S©MR

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

Том , №, стр. – () УДК 519.23
https://doi.org/ MSC 62F03, 62F05

АНАЛОГ КРИТЕРИЯ ХАРКЕ–БЕРА ДЛЯ СЛУЧАЯ, КОГДА ПРОВЕРЯЕТСЯ ГИПОТЕЗА О РАСПРЕДЕЛЕНИИ СТЬЮДЕНТА

А.В. Логачёв С.Е. Хрущев

Представлено Н.С. Аркашовым

Abstract: The Jarque–Bera test is commonly used in statistics and econometrics to test the hypothesis that sample elements adhere to a normal distribution with an unknown mean and variance. This paper proposes a modification of this criterion, which allows for testing hypotheses concerning the case where the sample comes from a skewed Student's distribution with an unknown number of degrees of freedom. The paper also provides power estimates for the constructed criterion, obtained through simulation of samples of various sizes.

Keywords: Jarque–Bera test; Student's distribution; Monte Carlo simulation.

1 Введение

Будем считать, что все рассматриваемые случайные элементы заданы на некотором вероятностном пространстве $(\Omega, \mathfrak{F}, \mathbf{P})$. Пусть случайные величины $X, X_1, \ldots, X_n$, $n \in \mathbb{N}$ независимы и одинаково распределены. Предполагается, что их распределение неизвестно.

Logachov, A.V., Khrushchev, S.E., Jarque–Bera Criterion Analog for Testing the Hypothesis of Student's Distribution.

Работа выполнена при поддержке РНФ, проект 24-21-00087.

Поступила ? декабря 2024 г., опубликована ? ? ?? г.

C. Jarque и A.K. Bera (см., например, [2]–[4]) предложили следующий критерий. Проверяется гипотеза

$H_0: \ X_i \sim N(a, \sigma^2)$, $1 \le i \le n$, $(a, \sigma^2$ неизвестны) против гипотезы

$H_1 : \ X_i$ имеет распределение из семейства распределений Пирсона отличное от нормального.

Используется статистика

$$JB = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right),$$

где n — объем выборки; S и K это выборочные асимметрия и эксцесс, соответственно, т.е.

$$S = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^3}{\hat{\sigma}^3}, \quad K = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^4}{\hat{\sigma}^4},$$

$$\hat{\sigma} = \left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right)^{\frac{1}{2}}, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Если верна гипотеза H_0, то при $n \to \infty$ распределение случайной величины JB сходится к $\chi^2(2)$. Поэтому при достаточно большом объеме выборки можно использовать следующее решающее правило, задаем p– уровень значимости, если $JB < \chi^2_{1-p}(2)$ (здесь $\chi^2_{1-p}(2)$ квантиль уровня $1-p$ распределения $\chi^2(2)$), то гипотеза H_0 принимается, иначе отвергается.

Заметим, что семейство распределений Пирсона весьма богато, в него входят, например, такие распределения, как показательное, гамма, бета, Стьюдента, нормальное. Если известно, что случайная величина имеет распределение из семейства Пирсона, то конкретный его тип определяется значениями асимметрии и эксцесса (см. [5]). Благодаря этому свойству критерий Харке – Бера является критерием согласия на этом семействе (т.е. если для элементов выборки справедлива гипотеза H_1, состоящая в том, что элементы выборки имеют отличное от нормального распределение из семейства Пирсона, то статистика JB стремится по вероятности к ∞ при $n \to \infty$). Отметим, что этот критерий на практике часто применяется эконометристами для проверки гипотезы о нормальности распределения случайных возмущений в моделях линейной регрессии. Также отметим работу [6], в которой рассмотрены модификации критерия Харке – Бера для случаев, когда: известно математическое ожидание и неизвестна дисперсия; неизвестно математическое ожидание и известна дисперсия; известны и математическое ожидание и дисперсия.

В данной работе мы рассмотрим аналогичный критерий, который позволяет проверять гипотезу о том, что элементы выборки имеют, вообще говоря, смещенное распределение Стьюдента. Как и в классическом критерии Харке – Бера, построенная статистика использует свойства третьего и четвертого моментов, а также равенство нулю нечетных моментов случайных величин, имеющих распределение Стьюдента.

Следует отметить, что предложенный критерий также может быть использован для проверки гипотезы о том, что в моделях линейной регрессии случайные возмущения имеют распределение Стьюдента. В частности, при проведении статистического анализа медицинских данных часто используется классическая линейная модель с нормальным распределением ошибок. Это объясняется простотой и удобством применения данной модели, а также наличием специализированных статистических пакетов, необходимых для выполнения расчетов. При этом проверка выполнения предпосылки о том, что случайные возмущения распределены нормально, зачастую игнорируется. В тех случаях, когда ошибки имеют распределение Стьюдента, диагностика модели будет значительно отличаться от диагностики классических линейных моделей с нормальными ошибками (см., например, [1]). Таким образом, обоснованный выбор модели при анализе реальных данных с учетом проверки распределения ошибок является важным этапом при проведении статистического анализа.

Оставшаяся часть работы состоит из четырех разделов. Во втором разделе приводятся основные результаты (предельная теорема и, как следствие из нее, критерий проверки соответствующих статистических гипотез). Третий раздел посвящен апробации построенного статистического критерия с помощью симуляций. Четвертый раздел посвящен доказательствам результатов раздела 2, пятый раздел содержит вспомогательные результаты.

Везде далее символ $\xrightarrow[n\to\infty]{d}$ будет означать сходимость по распределению; $\mathbf{E}$ и $\mathbf{D}$ будут означать математическое ожидание и дисперсию, соответственно. Также будем использовать обозначения $a := \mathbf{E}X$, $\sigma^2 := \mathbf{D}X$.

2 Основные результаты

Положим

$$A_{1,n} := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^3, \quad A_{2,n} := \frac{1}{n}\sum_{i=1}^{n}\left((2-\hat{\sigma}^2)(X_i - \overline{X})^4 - 3\hat{\sigma}^4\right),$$

Теорема 1. *Пусть случайные величины $X, X_1, \ldots, X_n$, $n \in \mathbb{N}$ независимы, одинаково распределены и удовлетворяют условиям*

1) $\mathbf{E}X^8 < \infty$*;*

2) $\mathbf{E}(X-a)^{2k+1} = 0,\ 1 \le k \le 3$*;*

3) $(2-\sigma^2)\mathbf{E}(X-a)^4 - 3\sigma^4 = 0$.
Тогда последовательность случайных векторов $\sqrt{n}(A_{1,n}, A_{2,n})$ *слабо сходится к случайному вектору* (W_1, W_2)*, который имеет нормальное распределение с* $\mathbf{E}W_1 = \mathbf{E}W_2 = 0$ *и ковариационной матрицей*

$$\Sigma = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix},$$

где

$$D_1 = \mathbf{E}\left((X-a)^3 - 3(X-a)\sigma^2\right)^2,$$

$$D_2 = \mathbf{E}\left((2-\sigma^2)(X-a)^4 - (6\sigma^2 + \mathbf{E}(X-a)^4)(X-a)^2 + \sigma^2\mathbf{E}(X-a)^4 + 3\sigma^4\right)^2.$$

Напрямую из теоремы 1 следует

Теорема 2. *Пусть выполнены условия теоремы 1, тогда*

$$n\left(\frac{A_{1,n}^2}{D_1} + \frac{A_{2,n}^2}{D_2}\right) \xrightarrow[n\to\infty]{d} \xi \sim \chi^2(2).$$

Заметим, что если случайная величина X имеет смещенное распределение Стьюдента с девятью и более степенями свободы (т.е. плотность распределения имеет вид $f(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{\pi k}\Gamma(\frac{k}{2})}(1 + \frac{(x-a)^2}{k})^{-\frac{k+1}{2}}$, $k \geq 9$), то выполнены условия теоремы 1 и

$$\mathbf{E}(X-a)^4 = \frac{3\sigma^4}{2-\sigma^2}, \quad \mathbf{E}(X-a)^6 = \frac{15\sigma^6}{(2-\sigma^2)(3-2\sigma^2)},$$

$$\mathbf{E}(X-a)^8 = \frac{105\sigma^8}{(2-\sigma^2)(3-2\sigma^2)(4-3\sigma^2)}.$$

В этом случае

$$D_1 = \mathbf{E}\left((X-a)^3 - 3(X-a)\sigma^2\right)^2 = \frac{15\sigma^6}{(2-\sigma^2)(3-2\sigma^2)} - \frac{18\sigma^6}{2-\sigma^2} + 9\sigma^2,$$

$$\begin{aligned} D_2 &= \mathbf{E}\left((2-\sigma^2)(X-a)^4 - (6\sigma^2 + \mathbf{E}(X-a)^4)(X-a)^2 + \sigma^2\mathbf{E}(X-a)^4 + 3\sigma^4\right)^2 \\ &= \mathbf{E}\left((2-\sigma^2)(X-a)^4 - \frac{12\sigma^2 - 3\sigma^4}{2-\sigma^2}(X-a)^2 + \frac{6\sigma^4}{2-\sigma^2}\right)^2 \\ &= \sigma^8\left(\frac{105(2-\sigma^2)}{(3-2\sigma^2)(4-3\sigma^2)} + \frac{27\sigma^4 - 180\sigma^2 + 360}{(2-\sigma^2)^3} - \frac{90(4-\sigma^2)}{(2-\sigma^2)(3-2\sigma^2)}\right) \end{aligned}$$

и справедливо следующее следствие из теоремы 2

Следствие 1. *Пусть случайные величины* $X, X_1, \ldots, X_n$, $n \in \mathbb{N}$ *независимы и имеют смещенное распределение Стьюдента* $T_a(k)$ *с* $k \geq 9$ *(возможно* $k = \infty$*, т.е. нормальное распределение* $N(a,1)$*) степенями свободы. Тогда*

$$n\left(\frac{A_{1,n}^2}{B_1} + \frac{A_{2,n}^2}{B_2}\right) \xrightarrow[n\to\infty]{d} \xi \sim \chi^2(2),$$

где

$$B_1 := \hat{\sigma}^6 \left(\frac{15}{(2-\hat{\sigma}^2)(3-2\hat{\sigma}^2)} + 9 - \frac{18}{2-\hat{\sigma}^2} \right),$$

$$B_2 := \hat{\sigma}^8 \left(\frac{105(2-\hat{\sigma}^2)}{(3-2\hat{\sigma}^2)(4-3\hat{\sigma}^2)} + \frac{27\hat{\sigma}^4 - 180\hat{\sigma}^2 + 360}{(2-\hat{\sigma}^2)^3} - \frac{90(4-\hat{\sigma}^2)}{(2-\hat{\sigma}^2)(3-2\hat{\sigma}^2)} \right).$$

Положим

$$\widehat{JB} := n \left(\frac{A_{1,n}^2}{|B_1|} + \frac{A_{2,n}^2}{|B_2|} \right).$$

Из следствия 1 следует следующий статистический критерий.

Следствие 2. *Пусть задан уровень значимости p.*

Проверяется гипотеза $H_0 : X_i \sim T_a(k)$, $1 \le i \le n$, $k \ge 9$ (математическое ожидание a неизвестно, число степеней k неизвестно) против альтернативы $H_1 : X_i$ имеет другое распределение.

Из следствия 1 следует, что при достаточно большом объеме выборки можно использовать правило: если $\widehat{JB} < \chi^2_{1-p}(2)$, то гипотеза H_0 принимается, иначе отвергается.

Замечание 1. *Отметим, что приведенное выше решающее правило является критерием согласия, в случае "более простой" конкурирующей гипотезы $H_1 : X_i$ имеет распределение такое, что или $\mathbf{E}(X_i - a)^3 \neq 0$ или $(2-\sigma^2)\mathbf{E}(X_i - a)^4 - 3\sigma^4 \neq 0$.*

3 Моделирование

В этом разделе мы проверяем эффективность работы построенного теста для проверки гипотезы о том, что выборка взята из несмещенного распределения Стьюдента $T(k)$. Эффективность проверялась с помощью моделирования альтернативных гипотез методом Монте-Карло. Моделирование проводилось в программном обеспечении R версии 4.2.3. Мы использовали выборки следующих размеров (малые, средние и большие): $n =$25, 50, 75, 100, 150, 200, 250, 500 и 1000 для случая сложной основной гипотезы ($H_0 : X_i \sim T(k)$, где число степеней свободы неизвестно). Всюду при тестировании мы использовали универсальное значение уровня значимости $p = 0,05$.

Для расчета оценки мощности (эмпирической мощности) мы генерировали 1000 выборок с указанным размером n, взятых из распределения, соответствующего альтернативной гипотезе. Эмпирическая мощность — это доля отклоненных критерием основных гипотез. В таблицах ниже приведены результаты симуляций, где всюду в столбцах H_a указано моделируемое распределение. Отметим, что в Таблице 1 в столбце "α" указана эмпирическая вероятность ошибки первого рода, в силу того, что моделируемое распределение соответствует основной гипотезе. Таким образом, в Таблице 1 в некотором смысле продемонстрирована скорость сходимости к предельному распределению тестовой статистики из Теоремы 2.

n	H_a	α	H_a	α	H_a	α	H_a	α
25	$T(10)$	0,086	$T(15)$	0,085	$T(20)$	0,072	$N(0,1)$	0,088
50	$T(10)$	0,053	$T(15)$	0,081	$T(20)$	0,080	$N(0,1)$	0,074
75	$T(10)$	0,055	$T(15)$	0,070	$T(20)$	0,067	$N(0,1)$	0,091
100	$T(10)$	0,059	$T(15)$	0,061	$T(20)$	0,074	$N(0,1)$	0,072
150	$T(10)$	0,043	$T(15)$	0,074	$T(20)$	0,071	$N(0,1)$	0,071
200	$T(10)$	0,056	$T(15)$	0,067	$T(20)$	0,054	$N(0,1)$	0,064
250	$T(10)$	0,057	$T(15)$	0,059	$T(20)$	0,050	$N(0,1)$	0,058
500	$T(10)$	0,054	$T(15)$	0,061	$T(20)$	0,068	$N(0,1)$	0,051
1000	$T(10)$	0,049	$T(15)$	0,051	$T(20)$	0,052	$N(0,1)$	0,053

Таблица 1. Эмпирическое значение вероятности ошибки первого рода α для сложной основной гипотезы $H_0 : X_i \sim T(k)$, где $N(0,1)$ — стандартное нормальное распределение.

Анализ Таблицы 2 говорит о том, что уже при малых объемах выборки построенный критерий обладает хорошей мощностью, особенно хорошо работает для альтернативных распределений $N(0,2)$ (из семейства распределения Пирсона) и равновесной смеси распределений $N(0,1)$ и $N(0,3)$ (близкое, но не принадлежащему к семейству распределений Пирсона). Равномерное распределение на интервале $(-3;3)$ в известном смысле также можно отнести к семейству распределений Пирсона (для него оно является предельным).

n	H_a	Мощность	H_a	Мощность	H_a	Мощность
25	$N(0,2)$	0,807	Mix	0,884	$U(-3;3)$	0,390
50	$N(0,2)$	0,939	Mix	0,980	$U(-3;3)$	0,636
75	$N(0,2)$	0,985	Mix	0,998	$U(-3;3)$	0,762
100	$N(0,2)$	0,999	Mix	1,000	$U(-3;3)$	0,867
150	$N(0,2)$	0,999	Mix	1,000	$U(-3;3)$	0,950
200	$N(0,2)$	1,000	Mix	1,000	$U(-3;3)$	0,986
250	$N(0,2)$	1,000	Mix	1,000	$U(-3;3)$	0,995
500	$N(0,2)$	1,000	Mix	1,000	$U(-3;3)$	1,000
1000	$N(0,2)$	1,000	Mix	1,000	$U(-3;3)$	1,000

Таблица 2. Оценка мощности для сложной основной гипотезы $H_0 : X_i \sim T(k)$, где $N(a,\sigma)$ — нормальное распределение со средним значением a и стандартным отклонением σ, Mix — равновесная смесь распределений $N(0,1)$ и $N(0,3)$, $U(-3;3)$ — равномерное распределение на интервале $(-3;3)$.

4 Доказательство теоремы 1

Доказательство. Из леммы 2 следует, что

$$\lim_{n\to\infty}\left(\sqrt{n}A_{1,n}-\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}((X_i-a)^3-3(X_i-a)\sigma^2)\right)\right)=0 \text{ п.н.}, \tag{1}$$

применяя лемму 3, получаем

$$\lim_{n\to\infty}\sqrt{n}\left(\hat\sigma^4-\frac{2\sigma^2}{n}\sum_{i=1}^{n}(X_i-a)^2+\sigma^4\right)=0 \text{ п.н.}, \tag{2}$$

используя лемму 4, получаем

$$\begin{gathered}\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big((2-\hat\sigma^2)(X_i-\overline{X})^4-\\ (2-\sigma^2)(X_i-a)^4+((X_i-a)^2-\sigma^2)\mathbf{E}\left((X-a)^4\right)\big)=0 \quad (\text{п.н.})\end{gathered} \tag{3}$$

Из формул (1)–(3) следует, что предельное распределение последовательности случайных векторов $\sqrt{n}(A_{1,n},A_{2,n})$ такое же как у последовательности

$$(A'_{1,n},A'_{2,n}):=\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}((X_i-a)^3-3(X_i-a)\sigma^2),\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left((2-\sigma^2)(X_i-a)^4\right.\right.$$

$$\left.\left.-(6\sigma^2+\mathbf{E}(X-a)^4)(X_i-a)^2+\sigma^2\mathbf{E}(X-a)^4+3\sigma^4\right)\right).$$

Из условий 2), 3) следует, что

$$\mathbf{E}((X-a)^3-3(X-a)\sigma^2)=0,$$

$$\mathbf{E}((2-\sigma^2)(X_i-a)^4-(6\sigma^2+\mathbf{E}(X-a)^4)(X_i-a)^2+\sigma^2\mathbf{E}(X-a)^4+3\sigma^4)=0.$$

Поэтому, в силу центральной предельной теоремы последовательность $(A'_{1,n},A'_{2,n})$ сходится к случайному вектору (W_1,W_2), координаты которого имеют совместное нормальное распределение.

Найдем параметры этого распределения. Из условий 1), 2) следует, что

$$\mathbf{D}W_1=\mathbf{E}((X-a)^3-3(X-a)\sigma^2)^2,$$

$$\mathbf{D}W_2=\mathbf{E}((2-\sigma^2)(X_i-a)^4-(6\sigma^2+\mathbf{E}(X-a)^4)(X_i-a)^2+\sigma^2\mathbf{E}(X-a)^4+3\sigma^4)^2,$$

$$\mathbf{cov}(W_1,W_2)=\mathbf{E}\Big(((X-a)^3-3(X-a)\sigma^2)$$

$$\times((2-\sigma^2)(X_i-a)^4-(6\sigma^2+\mathbf{E}(X-a)^4)(X_i-a)^2+\sigma^2\mathbf{E}(X-a)^4+3\sigma^4)\Big)=0.$$

5 Вспомогательные результаты

Лемма 1. *Пусть выполнены условия*

1) $\mathbf{E}X^{2k} < \infty$ *для некоторого фиксированного* $k \in \mathbb{N}$*;*

2) $\mathbf{E}(X-a)^{2r-1} = 0$ *для всех* $1 \le r \le k$*.*

Тогда

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_i - \overline{X})^{2k} - (X_i - a)^{2k}) = 0 \text{ п.н.}$$

Д о к а з а т е л ь с т в о. Легко видеть, что

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_i - \overline{X})^{2k} - (X_i - a)^{2k})$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\left(X_i - a - \frac{1}{n} \sum_{j=1}^{n} (X_j - a) \right)^{2k} - (X_i - a)^{2k} \right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{r=0}^{2k-1} C_{2k}^{r} (X_i - a)^r \left(\frac{-1}{n} \sum_{j=1}^{n} (X_j - a) \right)^{2k-r}$$

$$= \left(\frac{1}{n^{1-\frac{1}{4k}}} \sum_{j=1}^{n} (X_j - a) \right)^{2k}$$

$$+ \sum_{r=1}^{2k-1} C_{2k}^{r} \left(\frac{1}{\sqrt{n} n^{\frac{2k-r}{3}}} \sum_{i=1}^{n} (X_i - a)^r \right) \left(\frac{-1}{n^{\frac{2}{3}}} \sum_{j=1}^{n} (X_j - a) \right)^{2k-r}.$$

Из условия 2) и закона повторного логарифма следует, что случайные последовательности находящиеся в первой и третьей скобках, сходятся при $n \to \infty$ к нулю п.н. По этой же причине сходится к нулю п.н. при $n \to \infty$ и нечетных $r \le 2k-1$ последовательность стоящая во второй скобке. Для четных $r \le 2k-2$ выполнено неравенство $\sqrt{n} n^{\frac{2k-r}{3}} \ge n^{\frac{7}{6}}$. Поэтому, в силу усиленного закона больших чисел, случайная последовательность расположенная во второй скобке также сходится при $n \to \infty$ к нулю п.н. для четных r.

Лемма 2. *Пусть выполнены условия*

1) $\mathbf{E}|X|^{2k+1} < \infty$ *для некоторого фиксированного* $k \in \{0\} \cup \mathbb{N}$*;*

2) $\mathbf{E}(X-a)^{2r-1} = 0$ *для всех* $1 \le r \le k$*.*

Тогда

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_i - \overline{X})^{2k+1} - (X_i - a)^{2k+1} + (2k+1)(X_i - a)\mathbf{E}(X-a)^{2k}) = 0 \text{ п.н.}$$

Доказательство. Очевидно, что утверждение верно для $k = 0$. Пусть далее $k > 0$. Легко видеть, что

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}((X_i - \overline{X})^{2k+1} - (X_i - a)^{2k+1} + (2k+1)(X_i - a)\mathbf{E}(X-a)^{2k})$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\left(X_i - a - \frac{1}{n}\sum_{j=1}^{n}(X_j - a)\right)^{2k+1}\right.$$

$$\left.-(X_i - a)^{2k+1} + (2k+1)(X_i - a)\mathbf{E}(X-a)^{2k}\right)$$

$$= (2k+1)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - a)^{2k}\left(\frac{-1}{n}\sum_{j=1}^{n}(X_j - a)\right)$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{r=0}^{2k-1}C_{2k+1}^{r}(X_i - a)^{r}\left(\frac{-1}{n}\sum_{j=1}^{n}(X_j - a)\right)^{2k-r+1}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2k+1)(X_i - a)\mathbf{E}(X-a)^{2k}$$

$$= \left(\frac{-1}{n^{1-\frac{1}{4k+2}}}\sum_{j=1}^{n}(X_j - a)\right)^{2k+1}$$

$$+\sum_{r=1}^{2k-1}C_{2k+1}^{r}\left(\frac{1}{\sqrt{n}n^{\frac{2k-r+1}{3}}}\sum_{i=1}^{n}(X_i - a)^{r}\right)\left(\frac{-1}{n^{\frac{2}{3}}}\sum_{j=1}^{n}(X_j - a)\right)^{2k-r+1}$$

$$-(2k+1)\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - a)\right)\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - a)^{2k}\right)$$

$$+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(2k+1)(X_j - a)\mathbf{E}(X-a)^{2k}.$$

Первых два слагаемых сходятся при $n \to \infty$ к нулю п.н. по причинам полностью аналогичным тем, которые приведены в конце доказательства леммы 1. Запишем третье и четвертое слагаемое в следующем виде

$$(2k+1)\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - a)\right)\left(\frac{1}{n}\sum_{i=1}^{n}(\mathbf{E}(X-a)^{2k} - (X_i - a)^{2k})\right)$$

$$= (2k+1)\left(\frac{1}{n^{\frac{3}{4}}}\sum_{j=1}^{n}(X_j - a)\right)\left(\frac{1}{n^{\frac{3}{4}}}\sum_{i=1}^{n}(\mathbf{E}(X-a)^{2k} - (X_i - a)^{2k})\right).$$

Из условия 2) и закона повторного логарифма следует, что все находящиеся в скобках случайные последовательности сходятся при $n \to \infty$ к нулю п.н.

Лемма 3. *Пусть $\mathbf{E}X^2 < \infty$, тогда для любого $k \in \mathbb{N}$*

$$\lim_{n\to\infty} \sqrt{n}\left(\hat{\sigma}^{2k} - \frac{\sigma^{2k-2}k}{n}\sum_{i=1}^{n}(X_i-a)^2 + (k-1)\sigma^{2k}\right) = 0 \text{ п.н.}$$

Д о к а з а т е л ь с т в о. Положим

$$Y_n := \sqrt{n}\hat{\sigma}^{2k} - \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-a)^2\right)^k.$$

Будем иметь

$$Y_n = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)^k - \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-a)^2\right)^k$$
$$= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}((X_i-\overline{X})^2-(X_i-a)^2)\right)$$
$$\times\sum_{r=0}^{k-1}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)^{k-1-r}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-a)^2\right)^{r} =: g_{1,n}\sum_{r=0}^{k-1}f_{k,r,n}.$$

Из леммы 1 и усиленного закона больших чисел следует, что случайные последовательности $g_{1,n}$ и $f_{k,r,n}$, $0 \le r \le k-1$ сходятся п.н. к 0 и $\sigma^{2(k-1)}$, соответственно. Поэтому

$$\lim_{n\to\infty} Y_n = 0 \text{ п.н.}, \tag{4}$$

и нам достаточно показать, что

$$\lim_{n\to\infty}\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-a)^2\right)^k - \frac{\sigma^{2k-2}k}{n}\sum_{i=1}^{n}(X_i-a)^2 + (k-1)\sigma^{2k}\right) = 0 \text{ п.н.} \tag{5}$$

Легко видеть, что

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-a)^2\right)^k - \frac{\sigma^{2k-2}k}{n}\sum_{i=1}^{n}(X_i-a)^2 + (k-1)\sigma^{2k}\right)$$
$$= \sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n}((X_i-a)^2-\sigma^2)+\sigma^2\right)^k - \frac{\sigma^{2k-2}k}{n}\sum_{i=1}^{n}(X_i-a)^2 + (k-1)\sigma^{2k}\right)$$
$$= \sqrt{n}\sum_{r=0}^{k-2}C_k^r\left(\frac{1}{n}\sum_{i=1}^{n}((X_i-a)^2-\sigma^2)\right)^{k-r}\cdot\sigma^{2r}$$

$$+\sqrt{n}k\left(\frac{1}{n}\sum_{i=1}^{n}((X_i-a)^2-\sigma^2)\right)\sigma^{2(k-1)}$$

$$+\sqrt{n}\sigma^{2k}-\sqrt{n}\frac{\sigma^{2k-2}k}{n}\sum_{i=1}^{n}(X_i-a)^2+\sqrt{n}(k-1)\sigma^{2k}$$

$$=\sum_{r=0}^{k-2}C_k^r\left(\frac{1}{n^{1-\frac{1}{2(k-r)}}}\sum_{i=1}^{n}((X_i-a)^2-\sigma^2)\right)^{k-r}\cdot\sigma^{2r}.$$

В силу закона повторного логарифма случайные последовательности расположенные в скобках сходятся при $n\to\infty$ к нулю п.н. для всех $0\le r\le k-2$. Следовательно, справедлива формула (5).

Лемма 4. *Пусть* $\mathbf{E}X^4<\infty$, *тогда*

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big((2-\hat\sigma^2)(X_i-\overline{X})^4$$

$$-(2-\sigma^2)(X_i-a)^4+((X_i-a)^2-\sigma^2)\mathbf{E}(X-a)^4\big)=0 \text{ п.н.}$$

Д о к а з а т е л ь с т в о. Легко видеть, что

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2-\hat\sigma^2)(X_i-\overline{X})^4=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(2-\frac{1}{n}\sum_{j=1}^{n}((X_j-\overline{X})^2-\sigma^2)-\sigma^2\right)(X_i-\overline{X})^4$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2-\hat\sigma^2)(X_i-\overline{X})^4$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(2-\frac{1}{n}\sum_{j=1}^{n}((X_j-\overline{X})^2\pm(X_j-a)^2\pm\sigma^2)\right)(X_i-\overline{X})^4$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2-\sigma^2)(X_i-\overline{X})^4-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}((X_j-\overline{X})^2-(X_j-a)^2)\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^4$$

$$-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}((X_j-a)^2-\sigma^2)\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^4=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2-\sigma^2)(X_i-\overline{X})^4$$

$$-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}((X_j-\overline{X})^2-(X_j-a)^2)\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^4$$

$$-\frac{1}{n}\sum_{j=1}^{n}((X_j-a)^2-\sigma^2)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}((X_i-\overline{X})^4-(X_i-a)^4)$$

$$-\frac{1}{n^{\frac{3}{4}}}\sum_{j=1}^{n}((X_j-a)^2-\sigma^2)\frac{1}{n^{\frac{3}{4}}}\sum_{i=1}^{n}((X_i-a)^4-\mathbf{E}(X-a)^4)$$

$$-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}((X_j-a)^2-\sigma^2)\mathbf{E}(X-a)^4. \tag{6}$$

Из леммы 1 следует, что

$$\lim_{n\to\infty}\frac{2-\sigma^2}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\overline{X})^4-(X_i-a)^4)=0 \text{ п.н.}, \tag{7}$$

используя лемму 1 и усиленный закон больших чисел, получаем

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}((X_j-\overline{X})^2-(X_j-a)^2)\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^4=0 \text{ п.н.}, \tag{8}$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}((X_j-a)^2-\sigma^2)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}((X_i-\overline{X})^4-(X_i-a)^4)=0 \text{ п.н.}, \tag{9}$$

применяя закона повторного логарифма, получаем

$$\lim_{n\to\infty}\frac{1}{n^{\frac{3}{4}}}\sum_{j=1}^{n}((X_j-a)^2-\sigma^2)\frac{1}{n^{\frac{3}{4}}}\sum_{i=1}^{n}((X_i-a)^4-\mathbf{E}(X-a)^4)=0 \text{ п.н.} \tag{10}$$

Утверждение леммы 4 следует из (6)–(10).

References

[1] Cysneiros, F.; Paula, G. (2005). Restricted methods in symmetrical linear regression models. Computational Statistics and Data Analysis, 49, 689–708.

[2] Jarque, Carlos M.; Bera, Anil K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters 6 (3): 255-259. doi:10.1016/0165-1765(80)90024-5.

[3] Jarque, Carlos M.; Bera, Anil K. (1981). Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. Economics Letters 7 (4): 313-318. doi:10.1016/0165-1765(81)90035-5.

[4] Jarque, Carlos M.; Bera, Anil K. (1987). A test for normality of observations and regression residuals. International Statistical Review 55 (2): 163-172. JSTOR 1403192.

[5] Pearson, Karl (1916). Mathematical contributions to the theory of evolution, XIX: Second sup- plement to a memoir on skew variation. Philosophical Transactions of the Royal Society A. v. 216(538-548), pp. 429-457.

[6] Glinskiy, V.; Ismayilova, Y.; Khrushchev, S.; Logachov, A.; Logachova, O.; Serga, L.; Yambartsev, A.; Zaykov, K. (2024) Modifications to the Jarque–Bera Test. Mathematics, 12, 2523.

Artem Vasilhevich Logachov
Lab. of Probability Theory and Math. Statistics, Sobolev Institute of Mathematics,
pr. Koptyuga, 4,
630090, Novosibirsk, Russia
E-mail address: omboldovskaya@mail.ru

Sergey Evgenievich Khrushchev
Lab. of Applied Inverse Problems, Sobolev Institute of Mathematics,
pr. Koptyuga, 4,
630090, Novosibirsk, Russia
E-mail address: hrushew@rambler.ru