Записки научных
семинаров ПОМИ
Том 501, 2020 г.

М. С. Ермаков

КРИТЕРИЙ ХИ-КВАДРАТ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ ОДНОРОДНОСТИ

§1. Введение

Для задач проверки гипотезы согласия критерии хи-квадрат довольно хорошо изучены, даже когда число ячеек в критериях растет с ростом объема выборки [3–6, 12, 13, 15–17].

Пусть $X_1, \dots, X_n$– выборка независимых одинаково распределенных случайных величин на интервале $[0, 1]$, имеющих функцию распределения F_n. Обозначим $\widehat{F}_n$ – эмпирическую функцию распределения выборки. Обозначим $\Im$– множество всех функций распределения. Обозначим F_0 – функцию равномерного распределения на интервале $[0, 1]$. Задачу проверки гипотезы согласия мы будем рассматривать как задачу проверки гипотезы $\mathbb{H}_0 : F_n = F_0$ против альтернатив $\mathbb{H}_n : F_n \in \Psi_n \subset \Im$, где Ψ_n – некоторое непараметрическое множество альтернатив.

Обозначим $T_n(\widehat{F}_n)$ тестовую статистику критериев хи-квадрат и $T_n(F)$ – функционал на $\Im$, порождающий тестовую статистику $T_n(\widehat{F}_n)$

Для проверки гипотезы согласия нами показано в [5], что последовательности критериев хи-квадрат, имеющие увеличивающееся число зон с ростом объема выборки, равномерно состоятельны на множествах альтернатив $\Im(b_n) = \{F : T_n(F) > b_n, F \in \Im\}$, где $b_n > 0$ – последовательность констант, зависящих от числа ячеек и объема выборки n. Таким образом мы можем говорить о равномерной состоятельности последовательности множеств альтернатив $\Omega_n \subset \Im$, только если $\Omega_n \subset \Im(b_n)$ с последовательностями чисел b_n, удовлетворяющим определенным условиям. В [6] нами описаны все равномерно состоятельные последовательности множеств альтернатив Ω_n, заданных в терминах плотностей распределения, когда ячейки критерия хи-квадрат имеют одинаковую длину и и их число увеличивается с ростом объема выборки.

Ключевые слова: критерии согласия, состоятельность, критерий хи-квадрат, гипотеза однородности.

Исследование поддержано грантом РФФИ 20-01-00273.

Цель настоящей работы исследовать поведение критериев хи-квадрат для проверки гипотезы однородности, когда число ячеек растет вместе с ростом объема выборки. При этом также как и в [5, 6] основное внимание будет уделено описанию всего класса последовательностей множеств альтернатив таких, что на них критерий хи-квадрат равномерно состоятелен. В случае проверки гипотезы однородности ответ зависит от функций распределения уже двух выборок. Отметим, что задачам проверки гипотезы однородности в последнее время уделяется значительное внимание [7–10, 18].

Пусть интервал $[0,1]$ разбит на $m = m_n$ подынтервалов

$$I_{nj} = [e_{nj}, e_{n,j+1}), \quad p_{nj} = e_{n,j+1} - e_{nj} > 0, \quad e_{n0} = 0, \quad e_{nm} = 1,$$

$1 \leqslant j \leqslant m = m_n$, где $m_n \to \infty$ при $n \to \infty$. Функционал T_n, порождающий тестовую статистику хи–квадрат для проверки гипотезы согласия, имеет вид

$$T_n(F - F_0) = n \sum_{j=1}^{m} \frac{(r_{nj} - p_{nj})^2}{p_{nj}},$$

где $r_{nj} = F(e_{nj}) - F(e_{n,j-1})$ для всех $1 \leqslant j \leqslant m_n$ и $F_0(x) = x$, $x \in [0,1]$.

Тестовой статистикой критерия хи-квадрат является $T_n(\widehat{F}_n - F_0)$.

Для критерия проверки гипотез K_n мы будем обозначать $\alpha(K_n)$ – его вероятность ошибки первого рода и $\beta(K_n, F_n)$– его вероятность ошибки второго рода при альтернативе F_n.

Пусть S_n — последовательность тестовых статистик. Скажем что последовательность множеств альтернатив $\Psi_n \subset \Im$ равномерно состоятельно для тестовой статистики S_n, если для критериев K_n, порожденных тестовой статистикой S_n таких, что $\alpha(K_n) = \alpha(1 + o(1))$, $0 < \alpha < 1$, имеет место

$$\limsup_{n\to\infty} \sup_{F \in \Psi_n} \beta(K_n, F) < 1 - \alpha.$$

Аналогичные обозначения и терминологию мы будем использовать в задаче проверки гипотезы однородности.

Как говорилось, в случае задачи проверки гипотезы согласия критерий хи-квадрат является равномерно состоятельным в задаче проверки гипотез против множеств альтернатив $\Im(b_n)$. Более того [5], для любой последовательности простых альтернатив $F_n \in \Im$ для вероятностей ошибок второго рода $\beta(K_n, F_n)$ критериев K_n, $\alpha(K_n) = \alpha(1+o(1))$,

$0 < \alpha < 1$, порожденных тестовой статистикой $T_n(\widehat{F}_n - F_0)$, имеет место

$$\beta(K_n, F_n) = \Phi(x_\alpha - 2^{-1/2} m_n^{-1/2} T_n(F_n - F_0)) + o(1) \qquad (1.1)$$

при $n \to \infty$. Здесь критическое значение x_α задается уравнением $\alpha = 1 - \Phi(x_\alpha)$, где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\{-2\,t^2/2\}\,d\,t$ –функция стандартного нормального распределения, $x \in \mathbb{R}^1$.

Такая асимптотика вероятностей ошибок второго рода вместе с асимптотической минимаксностью [5] критерия хи-квадрат обосновывает аргументацию использования метода расстояний в непараметрической проверке гипотез применительно к критерию хи-квадрат.

В этой работе мы докажем аналогичные результаты для проверки гипотезы однородности, когда множества альтернатив образовано разностями функций распределения двух выборок. При этом на функцию распределения F_n одной из выборок налагается дополнительное условие, что $\mathbb{L}_2$ – нормы их плотностей распределения ограничены некоторой постоянной. Оказывается, что при этом условии состоятельность множеств альтернатив определяется значением функционала T_n на разностях функций распределения этих двух выборок. Такое утверждение о состоятельности множеств альтернатив позволяет распространить на данную постановку задачи результаты [6] о необходимых и достаточных условиях состоятельности последовательностей простых альтернатив, заданных в терминах плотностей распределения.

Условимся обозначать буквами C и c -различные постоянные. Для последовательностей a_n и b_n, $a_n = o(b_n)$ означает, что $a_n/b_n \to 0$ при $n \to \infty$ и $a_n = O(b_n)$ означает, что $0 < a_n/b_n < C < \infty$ для всех n. Символ $[a]$ означает целую часть числа a.

§2. Основные результаты

2.1. Постановка задачи. По сравнению с гипотезой согласия в случае проверки гипотезы однородности ситуация более сложная. Здесь у нас две выборки $X_1, \dots, X_n$ и $Y_1, \dots, Y_{l_n}$ независимых случайных величин, принимающих значения на интервале $[0, 1]$ и имеющих функции распределения F_n и G_{l_n} соответственно. Таким образом ответ о равномерной состоятельности множеств альтернатив приходится искать в терминах двух функций распределения.

Обозначим множество всех пар функций распределения (F, G) через $\Im \times \Im$.

На множестве $\Im \times \Im$ определим функционал

$$T_{1n}(F-G) = nm\sum_{j=1}^{m}(r_{nj}-s_{nj})^2, \qquad (F,G)\in\Im\times\Im,$$

где $s_{nj} = G(e_{nj}) - G(e_{n,j-1})$ для всех $1 \leqslant j \leqslant m_n$.

Обозначим $\widehat{G}_{l_n}(x)$ эмпирическую функцию распределения выборки $Y_1,\dots,Y_{l_n}$.

Обозначим $a_n = \frac{n}{l_n}$ и предположим, что $0 < c < a_n < C < \infty$, где c, C - постоянные.

Тогда тестовой статистикой критерия хи-квадрат для проверки гипотезы однородности является

$$T_{1n}(\widehat{F}_n - \widehat{G}_{l_n}) = nm\sum_{j=1}^{m}(\widehat{r}_{nj}-\widehat{s}_{nj})^2,$$

где $\widehat{s}_{nj} = \widehat{G}_{l_n}(e_{nj}) - \widehat{G}_{l_n}(e_{n,j-1})$ для всех $1 \leqslant j \leqslant m_n$.

Заметим, что $\mathbf{E}[T_{1n}(\widehat{F}_n - \widehat{G}_{l_n})]$ зависит только от разности функций распределения $F_n - G_{l_n}$ и нам не приходится добавлять в тестовую статистику дополнительные оценки слагаемых [1, 5, 7, 8], связанных с зависимостью от самих функций распределения F_n и G_{l_n}.

Для тестовой статистики

$$T_{2n}(\widehat{F}_n - \widehat{G}_{l_n}) = n\sum_{j=1}^{m} g_{nj}\frac{(\widehat{r}_{nj}-\widehat{s}_{nj})^2}{p_{nj}},$$

порожденной функционалом

$$T_{2n}(F-G) = n\sum_{j=1}^{m} g_{nj}\frac{(r_{nj}-s_{nj})^2}{p_{nj}}, \quad 0 < c < g_{jn} < C < \infty$$

будет доказана отдельная теорема.

Доказательства будут проведены для тестовой статистики

$$T_n(\widehat{F}_n - \widehat{G}_{l_n}) = n\sum_{j=1}^{m} \frac{(\widehat{r}_{nj}-\widehat{s}_{nj})^2}{p_{nj}},$$

порожденной функционалом

$$T_n(F-G) = n\sum_{j=1}^{m} \frac{(r_{nj}-s_{nj})^2}{p_{nj}}.$$

В случае тестовой статистики $T_{2n}(\widehat{F}_n - \widehat{G}_{l_n})$ рассуждения практически не отличаются и поэтому отличия не будут указаны.

Ответ будет получен при дополнительном предположении, что функции распределения F_n имеют плотности распределения $f_n(x) = \frac{dF_n(x)}{dx}$, $x \in [0,1]$, и дана априорная информация, что для некоторой положительной постоянной C

$$F_n \in \Xi(C) = \left\{F \,:\, \|f\|^2 < C,\ f(x) = \frac{dF(x)}{dx},\ F \in \Im\right\},$$

где $\|f\|^2 = \int\limits_0^1 f^2(x)\,d\,x$.

Вид главного члена асимптотики дисперсии тестовой статистики критерия хи-квадрат при справедливости альтернатив существенно упростится, если мы дополнительно предположим, что

$$F_n \in \Xi_{1n} = \left\{F \,:\, \sup_{x\in[0,1]} |f(x)| < c_n m_n^{1/2},\ f(x) = \frac{dF(x)}{dx},\ F \in \Im\right\},$$

где $c_n \to 0$ при $n \to \infty$.

Для последовательности $b_n > 0$, для $i = 1,2$, определим множества альтернатив $\Psi_i(b_n) = \{(F,G) \,:\, T_{in}(F-G) \geqslant b_n,\ (F,G) \in \Im \times \Im\}$.

Нами будет показана равномерная состоятельность тестовых статистик $T_{in}(\widehat{F}_n - \widehat{G}_{l_n})$, $i = 1,2$, критериев хи-квадрат в задачах проверки гипотез

$$\mathbb{H}_0 \ :\ F_n(x) = G_{l_n}(x), \qquad x \in [0,1]$$

против альтернатив

$$\mathbb{H}_n \ :\ (F_n, G_{l_n}) \in \Psi_i(b_n) \cap \Xi(C)$$

для последовательностей b_n, удовлетворяющих

$$0 < \liminf_{n\to\infty} m_n^{-1/2} b_n \leqslant \limsup_{n\to\infty} m_n^{-1/2} b_n < \infty. \tag{2.1}$$

Относительно длин ячеек нами предполагается, что для всех j, $1 \leqslant j \leqslant m_n$,

$$0 < c < m_n\, p_{nj} < C_1 < \infty \tag{2.2}$$

для некоторых положительных постоянных c и C_1.

Кроме того мы предполагаем $m_n = o(n)$ при $n \to \infty$.

Доказательство теорем базируется на методах предложенных в [5] для изучения критерия хи-квадрат для проверки гипотезы согласия.

Мы задаем функционал T_n на $\Im \times \Im$ равный

$$T_n(F-G) = n\sum_{j=1}^{m}\left(\int_0^1 \phi_{nj}(x)d(F(x)-G(x))\right)^2 p_{nj}^{-1},$$

где $\phi_{nj}(x) = \mathbf{1}_{\{x\in I_{nj}\}} - p_{nj}$, $x \in [0,1]$, $1 \leqslant j \leqslant m_n$ и $\mathbf{1}_{\{A\}}$– обозначает индикатор события A. После этого мы анализируем тестовые статистики, как тестовые статистики, порожденные этим функционалом.

Такой подход позволяет легко доказать результаты аналогичные многочисленным результатам [6,7,14,18], полученным для задач непараметрической проверки гипотез о плотности распределения, основанным на разложениях в ряды из ортогональных функций. Однако в нашем случае функции ϕ_{nj} не ортогональны.

Тестовая статистика $T_{2n}(\widehat{F}_n - \widehat{G}_{l_n})$ в этих обозначениях принимает вид

$$T_{2n}(\widehat{F}_n - \widehat{G}_{l_n}) = n\sum_{j=1}^{n} g_{nj}\left(\int_0^1 \phi_{nj}(x)d(\widehat{F}_n(x)-\widehat{G}_{l_n}(x))\right)^2 p_{nj}^{-1},$$

Заметим [1, 5], что математическое ожидание тестовой статистики $T_{2n}(\widehat{F}_n - \widehat{G}_{l_n})$ при справедливости гипотезы зависит от неизвестной функции распределения $F_n = G_{l_n}$, а в случае альтернативы зависит от обоих неизвестных функций распределения F_n и G_{l_n}. Это вызвано входящим в тестовую статистику слагаемым

$$W_n = n\sum_{j=1}^{n} g_{nj}\left(\int_0^1 \phi_{nj}^2(x)d\widehat{F}_n(x) + \int_0^1 \phi_{nj}^2(x)d\,\widehat{G}_{l_n}(x)\right)p_{nj}^{-1}$$

Поэтому в одной из постановок задач из тестовой статистики мы вычитаем часть этого слагаемого, устраняя эту зависимость. Для тестовой статистики $T_{1n}(\widehat{F}_n - \widehat{G}_{l_n})$ – это не так.

Не умаляя общности, мы можем предположить, что функции распределения F_n и G_{l_n} имеют плотности

$$f_n(x) = 1 + \sum_{j=1}^{m}\theta_{nj}\,\phi_{nj}(x), \quad x\in[0,1]$$

и

$$g_{l_n}(x) = 1 + \sum_{j=1}^{m} \tau_{nj}\, \phi_{nj}(x), \quad x \in [0,1]$$

соответственно, причем

$$\sum_{j=1}^{m} \theta_{nj} p_{nj} = 0, \qquad \sum_{j=1}^{m} \tau_{nj} p_{nj} = 0.$$

Обозначим $\eta_{nj} = \theta_{nj} - \tau_{nj}$.

2.2. Статистика T_{1n}. Обозначим $M_{1n}(\eta) = nm \sum\limits_{j=1}^{m} p_{nj}^2 \eta_{nj}^2$ и положим

$$\sigma_{1n}^2 = 2m^2 \sum_{j=1}^{m} p_{nj}^2 (1 + \theta_{nj} + a_n + a_n \tau_{nj})^2.$$

Лемма 2.1. *При $n \to \infty$ имеет место*

$$\mathbf{E}[T_{1n}(\widehat{F}_n - \widehat{G}_{l_n})] - (m-1)(1 + a_n) = nm \sum_{j=1}^{m} p_{nj}^2 \eta_{nj}^2 (1 + o(1)) \qquad (2.3)$$

и

$$\mathbf{Var}[T_{1n}(\widehat{F}_n - \widehat{G}_{l_n})] = \sigma_{1n}^2 (1 + o(1)) \\ + nm^2 \sum_{j=1}^{m} p_{nj}^3 (1 + \theta_{nj} + a_n + a_n \tau_{nj}) \eta_{nj}^2 (1 + o(1)) \doteq \sigma_{11n}^2 (1 + o(1)). \qquad (2.4)$$

Заметим, что второе слагаемое в правой части (2.4) при справедливости гипотезы отсутствует и возникает только при справедливости альтернативы. Таким образом имеет место интересная ситуация, когда множество альтернатив настолько богато, что асимптотическая дисперсия альтернатив, сближающихся с гипотезой, может быть существенно отлична от асимптотической дисперсии гипотезы. Как следует из (3.25), $\sigma_{11n}^2 - \sigma_{1n}^2 > 0$. Если $F_n \in \Xi_{2n}$, то $\sigma_{11n}^2 - \sigma_{1n}^2 = o(\sigma_{1n}^2)$ при $n \to \infty$.

Заметим, что мы можем подставить в (2.4) оценки

$$\widehat{\theta}_{nj} = \frac{\widehat{r}_{nj}}{p_{nj}} - 1, \quad \text{и} \quad \widehat{\tau}_{nj} = \frac{\widehat{s}_{nj}}{p_{nj}} - 1$$

параметров θ_{nj} и τ_{nj} и получить, как будет показано, состоятельные оценки

$$\widehat{\sigma}_{1n}^2 = 2m^2 \sum_{j=1}^{m} (\widehat{r}_{nj} + a_n \widehat{s}_{nj})^2$$

дисперсии σ_{1n}^2.

Другие методы оценивания дисперсии рассматриваются в [1, 7, 8].

Определим критерии

$$K_{1n} = \mathbf{1}_{\{\widehat{\sigma}_{1n}^{-1}(T_{1n}(\widehat{F}_n - \widehat{G}_{l_n}) - m(1+a_n)) > x_\alpha\}},$$

где x_α задается уравнением $1 - \alpha = \Phi(x_\alpha)$, $0 < \alpha < 1$.

Теорема 2.1. *Пусть выполнены условия* (2.1), (2.2) *и пусть* $m_n = o(n)$ *при* $n \to \infty$. *Тогда последовательность множеств альтернатив* $\Psi_{1n}(b_n) \cap \Xi(C)$*– равномерно состоятельна для последовательности критериев* K_{1n}*, порожденных тестовыми статистиками* $T_{1n}(\widehat{F}_n - \widehat{G}_{l_n})$.

При $n \to \infty$ *имеет место* $\alpha(K_{1n}) = \alpha(1 + o(1))$ *и*

$$\beta(K_{1n}, \Psi_{1n}(b_n)) = \Phi(\sigma_{11n}^{-1}(\sigma_{1n} x_\alpha - M_{1n}(\eta))) + o(1). \tag{2.5}$$

2.3. Статистики T_{2n} и T_{3n}. Обозначим $M_{2n}(\eta) = n \sum\limits_{j=1}^{m} g_{nj} p_{nj} \eta_{nj}^2$ и положим

$$\sigma_{2n}^2 = 2 \sum_{j=1}^{m} g_{nj}^2 (1 + \theta_{nj} + a_n + a_n \tau_{nj})^2.$$

Мы покажем, что состоятельной оценкой σ_{2n}^2 является

$$\widehat{\sigma}_{2n}^2 = 2 \sum_{j=1}^{m} g_{nj}^2 p_{nj}^{-2} (\widehat{r}_{nj} + a_n \widehat{s}_{nj})^2.$$

Критерии для тестовой статистики $T_{2n}(\widehat{F}_n - \widehat{G}_{l_n})$ строятся на основе следующих асимптотик.

Лемма 2.2. *При* $n \to \infty$ *имеет место*

$$\mathbf{E}[T_{2n}(\widehat{F}_n - \widehat{G}_{l_n})] = M_{2n}(\eta)(1 + o(1)) + \mathbf{E}[W_n], \tag{2.6}$$

$$\begin{aligned} \mathbf{E}[W_n] = \sum_{j=1}^{m} g_{nj}((1 - p_{nj} + \theta_{nj}(1 - p_{nj}) - p_{nj}\theta_{nj}^2) \\ + a_n(1 - p_{nj} + \tau_{nj}(1 - p_{nj}) - p_{nj}\tau_{nj}^2)) \doteq e_n, \end{aligned} \tag{2.7}$$

$$\begin{aligned}\mathbf{Var}[T_{2n}(\widehat F_n - \widehat G_n)] &= \sigma_{2n}^2(1+o(1))\\ &\quad + n\sum_{j=1}^m g_{nj}^2 p_{nj}(1+\theta_{nj}+a_n+a_n\tau_{nj})\eta_{nj}^2(1+o(1))\\ &\doteq \sigma_{21n}^2(1+o(1)).\end{aligned} \tag{2.8}$$

Как мы покажем, если $m_n = o(n^{2/3})$, то

$$e_n = \sum_{j=1}^m g_{nj}(1+a_n+\theta_{nj}+\tau_{nj}) + O(1). \tag{2.9}$$

Заметим, что мы можем подставить в (2.7) оценки $\widehat\theta_{nj}$ и $\widehat\tau_{nj}$ параметров θ_{nj} и τ_{nj} и получить, как будет показано, состоятельные оценки $\widehat e_n$ значения e_n.

Определим критерии

$$K_{2n} = \mathbf{1}_{\{\widehat\sigma_{2n}^{-1}(T_{2n}(\widehat F_n - \widehat G_{l_n}) - \widehat e_n) > x_\alpha\}},$$

где x_α задается уравнением $1-\alpha = \Phi(x_\alpha)$, $0<\alpha<1$.

Теорема 2.2. *Пусть выполнены условия* (2.1), (2.2) *и пусть* $m_n = o(n^{2/3})$ *при* $n\to\infty$. *Тогда последовательность множеств альтернатив* $\Psi_{2n}(b_n)\cap\Xi(C)$*– равномерно состоятельна для последовательности критериев* K_{2n}.

Пусть $m_n = o(n)$ *и дана априорная информация, что для некоторой постоянной* C, $\|g_n\| < C$, $g_n(x) = \frac{dG_n(x)}{dx}$, $x\in[0,1]$. *Тогда последовательность множеств альтернатив* $\Psi_{2n}(b_n)\cap\Xi(C)$*– равномерно состоятельна.*

При $n\to\infty$ *имеет место* $\alpha(K_{2n}) = \alpha(1+o(1))$ *и*

$$\beta(K_{2n}, F_n, G_n) = \Phi(\sigma_{21n}^{-1}(\sigma_{2n}x_\alpha - M_{1n}(\eta))) + o(1). \tag{2.10}$$

В [1, 5, 7, 8, 18] для несколько других задач проверки непараметрических гипотез из аналога тестовой статистики T_{2n} удаляется аналог слагаемого W_n и результаты получаются для такой модифицированной тестовой статистики.

Определим тестовую статистику

$$T_{3n}(\widehat F_n - \widehat G_{l_n}) = T_{2n}(\widehat F_n - \widehat G_{l_n}) - W_n.$$

Зададим порожденный ею критерий проверки гипотез

$$K_{3n} = \mathbf{1}_{\{\widehat\sigma_{2n}^{-1} T_{3n}(\widehat F_n - \widehat G_{l_n}) > x_\alpha\}},$$

где x_α задается уравнением $1-\alpha=\Phi(x_\alpha)$, $0<\alpha<1$.

Теорема 2.3. *Пусть выполнены условия* (2.1), (2.2) *и пусть* $m_n = o(n)$ *при* $n\to\infty$. *Тогда последовательность множеств альтернатив* $\Psi_{2n}(b_n)\cap\Xi(C)$*– равномерно состоятельна для последовательности критериев* K_{3n}.

При $n\to\infty$ *имеет место* $\alpha(K_{3n})=\alpha(1+o(1))$ *и*

$$\beta(K_{3n},F_n,G_n)=\Phi(\sigma_{21n}^{-1}(\sigma_{2n}x_\alpha-M_{2n}(\eta)))+o(1). \tag{2.11}$$

2.4. Проверка гипотезы однородности в терминах плотностей распределения. Утверждения об асимптотике вероятностей ошибок второго рода (2.5), (2.10) и (2.11) критериев хи-квадрат для проверки гипотезы однородности полностью аналогично утверждению [5, 6] об асимптотике вероятностей ошибок второго рода для проверки гипотезы согласия (1.1). Это позволяет перенести на задачи проверки гипотезы однородности необходимые и достаточные условия равномерной состоятельности, полученные в [6] для задач проверки гипотезы согласия, когда альтернативы задаются в терминах плотностей распределения.

Пусть функции распределения F_n и G_{l_n} имеют плотности распределения f_n, g_{l_n} соответственно и $F_n\in\Xi(C)$, $G_{l_n}\in\Xi(C)$. Обозначим $h_n=f_n-g_{l_n}$.

Рассмотрим задачу проверки гипотезы

$$\mathbb{H}_0 \,:\, h_n(x)=0, \qquad x\in[0,1],$$

против альтернатив

$$\mathbb{H}_n \,:\, h_n\in\Omega_n\subset\Gamma,$$

где $\Gamma=\{h\,:\,h=\frac{d(F-G)(x)}{d\,x}, \|h\|<\infty, F\in\Xi(C)\,\}$.

Для этой постановки задачи справедливы все утверждения теоремы 6.1 в [6], если в них заменить плотности распределения $1+f_n$ на функции h_n. При этом все требования, чтобы функции $1+f_n$ или функции, образованные из них в условии B, были плотностями распределения заменяются на требования, чтобы функции h_n и аналогично [6] образованные функции из h_n были разностями плотностей распределения. В частности это выполнено, если плотности функций распределения F_n и G_n удовлетворяют условию B.

Аналог теоремы 6.1 в [6] справедлив только для последовательностей простых альтернатив h_n, $\|h_n\|\asymp n^{-r}$, $\frac14<r<\frac12$, $m_n\asymp n^{2-4r}$.

Для этой постановки задачи, так же как и в [6], предполагается, что ячейки критериев хи-квадрат имеют одинаковую длину.

§3. Доказательство теорем

3.1. Оценка $\mathbf{E}\,[T_n]$. Как уже говорилось, рассуждения будут проведены для статистики T_n. В этом случае альтернативы удовлетворяют неравенству

$$T_n(F_n - G_n) = n\sum_{j=1}^{m} p_{nj}\eta_{nj}^2 \geqslant b_n.$$

Условие $f_n \in \Xi(C)$ означает, что

$$\sum_{j=1}^{m} p_{nj}\theta_{nj}^2 \leqslant \|f_n - 1\|^2 < C. \tag{3.1}$$

Утверждения приводимой ниже леммы получаются прямыми вычислениями.

Лемма 3.1. *Для $1 \leqslant j \leqslant m$ имеет место*

$$\mathbf{E}_\theta[\phi_{nj}(X_1)] = \theta_{nj}p_{nj}, \tag{3.2}$$

$$\mathbf{E}_\theta[\phi_{nj}^2(X_1)] = p_{nj}(1 - p_{nj} + \theta_{nj}(1 - 2p_{nj})), \tag{3.3}$$

$$\begin{aligned}\mathbf{E}[\bar\phi_{nj_1}^4(X_1)] &= p_{nj}(1+\theta_{nj})(1 - 4p_{nj}(1+\theta_{nj}) \\ &\quad + 6p_{nj}^2(1+\theta_{nj})^2 - 3p_{nj}^3(1+\theta_{nj})^3)\end{aligned} \tag{3.4}$$

и для $1 \leqslant j_1 < j_2 \leqslant m$ имеет место

$$\begin{aligned}&\mathbf{E}_\theta[\phi_{nj_1}(X_1)\,\phi_{nj_2}(X_1)] \\ &\qquad = -p_{nj_1}p_{nj_2}(1 + \theta_{nj}(1-2p_{nj}) + \theta_{nj_2}(1 - 2p_{nj_2})),\end{aligned} \tag{3.5}$$

$$\begin{aligned}\mathbf{E}_\theta[\bar\phi_{nj_1}^2(X_1)\,\bar\phi_{nj_2}^2(X_1)] \;=\; &p_{nj_1}p_{nj_2}(1+\theta_{nj_1})(1+\theta_{nj_2}) \\ &\times (p_{nj_1}(1+\theta_{nj_1}) + p_{nj_2}(1+\theta_{nj_2}) \\ &- 3p_{nj_1}p_{nj_2}(1+\theta_{nj_1})(1+\theta_{nj_2})).\end{aligned} \tag{3.6}$$

Доказательство Леммы 2.2. Начнем с доказательства (2.6). Для $x, y \in [0,1]$ обозначим

$$\bar\phi_{nj}(x) = \phi_{nj}(x) - \mathbf{E}_\theta\phi_{nj}(X_1) = \phi_{nj}(x) - \theta_{nj}p_{nj}$$

и

$$\widetilde\phi_{nj}(y) = \phi_{nj}(y) - \mathbf{E}_\tau\phi_{nj}(Y_1) = \phi_{nj}(y) - \tau_{nj}p_{nj}.$$

Тогда
$$T_n(\widehat{F}_n - \widehat{G}_n) = I_{1n} + I_{2n} + I_{3n} + W_n, \tag{3.7}$$
где
$$I_{1n} = 2\, I_{11n} + 2\, I_{12n} + 2\, I_{13n},$$
где
$$I_{11n} = \sum_{1\leqslant i_1<i_2\leqslant n} U_{1n}(X_{i_1}, X_{i_2}), \qquad I_{12n} = \sum_{1\leqslant i_1<i_2\leqslant l_n} U_{2n}(Y_{i_1}, Y_{i_2})$$
и
$$I_{13n} = \sum_{i_1=1}^{n} \sum_{i_2=1}^{l_n} U_{3n}(X_{i_1}, Y_{i_2}),$$
где
$$U_{1n}(X_{i_1}, X_{i_2}) = \sum_{j=1}^{m} \frac{\bar{\phi}_{nj}(X_{i_1})\bar{\phi}_{nj}(X_{i_2})}{np_{nj}},$$
$$U_{2n}(Y_{i_1}, Y_{i_2}) = \sum_{j=1}^{m} \frac{\widetilde{\phi}_{nj}(Y_{i_1})\widetilde{\phi}_{nj}(Y_{i_2})}{np_{nj}},$$
и
$$U_{3n}(X_{i_1}, Y_{i_2}) = \sum_{j=1}^{m} \frac{\bar{\phi}_{nj}(X_{i_1})\widetilde{\phi}_{nj}(Y_{i_2})}{np_{nj}}.$$
Наконец
$$I_{2n} = \sum_{j=1}^{m} \left(\frac{1}{n} \sum_{i=1}^{n} \bar{\phi}_{nj}(X_i) - \frac{1}{l_n} \sum_{i=1}^{l_n} \widetilde{\phi}_{nj}(Y_i) \right) \eta_{nj}, \tag{3.8}$$
$$I_{3n} = M_n(\eta) = n \sum_{j=1}^{n} p_{nj}\eta_{nj}^2 = T_n(F_n - G_{l_n}). \tag{3.9}$$
$$W_n = n^{-1} \sum_{j=1}^{m} \sum_{i=1}^{n} \bar{\phi}_{nj}^2(X_i) p_{nj}^{-1} + n l_n^{-2} \sum_{j=1}^{m} \sum_{i=1}^{l_n} \widetilde{\phi}_{nj}^2(Y_i) p_{nj}^{-1}. \tag{3.10}$$
Имеем
$$\mathbf{E} I_{1n} = 0, \qquad \mathbf{E} I_{2n} = 0, \tag{3.11}$$

$$
\begin{aligned}
\mathbf{E}[W_n] &= \sum_{j=1}^{m}(1 - p_{nj} + \theta_{nj}(1-2p_{nj}) - p_{nj}\theta_{nj}^2) \\
&\quad + n l_n^{-1}\sum_{j=1}^{m}(1 - p_{nj} + \tau_{nj}(1-2p_{nj}) - p_{nj}\tau_{nj}^2) \\
&= (1+a_n)\sum_{j=1}^{m}(1 - p_{nj} + \theta_{nj}(1-2p_{nj}) - p_{nj}\theta_{nj}^2) \\
&\quad + O(n^{-1/2} m M_{1n}^{1/2}(\eta))(1 + n^{-1}M_n(\eta))),
\end{aligned}
\tag{3.12}
$$

так как

$$
\begin{aligned}
\sum_{j=1}^{m}|\theta_{nj} - \tau_{nj}| &\leqslant \max_{1\leqslant j\leqslant m} p_{nj}^{-1}\sum_{j=1}^{m} p_{nj}|\eta_{nj}| \\
&\leqslant Cm\left(\sum_{j=1}^{m} p_{nj}\eta_{nj}^2\right)^{1/2}\left(\sum_{j=1}^{m} p_{nj}\right)^{1/2} \leqslant Cn^{-1/2} m M_{1n}^{1/2}(\eta)
\end{aligned}
\tag{3.13}
$$

и

$$
\begin{aligned}
\sum_{j=1}^{m}|\theta_{nj}^2 - \tau_{nj}^2| &\leqslant \max_{1\leqslant j\leqslant m} p_{nj}^{-1}\sum_{j=1}^{k} p_{nj}|\eta_{nj}|\,(|\theta_{nj}| + |\tau_{nj}|) \\
&\leqslant Cm^{-1}\left(\sum_{j=1}^{m} p_{nj}\eta_{nj}^2\right)^{1/2}\left(\sum_{j=1}^{m} p_{nj}(\theta_{nj}^2 + \tau_{nj}^2)\right)^{1/2} \\
&\leqslant Cn^{-1/2} m M_{1n}^{1/2}(\eta)(N_n(\theta) + N_n(\tau))^{1/2} \leqslant Cn^{-1/2} m M_{1n}^{1/2}(\eta),
\end{aligned}
\tag{3.14}
$$

так как

$$
|N_n^{1/2}(\tau) - N_n^{1/2}(\theta)| \leqslant n^{-1/2} M_n^{1/2}(\eta).
$$

Заметим, что остаточный член в правой части (3.12) есть $o(m_n)$ при $n \to \infty$, если $m_n = o(n^{2/3})$. □

3.2. Оценка $\mathbf{Var}[T_n]$. Имеем

$$
\mathbf{Var}[I_{11n}] = V_{11n} + V_{12n}, \tag{3.15}
$$

где

$$\begin{aligned} V_{11n} &= 2\sum_{j=1}^{m} p_{nj}^{-2}(\mathbf{Var}[\phi_j(X_1)])^2 \\ &= 2\sum_{j=1}^{m}(1-p_{nj}+\theta_{nj}(1-2p_{nj})-p_{nj}^2)^2 \\ &= 2\sum_{j=1}^{m}(1+\theta_{nj})^2(1+o(1)) \end{aligned} \tag{3.16}$$

и

$$\begin{aligned} V_{12n} &= 2\sum_{1\leqslant j_1<j_2\leqslant m} p_{nj_1}^{-1}p_{nj_2}^{-1}(\mathbf{Cov}[\phi_{j_1}(X_1),\phi_{j_2}(X_1)])^2 \\ &= 4\sum_{1\leqslant j_1<j_2\leqslant m} p_{nj_1}p_{nj_2}(1+\theta_{nj_1})^2(1+\theta_{nj_2})^2(1+o(1)) \\ &\leqslant (C+N_n^2(\theta_n))(1+o(1)). \end{aligned} \tag{3.17}$$

Следовательно

$$\mathbf{Var}[I_{11n}] = 2\sum_{j=1}^{m}(1+\theta_{nj})^2(1+o(1)). \tag{3.18}$$

Имеем

$$\begin{aligned} \mathbf{Var}[I_{12n}] &= 4a_n\sum_{j=1}^{m} p_{nj}^{-2}\mathbf{Var}[\phi_j(X_1)]\mathbf{Var}[\phi_j(Y_1)] \\ &= 4a_n\sum_{j=1}^{m}(1+\theta_{nj})(1+\tau_{nj})(1+o(1)). \end{aligned} \tag{3.19}$$

Аналогично (3.20) получаем

$$\mathbf{Var}[I_{13n}] = 2a_n^2\sum_{j=1}^{m}(1+\tau_{nj})^2(1+o(1)). \tag{3.20}$$

Нетрудно видеть, что

$$\mathbf{Cov}[I_{11n},I_{12n}]=0,\quad \mathbf{Cov}[I_{11n},I_{13n}]=0,\quad \mathbf{Cov}[I_{12n},I_{13n}]=0. \tag{3.21}$$

Таким образом из (3.20)– (3.21) следует

$$\mathbf{Var}[I_{1n}] = 2\sum_{j=1} m(1+a_n+\theta_{nj}+a_n\tau_{nj})^2(1+o(1)). \tag{3.22}$$

Оценка

$$\mathbf{Var}[I_{2n}] = J_{21n} + J_{22n} + J_{23n} + J_{24n}, \tag{3.23}$$

где

$$\begin{aligned} J_{21n} &= 2n^{-1}(n-1)^2 \\ &\times \sum_{1\leqslant j_1<j_2\leqslant m} \mathbf{Cov}[\phi_{j_1}(X_1), \phi_{j_2}(X_1)]\eta_{nj_1}\eta_{nj_2} \\ &= 2n^{-1}(n-1)^2 \\ &\times \sum_{1\leqslant j_1<j_2\leqslant m} p_{nj_1}p_{nj_2}(1+\theta_{nj_1})(1+\theta_{nj_2})\eta_{nj_1}\eta_{nj_2}(1+o(1)) \\ &\leqslant C\left(\sum_{j=1}^m p_{nj}(1+\theta_{nj_1})^2\right)\left(n\sum_{j=1}^m p_{nj}\eta_{nj}^2\right) \\ &\leqslant C\, M_{1n}(\eta)(1+N_n(\theta)), \end{aligned} \tag{3.24}$$

и

$$\begin{aligned} J_{22n} &= n^{-1}(n-1)^2 \sum_{j=1}^m \mathbf{Var}[\phi_{nj}(X_1)]\, \eta_{nj}^2 \\ &= n^{-1}(n-1)^2 \sum_{j=1}^m p_{nj}(1-p_{nj}+\theta_{nj}(1-2p_{nj}) - p_{nj}\theta_{nj}^2)\, \eta_{nj}^2 \\ &= n\sum_{j=1}^m p_{nj}(1+\theta_{nj})\, \eta_{nj}^2(1+o(1)) = O(m^{1/2}M_{1n}(\eta)), \end{aligned} \tag{3.25}$$

так как

$$\max_{1\leqslant j\leqslant m} |\theta_{nj}|^2 < C\, m\, N_n(\theta) < C\, m. \tag{3.26}$$

Слагаемые J_{23n} и J_{24n} получаются аналогично J_{21n} и J_{22n} соответственно только они образуются за счет параметров τ_{nj}, $1 \leqslant j \leqslant \infty$, а не θ_{nj}, $1 \leqslant j \leqslant \infty$. Они оцениваются аналогично и их оценка опускается.

Имеем

$$\mathbf{Var}[W_n] = A_{1n} + A_{2n} + A_{3n} + A_{4n}, \tag{3.27}$$

где

$$A_{1n} = n^{-1} \sum_{1\leqslant j_1<j_2\leqslant m} \mathbf{E}[\bar\phi_{nj_1}^2(X_1)\, \bar\phi_{nj_2}^2(X_1)]p_{nj_1}^{-1}p_{nj_2}^{-1} \tag{3.28}$$

и

$$A_{2n} = n^{-1} \sum_{j=1}^{m} \mathbf{E}[\bar{\phi}^4_{nj_1}(X_1)] p_{nj}^{-2}. \tag{3.29}$$

Слагаемые A_{3n} и A_{4n} задаются аналогично слагаемым A_{1n} и A_{2n} соответственно, только в них вместо случайных величин $X_1, \ldots, X_n$ входят случайные величины $Y_1, \ldots, Y_{l_n}$. Оцениваются они тоже аналогично и мы опустим их оценки.

Используя (3.4) и (3.26), имеем

$$\begin{aligned} A_{1n} &\leqslant n^{-1} \\ &\quad\times \sum_{1 \leqslant j_1 < j_2 \leqslant m} [p_{nj_1}(1+\theta_{nj_1})^2(1+\theta_{nj_2}) + p_{nj_2}(1+\theta_{nj_1})(1+\theta_{nj_2})^2] \\ &\leqslant Cn^{-1} \sum_{j=1}^{m} p_{nj}(1+|\theta_{nj}|)^2 \left(m + \sum_{j=1}^{m} p_{nj}|\theta_{nj}| \right) \\ &\leqslant Cn^{-1}(C + N_n(\theta))(m + m^{1/2}N^{1/2}(\theta)) \\ &\leqslant Cn^{-1}m + Cn^{-1}mN_n(\theta) + Cn^{-1}m^{1/2}N^{3/2}(\theta). \end{aligned} \tag{3.30}$$

Используя (3.6) и (3.26), имеем

$$\begin{aligned} A_{2n} = n^{-1} \sum_{j=1}^{m} p_{nj}^{-1}(1+\theta_{nj})[1 - 4p_{nj}(1+\theta_{nj}) \\ + 6p_{nj}^2(1+\theta_{nj})^2 - 3p_{nj}^3(1+\theta_{nj})^3]. \end{aligned} \tag{3.31}$$

Мы приведем оценку оценку только двух слагаемых, входящих в A_{2n}. Другие два оцениваются аналогично и имеют меньший порядок.

Имеем

$$\begin{aligned} n^{-1} \sum_{j=1}^{m} p_{nj}^{-1}(1+\theta_{nj}) &\leqslant Cn^{-1}m^2 \left(1 + \sum_{j=1}^{m} p_{nj}|\theta_{nj}| \right) \\ &\leqslant Cn^{-1}m^2 \left(1 + \left(\sum_{j=1}^{m} p_{nj}\theta_{nj}^2 \right)^{1/2} \right) \\ &\leqslant Cn^{-1}m^2(1 + N_n(\theta)) = o(m) \end{aligned} \tag{3.32}$$

и

$$n^{-1}\sum_{j=1}^{m} p_{nj}^2(1+\theta_{nj})^4 \leqslant Cn^{-1}m^{-1} + n^{-1}\sum_{j=1}^{m} p_{nj}^2\theta_{nj}^4 \leqslant Cn^{-1}(m^{-1}+N_n^2(\theta)). \tag{3.33}$$

Следовательно

$$A_{2n} \leqslant Cn^{-1}m^2(1+N_n^{1/2}(\theta)) + n^{-1}N_n^2(\theta). \tag{3.34}$$

3.3. Состоятельность оценок смещения и дисперсии статистики T_n. Покажем состоятельность оценок $\sum_{j=1}^{m} g_{nj}\theta_{nj}$ в (2.7) и (2.8). Имеем

$$\begin{aligned}
\mathbf{Var}\left[\sum_{j=1}^{m} g_{nj}\frac{1}{n}\sum_{i=1}^{n}\frac{\phi_{nj}(X_i)}{p_{nj}}\right] &= \frac{1}{n}\sum_{j=1}^{m} g_{nj}^2\frac{\mathbf{Var}[\phi_{nj}(X_1)]}{p_{nj}^2} \\
&+ \frac{1}{n}\sum_{1\leqslant j_1<j_2\leqslant m} g_{nj_1}g_{nj_2}\frac{\mathbf{Cov}[\phi_{nj_1}(X_1),\phi_{nj_2}(X_1)]}{p_{nj_1}p_{nj_2}} \\
&= \frac{1}{n}\sum_{j=1}^{m} g_{nj}^2\frac{1+\theta_{nj}}{p_{nj}}(1+o(1)) \\
&+ \frac{1}{n}\sum_{1\leqslant j_1<j_2\leqslant m} g_{nj_1}g_{nj_2}(1+\theta_{nj_1}+\theta_{nj_2})(1+o(1)) = o(m),
\end{aligned} \tag{3.35}$$

так как

$$n^{-1}\sum_{j=1}^{m}\frac{\theta_{nj}}{p_{nj}} \leqslant Cn^{-1}m^2\sum_{j=1}^{m} p_{nj}\theta_{nj} \leqslant Cn^{-1}m^2N_n^{1/2}(\theta) = o(m) \tag{3.36}$$

и

$$n^{-1}m\sum_{j=1}^{m} g_{nj}\theta_{nj} \leqslant Cn^{-1}m\max_{1\leqslant j\leqslant m} p_{nj}^{-1}\sum_{j=1}^{m} p_{nj}\theta_{nj} \leqslant Cn^{-1}m^2N_n^{1/2}(\theta) = o(m) \tag{3.37}$$

Из слагаемых возникающих при адаптивном оценивании дисперсии мы оценим только одно, так как оценки остальных аналогичны.

Имеем

$$n^{-4}\mathbf{Var}\left[\sum_{j=1}^{m} g_{nj}^2 p_{nj}^{-2} \sum_{1\leqslant i_1<i_2<n} \phi_{nj}(X_{i_1})\phi_{nj}(X_{i_2})\right] \leqslant B_{1n}+B_{2n}, \quad (3.38)$$

где

$$\begin{aligned} B_{1n} &= Cn^{-2}\sum_{j=1}^{m} p_{nj}^{-4}(\mathbf{Var}[\phi_{nj}(X_1)])^2 \\ &\leqslant Cn^{-2}\sum_{j=1}^{m} p_{nj}^{-2}(1+\theta_{nj})^2(1+o(1)) \\ &\leqslant Cn^{-2}\max_{1\leqslant j\leqslant m} p_{nj}^{-3}\sum_{j=1}^{m} p_{nj}(1+\theta_{nj})^2 \\ &= O(n^{-2}m^3(1+N_n(\theta)) = o(m) \end{aligned} \quad (3.39)$$

и

$$\begin{aligned} B_{2n} &= Cn^{-2}\sum_{1\leqslant j_1<j_2\leqslant m} \frac{(\mathbf{Cov}[\phi_{nj_1}(X_1),\phi_{nj_2}(X_1)])^2}{p_{nj_1}^2 p_{nj_2}^2} \\ &\leqslant Cn^{-2}\sum_{1\leqslant j_1<j_2\leqslant m}(1+\theta_{nj_1}+\theta_{nj_2})^2 \leqslant cn^{-2}m^2 \\ &+ Cn^{-2}m\max_{1\leqslant j\leqslant m} p_{nj}^{-1}\left(\left|\sum_{j=1}^{m} p_{nj}\theta_{nj}\right| + \sum_{j=1}^{m} p_{nj}\theta_{nj}^2\right) \\ &\leqslant Cn^{-2}m^2(1+N_n(\theta)) = o(1). \end{aligned} \quad (3.40)$$

Как изменятся оценки дисперсии и как это повлияет на адаптивные оценки, когда у нас вместо вместо наблюдений X_i появляются наблюдения Y_i и $N_n(\tau_n)\to\infty$ при $n\to\infty$. В окончательных оценках $N_n(\theta_n)$ заменится на $N_n(\tau_n)$. Но

$$N_n^{1/2}(\tau_n) \leqslant N_n^{1/2}(\theta_n) + n^{-1/2}M_n^{1/2}(\eta_n). \quad (3.41)$$

Так как $N_n^{1/2}(\theta_n) < C < \infty$, нам достаточно показать, что если в окончательных оценках мы заменим $N_n^{1/2}(\theta_n)$ на $n^{-1}M_n(\eta_n)$, то они будут меньше по порядку, чем $M_n^2(\eta_n)$.

Заметим, что в оценках (3.12)–(3.40) наибольшими порядками в окончательных оценках для функции распределения G_{l_n} являются

$M_{1n}(\eta_n)N_n(\tau_n)$ (аналог (3.24)), $n^{-1}m^2N_n^{1/2}(\tau_n)$ (аналог (3.30)) и $n^{-1}N_n^2(\tau_n)$ (аналог (3.34)).

Определенного обоснования заслуживает только оценка $n^{-1}m^2N_n^{1/2}(\tau_n)$. Ее вкладом мы можем пренебречь, так как

$$n^{-3/2}m^2M_n^{1/2}(\eta_n)M_n^{-2}(\eta_n) = O(n^{-3/2}m_n^2m_n^{-3/4}) = o(1), \tag{3.42}$$

когда $m_n^{-1/2}M_n(\eta_n) \to \infty$ при $n \to \infty$.

Таким образом

$$M_n(\eta_n)\widehat{\sigma}_n \to_P \infty, \tag{3.43}$$

когда $m_n^{-1/2}M_n(\eta_n) \to \infty$ при $n \to \infty$. Поэтому вероятности ошибок второго рода критериев K_n будут стремиться к нулю, если $N_n(\tau_n) \to \infty$ при $n \to \infty$.

3.4. Асимптотическая нормальность статистики T_n. Для получения асимптотик вероятностей ошибок первого и второго рода в теоремах 2.1 – 2.3 достаточно доказать асимптотическую нормальность статистики I_{1n}. При этом достаточно предположить, что $(F_n, G_{l_n}) \in \Xi_n(C)\times\Xi_n(C)$ для некоторого $C > 0$. Хотя статистика I_{1n} не является U–статистикой, для нее применима та же самая мартингальная техника доказательства [2, 5, 11, 14] и получается аналогичный результат, что и в случае критериев согласия [5,14]. Поскольку в [1] аналогичные рассуждения для проверки гипотезы однородности на основе тестовой статистики, построенной по $\mathbb{L}_2$ – норме ядерной оценки плотности, опущены, мы наметим их в этой работе.

Будем считать $l_n \leqslant n$.

Определим мартингал W_{ni}, $1 \leqslant i \leqslant n + l_n$, по индукции следующим образом. Положим

$$W_{n1} = U_{1n}(X_1, X_1), \quad \text{и} \quad W_{n2} = U_{2n}(Y_1, Y_1) + U_{3n}(X_1, Y_1).$$

Если i–нечетно, то положим $j = [i/2]$ и

$$W_{ni} = \sum_{s=1}^{j} U_{1n}(X_j, X_s) + \sum_{s=1}^{j-1} U_{3n}(X_j, Y_s).$$

Если i–четно, $i \leqslant 2l_n$, то положим $j = i/2$ и

$$W_{ni} = \sum_{s=1}^{j} U_{2n}(Y_j, Y_s) + \sum_{s=1}^{j-1} U_{3n}(X_s, Y_j).$$

Если $i \geqslant 2l_n$, то положим $j = i - l_n$ и

$$W_{ni} = \sum_{s=1}^{j} U_{1n}(X_j, X_s) + \sum_{s=1}^{l_n} U_{3n}(X_j, Y_s).$$

К этим мартингалам мы можем применить рассуждения работы [11] и получить следующий аналогичный результат.

Положим

$$V_{1n}(x,y) = \mathbf{E}[U_{1n}(x,X_1)U_{1n}(y,X_1)], \quad V_{2n}(x,y) = \mathbf{E}[U_{1n}(x,Y_1)U_{1n}(y,Y_1)],$$
$$V_{3n}(x,y) = \mathbf{E}[U_{3n}(X_1,x)U_{3n}(X_1,y)], \quad V_{4n}(x,y) = \mathbf{E}[U_{3n}(x,Y_1)U_{3n}(y,Y_1)].$$

Теорема 3.1. *Статистика I_{1n} асимптотически нормальна с нулевым средним и дисперсией σ_1^2, если*

$$\lim_{n\to\infty} m_n^{-1}[\mathbf{E}[V_{1n}^2(X_1,X_2) + V_{2n}^2(Y_1,Y_2) + V_{3n}^2(X_1,X_2) + V_{4n}^2(Y_1,Y_2)] + n^{-1}\mathbf{E}[U_{1n}^4(X_1,X_2) + U_{2n}^4(Y_1,Y_2) + U_{3n}^4(X_1,Y_1)]] = 0. \quad (3.44)$$

Доказательство теоремы по существу повторяет рассуждения доказательства аналогичного утверждения для проверки асимптотической нормальности в [11] и опускается.

Проверка условия (3.44) практически не отличается от проверки аналогичных условий в случае гипотезы согласия в [5]. Более того большинство необходимых оценок для доказательства (3.44) и оценок в [5] совпадают. Поэтому мы тоже опустим эти рассуждения.

Список литературы

1. N. Anderson, P. Hall, D. Titterington, *Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates.* — J. Multivariate Anal., **50** (1994), 41–54.
2. B. M. Brown, *Martingale central limit theorems.* — Ann. Math. Statist. **42** (1971), 59–66.
3. A. R. Barron, *Uniformly powerful goodness of fit tests.* — Ann. Statist., **17** (1989), 107–124.
4. D. M. Chibisov, *Asymptotic optimality of the chi-square test with large number of degrees of freedom within the class of symmetric tests.* — Math. Methods Statist., **1** (1992) 55–82.
5. М. С. Ермаков, *Асимптотическая минимаксность критериев хи-квадрат.* — Теория вероятн. и ее примен., **42** (1997), 668–695.
6. М. С. Ермаков, *О равномерной состоятельности непараметрических критериев. I.* — Зап. нучн. семин. ПОМИ **495** (2020), 147–176.

7. M. Fromont, B. Laurent, M. Lerasle, P. Reynaud-Bouret, *Kernels based tests with non-asymptotic bootstrap approaches for two-sample problem.* — JMLR: Workshop and Conference Proceedings, **23** (2012), 23–41.
8. M. Fromont, B. Laurent, P. Reynaud-Bouret, *The two-sample problem for poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach.*– The Annals of Statistics. **41** (2013), 1431–1461.
9. A. Gretton, K. Borgwardt, M. Rasch, B. Scholkopf, A. Smola, *A kernel two-sample test.* — J. Machine Learning Research. **13** (2012), 723–773.
10. A. Gretton, D. Sejdinovic, H. Strathmann, S. Balakrishnan, M. Pontil, K. Fukumizu, B. K. Sriperumbudur, *Optimal kernel choice for large-scale two-sample tests.* — Advances Neural Information Processing systems (2012) 1205–1213.
11. P. Hall, *Central limit theorem for integrated square error of multivariate nonparametric density estimators.* — J. Multivariate Analysis, **14** (1984), 1–16.
12. Г. И. Ивченко, Я. Ю. Медведев, *Разделимые статистики и проверка гипотез для группированных данных.* — Теория вероятн. и ее примен. bf 25 (1980), 549–560.
13. Ю. И. Ингстер, *О сравнении минимаксных свойств тестов Колмогорова, ω^2 и χ^2.* — Теор. вероятн. и ее примен., **32** (1987), 374–378.
14. Yu. I. Ingster, I. A. Suslina, *Nonparametric Goodness-of-fit Testing under Gaussian Models* — Lecture Notes in Statistics **169** Springer: N.Y. (2002).
15. H. B. Mann, A. Wald, *On the choice of the number of intervals in the application of chi-squared test.* — Ann. Math. Statist., **13** (1942), 306–318.
16. C. Morris, *Central limit theorems for multinomial sums.* — Ann. Statist., **3** (1975), 165–188.
17. J. Robins, L. Li, E. T. Tchetgen, Aad van der Vaart, *Asymptotic Normality of Quadratic Estimators.* — Stochastic Processes and their Applications. **126** (2015), 3733–3759.
18. T. Li, M. Yuan, *On the Optimality of Gaussian Kernel Based Nonparametric Tests against Smooth Alternatives.* `arXiv:1909.03302v1` (2019) 42p.

Ermakov M. S. Chi-squared test for testing of homogeneity.

We provide necessary and sufficient conditions of uniform consistency of nonparametric sets of alternatives of chi-squared test for testing of hypothesis of homogeneity. The number of cells of chi-squared test increases with sample size growth. Nonparametric sets of alternatives can be defined both in terms of densities and distribution functions.

Институт проблем машиноведения
РАН и Санкт-Петербургский государственный
университет, Санкт-Петербург, Россия
E-mail: erm2512@gmail.com

Поступило 7 сентября 2021 г.