S©MR

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ
Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru
ISSN 1813-3304

Том 22, № 2, стр. 1394–1400 (2025) УДК 519.214
https://doi.org/10.33048/semi.2025.22.084 MSC 60F05

ОБ ОДНОМ УСИЛЕНИИ ТЕОРЕМЫ ГАЕКА–ШИДАКА

И.С. Борисов, Ю.Ю. Линке

Представлено Н.С. Аркашовым

Abstract: An analogue of the Hajek–Sidak theorem is proved on asymptotic normality of the distributions of sums of weighted independent identically distributed centered random variables with a finite second moment in the case where the normalizing coefficients of these sums are not constants but random variables.
Keywords: Hajek-Sidak central limit theorem, series diagram, random weighting coefficient.

1 Введение и основной результат

В книге Гаека и Шидака [1] имеется раздел «Специальный случай центральной предельной теоремы» (см. главу 5, раздел 1.2), в котором приводится утверждение о слабой сходимости к нормальному закону суммы взвешенных независимых одинаково распределенных случайных величин в случае, когда так называемый треугольный массив коэффициентов, участвующий в указанной сумме, состоит из постоянных. Цель данной заметки — доказать аналог указанного утверждения

Borisov, I.S., Linke, Y.Y., On a strengthening of the Hajek–Sidak theorem.
© 2025 Борисов И.С., Линке Ю.Ю.
Работа выполнена при поддержке программы фундаментальных научных исследований СО РАН (проект FWNF-2024-0001).

Поступила 25 сентября 2025 г., опубликована 25 ноября 2025 г.

в ситуации, когда треугольный массив составляют случайные величины, т.е. этот массив взвешивающих коэффициентов образует «схему серий». Утверждение такого типа нам потребовалось в задачах регрессии. В частности, при доказательстве асимптотической нормальности некоторых явных оценок в задачах нелинейной регрессии и универсальных локально–постоянных оценок в непараметрической регрессии (см. [3] и [4]). Данный результат может быть использован и в других задачах регрессионного анализа.

Перейдем к точным формулировкам. Приведем прежде всего обсуждаемую теорему Гаека–Шидака (см. приводимую далее теорему 1). Нам удобнее использовать специализированную формулировку этой теоремы, предложенную в [2]. Условимся, что всюду в дальнейшем все пределы, если не оговорено иное, берутся при $n \to \infty$. Символ $\mathcal{N}(a, \sigma^2)$ обозначает гауссовскую случайную величину с параметрами a и σ^2, а запись вида $\zeta_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ означает слабую сходимость распределений.

Теорема 1. *Пусть ξ_1, $\xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с нулевым средним и конечным вторым моментом $\sigma^2 = \mathbb{E}\xi_1^2$. Кроме того, имеется такой треугольный массив коэффициентов c_{nk}, $k = 1, \ldots, n$, $n = 1, 2, \ldots$, что выполнены условия*

$$\max_{1 \le k \le n} |c_{nk}| \to 0, \qquad \sum_{k=1}^{n} c_{nk}^2 = 1 \quad \text{при всех} \quad n.$$

Тогда распределения последовательности сумм $\sum_{k=1}^{n} c_{nk}\xi_k$, $n = 1, 2, \ldots$, асимптотически нормальны:

$$\sum_{k=1}^{n} c_{nk}\xi_k \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Как отмечается в [2], случайные величины $c_{nk}\xi_k$, $k = 1, \ldots, n$, $n = 1, 2, \ldots$, образуют «схему серий» специального вида, в которой распределения участвующих в ней величин отличаются только параметрами масштаба.

Приведем теперь предлагаемый нами аналог теоремы 1 в ситуации случайных нормирующих коэффициентов. Справедлива

Теорема 2. *Пусть ξ_1, $\xi_2, \ldots$ — последовательность независимых одинаково распределенных центрированных случайных величин с конечным вторым моментом $\sigma^2 = \mathbb{E}\xi_1^2$, а случайные величины a_{nk}, $k = 1, \ldots, n$, $n = 1, 2, \ldots$, не зависящие от ξ_i, $i = 1, 2, \ldots$, удовлетворяют следующим двум условиям*:

$$\max_{1 \le k \le n} |a_{nk}| \xrightarrow{p} 0, \qquad \sum_{i=1}^{n} a_{nk}^2 = 1 \quad \text{при} \quad \text{всех} \quad n. \tag{1}$$

Тогда имеет место предельное соотношение

$$\sum_{k=1}^{n} a_{nk}\xi_k \xrightarrow{d} \mathcal{N}(0,\sigma^2).$$

Замечание 1. *Утверждение теоремы 2 останется справедливым при замене второго условия в (1) соотношением*

$$\sum_{i=1}^{n} a_{nk}^2 = 1 + \widetilde{o}_p(1),$$

где $\widetilde{o}_p(1)$ — случайная величина, сходящаяся к нулю по вероятности так, что $\sup|\widetilde{o}_p(1)| < \infty$ и $\sup$ берется по всем n и всем элементарным исходам.

Приведем один из примеров, относящийся к классической задаче оценивания регрессионной функции в непараметрической регрессии методом ядерного сглаживания, когда может возникнуть необходимость исследовать асимптотическое поведение рассматриваемых в теореме 2 сумм. Предположим, что наблюдения X_k, $k = 1,\dots,n$, представимы в виде

$$X_k = f(z_k) + g(z_k)\xi_k, \qquad k = 1,\dots,n,$$

где неизвестная функция $f(t)$, $t \in [0,1]$, непрерывна и подлежит оцениванию, детерминированные или случайные величины $\{z_k\}$ (регрессоры) нам известны, погрешности $\{\xi_k\}$ — ненаблюдаемые независимые одинаково распределенные случайные величины с нулевым средним и конечной дисперсией, не зависящие от регрессоров $\{z_k\}$, неизвестная функция $g(t)$, $t \in [0,1]$, непрерывна и положительна. В такой модели непараметрической регрессии весьма популярны методы ядерного сглаживания (см., например, [5]–[7]). Оценка Надарая–Ватсона для регрессионной функции f имеет следующую структуру:

$$f^*_{n,h}(t) = \frac{\sum_{k=1}^{n} X_k K_h(t-z_k)}{\sum_{k=1}^{n} K_h(t-z_k)},$$

где $K_h(x) = h^{-1}K(h^{-1}x)$ и $K(x)$ — некоторая ядерная функция (например, плотность симметричного распределения с носителем $[-1,1]$), а $h = h_n \to 0$ — размер окна; при этом знаменатель у приведенной оценки отличен от нуля при всех n. Нетрудно видеть, что справедливо следующее представление:

$$B_{n,h}^{-1}\bigl(f^*_{n,h}(t) - f(t) - r_{n,h}\bigr) =$$
$$= \left(\sum_{i=1}^{n} K_h^2\,(t-z_i)\,g^2(z_i)\right)^{-1/2} \sum_{k=1}^{n} K_h(t-z_k) g(z_k)\xi_k,$$

где

$$r_{n,h}(t) = \left(\sum_{i=1}^{n} K_h(t - z_i)\right)^{-1} \sum_{i=1}^{n} (f(z_i) - f(t)) K_h(t - z_i),$$
$$B^2_{n,h}(t) = \left(\sum_{i=1}^{n} K_h(t - z_i)\right)^{-2} \sum_{i=1}^{n} K_h^2\,(t - z_i)\, g^2(z_k).$$

Таким образом, доказательство асимптотической нормальности оценки Надарая–Ватсона эквивалентно доказательству аналогичного соотношения для суммы $\sum\limits_{k=1}^{n} a_{nk}\xi_k$ при

$$a_{nk} = \frac{K_h(t - z_k) g(z_k)}{\left(\sum_{i=1}^{n} K_h^2\,(t - z_i)\, g^2(z_k)\right)^{1/2}}.$$

Отметим, что в силу теоремы 2 для асимптотической нормальности оценки Надарая–Ватсона в точке t нужно требовать следующее условие равномерной малости:

$$\frac{\max_{k \le n} K_h^2(t - z_k) g^2(z_k)}{\sum_{i=1}^{n} K_h^2\,(t - z_i)\, g^2(z_k)} \xrightarrow{p} 0.$$

Подчеркнем, что это условие выполнено при весьма широких ограничениях на регрессоры и не требует, например, выполнения тех или иных форм слабой зависимости или регулярности $\{z_i\}$. Отметим, что если это условие выполнено для любого фиксированного $t \in [0, 1]$ (другими словами, оценка Надарая–Ватсона асимптотически нормальна для любой фиксированной точки из области определения f), то относительно регрессоров $\{z_i\}$ можно утверждать, что выполнено лишь условие плотного заполнения ими области задания регрессионной функции. Это по существу необходимое ограничение на регрессоры более слабое по сравнению с известными ранее в данной модели. В этой связи, не стремясь привести подробную библиографию, укажем, например, монографии [8]–[12] и работы [6], [7], [13]–[16], в которых исследуются вопросы асимптотической нормальности тех или иных ядерных оценок, включая оценки Надарая–Ватсона и их модификации.

2 Доказательство теоремы 2

Пусть $\mathcal{F}_n$ — σ-алгебра, порожденная набором случайных величин a_{nk}, $k = 1, \ldots, n$. Обозначим символом $\mathbb{E}_{\mathcal{F}_n}$ условное математическое ожидание при фиксации этой σ-алгебры и рассмотрим характеристическую функцию случайной величины $\eta_n = \sum_{k=1}^n a_{nk}\xi_k$. Для любого фиксированного действительного s имеем

$$\mathbb{E}e^{is\eta_n} = \mathbb{E}\mathbb{E}_{\mathcal{F}_n}\exp\left\{is\sum_{k=1}^n a_{nk}\xi_k\right\} = \mathbb{E}\mathbb{E}_{\mathcal{F}_n}\prod_{k=1}^n e^{isa_{nk}\xi_k} =$$

$$= \mathbb{E}\mathbb{E}_{\mathcal{F}_n}\prod_{k=1}^n\left[1 + isa_{nk}\xi_k - \frac{s^2a_{nk}^2\xi_k^2}{2}\left(1 + 2\int_0^1(1-u)\big(e^{iusa_{nk}\xi_k} - 1\big)du\right)\right] =$$

$$= \mathbb{E}\prod_{k=1}^n\left[1 - \frac{s^2\sigma^2a_{nk}^2}{2}\left(1 + \tau_s\left(a_{nk}\right)\right)\right], \qquad (2)$$

где

$$\tau_s(x) = 2\sigma^{-2}\mathbb{E}\left(\xi_1^2\int_0^1(1-u)(e^{isxu\xi_1} - 1)du\right). \qquad (3)$$

При выводе соотношения (2) мы использовали формулу Тейлора для функции e^{iz} при $z = sa_{nk}\xi_k$ с остаточным членом в интегральном виде, а именно

$$e^{iz} = 1 + iz - z^2\int_0^1(1-u)e^{izu}du =$$
$$= 1 + iz - \frac{z^2}{2}\left(1 + 2\int_0^1(1-u)\big(e^{izu} - 1\big)du\right).$$

Нам потребуется следующее неравенство для функции $\tau_s(x)$:

$$|\tau_s(x)| \le 2\sigma^{-2}\mathbb{E}\left\{\xi_1^2\int_0^1(1-u)\min\{|sxu\xi_1|, 2\}du\right\}.$$

Это соотношение очевидным образом следует из определения (3) и неравенства $\left|e^{ix} - 1\right| \le \min\{|x|,\ 2\}$ для любого $x \in \mathbb{R}$. В частности, $\tau_s(x) \to 0$ при $x \to 0$ в силу теоремы Лебега. Кроме того,

$$\sup_{x,s}|\tau_s(x)| \le 2, \qquad (4)$$

поскольку $\int_0^1(1-u)\min\{|sxu\xi_1|, 2\}du \le 1$.

Вернемся к изучению асимптотического поведения характеристической функции $\mathbb{E}e^{is\eta_n}$. Нам потребуется следующий хорошо известный вариант теоремы Лебега: если $X_n \xrightarrow{p} c_0$ и для любого $n \ge 1$ выполнено $|X_n| \le C$ для некоторой константы $C < \infty$, то $\mathbb{E}X_n \to c_0$. При этом случайные величины X_n и постоянная c_0 могут быть комплекснозначными. Обозначим через X_n произведение, стоящее под знаком математического ожидания в правой части (2), и воспользуемся следующим представлением:

$$\prod_{k=1}^{n}(1+b_{nk}) = \exp\left\{\sum_{k=1}^{n}\log(1+b_{nk})\right\} =$$
$$= \exp\left\{\sum_{k=1}^{n} b_{nk} - \sum_{k=1}^{n} b_{nk}^2 \int\limits_0^1 \frac{1-u}{(1+b_{nk}u)^2}du\right\} \tag{5}$$

при $b_{nk} = ca_{nk}^2(1+\tau_s(a_{nk}))$ и $c = -s^2\sigma^2/2$. Здесь мы опять использовали разложение Тейлора для функции $\varphi(z) = \log(1+z)$ с комплекснозначной переменной z:

$$\varphi(z) = \varphi(0) + \varphi'(0)z + z^2\int\limits_0^1 (1-u)\varphi''(uz)du.$$

Поскольку по условию теоремы $\max_k |a_{nk}| \xrightarrow{p} 0$, то в силу свойств функции $\tau_s(\cdot)$ выполнено

$$\max_{1\le k\le n} |\tau_s(a_{nk})| \xrightarrow{p} 0 \quad \text{и} \quad \max_{1\le k\le n} |b_{nk}| \xrightarrow{p} 0.$$

Стало быть, интегралы в правой части (5) определены корректно на множестве асимптотически полной меры. Учитывая теперь второе условие в (1), заключаем, что

$$\sum_{k=1}^{n} b_{nk} = c + o_p(1), \qquad \sum_{k=1}^{n} b_{nk}^2 = o_p(1).$$

Таким образом, в силу (5)

$$X_n = \prod_{k=1}^{n}\left[1 + ca_{nk}^2\bigl(1+\tau_s(a_{nk})\bigr)\right] \equiv \prod_{k=1}^{n}(1+b_{nk}) = \exp\left\{c + o_p(1)\right\}.$$

Легко видеть, что при всех n с учетом (1) и (4)

$$\left|\sum_{k=1}^{n} a_{nk}^2\tau_s(a_{nk})\right| \le 2. \tag{6}$$

Таким образом, с учетом (4), (6) и неравенства $|1+z| \le 1+|z| \le e^{|z|}$, справедливого при всех комплекснозначных z, при всех $n \ge 1$ мы имеем

$$|X_n| \le \exp\{3|c|\}.$$

Из полученных неравенств и предельных соотношений немедленно получаем, что $X_n \xrightarrow{p} c_0 \equiv e^{-s^2\sigma^2/2}$ и $\sup_n |X_n| \le C$ при некотором $C < \infty$. Таким образом, при любом фиксированном действительном s выполнено $\mathbb{E}e^{is\eta_n} \equiv \mathbb{E}X_n \to e^{-s^2\sigma^2/2}$. Теорема 2 доказана.

References

[1] J. Hájek, Z. Šidák, *Theory of rank tests*, Academic Press, New York-London; Academia, Publishing House of the Czechoslovak Academy of Sciences, Prague, 1967. Zbl 0161.38102

[2] D.M. Chibisov, *Lectures on the asymptotic theory of rank tests*, Lektsionnye Kursy NOTs, **14**, Matematicheskiĭ Institut im. V.A. Steklova, RAN, Moscow, 2009. Zbl 06235250

[3] Yu.Yu. Linke, I.S. Borisov, *Constructing explicit estimators in nonlinear regression models*, Theory Probab. Appl., **63**:1 (2018), 22–44. Zbl 1404.62071

[4] I.S. Borisov, Yu.Yu. Linke, P.S. Ruzankin, *Universal weighted kernel-type estimators for some class of regression models*, Metrika, **84**:2 (2021), 141–166. Zbl 1461.62046

[5] Y. Shen, C. Gao, D. Witten, F. Han, *Optimal estimation of variance in nonparametric regression with random design*, Ann. Stat., **48**:6 (2020), 3589–3618. Zbl 1460.62053

[6] Q. Xie, Q. Sun, J. Liu, *Local weighted composite quantile estimation and smoothing parameter selection for nonparametric derivative function*, Econom. Rev., **39**:3 (2020), 215–233. Zbl 1490.62103

[7] B. Kai, R. Li, H. Zou, *Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression*, J. R. Stat. Soc., Ser. B, Stat. Methodol., **72**:1 (2010), 49–69. Zbl 1411.62101

[8] W. Härdle, *Applied nonparametric regression*, Econometric Society Monographs, **19**, Cambridge University Press, Cambridge, 1990. Zbl 0714.62030

[9] J. Fan, *Local polynomial modelling and its applications*, Monographs on Statistics and Applied Probability, **66**, Chapman & Hall, London, 1996. Zbl 0873.62037

[10] J. Fan, Q. Yao, *Nonlinear time series. Nonparametric and parametric methods*, Springer, New York, 2003. Zbl 1014.62103

[11] W. Härdle, M. Müller, S. Sperlich, A. Werwatz, *Nonparametric and semiparametric models*, Springer, Berlin, 2004. Zbl 1059.62032

[12] I. Horová, J. Koláček, J. Zelinka, *Kernel smoothing in MATLAB. Theory and practice of kernel smoothing*, World Scientific, Hackensack, 2012. Zbl 1273.62019

[13] O.B. Linton, D.T. Jacho-Chávez, *On internally corrected and symmetrized kernel estimators for nonparametric regression*, TEST, **19**:1 (2010), 166–186. Zbl 1203.62070

[14] A.A. Georgiev, *Asymptotic properties of the multivariate Nadaraya-Watson regression function estimate: the fixed design case*, Stat. Probab. Lett., **7**:1 (1988), 35–40. Zbl 0662.62038

[15] Q. Zheng, C. Gallagher, K.B. Kulasekera, *Adaptively weighted kernel regression*, J. Nonparametric Stat., **25**:4 (2013), 855–872. Zbl 1416.62235

[16] S. Dhar, P. Jha, P. Rakshit, *The trimmed mean in non-parametric regression function estimation*, Theory Probab. Math. Stat., **107** (2022), 133–158. Zbl 7618876

Igor Semenovich Borisov
Sobolev Institute of Mathematics,
pr. Koptyuga, 4,
630090, Novosibirsk, Russia
Email address: sibam@math.nsc.ru

Yuliana Yurievna Linke
Sobolev Institute of Mathematics,
pr. Koptyuga, 4,
630090, Novosibirsk, Russia
Email address: linke@math.nsc.ru