotonicity of combinatorial sums for positive X's. Remember that monotonicity of sums of positive i.i.d. random variables are essentially used in the proof of the Kolmogorov SLLN. We also have no analogues of results on convergence of series of independent random variables. Moreover, we will not assume the independence of X's. This reduces our possibilities to prove strong limit theorems for combinatorial sums. Therefore, we obtain bounds for forth moments of combinatorial sums and apply the Borel–Cantelli lemma.

2. Combinatorial SLLN

Let $\{\|X_{nij}\|^n_{i,j=1}\}^\infty_{n=2}$ be a sequence of matrices of random variables with $EX_{nij} = c_{nij}$ for all $1 \leqslant i,j \leqslant n$ and $n \geqslant 2$ and $\{\pi_n\}^\infty_{n=2}$ be a sequence of random permutations of $1,2,\ldots,n$. Assume that for every $n \geqslant 2$, relation

$$c_{ni.} = \sum_{j=1}^{n} c_{nij} = 0, \quad c_{n.j} = \sum_{i=1}^{n} c_{nij} = 0, \quad \text{for all} \quad 1 \leqslant i,j \leqslant n, \tag{2.1}$$

holds, π_n has the uniform distribution on the set of permutations of $1,2,\ldots,n$ and $\|X_{nij}\|$ and π_n are independent. For all $n \geqslant 2$, put

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

where $\pi_n = (\pi_n(1), \pi_n(2), \ldots, \pi_n(n))$.

Note that if condition (2.1) is not satisfied, then one can center X's as follows. Put

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

It is not difficult to check that condition (2.1) holds with EX'_{nij} instead of c_{nij}.

The next result is the combinatorial SLLN.

THEOREM 1. *Suppose that the above assumptions hold and $EX_{nij}^4 < \infty$ for all i, j and n. For every n, put $C_n = \max_{1\leqslant i,j\leqslant n} EX_{nij}^4$ and $M_n = \max_{1\leqslant i\leqslant 4}\{m_{ni}\}$, where*

$$\begin{aligned}
m_{n1} &= \max_{1\leqslant i\neq j\neq k\neq l\leqslant n}\{|\sum_{p\neq q\neq r\neq s}(EX_{nip}X_{njq}X_{nkr}X_{nls} - c_{nip}c_{njq}c_{nkr}c_{nls})|\},\\
m_{n2} &= \max_{1\leqslant i\neq j\neq k\leqslant n}\{|\sum_{p\neq q\neq r}(EX_{nip}^2X_{njq}X_{nkr} - EX_{nip}^2c_{njq}c_{nkr})|\},\\
m_{n3} &= \max_{1\leqslant i\neq j\leqslant n}\{|\sum_{p\neq q}(EX_{nip}^2X_{njq}^2 - EX_{nip}^2EX_{njq}^2)|\},\\
m_{n4} &= \max_{1\leqslant i\neq j\leqslant n}\{|\sum_{p\neq q}(EX_{nip}^3X_{njq} - EX_{nip}^3c_{njq})|\}.
\end{aligned}$$

Let $\{b_n\}_{n=2}^{\infty}$ be a sequence of positive constants. Assume that the series $\sum_n(C_nn^2 + M_n)b_n^{-4}$ converges.

Then

$$\frac{S_n}{b_n} \to 0 \quad a.s. \tag{2.2}$$

PROOF. For all natural n and k, denote $(n)_k = n(n-1)\cdot\ldots\cdot(n-k+1)$.

Put $\xi_i = X_{ni\pi_n(i)}$ for $1\leqslant i\leqslant n$. We have

$$ES_n^4 = \sum_{i=1}^n E\xi_i^4 + 4\sum_{i\neq j}E\xi_i^3\xi_j + 3\sum_{i\neq j}E\xi_i^2\xi_j^2 + 6\sum_{i\neq j\neq k}E\xi_i^2\xi_j\xi_k + \sum_{i\neq j\neq k\neq l}E\xi_i\xi_j\xi_k\xi_l, \tag{2.3}$$

where $1\leqslant i,j,k,l\leqslant n$ in the last four sums. Since X's and π_n are independent and π_n is uniformly distributed, we get

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

Hence,

$$|E\xi_i\xi_j\xi_k\xi_l| \leqslant \frac{1}{(n)_4}M_n + \frac{1}{(n)_4}|T_0|, \quad \text{where} \quad T_0 = \sum_{p\neq q\neq r\neq s}c_{nip}c_{njq}c_{nkr}c_{nls}.$$

It is clear that

$$T_0 = \sum_{p=1}^n\sum_{q:q\neq p}\sum_{r:r\neq p,q}\sum_{s:s\neq p,q,r}c_{nip}c_{njq}c_{nkr}c_{nls}.$$

By condition (2.1), we have

$$\sum_{s:s\neq p,q,r}c_{nls} = -(c_{nlp}+c_{nlq}+c_{nlr}).$$

It follows that

$$\begin{aligned}
T_0 &= -\sum_{p=1}^n\sum_{q:q\neq p}\sum_{r:r\neq p,q}c_{nip}c_{njq}c_{nkr}\left(c_{nlp}+c_{nlq}+c_{nlr}\right)\\
&= -\sum_{p=1}^n\sum_{q:q\neq p}c_{nip}c_{njq}\left(c_{nlp}+c_{nlq}\right)\sum_{r:r\neq p,q}c_{nkr} - \sum_{p=1}^n\sum_{q:q\neq p}c_{nip}c_{njq}\sum_{r:r\neq p,q}c_{nkr}c_{nlr} = -T_1 - T_2
\end{aligned}$$

for all $i \neq j \neq k \neq l$. Using again condition (2.1), we have

$$T_1 = -\sum_{p=1}^{n} \sum_{q:q\neq p} c_{nip} c_{njq} \left(c_{nlp} + c_{nlq}\right) \left(c_{nkp} + c_{nkq}\right).$$

By the Lyapunov inequality, it follows that $|c_{nij}| \leqslant (E|X_{nij}|^4)^{1/4} \leqslant C_n^{1/4}$. It yields that

$$|T_1| \leqslant 4n^2 C_n.$$

Furthermore,

$$\begin{aligned} T_2 &= \sum_{p=1}^{n} \sum_{q:q\neq p} c_{nip} c_{njq} \left(\sum_{r=1}^{n} c_{nkr} c_{nlr} - (c_{nkp} c_{nlp} + c_{nkq} c_{nlq}) \right) \\ &= \left(\sum_{r=1}^{n} c_{nkr} c_{nlr} \right) \sum_{p=1}^{n} c_{nip} \sum_{q:q\neq p} c_{njq} - \sum_{p=1}^{n} \sum_{q:q\neq p} c_{nip} c_{njq} (c_{nkp} c_{nlp} + c_{nkq} c_{nlq}) \\ &= -\left(\sum_{r=1}^{n} c_{nkr} c_{nlr} \right) \sum_{p=1}^{n} c_{nip} c_{njp} - \sum_{p=1}^{n} \sum_{q:q\neq p} c_{nip} c_{njq} (c_{nkp} c_{nlp} + c_{nkq} c_{nlq}). \end{aligned}$$

In the last equality, we have applied condition (2.1). Using again inequalities $|c_{nij}| \leqslant C_n^{1/4}$, we get

$$|T_2| \leqslant 3n^2 C_n.$$

Therefore, for all $i \neq j \neq k \neq l$, inequalities

$$|E\xi_i \xi_j \xi_k \xi_l| \leqslant \frac{1}{(n)_4} (7n^2 C_n + M_n) \tag{2.4}$$

hold. For all $i \neq j \neq k$, we have

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

and

$$|E\xi_i^2 \xi_j \xi_k| \leqslant \frac{1}{(n)_3} M_n + \frac{1}{(n)_3} |T_3|, \quad \text{where} \quad T_3 = \frac{1}{(n)_3} \sum_{p\neq q\neq r} EX_{nip}^2 c_{njq} c_{nkr}.$$

We get by condition (2.1) that

$$T_3 = \frac{1}{(n)_3} \sum_{p=1}^{n} \sum_{q:q\neq p} EX_{nip}^2 c_{njq} \sum_{r:r\neq p,q} c_{nkr} = -\frac{1}{(n)_3} \sum_{p=1}^{n} \sum_{q:q\neq p} EX_{nip}^2 c_{njq} \left(c_{nkp} + c_{nkq}\right).$$

Since $EX_{nij}^2 \leqslant (EX_{nij}^4)^{1/2} \leqslant \sqrt{C_n}$ and $|c_{nij}| \leqslant C_n^{1/4}$ for all i and j, the latter implies that

$$|E\xi_i^2 \xi_j \xi_k| \leqslant \frac{1}{(n)_3} (2n^2 C_n + M_n) \tag{2.5}$$

for all $i \neq j \neq k$.

For all $i \neq j$, we get

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

and

$$E\xi_i^2\xi_j^2 \leqslant \frac{1}{(n)_2}M_n + \frac{1}{(n)_2}\sum_{p\neq q} EX_{nip}^2 EX_{njq}^2,$$
$$|E\xi_i^3\xi_j| \leqslant \frac{1}{(n)_2}M_n + \frac{1}{(n)_2}\sum_{p\neq q} EX_{nip}^3 EX_{njq}.$$

Applying inequalities $|c_{nij}| \leqslant C_n^{1/4}$, $EX_{nij}^2 \leqslant \sqrt{C_n}$ and $|EX_{nij}^3| \leqslant (EX_{nij}^4)^{4/3} \leqslant \sqrt{C_n}$ for all i and j, we have

$$E\xi_i^2\xi_j^2 \leqslant \frac{1}{(n)_2}(n^2C_n + M_n), \quad |E\xi_i^3\xi_j| \leqslant \frac{1}{(n)_2}(n^2C_n + M_n), \tag{2.6}$$

for all $i \neq j$.

Finally, for every i, we get

$$E\xi_i^4 = \frac{1}{n}\sum_{p=1}^{n} EX_{nip}^4 \leqslant C_n. \tag{2.7}$$

Substituting bounds (2.4)–(2.7) in equality (2.3), we have

$$ES_n^4 \leqslant nC_n + (4+3+12+7)n^2C_n + 4M_n \leqslant 27(n^2C_n + M_n).$$

It follows that

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

for all $\varepsilon > 0$. By the Borel–Cantelly lemma, we obtain

$$\frac{S_n}{b_n} \to 0 \quad \text{a.s.}$$

Note that $\sum_n DS_n b_n^{-2} < \infty$ is a sufficient condition for relation (2.2). For independent X's, $M_n = 0$ and, using formula for DS_n from Section 1, we see that DS_n has an order n in various partial cases. So, if $b_n = n$, the last series always diverges while the series from Theorem 1 can converge. For example, the latter holds for bounded (uniformly over n) random variables.

Remark 1. One can find further conditions sufficient for combinatorial SLLN by applications of bounds for ES_n^{2k} with $k \geqslant 3$ which may be derived in the same way as before.

Condition (2.1) is symmetric relatively to rows and columns of matrices of means. Substituting in (2.3) the formulae for the expectations, we can interchange sums over numbers of rows and columns. Further, we can apply the second equality in (2.1) instead of the first one. Hence, we arrive at the next remark.

Remark 2. In Theorem 1, one can interchange indices in maxima and sums in the definitions of $m_{n1}, \ldots, m_{n4}$.

Theorem 1 yields the following result.

Corollary 1. *If the conditions of Theorem 1 hold and series* $\sum_n (n^2C_n + M_n)n^{-4p}$ *converges for some* $p > 0$, *then*

$$\frac{S_n}{n^p} \to 0 \quad \textit{a.s.}$$

Note that if the series from Corollary 1 diverges, then its conclusion can fail. Indeed, let $\{\eta_n\}$ be a sequence of independent random variables such that $P(\eta_n = n^p) = P(\eta = -n^p) = 1/2$ for all n. Put $X_{nij} = \eta_i$ for all i, j and n. Then $C_n = E\eta_n^4 = n^{4p}$, $M_n = 0$ and

$$S_n = \eta_1 + \eta_2 + \cdots + \eta_n.$$

Assuming that $n^{-p} S_n \to 0$ a.s., we have

$$\frac{\eta_n}{n^p} = \frac{S_n}{n^p} - \frac{S_{n-1}}{(n-1)^p} \cdot \frac{(n-1)^p}{n^p} \to 0 \quad \text{a.s.}$$

that contradicts to relation $P(n^{-p}\eta_n = 1) = 1/2$ for all n.

It is clear that $M_n = 0$ provided every quadruple of different elements of matrices $\|X_{nij}\|$ is a set of independent random variables. Moreover, $M_n = 0$ when rows of $\|X_{nij}\|$ are independent while elements of one row may be dependent.

These conditions are much more less than mutual independence, but it is useful to have an example with positive M_n.

To this end, we consider matrices $\|X_{nij}\|$ with m-dependent rows, where m is a fixed natural number. The latter means that i-th and k-th rows are independent when $|i-k| > m$. (The case $m = 0$ correspond to independence.) At the same time, we do not assume that random variables of one rows are independent.

Note that there is a simple way to construct such matrix. Take matrix $\|X_{nij}\|$ of independent random variables and replace every even row by previous odd ones. Then rows will be 1-dependent. The construction for $m > 1$ follows the same pattern.

For simplicity, put $m = 1$ and assume that $C_n = C$ for all n. It is clear, that all items of sums in the definitions of m_{ni} are bounded by $2C$ and many of them equal to zero by independence of "far" rows. The number of zero items in m_{n1} is bounded from below by $n(n-2)(n-4)(n-6)$. Hence, the number of non-zero items is less than $(n)_4 - n(n-2)(n-4)(n-6) = O(n^3)$ as $n \to \infty$. It follows that $m_{n1} = O(n^3)$ as $n \to \infty$. Maxima m_{n2}, m_{n3} and m_{n4} have the same or smaller order. So, the series in Theorem 1 converges provided series $\sum_n n^3 b_n^{-4}$ converges. By Theorem 1, we have

$$\frac{S_n}{n(\ln n)^q} \to 0 \quad \text{a.s.}$$

for all $q > 1/4$.

Acknowledgement

The author thanks the anonymous referee for helpful remarks which improved the text of the paper.

References

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

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

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

[4] Goldstein L. (2005). Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing, *J. Appl. Probab.*, 42, 661-683.

[5] Neammanee K., Suntornchost J. (2005). A uniform bound on a combinatorial central limit theorem, *Stoch. Anal. Appl.*, 3, 559-578.

[6] Neammanee K., Rattanawong P. (2009). A constant on a uniform bound of a combinatorial central limit theorem, *J. Math. Research*, 1, 91-103.

[7] Chen L.H.Y., Fang X. (2015). 0n the error bound in a combinatorial central limit theorem, *Bernoulli*, 21 (1), 335-359.

[8] Frolov A.N. (2014). Esseen type bounds of the remainder in a combinatorial CLT, *J. Statist. Planning and Inference*, 149, 90-97.

[9] Frolov A.N. (2015a) Bounds of the remainder in a combinatorial central limit theorem, *Statist. Probab. Letters*, 105, 37-46.

[10] Frolov A.N. (2015b). On the probabilities of moderate deviations for combinatorial sums. Vestnik St. Petersburg University. Mathematics, 48(1), 23-28. Allerton Press, Inc., 2015.

[11] Frolov A.N. (2017). On Esseen type inequalities for combinatorial random sums. *Communications in Statistics -Theory and Methods*, 46(12), 5932-5940.