Math-Net.Ru

Общероссийский математический портал

М. С. Ермаков, Асимптотическая минимаксность критериев хи-квадрат, *Теория вероятн. и ее примен.*, 1997, том 42, выпуск 4, 668–695

DOI: <https://doi.org/10.4213/tvp2179>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.80.97.219

26 апреля 2023 г., 11:07:12

© 1997 г.

ЕРМАКОВ М. С.*

АСИМПТОТИЧЕСКАЯ МИНИМАКСНОСТЬ
КРИТЕРИЕВ ХИ-КВАДРАТ

Рассматривается асимптотическое поведение критериев хи-квадрат, когда число ячеек k_n растет вместе с ростом объема выборки n . Доказывается, что последовательность критериев хи-квадрат является асимптотически минимаксной, если $k_n = o(n^2)$ при $n \rightarrow \infty$. Доказательство использует новый вариант теоремы об асимптотической нормальности хи-квадрат статистик.

Ключевые слова и фразы: критерии хи-квадрат, асимптотическая эффективность, асимптотическая нормальность, асимптотическая минимаксность, критерии согласия.

1. Введение. Исследование задач проверки гипотез как задач с достаточно широким непараметрическим множеством альтернатив (см. [22], [26], [13]–[16], [4]–[10], [11] и др.) становится естественным дополнением к традиционной параметрической постановке задачи проверки гипотез. Особое внимание при этом обращается на задачи проверки гипотезы согласия. Исследования в этой области естественно разделяются на две части: проверки гипотез о плотности и функции распределения.

Первая из этих двух задач обычно рассматривается в следующей постановке. Пусть $X_1, \dots, X_n$ — независимые одинаково распределенные случайные величины с неизвестной плотностью распределения f . Предполагается, что имеется априорная информация $f \in U$, где U — множество функций, удовлетворяющих некоторым условиям гладкости. Необходимо проверить гипотезу $f = f_0$ против альтернатив $f \in \Gamma_n = \{f: \|f - f_0\| > b_n > 0, f \in U\}$. Здесь f_0 — некоторая известная плотность распределения и $\|\cdot\|$ — норма в некотором банаховом пространстве функций, содержащем множество U . Для этой постановки Ингстер [13]–[15] изучал скорость сходимости b_n к нулю при $n \rightarrow \infty$, позволяющую различить гипотезу и альтернативы $f \in \Gamma_n$. Для задачи проверки гипотезы в L_2 -норме асимптотически минимаксные критерии были построены Ингстером [16] в предположении, что плотность F бесконечно дифференцируема, и Ермаковым [4] при слабых условиях на гладкость плотности f .

*Институт проблем машиноведения РАН, Большой просп., 61, 199178 С.-Петербург, Россия.

Задачу проверки гипотезы о функции распределения естественно рассматривать в рамках теории статистических выводов о значении функционала. Такой подход был предложен в хорошо известной работе Стейна [26] и состоит в следующем. Пусть $X_1, \dots, X_n$ — независимые одинаково распределенные величины с функцией распределения F , и пусть T_n — функционал на множестве $\mathfrak{F}$ всех функций распределения. Задача состоит в проверке гипотез $T_n(F) = 0$, $F \in \mathfrak{F}$, против альтернатив $F \in \mathfrak{F}(T_n, b_n) = \{F: T_n(F) > b_n, F \in \mathfrak{F}\}$. В такой постановке обычно удается доказать асимптотическую минимаксность тестовой статистики $T_n(\hat{F}_n)$, где $\hat{F}_n$ — эмпирическая функция распределения $X_1, \dots, X_n$. Такие результаты были получены для функционалов T , порождающих статистики Колмогорова, ранговых критериев и омега-квадрат, а также были анонсированы для критериев хи-квадрат с растущим числом ячеек (см. [5], [6], [8], [10]). Отметим, что асимптотическая минимаксность критериев хи-квадрат с фиксированным числом ячеек следует из результатов Вальда [3] (см. также [2]). В практических приложениях число ячеек зависит от объема выборки n и растет вместе с ростом объема выборки.

В случае критерия хи-квадрат точная постановка задачи состоит в следующем. Пусть $X_1, \dots, X_n$ — независимые одинаково распределенные случайные величины с функцией распределения $F(x)$, $x \in [0, 1]$. Пусть интервал $[0, 1]$ разбит на $k = k_n$ подынтервалов $I_{jn} = [e_{jn}, e_{j+1,n})$, $p_{jn} = e_{j+1,n} - e_{jn} > 0$, $e_{0n} = 0$, $e_{kn} = 1$, $1 \leq j \leq k = k_n$. Для фиксированных чисел $g_{jn} > 0$, $1 \leq j \leq k$, определим функционал

$$T_n(F) = n \sum_{j=1}^k g_{jn} (r_{jn} - p_{jn})^2 p_{jn}^{-1},$$

где $r_{jn} = F(e_{jn}) - F(e_{j-1,n})$ для всех $1 \leq j \leq k_n$. Функционал T_n и порождает тестовую статистику критерия хи-квадрат $T_n(\hat{F}_n)$. Для данных последовательностей функционалов T_n мы показываем, что статистики критериев хи-квадрат $T_n(\hat{F}_n)$ являются асимптотически минимаксными для задач проверки гипотез $T_n(F) = 0$, $F \in \mathfrak{F}$, против альтернатив $F \in \mathfrak{F}(T_n, b_n)$.

Отметим, что такая постановка задачи может также интерпретироваться и как задача проверки гипотезы о плотности распределения. С этой точки зрения можно сказать, что проверка гипотез, основанная на критериях хи-квадрат, занимает промежуточную позицию между задачами проверки гипотез о плотности и функции распределения. Интерпретация в терминах проверки гипотез о плотности распределения использует представление тестовой статистики критерия хи-квадрат как L_2 -расстояния между гистограммой и плотностью распределения гипотезы и может быть описана следующим образом. Определим функцию $\hat{f}_n(x) = r_{jn}/p_{jn}$ для $x \in I_{jn}$, $1 \leq j \leq k$. Функция $\hat{f}_n$ может рассматри-

ваться как аппроксимация плотности f и в случае $F = \hat{F}_n$ совпадает с обычной гистограммой. Мы назовем $\bar{f}_n$ гистограммной функцией для F . При определенном выборе весов g_{jn} получаем

$$T_n(F) = n \int_0^1 (\bar{f}(x) - \bar{f}_0(x))^2 dx. \quad (1.1)$$

Здесь $\bar{f}_0$ — гистограммная функция для равномерного распределения, т.е., $\bar{f}_0(x) = 1$ для всех $x \in (0, 1)$. Таким образом, имеет место следующая ситуация. Мы не можем проверить гипотезу о плотности распределения для L_2 -нормы без какой-либо априорной информации о ее гладкости. В тоже время мы можем заменить плотность f ее простой аппроксимацией $\bar{f}_T$, рассмотреть задачу проверки гипотез в терминах функционала $T_n(F)$ и таким образом решить задачу.

Связь критериев хи-квадрат с задачей проверки гипотез о плотности распределения хорошо известна. На ее основе Ингстер [16] изучил задачу выбора числа ячеек для критериев хи-квадрат при априорной информации о гладкости плотности f . Он рассмотрел случай ячеек одинаковой длины, $p_{jn} = 1/k$, и показал, что если плотность f принадлежит шару в пространстве Соболева W_2^{2r} с $r > 0$, то оптимальный выбор числа ячеек есть $k_n = O(n^{2/(4r+1)})$. Таким образом, если f принадлежит шарам в W_2^{2r} для всех r , то $k_n = o(n^{2-\varepsilon})$ с произвольным $\varepsilon > 0$. Другие подходы к этой задаче развивались Манном и Вальдом [22], Холстом [29], Чибисовым и Гванцеладзе [31], Калленбергом, Оостерхофом и Шрайбером [19], Ивченко и Чукановым [18], Чибисовым [30] и др.

Результаты работы состоят в следующем. Если ячейки равной длины $p_{jn} = 1/k$, $1 \leq j \leq k = k(n)$, асимптотическая минимаксность критериев хи-квадрат доказана для $k_n = o(n^2)$. Для ячеек различной длины то же самое утверждение получено, если $n \min\{p_{jn}, 1 \leq j \leq k(n)\} \rightarrow \infty$ при $n \rightarrow \infty$. Это предположение означает, что число наблюдений в каждой ячейке стремится к бесконечности при $n \rightarrow \infty$. Если это предположение не выполнено и $k = o(n^2)$, мы показываем асимптотическую минимаксность некоторой модификации критерия хи-квадрат.

Доказательство результатов основано на следующем подходе. Мы задаем функционал T_n как сумму квадратов квазиортогональных функций

$$T_n(F) = n \sum_{j=1}^k g_{jn} \left(\int_0^1 \phi_{jn}(x) dF(x) \right)^2 p_{jn}^{-1},$$

где $\phi_{jn}(x) = \chi(x \in I_{jn}) - p_{jn}$, $x \in (0, 1)$, $1 \leq j \leq k(n)$, и $\chi(A)$ обозначает индикатор события A . Такое представление функционала $T_n(F)$ позволяет использовать технику, развитую для задач асимптотически минимаксной проверки гипотез о плотности распределения (см. [4], [7], [9] и [16]). В задаче проверки гипотез о плотности распределения соот-

ветствующий функционал $\bar{T}_n$ является L_2 -нормой отклонения плотности альтернативы от плотности гипотезы и может быть задан как

$$\bar{T}_n(F) = \sum_{j=1}^{\infty} \left(\int_0^1 \xi_j(x) d(F - F_0)(x) \right)^2,$$

где ξ_j , $1 \leq j < \infty$, — ортогональная система функций и F_0 — функция распределения гипотезы. Функции ϕ_{jn} , $1 \leq j \leq k$, не являются ортонормальными и не удовлетворяют каким-либо предположениям из [4], [7], [9], [16], что вызывает существенные отличия в доказательстве. Более того, прямое применение соответствующих методов, требующее серьезной модификации аналитической техники, позволяет показать асимптотическую минимаксность критериев хи-квадрат только в случае $k = o(n)$. Для того чтобы ослабить это предположение, мы существенно модифицировали метод доказательства. Если в [4], [7], [16] доказательство базируется на анализе асимптотических разложений логарифмов отношений правдоподобия, то в этой работе оно основано на непосредственном анализе асимптотических разложений отношений правдоподобия.

Верхние границы для вероятностей ошибок первого и второго рода следуют из теоремы об асимптотической нормальности статистик критерия хи-квадрат. Доказанная версия данной теоремы допускает следующую интерпретацию. Если вероятности ошибок первого и второго рода критериев хи-квадрат отделены от нуля и единицы, то соответствующая последовательность статистик критериев асимптотически нормальна как в случае гипотез, так и в случае последовательности альтернатив. Доказательство базируется на представлении хи-квадрат статистик как U -статистик и применении версии центральной предельной теоремы для мартингалов, предложенной Холлом [27]. Отметим, что асимптотическая нормальность статистик хи-квадрат доказывалась при различных предположениях в целом ряде работ (см. [27], [25], [29], [24], [23] и др.).

Расстояние (1.1) в метрике L_2 между гистограммой и плотностью не единственная мера отклонения гистограммы от плотности, используемая в проверке гипотез. Широкий класс статистик (см. [11]) может быть задан как

$$S_n(\hat{F}_n) = \sum_{j=1}^k np_{jn} h \left(\frac{\hat{r}_{jn} - np_{jn}}{np_{jn}} \right) - nh(0), \quad (1.2)$$

где $\hat{r}_{jn} = \hat{F}_n(e_{jn}) - \hat{F}_n(e_{j-1,n})$, $1 \leq j \leq k$, и функция h удовлетворяет некоторым условиям гладкости. В частности, класс (1.2) содержит статистики Кресси–Рида [20]

$$R^{(\lambda)}(\hat{F}_n) = \frac{2}{\lambda(\lambda-1)} \sum_{j=1}^k \hat{r}_{jn} \left(\left(\frac{\hat{r}_{jn}}{p_{jn}} \right)^{\lambda} - 1 \right), \quad (1.3)$$

определенные по непрерывности по λ для $\lambda = 0, -1$. Класс статистик R^λ содержит, в частности, статистики отношения правдоподобия ($\lambda = 0$), статистики Фримана-Тьюки ($\lambda = -\frac{1}{2}$) и статистики модифицированного критерия хи-квадрат Неймана ($\lambda = -2$).

Соответствующий функционал S_n на множестве всех функций распределения равен

$$S_n(F) = n \sum_{j=1}^k p_{jn} h \left(p_{jn}^{-1} \int_0^1 \phi_{jn} dF \right) - nh(0). \quad (1.4)$$

Мы будем предполагать, что функция $h(s)$ — возрастающая относительно $|s|$, $s \in \mathbf{R}^1$, дважды непрерывно дифференцируема в 0 и $h''(0) \neq 0$.

Как известно, во многих случаях статистики хи-квадрат $T_n(\hat{F}_n)$ и тестовые статистики $S_n(\hat{F}_n)$ асимптотически эквивалентны (имеют то же самое асимптотическое поведение). В данной работе мы указываем условия, когда такая эквивалентность имеет место, если число ячеек растет вместе с ростом объема выборки. Мы также рассматриваем задачу проверки гипотезы $S_n(F) = 0$ против $F \in \mathfrak{F}(S_n, b_n)$. Для этой задачи мы доказываем асимптотическую минимаксность как тестовой статистики $T_n(\hat{F}_n)$, так и тестовой статистики $S_n(\hat{F}_n)$. Последнее утверждение вытекает непосредственно из асимптотической эквивалентности этих статистик.

Заметим, что статистики хи-квадрат могут быть асимптотически минимаксны, даже если не выполнено условие, что функция h дважды непрерывно дифференцируема в нуле. В качестве примера мы доказываем такой результат для функционалов

$$S_{n\gamma}(F) = n^\gamma \sum_{j=1}^k g_{jn}^\gamma p_{jn}^\gamma \left(p_{jn}^{-1} \int_0^1 \phi_{jn} dF \right)^{2\gamma}, \quad 0 < \gamma \leq 1, \quad (1.5)$$

или, в терминах гистограммной функции, в случае ячеек равной длины и $g_j = 1$, $1 \leq j \leq k$,

$$S_{n\gamma}(F) = n^\gamma k^{1-\gamma} \int_0^1 |\bar{f}(x) - 1|^{2\gamma} dx. \quad (1.6)$$

Мы будем использовать следующие обозначения. Обозначим C, c произвольные постоянные, $\#\{i, i \in B\}$ — число элементов множества B , $[z]$ — целую часть числа $z \in \mathbf{R}^1$ и $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x \exp\{-s^2/2\} ds$ — функцию распределения стандартного нормального закона. Когда ясно из контекста, индекс n будет опускаться в обозначениях. Например, $p_j = p_{jn}$ и $g_j = g_{jn}$.

2. Основные результаты. Для любого критерия $L_n = L_n(X_1, \dots, X_n)$ обозначим $\alpha(L_n)$ его вероятность ошибки первого рода и $\beta(L_n, F)$ —

его вероятность ошибки второго рода при альтернативе $F \in \mathfrak{F}(T_n, b_n)$. Положим $\beta(L_n, b_n) = \sup\{\beta(L_n, F): F \in \mathfrak{F}(T_n, b_n)\}$.

Обозначим

$$d_n = \mathbf{E}_0[T_n(X_n)] = \sum_{j=1}^k g_j - \sum_{j=1}^k g_j p_j,$$

$$\sigma_n^2 = \text{Var}_0[T_n(X_n)] = 2k^{-1} \sum_{j=1}^k g_j^2.$$

Определим последовательность критериев хи-квадрат

$$K_n = \chi\left\{k^{-1/2} \sigma_n^{-1} (T_n(X^{(n)}) - d_n) > x_{\alpha_n} (1 + o(1))\right\},$$

где x_{α_n} определяется значением вероятности ошибки первого рода $\alpha_n = \alpha(K_n)$.

Предположим, что выполнено одно из следующих трех условий.

A. $k = k_n = o(n^2)$, $k_n \rightarrow \infty$ при $n \rightarrow \infty$ и $p_{jn} = 1/k$ для всех $1 \leq j \leq k$.

B. $n \min\{p_{jn}, 1 \leq j \leq k\} \rightarrow \infty$, $\max\{p_{jn}, 1 \leq j \leq k\} \rightarrow 0$ при $n \rightarrow \infty$.

C. $nk_n^{1/2} \min\{p_{jn}, 1 \leq j \leq k_n\} \rightarrow \infty$, $\max\{p_{jn}, 1 \leq j \leq k_n\} \rightarrow 0$ при $n \rightarrow \infty$.

Если $p_{jn} \asymp 1/k_n$ для всех $1 \leq j \leq k(n)$, то условие C означает $k_n = o(n^2)$ и $p_{jn} = o(1)$ при $n \rightarrow \infty$.

Мы скажем, что последовательность критериев N_n является асимптотически минимаксной, если для любой последовательности критериев L_n , $\alpha(L_n) < \alpha(N_n)$,

$$\limsup_{n \rightarrow \infty} (\beta(N_n, b_n) - \beta(L_n, b_n)) \leq 0.$$

Теорема 2.1. Пусть выполнены условия A или B. Пусть

$$0 < \liminf_{n \rightarrow \infty} b_n k_n^{-1/2} \leq \limsup_{n \rightarrow \infty} b_n k_n^{-1/2} < \infty. \quad (2.1)$$

Тогда последовательность критериев хи-квадрат K_n является асимптотически минимаксной. Пусть $0 < c_1 \leq \alpha_n = \alpha(K_n) \leq c_2 < 1$. Тогда x_{α_n} может быть определено из соотношения $\alpha_n = 1 - \Phi(x_{\alpha_n})$ и

$$\beta(K_n, b_n) = \Phi(x_{\alpha_n} - b_n k_n^{-1/2}) (1 + o(1)) \quad \text{при } n \rightarrow \infty. \quad (2.2)$$

З а м е ч а н и е. Из теоремы 2.1 следует, что если $\alpha(K_n) > \alpha > 0$ и $b_n k_n^{-1/2} \rightarrow 0$ при $n \rightarrow \infty$, то $\beta(K_n, b_n) = 1 + o(1)$ при $n \rightarrow \infty$. Если $\alpha(K_n) > \alpha > 0$ и $b_n k_n^{-1/2} \rightarrow \infty$ при $n \rightarrow \infty$, то $\beta(K_n, b_n) = o(1)$ при $n \rightarrow \infty$.

В случае выполнения более общего условия С мы можем доказать асимптотическую минимаксность только лишь некоторой модификации статистики хи-квадрат

$$T_{1n}(\hat{F}_n) = T_n(\hat{F}_n) - \sum_{j=1}^k g_{jn} \hat{r}_{jn} \frac{1 - 2p_{jn}}{p_{jn}n} - \sum_{j=1}^k g_{jn} p_{jn}. \quad (2.3)$$

Рассмотрение статистики T_{1n} может быть объяснено на основе следующих соображений. Мы имеем

$$T_n(\hat{F}_n) = n^{-1} \sum_{j=1}^k g_{jn} \left(\sum_{s_1=1}^n \phi_j(X_{s_1}) \right)^2 p_{jn}^{-1}, \quad (2.4)$$

$$\begin{aligned} T_{1n}(\hat{F}_n) &= T_n(\hat{F}_n) - W_n(\hat{F}_n) \\ &= 2n^{-1} \sum_{j=1}^k g_{jn} \sum_{1 \leq s_1 < s_2 \leq n} \phi_j(X_{s_1}) \phi_j(X_{s_2}) p_{jn}^{-1}, \end{aligned} \quad (2.5)$$

где

$$W_n(\hat{F}_n) = n^{-1} \sum_{j=1}^k g_{jn} \sum_{s=1}^n \phi_j^2(X_s) p_{jn}^{-1}.$$

Если выполнено условие А, то $W_n(\hat{F}_n) = d_n = \text{const.}$

Если выполнено условие В, то

$$\text{Var}_F[W_n(\hat{F}_n)] = o(\text{Var}_F[T_n(\hat{F}_n)]) \quad (2.6)$$

для всех $F \in \mathfrak{F}(T_n, b_n)$.

Следовательно, влияние слагаемых W_n в этих случаях пренебрежимо. Если выполнено условие С, то в большинстве случаев

$$\text{Var}_F[T_{1n}(\hat{F}_n)] = o(\text{Var}_F[T_n(\hat{F}_n)]). \quad (2.7)$$

Это означает, что влияние W_n весьма существенно. По этой причине и была введена модификация $T_{1n}(\hat{F}_n)$ хи-квадрат статистики $T_n(\hat{F}_n)$. Отметим, что аналогичная модификация асимптотически минимаксных тестовых статистик также рассматривалась и для задач проверки гипотез о плотности распределения (см. [4], [7], [16]).

Пусть $g_{jn} = 1$, $p_{jn} = 1/k$ для всех $1 \leq j \leq k$. Тогда тестовая статистика

$$\begin{aligned} T_{1n}(X^{(n)}) &= 2n^{-1} \left(\sum_{j=1}^k \sum_{1 \leq s_1 < s_2 \leq n} \chi(X_{s_1} \in I_{jn}) \right. \\ &\quad \left. \times \chi(X_{s_2} \in I_{jn}) - \frac{n(n-1)}{2k} \right) \end{aligned} \quad (2.8)$$

может интерпретироваться как нормированная сумма числа пар наблюдений X_{s_1}, X_{s_2} , содержащихся в ячейках. Такое представление статистики T_{1n} позволяет дать следующую интерпретацию результатов, если $k_n = o(n^2)$, т.е. выполнено условие С. Гипотезы и множества альтернатив различимы на основе тестовой статистики $T_{1n}(X^{(n)})$, если число ячеек, содержащих по крайней мере два наблюдения, стремится к бесконечности с ростом объема выборки.

Определим последовательность критериев

$$K_{1n}(X^{(n)}) = \chi \left\{ k^{-1/2} \sigma_n^{-1} T_{1n}(\hat{F}_n) > x_{\alpha_n} (1 + o(1)) \right\},$$

где x_{α_n} определяется вероятностью ошибки первого рода $\alpha_n = \alpha(K_{1n})$.

Теорема 2.2. Пусть выполнено условие С и имеет место (2.1). Тогда последовательность критериев K_{1n} асимптотически минимаксна. Если $0 < c_1 < \alpha_n < c_2 < 1$, то

$$\beta(K_{1n}, b_n) = \Phi(x_{\alpha_n} - b_n k_n^{-1/2}) (1 + o(1)) \quad (2.9)$$

при $n \rightarrow \infty$ и x_{α_n} может быть задано уравнением $\alpha_n = 1 - \Phi(x_{\alpha_n})$.

З а м е ч а н и е. Пусть выполнено условие А и имеет место (2.1). Тогда множество альтернатив определяется соотношением

$$T_n(F) = n \int_0^1 (\bar{f}_n(x) - 1)^2 dx \geq b_n \asymp k_n^{1/2}. \quad (2.10)$$

Таким образом, условие $k_n = o(n^2)$ позволяет охватить все возможные отклонения L_2 -нормы.

В практических приложениях при проверке гипотезы согласия заданное множества альтернатив не известно. В то же время статистик обычно имеет грубую априорную информацию о потерях $\rho(F)$, если верна альтернатива F и принимается гипотеза. Часто эти потери могут быть выражены как функция от окончательного решения $d(F)$, которое принимается на основе статистического анализа выборки. Таким образом, выбор тестовой статистики $T(\hat{F}_n)$ (обычно $T(F)$ является полунормой на множестве всех функций распределения $\mathfrak{F}$) непосредственно связан с непрерывностью функционалов $\rho(F)$ или $d(F)$ относительно полунормы $T(F)$ (отсутствие непрерывности влекло бы несоизмеримость значений функции потерь при принятии или отклонении гипотезы с реальными потерями). Наиболее часто при проверке гипотезы согласия выбираются критерии Колмогорова и омега-квадрат, поскольку функционалы $\rho(F)$ и $d(F)$ обычно являются непрерывными в соответствующих нормах $T(F) = \max\{|F(x) - x| : x \in (0, 1)\}$ и $T(F) = (\int_0^1 (F(x) - x)^2 dx)^{1/2}$ и эти нормы являются достаточно сильными, чтобы в них были непрерывны функционалы $d(F)$ и $\rho(F)$. Хорошо согласуется с соображениями непрерывности $\rho(F)$ и $d(F)$ и выбор критерия хи-квадрат, что объясняет его

частое применение на практике. В то же время при большом числе зон критерий хи-квадрат дает существенно более жесткую оценку близости функций распределения, чем статистики Колмогорова и омега-квадрат.

Критерии, связанные с проверкой гипотез о плотности распределения типа критериев Неймана (см. [14]–[16], [4], [7], [9], [12]), связаны с существенно более жесткой мерой близости гипотез и альтернатив, что и объясняет их существенно меньшую применимость. В [12] показано, что эффективность критериев Неймана совпадает с эффективностью критериев отношения правдоподобия (в смысле эффективности по Калленбергу). Однако альтернативы, рассматриваемые в [12], значительно более удалены от гипотезы по сравнению с порядком различимости b_n (см. [13]–[16], [4], [7], [9]).

Результаты об асимптотической минимаксности статистик $T_n(\hat{F}_n)$ (см. [5], [6], [8], [10]) допускают следующее обобщение в рамках рассматриваемой модели. Пусть множество альтернатив не известно точно статистику и он имеет только представление о потерях $\rho_n(F)$ от принятия гипотезы F_0 , когда справедлива альтернатива F , и пусть $\rho_n(F)$ допускает представление $\rho_n(F) = \rho_n(k_n^{-1/2}T_n(F)) > 0$. Пусть потери статистика от принятия альтернативы, когда верна гипотеза, равны κ_n , $0 < c_1 < \kappa_n < c_2 < 1$. Для любого критерия L_n риск равен, соответственно, $R_0(K_n) = \kappa_n \alpha(L_n)$ при справедливости гипотезы и $R_n(L_n, F) = \rho_n(k_n^{-1/2}T_n(F))\beta(L_n, F)$ при справедливости альтернативы F . Обозначим $R_n(K_n) = \sup\{R_n(K_n, F): F \in \mathfrak{F}, F \neq F_0\}$.

Скажем, что последовательность тестовых статистик $T_n(\hat{F}_n)$ сильно асимптотически минимаксна, если для построенных по ней критериев K_n таких, что $0 < c_1 R_0(K_n) \kappa_n^{-1} < c_2 < 1$, имеет место неравенство:

$$\limsup_{n \rightarrow \infty} \frac{R_n(K_n) - R_n(L_n)}{R_n(K_n)} < 0$$

для любой последовательности критериев L_n такой, что $R_0(L_n) \leq R_0(K_n)$.

Пусть $0 < \sup\{\rho_n(t), t > 0\} = \sup\{\rho_n(t), 0 < c_1 < t < c_2 < \infty\} < \infty$ для некоторых $0 < c_1 < c_2$. Тогда статистика $T_n(\hat{F}_n)$ (соответственно, $T_{1n}(\hat{F}_n)$) сильно асимптотически минимаксна, если выполнено условие А или В (соответственно, выполнено условие С). Данное утверждение непосредственно вытекает из того, что $R_n(K_n) \geq \beta(K_n, b_n) \rho_n(b_n k_n^{-1/2})(1 + o(1))$ для любой последовательности b_n , удовлетворяющей (2.1). Аналогичные результаты справедливы и для постановок задач [6], [8], [10].

Утверждения, касающиеся асимптотической нормальности статистик $T_n(X^{(n)})$ и $T_{1n}(X^{(n)})$, будут даны в терминах параметров $\theta_n = \theta_n(F) = \{\theta_{jn}\}_1^k$, $\theta_{jn} = (r_{jn} - p_{jn})/p_{jn}$. Ясно, что распределения статистик $T_n(X^{(n)})$, $T_{1n}(X^{(n)})$ однозначно определяются значениями этих параметров.

Обозначим

$$M_n(\theta) = (n-1) \sum_{j=1}^k g_j \theta_{jn}^2 p_{jn}.$$

Теорема 2.3. Пусть выполнено условие А или В. Тогда

$$E_{\theta}[T_n(\hat{F}_n)] - d_n = \sum_{j=1}^k g_j \theta_{jn} + M_n(\theta_n) (1 + o(1)) + o(1), \quad (2.11)$$

$$\text{Var}_{\theta}[T_n(\hat{F}_n)] = k\sigma_n^2 + o(M_n^2(\theta_n) + k). \quad (2.12)$$

Пусть выполнено условие С. Тогда

$$E_{\theta}[T_{1n}(\hat{F}_n)] = M_n(\theta_n) (1 + o(1)), \quad (2.13)$$

$$\text{Var}_{\theta}[T_{1n}(\hat{F}_n)] = k\sigma_n^2 + o(M_n^2(\theta_n) + k). \quad (2.14)$$

З а м е ч а н и е. Из (2.11) следует, что статистика хи-квадрат $T_n(\hat{F}_n)$ существенно лучше различает альтернативы в направлении $(g_1, g_2, \dots, g_k)$. Это явление было замечено Боровковым [1] и детально изучено Ивченко и Медведевым [19]. Можно сказать, что в направлении $(g_1, g_2, \dots, g_k)$ имеет место эффект суперэффективности. Этот эффект обусловлен вхождением в статистику $T_n(\hat{F}_n)$ слагаемого $W_n(\hat{F}_n)$, так как $E_{\theta_n}[W_n(\hat{F}_n)] = d_n + \sum_{j=1}^k g_j \theta_{jn} + o(M_n(\theta_n))$. Заметим, что (см. также [1], [19]) вклад $\text{Var}_{\theta}[W_n(\hat{F}_n)]$ в $\text{Var}_{\theta}[T_n(\hat{F}_n)]$ существенно больше, чем вклад дисперсий в других направлениях. При доказательстве мы показываем асимптотическую минимаксность тестовой статистики $T_{1n}(\hat{F}_n) = T_n(\hat{F}_n) - W_n(\hat{F}_n)$ и присоединяем $W_n(\hat{F}_n)$, чтобы получить статистику хи-квадрат. Таким образом, можно сказать, что в случае, когда не имеет места равенство $g_1/p_1 = g_2/p_2 = \dots = g_k/p_k$, статистика хи-квадрат может быть представлена как сумма двух слагаемых. Первое $T_{1n}(\hat{F}_n)$ асимптотически минимаксно для альтернатив

$$n \int_0^1 |\bar{f}_n(x) - 1|^2 r(x) dx > b_n$$

с некоторой весовой функцией $r(x)$. Второе асимптотически эффективно в направлении $(g_1, g_2, \dots, g_k)$. Таким образом, в приложениях представляется более естественным изучать поведение тестовых статистик $T_{1n}(\hat{F}_n)$ и $W_n(\hat{F}_n)$ по отдельности. Отметим, что в задаче проверки гипотез на основе статистик L, M и R мы также имеем эффективную различимость в одном направлении и другие скорости различимости в других.

По неравенству Чебышева из (2.11)–(2.14) следует, что если $k_n^{-1/2} M_n(\theta_n) \rightarrow \infty$, то $\beta(K_n, \theta_n) \rightarrow 0$, $\beta(K_{1n}, \theta_n) \rightarrow 0$ при $n \rightarrow \infty$. Следовательно, достаточно изучить асимптотическое поведение $\beta(K_n, F)$,

$\beta(K_{1n}, F)$ только для значений параметра

$$\theta_n(F) \in \Gamma_n(C) = \left\{ \theta(F): M_n(\theta(F)) < Ck_n^{1/2} \right\}.$$

Теорема 2.4. Пусть выполнено условие А или В. Тогда P_θ -распределения $k^{-1/2}\sigma_n^{-1}(T_n(\hat{F}_n) - E_\theta[T_n(\hat{F}_n)])$ сходятся к нормальному равномерно по $\theta = \theta_n(F) \in \Gamma_n(C)$.

Пусть выполнено условие С. Тогда P_θ -распределения $k_n^{-1/2}\sigma_n^{-1} \times T_{1n}(\hat{F}_n)$ сходятся к нормальному равномерно по $\theta = \theta_n(F) \in \Gamma_n(C)$.

З а м е ч а н и е. Если справедлива гипотеза (распределение F_0 фиксировано и равномерно), асимптотическая нормальность статистик хи-квадрат при выполнении условий А или В хорошо известна (см. [27], [24], [23] и др). Для альтернатив, сходящихся к F_0 , такой результат был доказан при существенно более сильных предположениях.

В задаче проверки гипотез с множеством альтернатив $F \in \mathfrak{S}(S_n, b_n)$ результаты доказаны при более сильных предположениях.

Д. $nk_n^{-1/2} \min_{1 \leq j \leq k} p_{jn} \rightarrow \infty$, $\max_{1 \leq j \leq k} p_{jn} \rightarrow 0$ при $n \rightarrow \infty$.
Условие Д может быть заменено условием В, если дополнительно предположить, что имеется априорная информация $F \in \mathfrak{R}(c_n) = \{F: F(x) \text{ имеет плотность } f(x) = dF(x)/dx \text{ и } \sup_{0 \leq x \leq 1} |f(x) - 1| < c_n\}$, где $c_n \rightarrow 0$ при $n \rightarrow \infty$.

Для любого критерия L_n обозначим

$$\beta_S(L_n, b_n) = \sup \{ \beta(L_n, F): F \in \mathfrak{S}(S_n, b_n) \}$$

и

$$\begin{aligned} \bar{\beta}(L_n, b_n) &= \sup \{ \beta(L_n, F): F \in \mathfrak{S}(T_n, b_n) \cap \mathfrak{R}(c_n) \}, \\ \bar{\beta}_S(L_n, b_n) &= \sup \{ \beta(L_n, F): F \in \mathfrak{S}(S_n, b_n) \cap \mathfrak{R}(c_n) \}. \end{aligned}$$

Определим последовательность критериев

$$K_{S_n} = \chi \{ S_n(\hat{F}_n) > k_n^{1/2} \bar{h} x_{\alpha_n} \},$$

где x_{α_n} задается вероятностью ошибки первого рода $\alpha_n = \alpha(K_{S_n})$ и $\bar{h} = \frac{1}{2} d^2 h / dx^2(0)$.

Теорема 2.5. Пусть выполнено условие Д и имеет место (2.1). Тогда для задачи проверки гипотез $S_n(F) = 0$ против альтернатив $F \in \mathfrak{S}(S_n, b_n)$ последовательность критериев хи-квадрат K_n и критериев K_{S_n} является асимптотически минимаксной. Если $0 < c_1 < \alpha_n = \alpha(K_n) = \alpha(K_{S_n})(1 + o(1)) < c_2 < 1$, то

$$\begin{aligned} \beta_S(K_n, b_n) &= \beta_S(K_{S_n}, b_n) (1 + o(1)) \\ &= \Phi(x_{\alpha_n} - b_n k_n^{-1/2}) (1 + o(1)), \end{aligned} \quad (2.15)$$

$$\begin{aligned} \beta(K_n, b_n) &= \beta(K_{S_n}, b_n) (1 + o(1)) \\ &= \Phi(x_{\alpha_n} - b_n k_n^{-1/2}) (1 + o(1)) \end{aligned} \quad (2.16)$$

при $n \rightarrow \infty$ и x_{α_n} задается уравнением $\alpha_n = 1 - \Phi(x_{\alpha_n})$.

Для всех $\varepsilon > 0$

$$P_F(k_n^{-1/2} |T_n(\hat{F}_n) - \bar{h}^{-1} S_n(\hat{F}_n)| > \varepsilon) = o(1) \quad (2.17)$$

равномерно по $F \in \Gamma_n(C)$.

Пусть выполнено условие В, имеет место (2.1), и пусть $c_n \rightarrow 0$ при $n \rightarrow \infty$. Тогда последовательность критериев K_n и K_{S_n} асимптотически минимаксна для задачи проверки гипотез $S_n(F) = 0$ против альтернатив $F \in \mathfrak{S}(S_n, b_n) \cap \mathfrak{R}(c_n)$. Если $0 < c_1 < \alpha_n = \alpha(K_n) = \alpha(K_{S_n})(1 + o(1)) < c_2 < 1$, то

$$\begin{aligned} \bar{\beta}_S(K_n, b_n) &= \bar{\beta}_S(K_{S_n}, b_n) (1 + o(1)) \\ &= \Phi(x_{\alpha_n} - b_n k_n^{-1/2}) (1 + o(1)), \end{aligned} \quad (2.18)$$

$$\begin{aligned} \bar{\beta}(K_n, b_n) &= \bar{\beta}(K_{S_n}, b_n) (1 + o(1)) \\ &= \Phi(x_{\alpha_n} - b_n k_n^{-1/2}) (1 + o(1)) \end{aligned} \quad (2.19)$$

при $n \rightarrow \infty$ и x_{α_n} задается тем же самым уравнением $\alpha_n = 1 - \Phi(x_{\alpha_n})$. В этих предположениях (2.17) имеет место равномерно по $F \in \Gamma_n(C) \cap \mathfrak{R}(c_n)$ для любых $C > 0$ при $n \rightarrow \infty$.

Для задачи проверки гипотез $S_{n\gamma}(F) = 0$ против альтернатив $F \in \mathfrak{S}(S_{n\gamma}, b_{n\gamma})$ результаты полностью аналогичны теоремам 2.1 и 2.2.

Обозначим

$$b_{n\gamma} = k^{1-\gamma} b_n^\gamma \sum_{j=1}^k g_j^{2\gamma} \left(\sum_{j=1}^k g_j^2 \right)^{-1}.$$

Для любого критерия L_n положим

$$\beta_\gamma(L_n, b_{n\gamma}) = \sup \{ \beta(L_n, F) : F \in \mathfrak{S}(S_{n\gamma}, b_{n\gamma}) \}.$$

Теорема 2.6. Пусть выполнено условие А или В и имеет место (2.1). Тогда в задаче проверки гипотез $S_{n\gamma}(F) = 0$ против альтернатив $F \in \mathfrak{S}(S_{n\gamma}, b_{n\gamma})$ последовательность критериев хи-квадрат K_n является асимптотически минимаксной. Если $0 < c_1 < \alpha_n < c_2 < 1$, то

$$\beta(K_n, b_n) = \Phi(x_{\alpha_n} - b_n k_n^{-1/2}) (1 + o(1)) \quad (2.20)$$

при $n \rightarrow \infty$ и x_{α_n} определяется из уравнения $\alpha_n = 1 - \Phi(x_{\alpha_n})$.

Если выполнено условие С и имеет место (2.1), то последовательность критериев K_{1n} асимптотически минимаксна и

$$\beta(K_{1n}, b_n) = \Phi(x_{\alpha_n} - b_n k_n^{-1/2}) (1 + o(1)) \quad (2.21)$$

для любых $\alpha_n = \alpha(K_n)$, $0 < c_1 < \alpha_n < c_2 < 1$ и для x_{α_n} , определяемых уравнением $\alpha_n = 1 - \Phi(x_{\alpha_n})$.

З а м е ч а н и е. Аналогичные результаты могут быть получены и для критериев Неймана. Развитый метод доказательства позволяет также усилить результаты в задаче проверки гипотез о плотности распределения (эти задачи будут рассмотрены в другой работе).

3. Доказательства теорем 2.1, 2.2 и 2.6. Нижние границы. Основные леммы. Доказательство нижних границ основано на определенной асимптотической версии леммы о наименее благоприятном распределении (см. [21]). Нами задается последовательность байесовских априорных распределений и показывается, что байесовское отношение правдоподобия может быть представлено в виде произведения двух сомножителей: первый является функцией статистики $T_{1n}(\hat{F}_n)$, а второй имеет порядок $1 + o(1)$. Затем мы доказываем, что асимптотика байесовских вероятностей ошибок второго рода критериев $K_{1n}(X^{(n)})$ равна $\beta(K_{1n}, b_n)(1 + o(1))$. Используя асимптотическую версию леммы о наименее благоприятном распределении и теоремы 2.3, 2.4, мы выводим отсюда теорему 2.2. В силу (2.11)–(2.14), если выполнено условие А или В, то разность статистик $T_n(X^{(n)})$ и $T_{1n}(X^{(n)})$ не играет существенной роли. Поэтому теорема 2.1 следует из теоремы 2.2. Отличия в доказательстве теорем 2.2 и 2.6 возникают только при доказательстве леммы 3.2. Аналогичный факт для задачи проверки гипотезы о плотности распределения был замечен Ингстером [16].

Обозначим U_n множество всех функций распределения G_τ , имеющих плотности

$$g_\tau(x) = 1 + \sum_{j=1}^k \tau_j \phi_j(x); \quad \tau = \{\tau_j\}_1^k, \tau = \tau(G_\tau) \in \mathbf{R}^k.$$

Определим множества

$$Q_n = \left\{ \tau: n \sum_{j=1}^k g_j p_j \left(\tau_j - \sum_{t=1}^k \tau_t p_t \right)^2 > b_n, \tau = \{\tau_j\}_1^k, G_\tau \in U_n \right\}.$$

Для любой функции распределения $F \in \mathfrak{F}(T_n, b_n)$ найдется функция распределения G_τ , $\tau \in Q_n$, такая, что $\theta(F) = \theta(G_\tau)$ и $T_n(F) = T_n(G_\tau) = n \sum_{j=1}^k g_j p_j (\tau_j - \sum_{t=1}^k \tau_t p_t)^2$. Таким образом, достаточно доказать теорему 2.2 для параметрического множества альтернатив Q_n вместо множества $\mathfrak{F}(T_n, b_n)$.

Лемма 3.1. Для любых j, t , $1 \leq j < t \leq k$,

$$\mathbf{E}_\tau [\phi_j(X_1)] = p_j \left(\tau_j - \sum_{i=1}^k \tau_i p_i \right), \quad (3.1)$$

$$\mathbf{E}_\tau [\phi_j^2(X_1)] = p_j \left(1 - p_j - (1 - 2p_j) \sum_{i=1}^k \tau_i p_i + \tau_j (1 - 2p_j) \right), \quad (3.2)$$

$$E_{\tau}[\phi_j(X_1)\phi_t(X_1)] = -p_j p_t \left(1 - 2 \sum_{i=1}^k \tau_i p_i + \tau_j + \tau_t\right). \quad (3.3)$$

Пусть $1 \leq j_1 < j_2 < \dots < j_l \leq k$ и $\tau_{j_i} = \tau_{j_i}^{(n)} = o(1)$ при $n \rightarrow \infty$ для всех i , $1 \leq i \leq l$. Тогда для любых натуральных чисел $t_1, \dots, t_l$

$$\left| E_{\tau}[\phi_{j_1}^{t_1}(X_1) \dots \phi_{j_l}^{t_l}(X_1)] \right| \leq C p_{j_1} \dots p_{j_l} (1 + o(1)). \quad (3.4)$$

Заметим, что $\sum_{i=1}^k \theta_i p_i = 0$. Положим

$$\rho_{n\delta} = k(1 + \delta) \left(\sum_{j=1}^k g_j^2 \right)^{-1}.$$

Обозначим $\kappa_j^2 = \kappa_{jn\delta}^2 = n^{-1} k^{-1} g_j \rho_{n\delta} b_n p_j^{-1}$ для всех j , $1 \leq j \leq k$. Определим случайный вектор $\eta^{(n)} = \eta_{\delta}^{(n)} = \{\eta_j\}_1^k$ с независимыми компонентами η_j , имеющими биномиальное распределение $P\{\eta_j = \kappa_j\} = P\{\eta_j = -\kappa_j\} = \frac{1}{2}$. Обозначим μ_n вероятностную меру случайного вектора $\eta^{(n)}$. Тогда байесовская априорная мера ν_n является условной мерой μ_n при условии $\eta^{(n)} \in Q_n$.

Заметим, что выбор биномиального распределения, как наименее благоприятного распределения, не играет существенной роли в доказательстве. Например, вместо η_j можно взять произвольные ограниченные или гауссовские случайные величины с теми же самыми первыми и вторыми моментами, что и у $\eta_j^{(n)}$, $1 \leq j \leq k$.

Определим множества $W_n = \{\theta: \theta_j = \pm \kappa_j, \theta = \{\theta_j\}_1^k, 1 \leq j \leq k\}$, $W_{1n} = \{\theta: \theta \in W_n, T_n(G_{\theta}) > b_n\}$ и $W_{2n} = W_n \setminus W_{1n}$.

Так как $\kappa_n = \max\{\kappa_{jn\delta}, 1 \leq j \leq k\} = o(1)$, то $g_w(x) \geq 0$ для всех $x \in (0, 1)$ и $w \in W_n$. Следовательно, g_w является плотностью для всех $w \in W_n$.

Лемма 3.2. Для любого $0 < \gamma \leq 1$

$$\lim_{n \rightarrow \infty} P_{\eta}\{S_{n\gamma}(G_{\eta}) \geq b_{n\gamma}\} = 1. \quad (3.5)$$

Доказательство леммы 3.2 и последующих лемм 3.3–3.5 будет дано в разделе 4. Лемма 3.7 будет доказана в разделе 5.

Для любого критерия L_n обозначим

$$\beta(L_n, \lambda_n) = \int_{W_n} \beta(L_n, \theta) d\lambda_n, \quad (3.6)$$

где $\lambda_n = \mu_n$ или $\lambda_n = \nu_n$.

По лемме 3.2, $\beta(L_n, \mu_n) = \beta(L_n, \nu_n)(1 + o(1))$ для любой последовательности критериев L_n такой, что $\beta(L_n, \mu_n) > c > 0$. Следовательно,

теорема 2.2 будет следовать из

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \beta(K_{1n}, \mu_n) = \lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \beta(K_{1n}, b_n) = \lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \beta(S_n, \mu_n). \quad (3.7)$$

Здесь и в дальнейшем $S_n = S_n(X^{(n)})$ обозначает байесовский критерий отношения правдоподобия, имеющий тот же самый уровень значимости, что и K_n ,

$$S_n(X^{(n)}) = \chi \left\{ E_\eta \left[\prod_{s=1}^n g_\eta(X_s) \right] > c_{\alpha_n} \right\}. \quad (3.8)$$

В оставшейся части раздела мы покажем, что байесовское отношение правдоподобия может быть представлено как функция $g_n(T_{1n}(X^{(n)}))(1 + o_P(1))$. Это будет означать равенство левой и правой частей (3.7).

Имеем

$$\prod_{s=1}^n \left(1 + \sum_{j=1}^k \eta_j \phi_j(X_s) \right) = 1 + \sum_{l=1}^n \frac{1}{l!} R_l. \quad (3.9)$$

Здесь R_l обозначает сумму всех слагаемых

$$\eta_{j_1} \cdots \eta_{j_l} \phi_{j_1}(X_{s_1}) \cdots \phi_{j_l}(X_{s_l}) \quad (3.10)$$

таких, что индексы $s_1, \dots, s_l$ различны, т.е. $R_l = R_{1l} - R_{2l}$, где

$$R_{1l} = \left(\sum_{s=1}^n \sum_{j=1}^k \eta_j \phi_j(X_s) \right)^l$$

и R_{2l} содержит все слагаемые R_{1l} , имеющие одним из сомножителей $\phi_{j_1}(X_{s_1}) \phi_{j_2}(X_{s_2})$ с $s_1 = s_2$. Множитель $l!$ возникает в (3.9), так как каждое слагаемое из (3.10) содержится в $R_l l!$ раз.

Ясно, что $E_\eta[R_l] = 0$ для нечетных l .

Если $l = 2$, то $E_\eta[R_2] = T_{1n}(X^{(n)})$.

Пусть $l = 2r$, $r > 1$. Положим

$$J_l = J_l(X^n) = (2r-1)!! \sum' \kappa_{j_1}^2 \cdots \kappa_{j_r}^2 \phi_{j_1}(X_{s_1}) \phi_{j_1}(X_{s_2}) \phi_{j_2}(X_{s_3}) \\ \times \phi_{j_2}(X_{s_4}) \cdots \phi_{j_r}(X_{s_{l-1}}) \phi_{j_r}(X_{s_l}). \quad (3.11)$$

Здесь $\sum'$ обозначает суммирование по всем различным индексам $1 \leq j_1, \dots, j_r \leq k$ и $1 \leq s_1, \dots, s_l \leq k$, т.е. $j_{i_1} \neq j_{i_2}$ и $s_{t_1} \neq s_{t_2}$, если $1 \leq i_1 < i_2 \leq r$ и $1 \leq t_1 < t_2 \leq l$, соответственно.

Лемма 3.3. Пусть $l = 2r$, $r \geq 1$. Тогда

$$E_\eta[R_l] = J_l + o_P(1). \quad (3.12)$$

Таким образом, остается только изучить структуру J_l . Обозначим

$$\Lambda_n(X^{(n)}) = 2 \sum_{j=1}^k \sum_{1 \leq s_1 < s_2 \leq n} \kappa_j^4 \phi_j^2(X_{s_1}) \phi_j^2(X_{s_2}),$$

$$A_n = E[\Lambda_n(X^{(n)})] = n^2 \sum_{j=1}^k \kappa_j^4 p_j^2 (1 + o(1)).$$

Для любого $A > 0$ обозначим

$$\begin{aligned} H_l(A, X^{(n)}) &= \frac{1}{2^r r!} T_{1n}^r(X^{(n)}) - \frac{1}{2(r+e)e!} A^e \\ &\quad - \frac{1}{2^{r+e-1}(e-1)!2!} J_4(X^{(n)}) A^{e-1} - \dots \\ &\quad - \frac{1}{2^{r+e-i}(e-i)!(2i)!(4i-1)!!} A^{e-i} J_{4i}(X^{(n)}) - \dots \\ &\quad - \frac{1}{2^{e+1}(r-2)!(l-5)!!} J_{l-4}(X^{(n)}) A \end{aligned}$$

для четных r и

$$\begin{aligned} H_l(A, X^{(n)}) &= \frac{1}{2^r r!} T_{1n}^r(X^{(n)}) - \frac{1}{2(r+e)e!} A^e J_2(X^{(n)}) \\ &\quad - \frac{1}{2^{r+e-1}(e-1)!2!5!!} A^{e-1} J_6(X^{(n)}) - \dots \\ &\quad - \frac{r!}{2^{r+1}(r-2)!(l-5)!!} J_{l-4}(X^{(n)}) A \end{aligned}$$

для нечетных r . Здесь $e = [r/2]$.

Если положить $A = \Lambda_n(X^{(n)})$, то $H_l(\Lambda_n(X^{(n)}), X^{(n)})$ получается из $(2^r r!)^{-1} T_{1n}^r(X^{(n)})$ удалением всех слагаемых, входящих в произведения $\Lambda_n^i(X^{(n)}) J_{l-4i}(X^{(n)})$, $1 \leq i \leq l$. Мы удалили такие слагаемые, так как $\text{Var}(\Lambda_n^i(X^{(n)}) J_{l-4i}) = O(1)$.

Лемма 3.4.

$$J_l(X^{(n)}) = l! H_l(A_n, X^{(n)}) + o_P(1) = l! H_l(\Lambda_n(X^{(n)}), X^{(n)}) + o_P(1). \quad (3.13)$$

Лемма 3.5.

$$E \left(\sum_{i=l}^{\infty} \frac{1}{2^i i!} J_i(X^{(n)}) \right)^2 = o(1) \quad (3.14)$$

при $l \rightarrow \infty$ равномерно по n .

Лемма 3.6.

$$\beta(K_{1n}, \mu_n) = \beta(S_n, \mu_n) (1 + o(1)). \quad (3.15)$$

Доказательство. Из (3.9) и лемм 3.3–3.5 непосредственно следует, что байесовский критерий отношения правдоподобия может быть представлен в виде $g_n(T_{1n}(X^{(n)})) + o_P(1)$, что влечет (3.15).

Лемма 3.7. *Как в случае справедливости гипотезы, так и байесовских альтернатив распределение статистик хи-квадрат $k^{-1/2}\sigma_n^{-1/2}T_{1n}(X^{(n)})$ асимптотически нормально. Имеют место соотношения*

$$E_{\nu_n}[T_{1n}(X^{(n)})] = n \sum_{j=1}^k g_j \kappa_j^2 p_j (1 + o(1)), \quad (3.16)$$

$$\text{Var}_{\nu_n}[T_{1n}(X^{(n)})] = k \sigma_n^2 (1 + o(1)) \quad (3.17)$$

при $n \rightarrow \infty$.

Теоремы 2.1, 2.2 и 2.6 следуют из лемм 3.6, 3.7 и теорем 2.3, 2.4.

4. Доказательство лемм 3.2–3.5.

Доказательство леммы 3.2. Так как

$$\text{Var} \left(\sum_{i=1}^k \eta_i p_i \right) < C n^{-1} k^{-1/2}, \quad (4.1)$$

то по неравенству Чебышева

$$P \left\{ \left| \sum_{i=1}^k \eta_i p_i \right| > C n^{-1/2} k^{-1/4} \right\} < C C n^{-2}. \quad (4.2)$$

Отсюда, поскольку $n^{1/2} k^{1/4} \min_{1 \leq i \leq k} |\eta_i| \rightarrow \infty$ при $n \rightarrow \infty$, получаем

$$S_{n\gamma}(G_\eta) = b_{n\gamma} (1 + o_P(1)) \quad \text{при } n \rightarrow \infty. \quad (4.3)$$

Доказательство леммы 3.3. Обозначим $J_{0l} = E_\eta[R_l] - J_l$. Имеем

$$J_{0l} = \sum_{u=1}^r \sum^* V(t_1, \dots, t_u), \quad (4.4)$$

где

$$\begin{aligned} V(t_1, \dots, t_u) &= \frac{l!}{t_1! \dots t_u! m_1! \dots m_v!} \\ &\times \sum' E_\eta[\eta_{j_1}^{t_1} \dots \eta_{j_u}^{t_u}] \phi_{j_1}(X_{s_1}) \dots \phi_{j_1}(X_{s_{t_1}}) \\ &\times \phi_{j_2}(X_{s_{t_1+1}}) \dots \phi_{j_2}(X_{s_{t_1+t_2}}) \\ &\times \phi_{j_3}(X_{s_{t_1+t_2+1}}) \dots \phi_{j_u}(X_{s_l}). \end{aligned} \quad (4.5)$$

Здесь суммирование $\sum^*$ проводится по всем целым $t_1, \dots, t_u$ таким, что $t_1 + \dots + t_u = l$ и по крайней мере одно значение $t_i > 2$ для $1 \leq i \leq u$. Суммирование $\sum'$ проводится по всем различным индексам $1 \leq j_1, \dots, j_u \leq k$

и $1 \leq s_1, \dots, s_l \leq k$. Значения $m_1, \dots, m_v$ определяются соотношениями $m_1 = \#\{t_i: t_i = d_1\}$, где $d_1 = \min\{t_s, 1 \leq s \leq k\}$, $m_2 = \#\{t_i: t_i = d_2\}$, где $d_2 = \min\{t_s, 1 \leq s \leq u, t_s > d_1\}$, ..., $m_v = \#\{t_i: t_i = d_v\}$, где $d_v = \min\{t_s, 1 \leq s \leq u, t_s > d_{v-1}\}$.

Ясно, что $E[V(t_1, \dots, t_u)] = 0$ и

$$\begin{aligned} \sigma^2(t_1, \dots, t_u) &= \text{Var}[V(t_1, \dots, t_u)] = \frac{(l!)^2}{(t_1! \dots t_u! m_1! \dots m_v!)^2} \\ &\times \sum' \kappa_{j_1}^{2t_1} \dots \kappa_{j_u}^{2t_u} E[\phi_{j_1}^2(X_{s_1}) \dots \phi_{j_1}^2(X_{s_{i_1}}) \\ &\quad \times \phi_{j_2}^2(X_{s_{i_1+1}}) \dots \phi_{j_u}^2(X_{s_l})] \\ &< \frac{C n^l (l!)^2}{(t_1! \dots t_u! m_1! \dots m_v!)^2} \sum_{j_1, \dots, j_u=1}^k \kappa_{j_1}^{2t_1} \dots \kappa_{j_u}^{2t_u} p_{j_1}^{t_1} \dots p_{j_u}^{t_u} \\ &\leq C k^{-(r-u)} (l!)^2 (t_1! \dots t_u! m_1! \dots m_v!)^{-2} \end{aligned} \quad (4.6)$$

для четных значений $t_1, \dots, t_u$.

Заметим, что

$$\omega(l, u) = \sum_{t_1 + \dots + t_u = l} \frac{l!}{t_1! \dots t_u! m_1! \dots m_v!}$$

являются числами Стирлинга второго рода со следующей воспроизводящей функцией:

$$\frac{1}{u!} (e^z - 1)^u = \sum_{t=u}^{\infty} \omega(t, u) \frac{z^t}{t!}.$$

Полагая $z = 1$, получаем

$$\omega(l, u) < \frac{l! \exp\{u\}}{u!}. \quad (4.7)$$

Из (4.6), (4.7), применяя формулу Стирлинга, получаем

$$\begin{aligned} (l!)^{-2} \text{Var}[J_{0l}] &\leq (l!)^{-2} \left(\sum_{u=1}^{r-1} \sum_{t_1 + \dots + t_u = l} \sigma(t_1, \dots, t_u) \right)^2 \\ &< (l!)^{-2} \left(\sum_{u=1}^{r-1} \frac{l! \exp\{u\}}{u!} k^{-(r-u)/2} \right)^2 = o(1). \end{aligned} \quad (4.8)$$

Доказательство леммы 3.4. Начнем со следующей вспомогательной леммы.

Лемма 4.1. *Имеет место равенство*

$$J_{1l}(X^{(n)}) = l! H_l(\Lambda_n(X^{(n)}), X^{(n)}) - J_l(X^{(n)}) = o_P(1) \quad \text{при } n \rightarrow \infty. \quad (4.9)$$

Доказательство. Левая часть (4.9) является суммой следующих слагаемых:

$$B_n = \sum' \kappa_{j_1}^{2t_1} \cdots \kappa_{j_u}^{2t_u} \phi_{j_1}(X_{s_1}) \cdots \phi_{j_u}(X_{s_u}), \quad (4.10)$$

где суммирование $\sum'$ проводится по всем натуральным числам $t_1, \dots, t_u$ таким, что $t_1 + \dots + t_u = r$, произведение $\phi_{j_1}(X_{s_1}) \cdots \phi_{j_u}(X_{s_u})$ имеет $2r$ сомножителей, $2u < v < l$ и B_n не содержит слагаемых $CL_n^{e-i}(X^{(n)})J_{4i}$, $1 \leq i \leq e$.

Для того чтобы доказать, что $E[B_n] = o(1)$, запишем произведение (4.10) следующим образом:

$$\kappa_{j_1} \phi_{j_1}(X_{s_1}) \kappa_{j_1} \phi_{j_1}(X_{s_2}) \cdots \kappa_{j_u} \phi_{j_u}(X_{s_{v-1}}) \kappa_{j_u} \phi_{j_u}(X_{s_v}). \quad (4.11)$$

Предположим, что (4.10) содержит элементами $\kappa_{j_\alpha}^d \phi_{j_\alpha}^d(X_{s_t})$, где $d \geq 2$, и индекс s_t не появляется в других сомножителях. Тогда

$$E \left[\sum_{s_t=1}^n \kappa_{j_\alpha}^d \phi_{j_\alpha}^d(X_{s_t}) \right] < Ck^{-1/2}. \quad (4.12)$$

Пусть $t_1 = 1$. Выделяя в (4.10) множители $\kappa_{j_1} \phi_{j_1}(X_{s_1})$, $\kappa_{j_1} \phi_{j_1}(X_{s_2})$ и все множители со случайными величинами X_{s_1}, X_{s_2} , получаем следующую сумму:

$$Y_n = 2 \sum_{j_1}^k E \left[\sum_{1 \leq s_1 < s_2 \leq n} \kappa_{j_1} \phi_{j_1}(X_{s_1}) \kappa_{j_{i_1}}^{a_1} \phi_{j_{i_1}}^{a_1}(X_{s_1}) \cdots \kappa_{j_{i_t}}^{a_t} \phi_{j_{i_t}}^{a_t}(X_{s_1}) \kappa_{j_1} \phi_{j_1}^{b_1}(X_{s_2}) \times \kappa_{j_{r_1}}^{b_1} \phi_{j_{r_1}}^{b_1}(X_{s_2}) \cdots \kappa_{j_{r_u}}^{b_u} \phi_{j_{r_u}}^{b_u}(X_{s_2}) \right], \quad (4.13)$$

которая не зависит от суммирования по другим индексам в (4.10).

Непосредственными вычислениями получаем

$$Y_n = O(nk^{-1/2} \gamma_{j_{i_1}}(a_1) \cdots \gamma_{j_{i_t}}(a_t) \gamma_{j_{r_1}}(b_1) \cdots \gamma_{j_{r_u}}(b_u)), \quad (4.14)$$

где $\gamma_j(a) = n^{-a/2} p_j^{-a/2+1} k^{-a/4} = O(n^{-1/2} k^{-1/4} p_j^{1/2})$ для всех натуральных a .

Пусть $t_1 > 1$, и пусть $\kappa_{j_1}^{2t_1} \phi_{j_1}^{d_1}(X_{s_1}) \cdots \phi_{j_1}^{d_w}(X_{s_w})$ — все множители с индексом j_1 . Выделяя в B_n все слагаемые $\kappa_{j_{i_1}}^{a_1} \phi_{j_{i_1}}^{a_1}(X_{s_1}) \cdots \kappa_{j_{i_t}}^{a_t} \phi_{j_{i_t}}^{a_t}(X_{s_w})$ со случайными величинами X_{s_m} , $1 \leq m \leq w$, получаем

$$\sum_{s_1, \dots, s_w=1}^n E \left[\sum_{j_1}^k \kappa_{j_1}^{2t_1} \phi_{j_1}^{d_1}(X_{s_1}) \cdots \phi_{j_1}^{d_w}(X_{s_w}) \kappa_{j_{i_1}}^{a_1} \phi_{j_{i_1}}^{a_1}(X_{s_1}) \cdots \kappa_{j_{i_t}}^{a_t} \phi_{j_{i_t}}^{a_t}(X_{s_w}) \right]$$

$$\begin{aligned}
&< Cn^w \mathbf{E} \left[\sum_{j_1=1}^k \kappa_{j_1}^{2t_1} \phi_{j_1}^{d_1}(X_{s_1}) \cdots \phi_{j_1}^{d_w}(X_{s_w}) \right] \gamma_{j_{i_1}}(a_1) \cdots \gamma_{j_{i_t}}(a_t) \\
&< D(t_1, w) \gamma_{j_{i_1}}(a_1) \cdots \gamma_{j_{i_t}}(a_t),
\end{aligned} \quad (4.15)$$

где $D(t_1, w) = O(1)$, если $t_1 = w = d_1 = d_2 = 2$;

$$D(t_1, w) = \sum_{j_1=1}^k n^{w-t_1} k^{-t_1/2} p_{j_1}^{w-t_1} = o(k^{-w/2+1}), \quad (4.16)$$

если $w \leq t_1$ и равенство $t_1 = w = d_1 = d_2 = 2$ не имеет места;

$$D(t_1, w) = o(n^{w-t_1} k^{-t_1/2}), \quad (4.17)$$

если $w > t_1$. Заметим, что в последнем случае $2(w - t_1)$ не превосходит числа случайных величин X_{s_i} , входящих в $\phi_{j_1}^{d_1}(X_{s_1}) \cdots \phi_{j_1}^{d_w}(X_{s_w})$ с $d_i = 1$. Каждое X_{s_i} входит в произведение (4.15) по крайней мере дважды. Отсюда, используя задание $\gamma_j(a)$, получаем, что множители n в (4.14) и n^{w-t_1} в (4.17) сокращаются за счет других сомножителей.

Повторяя те же самые оценки для $t_2, \dots, t_u$, получаем, что из оценок типа (4.14)–(4.17) следует $\mathbf{E}(J_l - l!H_l) = o(1)$.

Оценки $\mathbf{E}[B_n^2]$ аналогичны. Заметим, что слагаемые B_n^2 не содержат слагаемых $\Lambda_n^{e-i}(X^{(n)})J_{4i}$, $1 \leq i \leq u$, и допускают те же задания, что и слагаемые, входящие в B_n . Отсюда следует, что $\mathbf{E}[B_n^2] = o(1)$.

Чтобы доказать второе равенство в (3.13), мы покажем, что

$$H_l(\Lambda_n(X^{(n)}), X^{(n)}) - H_l(A_n, X^{(n)}) = o_P(1). \quad (4.18)$$

Отметим, что (4.18) будет следовать из

$$\begin{aligned}
&\mathbf{E}[\Lambda_n^m(X^{(n)})] - \left(\mathbf{E}[\Lambda_n(X^{(n)})] \right)^m \\
&= \mathbf{E}[\Lambda_n^m(X^{(n)})] - A_n^m + o(1) = o(1)
\end{aligned} \quad (4.19)$$

для четных m , так как $\text{Var}[H_l(\Lambda_n(X^{(n)}), X^{(n)}) - H_l(A_n, X^{(n)})]$ является суммой разностей таких слагаемых и слагаемых порядка $o(1)$. Оценки последних аналогичны оценкам доказательства леммы 4.1 и опускаются.

Левая часть (4.19) может быть записана как сумма следующих слагаемых:

$$\begin{aligned}
Z_m = u! \sum_{1 \leq j_1 < \dots < j_u \leq k} \sum_{s_1, \dots, s_u=1}^n \kappa_{j_1}^{4t_1} \cdots \kappa_{j_u}^{4t_u} \\
\times \mathbf{E}[\phi_{j_1}^2(X_{s_1}) \cdots \phi_{j_u}^2(X_{s_u})].
\end{aligned} \quad (4.20)$$

Здесь $t_1 + \dots + t_u = m$, каждая функция $\phi_{j_i}^2$, $1 \leq i \leq u$, входит в произведение точно $2t_i$ раз и $2u < v < 2r$.

Непосредственными вычислениями получаем

$$Z_m < C n^{-2m+v} k^{-m+u} \left(\min_{1 \leq j \leq k} p_j \right)^{-2m+v} = o(1), \quad (4.21)$$

откуда следует (4.19).

Доказательство леммы 3.5. Имеем

$$\begin{aligned} E[J_l]^2 &= \frac{1}{2^{2l}(l!)^2} \sum_{u=0}^l \sum_{v=0}^u \sum'_{j_1, \dots, j_{2l-u}} \sum'_{s_1, \dots, s_{2l-2v}} \kappa_{j_1}^4 \cdots \kappa_{j_u}^4 \kappa_{j_{u+1}}^2 \cdots \kappa_{j_{2l-u}}^2 \\ &\times E[\phi_{j_1}^2(X_{s_1}) \phi_{j_1}^2(X_{s_2}) \phi_{j_2}^2(X_{s_3}) \phi_{j_2}^2(X_{s_4}) \cdots \phi_{j_v}^2(X_{s_{2v-1}}) \phi_{j_v}^2(X_{s_{2v}}) \\ &\times \phi_{j_{v+1}}(X_{s_{2v+1}}) \phi_{j_{v+1}}(X_{s_{2v+2}}) \cdots \\ &\times \phi_{j_{2l-u}}(X_{s_{2l-2v-1}}) \phi_{j_{2l-u}}(X_{s_{2l-2v}})] . \end{aligned} \quad (4.22)$$

Здесь индексы $j_1, \dots, j_{2l-u}, s_1, \dots, s_{2l-2v}$ принимают различные значения $1 \leq j_1, \dots, j_{2l-u} \leq k$, $1 \leq s_1, \dots, s_{2l-2v} \leq n$, т.е. $j_{i_1} \neq j_{i_2}$, $s_{t_1} \neq s_{t_2}$ для $i_1 \neq i_2$, $1 \leq i_1, i_2 \leq 2l-u$, и $t_1 \neq t_2$, $1 \leq t_1, t_2 \leq 2l-2v$, соответственно. Каждая случайная величина X_{s_i} , $2v < i \leq 2l-2v$, входит в произведение $\phi_{j_{v+1}}(X_{s_{2v+1}}) \cdots \phi_{j_{2l-u}}(X_{s_{2l-2v}})$ дважды, и каждая функция ϕ_{j_i} , $v < i \leq 2l-u$, также присутствует в произведении дважды.

Непосредственными вычислениями получаем

$$E[J_l^2] \leq \frac{(2l)!}{2^{2l}(l!)^2 (2l-2u)! u! 2^u} \frac{u!}{(u-v)! v!} (2l-2v)! k^{v-l}. \quad (4.23)$$

Обозначим $x = (l-u)/l$, $y = v/l$. Применяя формулу Стирлинга, имеем

$$\begin{aligned} \log E[J_l^2] &\leq -(2l+u) \log 2 - 2l \log l + 2l \log(2l) - (2l-2u) \log(2l-2u) \\ &+ (2v-u) - (u-v) \log(u-v) - v \log v \\ &+ (2l-2v) \log(2l-2v) - (l-v) \log k. \end{aligned} \quad (4.24)$$

Супремум в правой части достигается при $8x^2 = (1-x-y)/l^2$, $y = \frac{1}{2}(1 - (1 - k/l^2)^{1/2})$ для $k/l^2 < 1$ и $y = 0$, $y = 1-x$ для $k/l^2 > 1$.

Пусть $x = 0$, $y = 1$. Тогда

$$E[J_l^2] < \frac{(2l)!}{2^{3l}(l!)^3} = o(1). \quad (4.25)$$

Пусть $y = 0$. Тогда $x = (8l)^{-1/2}$ и

$$E[J_l^2] \leq C \left(\frac{2l}{ek} \right)^l = o(1). \quad (4.26)$$

Пусть $k/l^2 < 1$. Тогда

$$\begin{aligned} \mathbf{E}[J_l^2] &\leq \exp \{l(1-x-2y)\} \left(\frac{ly}{k}\right)^l \\ &\leq C \left\{ \frac{1}{2} \exp \left\{ \left(1 - \frac{k}{l^2}\right)^{1/2} \right\} \left(1 - \left(1 - \frac{k}{l^2}\right)^{1/2}\right)^l \right\} = o(1), \end{aligned} \quad (4.27)$$

так как $e^z(1-z) < 2$ для всех $0 \leq z \leq 1$.

Из (4.22)–(4.27) следует (3.14).

5. Доказательство теорем 2.3 и 2.4.

Доказательство теоремы 2.3. Обозначим

$$\bar{\phi}_j(x) = \phi_j(x) - \mathbf{E}_\theta[\phi_j(X_1)] = \phi_j(x) - \theta_j p_j.$$

Тогда

$$T_n(X^{(n)}) = I_{1n} + I_{2n} + I_{3n} + W_n, \quad (5.1)$$

где

$$I_{1n} = 2 \sum_{1 \leq i_1 < i_2 \leq n} U_n(X_{i_1}, X_{i_2}), \quad (5.2)$$

$$U_n(x, y) = \sum_{j=1}^k g_j \frac{\bar{\phi}_j(x) \bar{\phi}_j(y)}{np_j}, \quad (5.3)$$

$$I_{2n} = (n-1) \sum_{j=1}^k g_j (np_j)^{-1} \sum_{s=1}^n \bar{\phi}_j(X_s) \mathbf{E}_\theta[\phi_j(X_s)], \quad (5.4)$$

$$I_{3n} = (n-1) \sum_{j=1}^k g_j p_j^{-1} \left(\mathbf{E}_\theta[\phi_j(X_1)] \right)^2 = (n-1) \sum_{j=1}^k g_j p_j \theta_j^2. \quad (5.5)$$

Непосредственными вычислениями получаем

$$\mathbf{E}_\theta[I_{1n}] = \mathbf{E}_\theta[I_{2n}] = 0, \quad (5.6)$$

$$\mathbf{E}_\theta[W_n] = \sum_{j=1}^k g_j (1 - p_j + \theta_j(1 - 2p_j)). \quad (5.7)$$

Применяя неравенство Шварца, имеем

$$\begin{aligned} \sum_{j=1}^k g_j p_j |\theta_j| &\leq \left(\sum_{j=1}^k g_j p_j \right)^{1/2} (n^{-1} M(\theta))^{1/2} \\ &< C n^{-1/2} M_n^{1/2}(\theta) = o(1 + M_n(\theta_n)). \end{aligned} \quad (5.8)$$

Из (5.1)–(5.8) следует (2.11), (2.13).

Для доказательства (2.14) запишем

$$\text{Var}_\theta[I_{1n}] = V_{1n} + V_{2n},$$

где

$$\begin{aligned} V_{1n} &= 2 \sum_{j=1}^k g_j^2 p_j^{-2} \text{Var}_\theta^2[\phi_j(X_1)] \\ &= 2 \sum_{j=1}^k g_j^2 (1 - p_j + \theta_j(1 - 2p_j) - p_j \theta_j^2)^2, \end{aligned} \quad (5.9)$$

$$\begin{aligned} V_{2n} &= 2 \sum_{1 \leq j_1 < j_2 \leq k} g_{j_1} g_{j_2} p_{j_1}^{-1} p_{j_2}^{-1} \text{Cov}_\theta^2[\phi_{j_1}(X_1), \phi_{j_2}(X_1)] \\ &= 4 \sum_{1 \leq j_1 < j_2 \leq k} g_{j_1} g_{j_2} p_{j_1} p_{j_2} (1 + \theta_{j_1})^2 (1 + \theta_{j_2})^2. \end{aligned} \quad (5.10)$$

Используя условие С и неравенство Шварца, имеем

$$\begin{aligned} \sum_{j=1}^k g_j^2 |\theta_j| &\leq \left(\sum_{j=1}^k g_j^3 \right)^{1/2} \left(\sum_{j=1}^k g_j \theta_j^2 p_j \right)^{1/2} \max_{1 \leq j \leq k} p_j^{-1/2} \\ &= o(k^{3/4} M_n^{1/2}(\theta)) = o(k + M_n^2(\theta)), \end{aligned} \quad (5.11)$$

$$\begin{aligned} \sum_{j=1}^k g_j^2 \theta_j^2 &< n^{-1} M(\theta) \max_{1 \leq j \leq k} p_j^{-1} = o(k^{1/2} M(\theta)) \\ &= o(k + M_n^2(\theta)). \end{aligned} \quad (5.12)$$

Из (5.8), (5.11), (5.12) получаем

$$\text{Var}_\theta[I_{1n}] = V_{1n}(1 + o(1)) = k\sigma_n^2 + o(k + M_n^2(\theta)). \quad (5.13)$$

Имеем

$$\text{Var}_\theta[I_{2n}] = I_{21n} + I_{22n}, \quad (5.14)$$

где

$$\begin{aligned} I_{21n} &= 2n^{-1}(n-1)^2 \sum_{1 \leq j_1 < j_2 \leq k} g_{j_1} g_{j_2} \text{Cov}_\theta(\phi_{j_1}(X_1), \phi_{j_2}(X_1)) \\ &\quad \times \mathbf{E}_\theta[\phi_{j_1}(X_1)] \mathbf{E}_\theta[\phi_{j_2}(X_1)] p_{j_1}^{-1} p_{j_2}^{-1} \\ &= 2n^{-1}(n-1)^2 \sum_{1 \leq j_1 < j_2 \leq k} g_{j_1} g_{j_2} p_{j_1} p_{j_2} (1 + \theta_{j_1})(1 + \theta_{j_2}) \theta_{j_1} \theta_{j_2}, \end{aligned} \quad (5.15)$$

$$I_{22n} = n^{-1}(n-1)^2 \sum_{j=1}^k g_j^2 \text{Var}_\theta[\phi_j(X_1)] \mathbf{E}_\theta^2[\phi_j(X_1)] p_j^{-2}$$

$$= n^{-1}(n-1)^2 \sum_{j=1}^k g_j^2 p_j (1-p_j + \theta_j(1-2p_j) - p_j \theta_j^2) \theta_j^2. \quad (5.16)$$

Мы оценим только часть слагаемых в I_{21n} , I_{22n} . Оценка остальных слагаемых аналогична.

Применяя неравенство Шварца, условие С и (5.8), имеем

$$\begin{aligned} n \sum_{j=1}^k g_j^2 p_j \theta_j^3 &\leq C M_n(\theta) \max \{|\theta_j|, 1 \leq j \leq k\} \\ &\leq C n^{-1/2} M_n^{3/2}(\theta) \max \{p_j^{-1/2}; 1 \leq j \leq k\} \\ &= o(k^{1/4} M_n^{3/2}(\theta)), \end{aligned} \quad (5.17)$$

$$n \sum_{1 \leq j_1 < j_2 \leq k} g_{j_1} g_{j_2} p_{j_1} p_{j_2} \theta_{j_1}^2 \theta_{j_2}^2 \leq n^{-1/2} M_n^{3/2}(\theta), \quad (5.18)$$

$$n \sum_{1 \leq j < j_2 \leq k} g_j g_{j_2} p_j p_{j_2} \theta_j^2 \theta_{j_2}^2 < n^{-1} M_n^2(\theta), \quad (5.19)$$

$$n \sum_{j=1}^k g_j^2 p_j^2 \theta_j^4 < n^{-1} M^2(\theta). \quad (5.20)$$

Из (5.15)–(5.20) получаем

$$\text{Var}_\theta[I_{2n}] = o(k + M_n^2(\theta)), \quad (5.21)$$

что завершает доказательство соотношения (2.14).

Предположим, что выполнено условие В. Тогда

$$\begin{aligned} \text{Var}_\theta[W_n] &= n^{-1} \sum_{j=1}^k g_j^2 p_j^{-2} \text{Var}_\theta[\phi_j^2(X_1)] \\ &< C n^{-1} \sum_{j=1}^k g_j^2 p_j^{-1} (1 + |\theta_j|) (1 + o(1)) = o(k), \end{aligned} \quad (5.22)$$

так как

$$\sum_{j=1}^k (n p_j)^{-3/2} (n p_j)^{1/2} |\theta_j| < C \left(\sum_{j=1}^k (n^3 p_j^3)^{-1} \right)^{1/2} M_n^{1/2}(\theta) = o(k). \quad (5.23)$$

Пусть выполнено условие А. Тогда

$$W_n = \frac{k}{n} \sum_{j=1}^k \sum_{s=1}^n \left(\chi(X_s \in (e_{j-1}, e_j)) - \frac{1}{k} \right)^2 = \left(1 - \frac{1}{k} \right) D_n + \frac{1}{k}, \quad (5.24)$$

где

$$D_n = \frac{k}{n} \sum_{j=1}^k \sum_{s=1}^n \chi(X_s \in (e_{j-1}, e_j)) = \frac{k^2}{n}. \quad (5.25)$$

Имеем

$$\text{Cov}_\theta [I_{1n}, W_n] \leq \text{Var}_\theta^{1/2} [I_{1n}] \text{Var}_\theta^{1/2} [W_n], \quad (5.26)$$

$$\text{Cov}_\theta [W_n, I_{2n}] \leq \text{Var}_\theta^{1/2} [W_n] \text{Var}_\theta^{1/2} [I_{2n}], \quad (5.27)$$

$$\text{Cov}_\theta [I_{1n}, I_{2n}] \leq \text{Var}_\theta^{1/2} [I_{1n}] \text{Var}_\theta^{1/2} [I_{2n}]. \quad (5.28)$$

Из (5.13), (5.21), (5.22), (5.24)–(5.27) следует (2.12), что завершает доказательство теоремы 2.3.

Доказательство теоремы 2.4. Как следует из (5.6), (5.13)–(5.20), достаточно доказать асимптотическую нормальность U -статистик I_{1n} . Доказательство основано на версии Ингстера [13] центральной предельной теоремы для вырожденных U -статистик (см. [28]). Согласно этой теореме, достаточно показать, что

$$\lim_{n \rightarrow \infty} \left(\frac{n^4 H_{2n}}{V_{1n}^2} + \frac{n^2 H_{4n}}{V_{1n}} \right) = 0, \quad (5.29)$$

где

$$H_{4n} = \mathbb{E}_\theta [U_n^4(X_1, X_2)] \quad \text{и} \quad H_{2n} = \mathbb{E}_\theta [q_n^2(X_1, X_2)],$$

где

$$q_n(x, y) = \mathbb{E}_\theta [U_n(x, X_3) U_n(X_3, y)].$$

Непосредственными вычислениями, используя (5.12), получаем

$$\begin{aligned} H_{2n} &< C \sum_{j=1}^k g_j^4 (1 + \theta_j + o(1)) + C \left(\sum_{j=1}^k g_j p_j \theta_j^2 \right)^4 \\ &< Ck + C \sum_{j=1}^k g_j^4 \theta_j^4 + C \left(\sum_{j=1}^k g_j p_j \theta_j^2 \right)^4 \\ &< Ck + C \left(\sum_{j=1}^k g_j^2 \theta_j^2 \right)^2 + C \frac{k^2}{n^4} = o(k^2). \end{aligned} \quad (5.30)$$

В силу условия С получаем

$$\begin{aligned} n^2 H_{4n} &\leq C n^{-2} \sum_{j=1}^k g_j^4 (1 + \theta_j (1 + o(1)) + o(1))^2 p_j^{-2} \\ &\quad + C n^{-2} \left(\sum_{j=1}^k g_j p_j (1 + \theta_j + o(1)) \right)^4 \end{aligned}$$

$$\begin{aligned}
&\leq Ck + Cn^{-3} M(\theta) \max_{1 \leq j \leq k} p_j^{-3} + o(n^{-4} M^2(\theta)) \\
&= o(k^{3/2} M(\theta) + n^{-4} k) = o(k^2).
\end{aligned} \tag{5.31}$$

Из (5.13), (5.30), (5.31) следует (5.29).

6. Доказательство теоремы 2.5. Начнем с доказательства верхней границы. Пусть $\theta \in \Gamma_n(C)$. Тогда, в силу условия D,

$$\max_{1 \leq j \leq k} |\theta_{jn}| = o(1) \quad \text{при } n \rightarrow \infty.$$

Зафиксируем $\varepsilon > 0$ и пусть $t \leq \varepsilon/2$. Тогда, применяя неравенство Чернова, получаем

$$\begin{aligned}
&P_\theta \left\{ \max_{1 \leq j \leq k} \left| \sum_{s=1}^n \bar{\phi}_j(X_s) \right| > \varepsilon n p_j \right\} < \sum_{j=1}^k P_\theta \left\{ \left| \sum_{s=1}^n \bar{\phi}_j(X_s) \right| > \varepsilon n p_j \right\} \\
&< \sum_{j=1}^k \exp \{ -\varepsilon t n p_j - t p_j (1 + \theta_j) \} \\
&\quad \times (p_j (1 + \theta_j) \exp\{t\} + 1 - p_j (1 + \theta_j))^n \\
&= \sum_{j=1}^k \exp \left\{ -\varepsilon t n p_j + \frac{n t^2}{2} p_j (1 + \theta_j) (1 + o(1)) \right\} \\
&< Ck \exp \{ -\varepsilon t k^{1/2} \} = o(1).
\end{aligned} \tag{6.1}$$

Отсюда, применяя формулу Тейлора, получаем

$$\begin{aligned}
S_n(\hat{F}_n) &= n \sum_{j=1}^k p_j h \left(\sum_{s=1}^n \bar{\phi}_j(X_s) (n p_j)^{-1} + \theta_j \right) \\
&= n h''(\xi_n) \sum_{j=1}^k p_j \left(\sum_{s=1}^n \bar{\phi}_j(X_s) (n p_j)^{-1} + \theta_j \right)^2 \\
&= h''(\xi_n) T_n(\hat{F}_n),
\end{aligned} \tag{6.2}$$

где $\xi_n \rightarrow_P 0$ при $n \rightarrow \infty$, что завершает доказательство границы сверху.

Нижняя граница следует из (6.1), (6.2) и

$$\max_{1 \leq j \leq k} \kappa_j^2 = o(1) \quad \text{при } n \rightarrow \infty.$$

Данные рассуждения аналогичны [9].

СПИСОК ЛИТЕРАТУРЫ

1. Боровков А. А. О мощности критерия хи-квадрат при увеличении числа групп. — Теория вероятн. и ее примен., 1977, т. XXII, в. 2, с. 375–378.
2. Боровков А. А. Математическая статистика. М.: Наука, 1984.
3. Wald A. Tests of statistical hypothesis concerning several parameters when the number of observation is large. — Trans. Amer. Math. Soc., 1943, v. 54, № 2, p. 426–482.
4. Ермаков М. С. Асимптотически минимаксные критерии проверки непараметрических гипотез о плотности распределения. — Зап. научн. семин. ЛОМИ, 1988, т. 166, с. 44–53.
5. Ermakov M. S. Asymptotic minimaxity of usual goodness-of-fit tests. — Probability Theory and Mathematical Statistics (Proceedings of the Fifth Vilnius Conference). Ed. by B. Grigelionis et al. Utrecht/Vilnius: VSP/Mokslas, 1990, p. 323–331.
6. Ermakov M. S. On asymptotic minimaxity of rank tests. — Statist. Probab. Lett., 1992, v. 15, p. 191–196.
7. Ермаков М. С. Минимаксная непараметрическая проверка гипотез о плотности распределения. — Теория вероятн. и ее примен., 1994, т. 39, № 3, с. 488–512.
8. Ермаков М. С. Асимптотическая минимаксность критериев типа Колмогорова и омега-квадрат. — Теория вероятн. и ее примен., 1995, т. 40, № 1, с. 54–67.
9. Ermakov M. S. Minimax nonparametric hypothesis testing for Hellinger metrics and Kullback–Leibler information. — Proceedings of the Kolmogorov semester at the Euler International Mathematical Institute (St. Petersburg). Ed. by I. A. Ibragimov and A. Yu. Zaitsev. New York: Gordon and Breach, 1996, p. 187–193.
10. Ermakov M. S. On asymptotic minimaxity of Kolmogorov and omega-square tests. — Statist. Probab. Lett., 1996, v. 30, № 3, p. 227–233.
11. Инглот Т., Юрлевич Т., Ледвина Т. Асимптотики для мультиномиальных критериев согласия при простой гипотезе. — Теория вероятн. и ее примен., 1990, т. 37, в. 3, с. 797–803.
12. Inglot T., Ledwina T. Asymptotic optimality of data-driven Neyman's tests for uniformity. — Ann. Statist., 1996, v. 24, № 5, p. 1982–2020.
13. Ингстер Ю. И. Минимаксная проверка непараметрических гипотез в метриках L_p . — Теория вероятн. и ее примен., 1987, т. XXXII, в. 2, с. 333–337.
14. Ингстер Ю. И. Минимаксная проверка простой гипотезы против слабо непараметрического множества альтернатив. — Теория вероятн. и ее примен., 1988, т. XXXIII, в. 3, с. 592–596.
15. Ingster Yu. I. Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. — Math. Meth. Statist., 1993, v. 2, № 2, p. 85–114; № 3, p. 171–189; № 4, p. 249–268.
16. Ингстер Ю. И. Минимаксная проверка гипотез о плотности распределения для эллипсоидов в L_p . — Теория вероятн. и ее примен., 1994, т. 39, в. 3, с. 530–554.
17. Ивченко Г. И., Медведев Я. Ю. Разделимые статистики и проверка гипотез для группированных данных. — Теория вероятн. и ее примен., 1980, т. XXV, в. 3, с. 549–560.
18. Ивченко Г. И., Чуканов С. Б. Новые статистические критерии в полиномиальной схеме и их асимптотические свойства. — Теория вероятн. и ее примен., 1988, т. XXXIII, в. 2, с. 201–205.
19. Kallenberg W. C. M., Oosterhoff J., Schriever B. F. The number of classes in chi-squared goodness-of-fit tests. — J. Amer. Statist. Assoc., 1985, v. 80, № 4, p. 959–968.
20. Cressie N., Read T. R. C. Multinomial goodness-of-fit tests. — J. Roy. Statist. Soc. B, 1984, v. 46, в. 2, p. 440–464.
21. Lehmann E. L. Testing Statistical Hypothesis. New York: Wiley, 1986.

22. *Mann H. B., Wald A.* On the choice of the number of intervals in the application of chi-squared test. — *Ann. Math. Statist.*, 1942, v. 13, № 2, p. 306–318.
23. *Медведев Я. Ю.* Разделимые статистики в полиномиальной схеме. I, II. — *Теория вероятн. и ее примен.*, 1977, т. XXII, в. 1, с. 3–17, в. 4, с. 623–631.
24. *Morris C.* Central limit theorems for multinomial sums. — *Ann. Statist.*, 1975, v. 3, № 1, p. 165–188.
25. *Steck G. P.* Limit theorems for conditional distributions. — *Univ. California Publ. Statist.*, 1957, v. 2, p. 237–284.
26. *Stein C.* Efficient nonparametric testing and estimation. — *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, v. 1. Univ. California Press, 1956, p. 187–195.
27. *Туманян С. Н.* Асимптотическое поведение критериев хи-квадрат при одновременном возрастании объема наблюдений и числа ячеек. — *Теория вероятн. и ее примен.*, 1956, т. I, в. 1, с. 131–145.
28. *Hall P.* Central limit theorem for integrated squared error of multivariate nonparametric density estimators. — *J. Multivariate Anal.*, 1984, v. 14, № 1, p. 1–16.
29. *Holst L.* Asymptotic normality and efficiency for certain goodness-of-fit tests. — *Biometrika*, 1972, v. 59, № 2, p. 137–145.
30. *Chibisov D. M.* Asymptotic optimality of the chi-squared tests with large number of degrees of freedom within the class of symmetric tests. — *Math. Meth. Statist.*, 1992, v. 1, № 1, p. 55–82.
31. *Chibisov D. M., Gvanceladze L. G.* On test of fit based on grouped data. — *Contribution to Statistics (Hajek Memorial Volume)*. Ed. by J. Jureckova. Prague: Academia, 1979, p. 79–90.

Поступила в редакцию
11.XI.1996