УДК 519.2 Вестник СПбГУ. Математика. Механика. Астрономия. 2020. Т. 7 (65). Вып. 3
MSC 60F15, 62G30, 62G20; 60F05, 62G05

Об усиленном законе больших чисел для линейных комбинаций конкомитантов

О. И. Дудкина[1], *Н. В. Грибкова*[1,2]

[1] Санкт-Петербургский государственный университет,
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9
[2] Петербургский государственный университет путей сообщения Императора Александра I,
Российская Федерация, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: *Дудкина О. И., Грибкова Н. В.* Об усиленном законе больших чисел для линейных комбинаций конкомитантов // Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия. 2020. Т. 7 (65). Вып. 3. С. 418–424.
https://doi.org/10.21638/spbu01.2020.305

В статье доказана теорема об усиленном законе больших чисел для линейных функций конкомитантов (индуцированных порядковых статистик) для последовательностей независимых одинаково распределенных двумерных случайных векторов. Результат дополняет предшествующие работы Янга (1981), Грибковой и Зитикиса (2017, 2019). Доказательство базируется на свойстве условной независимости конкомитантов Бхаттачариа (1974), применяется усиленный закон больших чисел для функций порядковых статистик ван Цвета (1980), классические неравенства, в том числе неравенство Розенталя (1970).

Ключевые слова: конкомитанты, индуцированные порядковые статистики, линейные комбинации, усиленный закон больших чисел.

1. Введение. Пусть $\{(X_n, Y_n)\}_{n=1}^{\infty}$ — последовательность независимых одинаково распределенных двумерных случайных векторов, $X_{1:n} \le \cdots \le X_{n:n}$ — порядковые статистики, соответствующие первым n элементам первой компоненты последовательности. Мы будем предполагать, что первая компонента вектора имеет непрерывное распределение, тогда с вероятностью 1 среди значений $X_{r:n}$, $r = 1, \ldots, n$, нет совпадающих, и мы можем однозначно определить конкомитант $Y_{[r:n]}$, соответствующий r-й порядковой статистике $X_{r:n}$, $r = 1, \ldots, n$, равенством $Y_{[r:n]} = Y_j$, если $X_{r:n} = X_j$.

Термин *конкомитант r-й порядковой статистики* был введен Г. Дэйвидом [1]. Ряд основополагающих для теории конкомитантов результатов был получен в работах Бхаттачариа [2, 3] (который предложил для $Y_{[r:n]}$ альтернативное название — *индуцированные порядковые статистики*), а также в работах [4, 5]. Многими авторами изучались свойства многомерных конкомитантов (см., напр., [6, 7]) и конкомитантов в последовательностях зависимых случайных величин (напр., [8]). В [9, 10] Егоровым и Невзоровым было предложено обобщающее определение индуцированных порядковых статистик, при котором конкомитанты $Y_{[r:n]}$ нумеруются в порядке возрастания значений $f(X_r)$, принимаемых некоторой измеримой функцией f. В этих работах были изучены свойства обобщенных конкомитантов, в частности доказана асимптотическая нормальность их нормированных сумм. Отметим, что частным случаем таких обобщенных конкомитантов являются, например, абсолютные порядковые статистики [9]. В завершение этого короткого обзора, упомянем работы Давыдова и Егорова [11, 12], в которых для последовательностей индуцированных порядковых статистик доказан ряд функциональных предельных теорем.

В этой заметке мы рассматриваем линейные комбинации индуцированных порядковых статистик

$$T_n = \frac{1}{n}\sum_{r=1}^{n} Y_{[r:n]} w\left(\frac{r}{n+1}\right) \tag{1.1}$$

для последовательности (X_n, Y_n), $n = 1, 2\ldots$, независимых наблюдений, где $w : [0,1] \to \mathbb{R}$ — борелевская функция весов. Мы доказываем результат о сходимости T_n к постоянному пределу с вероятностью 1. Наше исследование мотивировано предшествующими работами Янга [13–15] и Грибковой и Зитикиса [16, 17]. Линейные комбинации конкомитантов вида (1.1) находят применение в прикладной статистике. Так, в статьях [6, 14, 15] обсуждались различные приложения, например для оценки параметров регрессии. В работах Янга [14, 15] были доказаны варианты центральной предельной теоремы для нормированных статистик вида (1.1), однако закон больших чисел, т. е. сходимость статистики T_n к постоянному пределу (в том или ином вероятностном смысле), Янгом не рассматривался. В статьях Грибковой и Зитикиса [16, 17] линейные комбинации конкомитантов появлялись как составные части оценок ряда экономических и финансовых показателей (например, взвешенного бета Джини). При изучении асимптотических свойств этих оценок в [16, 17] был попутно доказан ряд утверждений о сходимости линейных комбинаций конкомитантов к пределу в смысле сходимости по вероятности, то есть слабый закон больших чисел. В этой работе мы доказываем сходимость с вероятностью 1. Насколько мы знаем, усиленный закон больших чисел для статистик T_n вида (1.1) никем ранее не изучался. Наше доказательство базируется на свойстве условной независимости конкомитантов [2], применяются классические неравенства, в частности неравенство Розенталя [18].

2. Основной результат. Введем обозначение $F_X(x) = \mathbf{P}(X \le x)$ для функции распределения случайной величины X и $F_X^{-1}(u) = \inf\{x : F_X(x) \ge u\}$, $u \in (0,1)$, — для ее непрерывной слева инверсии. Определим функции условного математического ожидания

$$g_{Y|X}(x) = \mathbf{E}[Y|X = x] \tag{2.1}$$

и условного абсолютного центрального момента порядка $\alpha > 0$

$$v^{\alpha}_{Y|X}(x) = \mathbf{E}\left[|Y - g_{Y|X}(x)|^{\alpha}|X = x\right], \tag{2.2}$$

которые будут корректно определены при всех $x \in \mathbb{R}$ в условиях представленной ниже теоремы. Будем использовать обозначение $\mathcal{L}^p[a,b]$ для класса функций, абсолютно интегрируемых в степени p в интервале $[a,b]$, $a < b$, и обозначение $f \circ h$ — для композиции функций f и h.

Основным результатом работы является следующая теорема.

Теорема. *Предположим, что функция F_X непрерывна и существуют $p, q \in [1,\infty]$, $\frac{1}{p} + \frac{1}{q} = 1$, такие что $v^{2+\delta}_{Y|X} \circ F_X^{-1} \in \mathcal{L}^p[0,1]$ и $|w|^{2+\delta} \in \mathcal{L}^q[0,1]$ для некоторого $\delta > 0$. Тогда*

$$T_n = \frac{1}{n}\sum_{r=1}^{n} Y_{[r:n]} w\left(\frac{r}{n+1}\right) \xrightarrow[n\to\infty]{п.н.} \mathbf{E}[Y\, w \circ F_X(X)], \tag{2.3}$$

где $F_X(X)$ — случайная величина (значение функции распределения F_X от случайного аргумента X), имеющая равномерное на $[0,1]$ распределение.

Доказательство. Представим T_n в виде суммы двух слагаемых:

$$T_n = T_{n,1} + T_{n,2}, \tag{2.4}$$

где

$$T_{n,1} = \frac{1}{n}\sum_{r=1}^{n} g_{Y|X}(X_{r:n}) w_{r,n},$$
$$T_{n,2} = \frac{1}{n}\sum_{r=1}^{n}(Y_{[r:n]} - g_{Y|X}(X_{r:n})) w_{r,n}.$$

Здесь и далее мы используем сокращенное обозначение $w_{r,n} := w\left(\frac{r}{n+1}\right)$. Первое слагаемое в правой части (2.4) (т. е. $T_{n,1}$) — это линейная комбинация функций порядковых статистик, и для нее выполнены условия усиленного закона больших чисел ван Цвета [19, следствие 2.1]. Действительно, из того, что $v^{2+\delta}_{Y|X} \circ F_X^{-1} \in \mathcal{L}^p[0,1]$ и $|w|^{2+\delta} \in \mathcal{L}^q[0,1]$, следует, что $g_{Y|X} \circ F_X^{-1} \in \mathcal{L}^p[0,1]$ и $w \in \mathcal{L}^q[0,1]$ с теми же $p, q \in [1,\infty]$, $\frac{1}{p} + \frac{1}{q} = 1$ (более того, нетрудно видеть, что $g_{Y|X} \circ F_X^{-1} \in \mathcal{L}^2[0,1]$ и $w \in \mathcal{L}^2[0,1]$, поэтому условия из [19] выполнены и с $p = q = 2$). Следовательно, по усиленному закону больших чисел ван Цвета имеем

$$T_{n,1} \xrightarrow[n\to\infty]{\text{п.н.}} \int_0^1 g_{Y|X} \circ F_X^{-1}(t) w(t) dt = \int_{-\infty}^{\infty} g_{Y|X}(x) w(F_X(x))\, dF_X(x) =$$
$$= \int_{-\infty}^{\infty} \mathbf{E}[Y w(F_X(X))|X = x]\, dF_X(x) = \mathbf{E}[Y w \circ F_X(X)]. \tag{2.5}$$

Для доказательства теоремы остается показать, что

$$T_{n,2} \xrightarrow[n\to\infty]{\text{п.н.}} 0. \tag{2.6}$$

Для этого достаточно проверить, что для любого $\varepsilon > 0$ выполняется неравенство

$$\sum_{n=1}^{\infty} \mathbf{P}\big(|T_{n,2}| > \varepsilon\big) < \infty. \tag{2.7}$$

Имеем

$$\mathbf{P}\big(|T_{n,2}| > \varepsilon\big) = \mathbf{E}\left[\mathbf{P}\left(\Big|\sum_{r=1}^{n}\big(Y_{[r:n]} - g_{Y|X}(X_{r:n})\big) w_{r,n}\Big|^{2+\delta} > (\varepsilon n)^{2+\delta} \Big| X_1, \ldots, X_n\right)\right]. \tag{2.8}$$

Оценим вероятность под знаком математического ожидания, применив неравенство Маркова:

$$\mathbf{P}\left(\Big|\sum_{r=1}^{n}\big(Y_{[r:n]} - g_{Y|X}(X_{r:n})\big) w_{r,n}\Big|^{2+\delta} > (\varepsilon n)^{2+\delta} \Big| X_1, \ldots, X_n\right) \le$$
$$\le \frac{\mathbf{E}\left[\big|\sum_{r=1}^{n}\big(Y_{[r:n]} - g_{Y|X}(X_{r:n})\big) w_{r,n}\big|^{2+\delta} \big| X_1, \ldots, X_n\right]}{(\varepsilon n)^{2+\delta}}. \tag{2.9}$$

Под знаком модуля в условном математическом ожидании стоит сумма условно независимых случайных величин [2, лемма 1]. Применим последовательно следствие неравенства Розенталя [18] и неравенство Гёльдера. Имеем

$$\frac{\mathbf{E}\left[\left|\sum_{r=1}^{n}\left(Y_{[r:n]}-g_{Y|X}(X_{r:n})\right)w_{r,n}\right|^{2+\delta}\Big|X_1,\ldots,X_n\right]}{(\varepsilon n)^{2+\delta}}\leq$$

$$\leq c(\varepsilon,\delta)\frac{n^{\frac{\delta}{2}}\sum_{r=1}^{n}v_{Y|X}^{2+\delta}(X_{r:n})|w_{r,n}|^{2+\delta}}{n^{2+\delta}}=$$

$$=c(\varepsilon,\delta)\frac{\frac{1}{n}\sum_{r=1}^{n}v_{Y|X}^{2+\delta}(X_{r:n})|w_{r,n}|^{2+\delta}}{n^{1+\frac{\delta}{2}}}\leq$$

$$\leq\frac{c(\varepsilon,\delta)}{n^{1+\frac{\delta}{2}}}\left[\frac{1}{n}\sum_{r=1}^{n}\left(v_{Y|X}^{2+\delta}(X_r)\right)^p\right]^{\frac{1}{p}}\left[\frac{1}{n}\sum_{r=1}^{n}\left(|w_{r,n}|^{2+\delta}\right)^q\right]^{\frac{1}{q}}. \quad (2.10)$$

Здесь и далее $c(\varepsilon,\delta)$ обозначает неотрицательную величину, зависящую от ε и δ и не зависящую от n, значение которой в последующем может изменяться. Из (2.8)–(2.10) вытекает, что для оценки сверху общего члена ряда (2.7) остается оценить математическое ожидание величины, стоящей в правой части (2.10).

Поскольку $1/p\leq 1$, то по неравенству Ляпунова имеем

$$\mathbf{E}\left[\frac{1}{n}\sum_{r=1}^{n}\left(v_{Y|X}^{2+\delta}(X_r)\right)^p\right]^{\frac{1}{p}}\leq\left[\mathbf{E}\left(\frac{1}{n}\sum_{r=1}^{n}\left(v_{Y|X}^{2+\delta}(X_r)\right)^p\right)\right]^{\frac{1}{p}}=$$

$$=\left[\mathbf{E}\left(v_{Y|X}^{2+\delta}(X_1)\right)^p\right]^{\frac{1}{p}}=\left[\mathbf{E}\left[v_{Y|X}^{2+\delta}\circ F_X^{-1}(U)\right]^p\right]^{\frac{1}{p}}=\parallel v_{Y|X}^{2+\delta}\circ F_X^{-1}\parallel_{\mathcal{L}^p[0,1]}, \quad (2.11)$$

где U — случайная величина, равномерно распределенная на $[0,1]$, и поскольку по условию $v_{Y|X}^{2+\delta}\circ F_X^{-1}\in\mathcal{L}^p[0,1]$, правая часть (2.11) конечна. Теперь заметим, что величина, стоящая в квадратных скобках в третьем сомножителе справа в (2.10), не случайна и является римановой суммой для интеграла $\int_0^1|w(t)|^{(2+\delta)q}\,dt$, который конечен, поскольку по условию теоремы $|w|^{2+\delta}\in\mathcal{L}^q[0,1]$. Таким образом, из наших оценок (2.8)–(2.11) следует неравенство

$$\mathbf{P}\big(|T_{n,2}|>\varepsilon\big)\leq\frac{c(\varepsilon,\delta)}{n^{1+\frac{\delta}{2}}},$$

которое означает, что ряд в (2.7) мажорируется сходящимся рядом

$$\sum_{n=1}^{\infty}\frac{c(\varepsilon,\delta)}{n^{1+\frac{\delta}{2}}}.$$

Следовательно, ряд в (2.7) сходится, что доказывает (2.6). Из (2.4)–(2.6) следует (2.3). Теорема доказана. □

Замечание. В заключение отметим, что хотя условия доказанной выше теоремы достаточны для большинства приложений, они не оптимальны. Действительно, если весовая функция $w\equiv 1$ на $[0,1]$, то линейная комбинация T_n становится средним $\overline{Y}$ вторых компонент вектора, и в этом случае моментные условия теоремы, очевидно, избыточны. Авторы планируют ослабить моментные предположения в следующих работах.

Мы благодарим двух анонимных рецензентов за внимание к работе и ряд полезных конструктивных замечаний.

Литература

1. *David H. A.* Concomitants of order statistics // Bull. Int. Statist. Inst. 1973. Vol. 45. P. 295–300.

2. *Bhattacharya P. K.* Convergence of sample paths of normalized sums of induced order statistics // Ann. Statist. 1974. Vol. 2, no. 5. P. 1034–1039.

3. *Bhattacharya P. K.* Induced order statistics: Theory and applications. In: Handbook of Statistics / Eds. by P. R. Krishnaiah, P. K. Sen. Elsevier, 1984. Vol. 4. Nonparametric Methods. P. 383–403. https://doi.org/10.1016/S0169-7161(84)04020-7

4. *David H. A., Galambos J.* The asymptotic theory of concomitants of order statistics // J. Appl. Probab. 1974. Vol. 11. P. 762–770.

5. *David H. A., Nagaraja H. N.* Concomitants of order statistics. In: Handbook of Statistics / Eds. by N. Balakrishnan, C. R. Rao. Elsevier, 1998. Vol. 16. Order Statistics: Theory and Methods. P. 487–513. https://doi.org/10.1016/S0169-7161(98)16020-0

6. *Barnett V., Green P. J., Robinson A.* Concomitants and correlation estimates // Biometrika. 1976. Vol. 63. P. 323–328.

7. *Wang X., Stokes S. L., Lim J., Chen M.* Concomitants of multivariate order statistics with application to judgment post-stratification // J. Amer. Statist. Assoc. 2006. Vol. 101. P. 1693–1704.

8. *Ke Wang M. S.* On concomitants of order statistics. PhD thesis. The Ohio State University, 2008.

9. *Егоров В. А., Невзоров В. Б.* Некоторые теоремы для индуцированных порядковых статистик // Теория вероятн. и ее примен. 1982. Т. 27, № 3. С. 592–599.

10. *Егоров В. А., Невзоров В. Б.* О скорости сходимости к нормальному закону сумм индуцированных порядковых статистик // Зап. научн. сем. ЛОМИ. 1981. Т. 108. Р. 45–56.

11. *Davydov Y., Egorov V.* Functional limit theorems for induced order statistics // Math. Meth. Stat. 2000. Vol. 9. P. 297–313.

12. *Davydov Y., Egorov V.* Functional CLT and LIL for induced order statistics. In: Asymptotic Methods in Probability and Statistics with Applications / Eds. by N. Balakrishnan, I. A. Ibragimov, V. B. Nevzorov. Boston, MA: Birkhäuser, 2001. P. 333–349 (Statistics for Industry and Technology). https://doi.org/10.1007/978-1-4612-0209-7_24

13. *Yang S.-S.* General distribution theory of the concomitants of order statistics // Ann. Statist. 1977. Vol. 5. P. 996–1002.

14. *Yang S.-S.* Linear combination of concomitants of order statistics with application to testing and estimation // Ann. Inst. Statist. Math. 1981. Vol. 33, no. 3. P. 463–470.

15. *Yang S.-S.* Linear functions of concomitants of order statistics with application to nonparametric estimation of a regression function // J. Amer. Statist. Assoc. 1981. Vol. 76. P. 658–662.

16. *Gribkova N., Zitikis R.* Statistical foundations for assessing the difference between the classical and weighted-Gini betas // Math. Meth. Stat. 2017. Vol. 26, no. 4. P. 267–281. https://doi.org/10.3103/S1066530717040020

17. *Gribkova N., Zitikis R.* Weighted allocations, their concomitant-based estimators, and asymptotics // Ann. Inst. Statist. Math. 2019. Vol. 71, no. 4. P. 811–835. https://doi.org/10.1007/s10463-018-0660-2

18. *Rosenthal H. P.* On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables // Israel J. Math. 1970. Vol. 8, no. 3. P. 273–303.

19. *van Zwet W. R.* A strong law for linear functions of order statistics // Ann. Probab. 1980. Vol. 8, no. 5. P. 986–990.

Статья поступила в редакцию 17 ноября 2019 г.;
после доработки 27 февраля 2020 г.;
рекомендована в печать 19 марта 2020 г.

Контактная информация:

Дудкина Ольга Игоревна — исследователь; ola.dudkina@gmail.com
Грибкова Надежда Викторовна — д-р физ.-мат. наук, проф.; nv.gribkova@gmail.com; n.gribkova@spbu.ru

On the strong law of large numbers for linear combinations of concomitants

O. I. Dudkina[1], *N. V. Gribkova*[1,2]

[1] St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation
[2] St. Petersburg State Transport University, 9, Moskovskiy pr., St. Petersburg, 190031, Russian Federation

For citation: Dudkina O. I., Gribkova N. V. On the strong law of large numbers for linear combinations of concomitants. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy*, 2020, vol. 7 (65), issue 3, pp. 418–424. https://doi.org/10.21638/spbu01.2020.305 (In Russian)

The article proves a theorem on the strong law of large numbers for linear functions of concomitants (induced order statistics) for sequences of independent identically distributed two-dimensional random vectors. The result complements previous work by Yang (1981), Gribkova and Zitikis (2017, 2019). The proof is based on the conditional independence property of concomitants Bhattacharya (1974), the strong law of large numbers for functions of order statistics by van Zwet (1980) is used, classical inequalities apply, including Rosenthal's (1970).

Keywords: concomitants, induced order statistics, linear combinations, strong law of large numbers.

References

1. David H. A., "Concomitants of order statistics", *Bull. Int. Statist. Inst.* **45**, 295–300 (1973).

2. Bhattacharya P. K., "Convergence of sample paths of normalized sums of induced order statistics", *Ann. Statist.* **2** (5), 1034–1039 (1974).

3. Bhattacharya P. K., "Induced order statistics: Theory and applications", in: *Handbook of Statistics* **4**, 383–403 (P. R. Krishnaiah, P. K. Sen (eds.), Elsevier, 1984). https://doi.org/10.1016/S0169-7161(84)04020-7

4. David H. A., Galambos J., "The asymptotic theory of concomitants of order statistics", *J. Appl. Probab.* **11**, 762–770 (1974).

5. David H. A., Nagaraja H. N., "Concomitants of order statistics", in: *Handbook of Statistics* **16**, 487–513 (N. Balakrishnan, C. R. Rao (eds.), Elsevier, 1998). https://doi.org/10.1016/S0169-7161(98)16020-0

6. Barnett V., Green P. J., Robinson A., "Concomitants and correlation estimates", *Biometrika* **63**, 323–328 (1976).

7. Wang X., Stokes S. L., Lim J., Chen M., "Concomitants of multivariate order statistics with application to judgment post-stratification", *J. Amer. Statist. Assoc.* **101**, 1693–1704 (2006).

8. Ke Wang M. S., *On concomitants of order statistics* (PhD thesis, The Ohio State University, 2008).

9. Egorov V. A., Nevzorov V. B., "Some theorems for induced order statistics", *Theory Probab. Appl.* **27** (3), 633–639 (1983). https://doi.org/10.1137/1127074

10. Egorov V. A., Nevzorov V. B., "Rate of convergence to the normal law of sums of induced order statistics", *J. Math. Sci.* **25** (3), 1139–1146 (1984). https://doi.org/10.1007/BF01084792

11. Davydov Y., Egorov V., "Functional limit theorems for induced order statistics", *Math. Meth. Stat.* **9**, 297–313 (2000).

12. Davydov Y., Egorov V., "Functional CLT and LIL for induced order statistics", in: *Asymptotic Methods in Probability and Statistics with Applications*, 333–349 (N. Balakrishnan, I. A. Ibragimov,

V. B. Nevzorov (eds.), Birkhäuser, Boston, MA, 2001, Statistics for Industry and Technology). https://doi.org/10.1007/978-1-4612-0209-7_24

13. Yang S.-S., "General distribution theory of the concomitants of order statistics", *Ann. Statist.* **5**, 996–1002 (1977).

14. Yang S.-S., "Linear combination of concomitants of order statistics with application to testing and estimation", *Ann. Inst. Statist. Math.* **33** (3), 463–470 (1981).

15. Yang S.-S., "Linear functions of concomitants of order statistics with application to nonparametric estimation of a regression function", *J. Amer. Statist. Assoc.* **76**, 658–662 (1981).

16. Gribkova N., Zitikis R., "Statistical foundations for assessing the difference between the classical and weighted-Gini betas", *Math. Meth. Stat.* **26** (4), 267–281 (2017). https://doi.org/10.3103/S1066530717040020

17. Gribkova N., Zitikis R., "Weighted allocations, their concomitant-based estimators, and asymptotics", *Ann. Inst. Statist. Math.* **71** (4), 811–835 (2019). https://doi.org/10.1007/s10463-018-0660-2

18. Rosenthal H. P., "On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables", *Israel J. Math.* **8** (3), 273–303 (1970).

19. van Zwet W. R., "A strong law for linear functions of order statistics", *Ann. Probab.* **8** (5), 986–990 (1980).

Received: November 17, 2019
Revised: February 27, 2020
Accepted: March 19, 2020

Authors' information:

Olga I. Dudkina — ola.dudkina@gmail.com
Nadezhda V. Gribkova — nv.gribkova@gmail.com; n.gribkova@spbu.ru