Ю. И. Ингстер

АДАПТИВНЫЕ ТЕСТЫ ХИ-КВАДРАТ

1. Введение

1.1 Минимаксная различимость

Пусть $X^{(n)} = X_1, ...X_n, \ X_i \in [0,1]$ – независимая однородная выборка из распределения с плотностью $f \in \Phi$, где Φ есть множество плотностей распределения относительно меры Лебега на $(B, [0,1])$. Будем рассматривать задачу проверки гипотезы $H_0 : f = f_0; \quad f_0(x) \equiv 1$ о равномерности плотности распределения выборки $X^{(n)}$ в асимптотическом минимаксном варианте при $n \to \infty$. Именно пусть $F_n \subset \Phi$ и альтернатива имеет вид $H_0 : f \in F_n$. Обозначим Ψ_n множество (рандомизированных) тестов $\psi_n = \psi_n(X^{(n)}) \in [0,1]$, т.е. измеримых функций выборки со значениями в $[0,1]$. Качество теста $\psi_n \in \Psi_n$ будем характеризовать величинами вероятностей ошибок первого рода $\alpha(\psi_n) = E_{f_0,n}(\psi_n)$ и максимумом вероятностей ошибок второго рода

$$\beta(\psi_n, F_n) = \sup_{f \in F_n} \beta(\psi_n, f); \ \beta(\psi_n, f) = E_{f,n}(1 - \psi_n).$$

Здесь и ниже $E_{f,n}(\cdot)$ есть среднее по мере P_f^n, соответствующей выборке $X^{(n)}$ с плотностью f.

Достижимое качество различения гипотезы и альтернативы характеризуется величинами минимаксных вероятностей ошибок:

$$\beta_n(\alpha, F_n) = \inf_{\psi \in \Psi_{n,\alpha}} \beta(\psi_n, F_n), \ \alpha \in (0,1),$$

где $\Psi_{n,\alpha} = \{\psi \in \Psi_n : \ \alpha(\psi) \le \alpha\}$. При этом имеют место неравенства:

$$0 \le \beta_n(\alpha, F_n) \le 1 - \alpha.$$

Работа частично поддержана грантом РФФИ No. 96-01-00684 и грантом ведущих научных школ No. 96-15-96199.

В этой работе нас будет интересовать грубая асимптотика минимаксных вероятностей ошибок, т.е. условия асимптотической различимости. Пусть $n \to \infty$. Будем говорить, что гипотеза и альтернатива *асимптотически неразличимы*, если $\beta_n(\alpha, F_n) \to 1 - \alpha$ для любого $\alpha \in (0,1)$. Будем говорить, что гипотеза и альтернатива *состоятельно различимы*, если $\beta_n(\alpha, F_n) \to 0$ для любого $\alpha \in (0,1)$; ясно, что при этом найдется *состоятельная* последовательность тестов ψ_n, такая что $\alpha(\psi_n) \to 0,\ \beta(\psi_n, F_n) \to 0$. Здесь и ниже, если не оговорено противное, пределы имеются в виду при $n \to \infty$.

Ясно, что для различимости множество F_n не должно содержать плотностей, слишком близких к f_0. Поэтому естественно рассматривать альтернативы $F_n = F_n(\tau, \rho_n)$, получающиеся удалением из некоторых функциональных множеств $F = F(\tau_1) \subset \Phi$ некоторых окрестностей $U_n = U(\tau_2, \rho_n)$ плотности f_0; здесь параметры τ_1 характеризуют множество F, а τ_2, ρ_n – “форму” и “размеры” окрестностей U_n. Представляет интерес изучение условий асимптотической различимости в зависимости от параметров $\tau = (\tau_1, \tau_2)$ и скорости убывания ρ_n.

Подобная постановка задачи при непараметрических альтернативах (множества F бесконечномерны) восходит, по-видимому, к работам Манна и Вальда, Стейна; более детально она изучалась в работах автора [1–3], Ермакова [4, 5]; для аналогичных задач в моделе гауссовского белого шума – также в недавних работах Лепского и Спокойного [6], Спокойного [7, 8], Лепского и Цыбакова [9]; см. обзор [2] и [8] для более детального обсуждения.

1.2. Гладкие альтернативы

Наиболее важные и естественные альтернативы соответствуют множествам F плотностей распределения некоторого уровня гладкости и окрестностям U_n типа шаров радиуса ρ_n в некоторой норме. Пусть

$$F(\tau_1) = \{f = 1 + g \in \Phi:\ g \in S_q^\sigma(R),\ \tau_1 = (\sigma, q, R)\} \qquad (1.1)$$

или

$$F(\tau_1) = \{f = 1 + g \in \Phi:\ g \in B_{q,h}^\sigma(R),\ \tau_1 = (\sigma, q, h, R)\}, \qquad (1.2)$$

где $S_q^\sigma(R)$ есть шары Соболева или Никольского, а $B_{q,h}^\sigma(R)$ есть шар Бесова радиуса $R > 0$; здесь параметр $\sigma > 0$ характеризует гладкость функций, а параметры $1 \le q, h \le \infty$ – норму, используемую для измерения гладкости (определения см. [10, 11], а также, например, [2, 6]). Пусть также

$$F(\tau, \rho_n) = F(\tau_1) \setminus U(p, \rho_n),$$
$$U(p, \rho_n) = \{f : \ \|f - 1\|_p < \rho_n\}, \ \tau_2 = p, \tag{1.3}$$

где $\|\cdot\|_p$ есть обычная L_p-норма.

Приведем здесь лишь один результат (см. [1, 2]) для шаров Соболева или Никольского; на шары Бесова методы [1, 2] легко переносятся. Пусть $1 \le p \le 2$; $p \le q \le \infty$. Тогда условия различимости можно выразить в терминах *критических радиусов*:

$$\rho_n^*(\sigma) = n^{-2\sigma/(4\sigma+1)}. \tag{1.4}$$

Именно, для любого $\alpha \in (0,1)$

$$\beta_n(\alpha, F_n(\tau, \rho_n)) \to 1 - \alpha, \quad \text{если } \rho_n/\rho_n^*(\sigma) \to 0;$$
$$\beta_n(\alpha, F_n(\tau, \rho_n)) \to 0, \quad \text{если } \rho_n/\rho_n^*(\sigma) \to \infty$$

и при выполнении второго условия состоятельной является последовательность $\psi_{n,k;T_n} = \mathbf{1}_{L_{n,k} > T_n}$ тестов хи-квадрат, соответствующая последовательности равномерных разбиений $\Delta_k = \{\delta_{1,k}, \dots \delta_{k,k}\}$, $\delta_{k,j} = ((j-1)/k, j/k]$ интервала $(0,1]$ на $k = k_n \asymp n^{2/(4\sigma+1)}$ интервалов. Здесь

$$L_{n,k} = (2(k-1))^{-1/2} \sum_{i=1}^{k} \left(\frac{(n_l - np_k)^2}{np_k(1-p_k)} - 1 \right),$$
$$p_k = 1/k; \ n_l = \#\{X_i \in \delta_{k,l}\} \tag{1.5}$$

есть последовательность центрированных и нормированных статистик хи-квадрат, соответствующих разбиениям Δ_k, а $T_n \to \infty$ – некоторая последовательность, скорость возрастания которой зависит от скорости возрастания $\rho_n/\rho_n^*(\sigma)$.

1.3. Адаптивная постановка задачи

Структура состоятельных последовательностей тестов хи-квадрат, описанных выше, существенно зависит от параметра гладкости $\sigma \in (0, \infty)$: скорость степенного роста $2/(4\sigma+1)$ числа

$k = k_n$ интервалов разбиений может меняться от 0 до 2; в частности, $k_n \gg n$ при $\sigma < 1/4$. В практических задачах статистик обычно может рассчитывать на некоторую гладкость альтернатив, однако параметр гладкости ему, как правило, неизвестен; более того, часто нужно одновременно рассматривать альтернативы разной гладкости. Можно, конечно, задаться некоторой минимальной гладкостью $\sigma_0 > 0$ и использовать соответствующие тесты. Однако для альтернатив большей гладкости $\sigma > \sigma_0$ критические радиусы будут существенно завышены: $\rho_n^*(\sigma) \gg \rho_n^*(\sigma_0)$, причем различие имеет место в показателях степени! Это не позволяет отличать от гипотезы существенную часть альтернатив, которые можно было бы отличить при использовании тестов, рассчитанных на большую гладкость.

Поэтому большой интерес представляет вопрос: можно ли построить такие последовательности тестов, которые бы обеспечивали бы состоятельную различимость при всевозможных показателях гладкости (или, по крайней мере, для достаточно широкого интервала гладкостей)? Разумеется, при этом допустимая близость альтернатив к гипотезе должна зависеть от гладкости альтернативы: $\|f-1\|_p \gg \rho_n^*(\sigma)$ при $f - 1 \in S_q^\sigma$ или $f - 1 \in B_{q,h}^\sigma$.

Впервые подобная задача рассматривалась Спокойным [7] для модели гауссовского белого шума интенсивности $\epsilon = n^{-1/2}$ (для которой также справедливы результаты п. 1.2) и была названа "адаптивной" (по-видимому, аналогии с аналогичными задачами оценивания функций неизвестной гладкости, см. Лепский, [12]). В [7] показано, что положительный ответ на сформулированный выше вопрос невозможен без некоторого увеличения критических радиусов. Именно, введем *адаптивные критические радиусы*:

$$\rho_n^{ad}(\sigma) = (n/\sqrt{\log\log n})^{-2\sigma/(4\sigma+1)} \tag{1.6}$$

(эти величины соответствуют критическим радиусам (1.4) для объема выборки, уменьшенного в $\sqrt{\log\log n}$ раз). В [7] показано, что если множество неизвестных параметров σ содержит некоторый интервал (σ_0, σ_1), $0 < \sigma_0 < \sigma_1$, то существует такая постоянная $L_0 > 0$, что неразличимость имеет место при выполнении условия

$$\lim \sup_{\sigma_0<\sigma<\sigma_1} \rho_n(\sigma)/\rho_n^{ad}(\sigma) < L_0. \tag{1.7}$$

В работе [8] были построены последовательности тестов, основанные на wavelet-преобразовании и обеспечивающие состоятельную различимость для более широкого класса неизвестных параметров τ. В частности, для альтернатив рассматриваемого типа состоятельная различимость для всех $\sigma \in (\sigma_0, \sigma_1)$ при любых $0 < \sigma_0 < \sigma_1$ обеспечивается при

$$\lim \inf_{\sigma_0 < \sigma < \sigma_1} \rho_n(\sigma)/\rho_n^{ad}(\sigma) > L_1 \tag{1.8}$$

для некоторой положительной постоянной L_1. Соотношения (1.7), (1.8) показывают, что (1.6) дает точный порядок адаптивных критических радиусов в рассматриваемом случае.

Целью данной работы является получение аналогичных результатов для задачи проверки гипотезы равномерности распределения независимой выборки для гладких альтернатив рассматриваемого типа. При этом мы покажем, что состоятельную различимость обеспечивают тесты, основанные на одновременном использовании наборов классических тестов хи-квадрат (такие "объединенные" тесты мы называем *адаптивными тестами хи-квадрат*). Отметим, что предположение о равномерности плотности распределения f_0 при нулевой гипотезе не является существенным (ср. с [13]): для любой простой гипотезы с непрерывной функцией распределения $F_0(t)$ ц помощью преобразования Смирнова $t \to F_0(t) \in [0,1]$ можно перейти к равномерному распределению, при этом "гладкость" альтернатив f и их "удаленность" от H_0 понимается в смысле ограничений на функции $\phi(t) = f(F_0^{-1}(t))$ типа $\phi - 1 \in B^{\sigma}_{q,h}$, $\|\phi - 1\|_p > \rho_n$.

2. Основные результаты

Формальная постановка задачи "адаптивной" проверки гипотез в рамках п. 1.1 состоит в следующем. Обозначим через Ξ множество параметров

$$\Xi = \{\tau = (\sigma, p, q, h, R) : \sigma \in (0, \infty), R \in (0, \infty),$$
$$h \in [1, \infty], 1 \le p \le 2, p \le q \le \infty\},$$

через $\sigma(\tau)$ обозначим соответствующую компоненту параметра τ. Пусть заданы подмножество $K \subset \Xi$ и последовательность положительных функций $\rho_n = \rho_n(\tau)$, $\tau \in K$. Рассмотрим последовательность "объединенных" альтернатив, сооветствующих множеству

$$F_n(K, \rho_n) = \bigcup_{\tau \in K} F_n(\tau, \rho_n(\tau)),$$

где $F_n(\tau, \rho_n(\tau))$ определяется (1.3) (мы рассмотрим случай шаров Бесова; поскольку сформулированные ниже результаты не зависят от параметра h, аналогичные результаты для шаров Соболева и Никольского вытекают из известных теорем вложения). Задача состоит в определении условий асимптотической различимости и неразличимости в зависимости от свойств множества K и последовательности функций ρ_n. Ниже будем предполагать, что для рассматриваемых множеств K множество значений параметра R для $\tau \in K$ заключено в некотором фиксированном интервале $R_{\min},\ R_{\max}:\ 0 < R_{\min} \leqslant R \leqslant R_{\max}$

Для подмножества $\Delta \subset \Xi$ через $\sigma(\Delta)$ обозначим множество $\{\sigma(\tau), \tau \in \Delta\} \subset (0, \infty)$. Назовем подмножество $\Delta \subset \Xi$ *существенным*, если $\sigma(\Delta)$ содержит нетривиальный интервал: $\sigma(\Delta) \supset (\sigma_0, \sigma_1);\ 0 < \sigma_0 < \sigma_1$.

Условия асимптотической неразличимости описываются следующей теоремой.

Теорема 1. *Существует такая постоянная* $L_0 > 0$, *что* $\beta_n(K, \rho_n) \to 1 - \alpha$ *для любого* $\alpha \in (0,1)$, *если для некоторого существенного подмножества* $\Delta \subset K$ *выполнено предельное соотношение*

$$\limsup \sup_{\tau \in \Delta} \rho_n(\tau)/\rho_n^{ad}(\sigma(\tau)) < L_0,$$

где величины $\rho_n^{ad}(\tau)$ *определяются* (1.6).

Замечание 1. Условие существенности множества $\Delta \subset K$ в Теореме 1 может быть ослаблено: множество Δ в предельном соотношении может быть заменено последовательностью $\Delta_n \subset K$ таких подмножеств, что $\sigma(\Delta_n)$ содержат интервалы $(\sigma_{0,n}, \sigma_{1,n})$ длины $l_n = \sigma_{1,n} - \sigma_{0,n} \to 0,\ l_n(\log n)^b \to \infty,\ b \in (0,1)$, или даже последовательностью дискретных подмножеств $\Delta_n = \{\tau_{1,n}, \dots, \tau_{k_n,n}\} \subset K$, таких что $k_n \gg (\log n)^\delta$ для некоторого $\delta > 0$ соответствующие компоненты $(\log n)^{-1} \ll \sigma_{1,n} < \dots < \sigma_{k_n,n} = O(1)$ достаточно “разнесены”: $\min_i(\sigma_{i+1,n} - \sigma_{i,n}) \gg (\log n)^{-1}$. Однако в отличии от задачи адаптивного оценивания функции неизвестной гладкости, где уже для двухточечного множества $\Sigma(\Delta)$ невозможно получить тот же порядок точности оценивания, что и при известном параметре гладкости σ (см. [12]), нельзя полностью отказаться от этого условия. Действительно, для любого конечного множества

$K = \{\tau_1, ..., \tau_M\}$ условие $\min_{1\le i\le M}\ \rho_n(\tau_i)/\rho_n(\sigma(\tau_i)) \to \infty$ с обычными критическими радиусами (1.4) достаточно для различимости: легко видеть, состоятельной является последовательность "объединений" $\psi_n = \max_{1\le i\le M}\ \psi_{n,i}$ соответствующих состоятельных тестов $\psi_{n,i} = \psi_n(\tau_i)$.

Опишем теперь "адаптивные тесты хи-квадрат". Рассмотрим последовательность наборов $\bar\Delta_n = (\Delta^{(r_n)}, ..., \Delta^{(m_n)})$; здесь $\Delta^{(l)} = \Delta_{2^l}$ есть равномерное разбиение интервала $(0,1]$ на 2^l интервалов, $l = r_n, ..., m_n$, причем $r_n \sim B_1 \log\log n,\ 2^{m_n} \asymp n^2/(\log n)^{B_2}$ при достаточно больших $B_1 > 0,\ B_2 > 0$. Для каждого $l = r_n, ..., m_n$ рассмотрим соответствующие статистики $L_{n,2^l}$ вида (1.5) и положим

$$\psi_n = \max_{r_n\le l\le m_n} \psi_{n,2^l,T_n} = \mathbf{1}_{\max_{r_n\le l\le m_n} L_{n,2^l} > T_n};\ T_n = \sqrt{2\log m_n}$$

(т.е. мы рассматриваем "объединения" тестов хи-квадрат $\psi_{n,2^l,T_n}$ для разбиений из $\bar\Delta_n$ и порогов $T_n = \sqrt{2\log m_n} \sim \sqrt{C\log\log n};\ C = 2\log 2$: гипотеза отвергается, если она отвергается каким-либо из тестов $\psi_{n,2^l,T_n},\ l = r_n, ..., m_n$).

Теорема 2. *Справедливо соотношение: $\alpha(\psi_n) \to 0$, и для любого множества $K \subset \Xi$ с компактным $\sigma(K) \subset R^1_+$ существует такая постоянная $L_1 > 0$, что при выполнении условия*

$$\liminf \inf_{\tau\in K}\ \rho_n(\tau)/\rho_n^{ad}(\sigma(\tau)) > L_1$$

выполнено соотношение: $\beta_n(\psi_n, F_n(K, \rho_n)) \to 0$.

Другими словами, при выполнении условий теоремы последовательность тестов ψ_n состоятельна.

Замечание 2. Множество K в Теореме 2 может быть заменено последовательностью таких множеств $K_n \subset \Xi$, что множества $\sigma(K_n)$ лежат в интервалах $[(\log\log n)^{1+\delta}/\log n, \sigma_{\max}]$ для некоторых $\delta > 0,\ \sigma_{\max} > 0$.

Следствие. *Пусть множество $K \subset \Xi$ существенно, $\sigma(K)$ – компактно и $\rho_n(\tau) = C\rho_n^{ad}(\sigma(\tau))$. Тогда для некоторых постоянных $0 < L_0 \le L_1$ при $C < L_0$ гипотеза и альтернатива, соответствующая $F_n(K, \rho_n)$, асимптотически неразличимы, а при $C > L_0$ – состоятельно различимы с помощью адаптивных тестов хи-квадрат.*

Утверждение следствия аналогично результатам Спокойного [7, 8] для рассматриваемого множества параметров Ξ. Отметим, что для других областей параметров $\{\tau\}$ вид адаптивных критических радиусов и структура "адаптивных" тестов существенно отличаются.

3. Доказательство теоремы 1

Доказательство теоремы проводится по схеме, аналогичной [7], основанной на построении большого числа "ортогональных" априорных продакт-мер, сосредоточенных на альтернативе, идея которой (для ортогональных сигналов) восходит к работе Бурнашева [14].

Из теории вейвлетов (wavelets) хорошо известно (см., например, [11]), что для заданного $\sigma_{\max}$ и достаточно большого $j_0 > 0$ можно построить такую ортонормальную систему $\{\phi_\eta;\ \eta = (i,j), j \ge j_0, i = 1,...2^j\}$ в $L_2(0,1)$, что $(\phi_\eta, 1) = 0$ и обладающую следующим свойством. Обозначим G и G_j линейные оболочки систем $\{\phi_\eta\}$ и $\{\phi_{i,j}; i = 1,...2^j\}$. Тогда найдутся такие постоянные $C_1 \ge C_0 > 0$, что при всех $p \ge 1, h \ge 1, \sigma \in (0, \sigma_{\max}), j \ge j_0$ справедливы неравенства:

$$C_0 \|g_{\theta,j}\|_{\sigma,p,h} \ge \|\theta\|_p 2^{j(\sigma+1/2-1/p)} \ge C_1 \|g_{\theta,j}\|_{\sigma,p,h},\ \theta \in R^{2^j}. \tag{3.9}$$

Здесь

$$g_{\theta,j} = \sum_i \theta_i \phi_{i,j} \in G_j,\ \|\theta\|_p = \left(\sum_i |\theta_i|^p\right)^{1/p}$$

с очевидным обобщением на $p = \infty$, $\|g\|_{\sigma,p,h}$ есть норма Бесова; аналогичные неравенства при $\sigma = 0$ справедливы для L_p-нормы $\|g\|_p$.

Для доказательства теоремы достаточно построить такую последовательность π_n априорных распределений на F, что $\pi_n(F_n(K,\rho_n)) = 1$ (или, по крайней мере, $\pi_n(F_n(K,\rho_n)) \to 1$) и справедливо соотношение

$$l_n = E_{f_0,n}\left(\frac{dP_{\pi_n}}{dP^n_{f_0}} - 1\right)^2 \to 0. \tag{3.10}$$

Здесь P_{π_n} есть смесь мер P^n_f по мере π_n. Мы будем рассматривать априорные меры на функциях вида $f = 1 + g,\ g \in G$, так что меры π_n можно рассматривать на множестве параметров $\{\theta_\eta;\ \eta=(i,j), j\geq j_0, i=1,...2^j\}$. При этом

$$dP_{\pi_n}/dP_{f_0}^n = E_{\pi_n}\prod_{\alpha=1}^{n}\left(1+\sum_\eta \theta_\eta\phi_\eta(x_\alpha)\right). \qquad (3.11)$$

Выберем меры π_n в виде

$$\pi_n = k^{-1}\sum_{l=1}^{k}\pi_{n,l};\ \ \pi_{n,l}(d\theta)=\prod_\eta \pi_{n,l,\eta}(\sqrt{n}d\theta_\eta),$$

т.е. $\pi_{n,l}$ есть продакт-меры, соответствующие последовательностям $\bar\pi_{n,l}=\{\pi_{n,l,\eta}\}$ вероятностных мер на прямой с множителем $\sqrt{n}$; вид этих последовательностей и величины $k=k_n$ будут уточнены ниже. Легко видеть, что $E_{f_0,n}\left(dP_{\pi_n}/dP_{f_0}^n\right)=1$ и соотношение (3.10) равносильно следующему:

$$\limsup\ \frac{1}{k^2}\sum_{l_1,l_2=1}^{k}E_{f_0,n}\frac{dP_{\pi_{n,l_1}}}{dP_{f_0}^n}\frac{dP_{\pi_{n,l_2}}}{dP_{f_0}^n}\leq 1. \qquad (3.12)$$

Будем использовать следующее утверждение. Рассмотрим неотрицательно определенную билинейную форму

$$(\bar\pi_{n,l_1},\bar\pi_{n,l_2})=\sum_\eta(\pi_{n,l_1,\eta},\pi_{n,l_2,\eta})=$$
$$\sum_\eta\int_{R^1}\int_{R^1}(e^{uv}-1)\pi_{n,l_1,\eta}(du)\pi_{n,l_2,\eta}(dv).$$

Лемма 1. *Справедливо неравенство:*

$$E_{f_0,n}\frac{dP_{\pi_{n,l_1}}}{dP_{f_0}^n}\frac{dP_{\pi_{n,l_2}}}{dP_{f_0}^n}\leq \exp(\bar\pi_{n,l_1},\bar\pi_{n,l_2}).$$

Доказательство. Используя неравенство $1+a\leq e^a$, (3.11), ортонормальность системы $\{f_0,\phi_\eta\}$ и теорему Фубини, имеем (здесь δ_{η_1,η_2} есть символ Кронекера):

$$E_{f_0,n}\frac{dP_{\pi_{n,l_1}}}{dP_{f_0}^n}\frac{dP_{\pi_{n,l_2}}}{dP_{f_0}^n}=E_{\pi_{n,l_1}\times\pi_{n,l_2}}E_{f_0,n}\prod_{\alpha=1}^{n}(1+\sum_{\eta_1}\theta_{\eta_1,l_1}\phi_{\eta_1}(x_\alpha)+$$

$$+\sum_{\eta_2}\theta_{\eta_2,l_2}\phi_{\eta_2}(x_\alpha)+\sum_{\eta_1,\eta_2}\theta_{\eta_1,l_1}\theta_{\eta_2,l_2}\phi_{\eta_1}(x_\alpha)\phi_{\eta_2}(x_\alpha))=$$

$$=E_{\pi_{n,l_1}\times\pi_{n,l_2}}\left(1+\sum_{\eta_1,\eta_2}\theta_{\eta_1,l_1}\theta_{\eta_2,l_2}\delta_{\eta_1,\eta_2}\right)^n=$$

$$=\prod_\eta E_{\pi_{n,l_1,\eta}\times\pi_{n,l_2,\eta}}\left(1+\sum_\eta\theta_{\eta,l_1}\theta_{\eta,l_2}\right)^n\le$$

$$\le\prod_\eta E_{\pi_{n,l_1,\eta}\times\pi_{n,l_2,\eta}}\exp\left(n\sum_\eta\theta_{\eta,l_1}\theta_{\eta,l_2}\right)=$$

$$=\prod_\eta\int\limits_{R^1}\int\limits_{R^1}e^{uv}\pi_{n,l_1,\eta}(du)\pi_{n,l_2,\eta}(dv)=$$

$$=\prod_\eta((\pi_{n,l_1,\eta},\pi_{n,l_2,\eta})+1)\le\exp(\bar\pi_{n,l_1},\bar\pi_{n,l_2}).$$

Лемма доказана.

Для доказательства утверждений Теоремы 1 и Замечания 1 достаточно рассмотреть случай, когда множество K содержит такую последовательность подмножеств $K_n=\{\tau_{n,1},...,\tau_{n,k_n}\}$, что $k_n\asymp(\log n)^\delta$, $\delta>0$ и для параметров $\sigma_{n,l}=\sigma(\tau_{n,l})$ имеют место соотношения: $\rho_n(\tau_{n,l})\le L_0\rho_n^{ad}(\sigma_{n,l})$,

$$(\log n)^{-1}\ll\sigma_{n,1}<...<\sigma_{n,k_n}\le\sigma_{\max};\ \min_l(\sigma_{n,l+1}-\sigma_{n,l})\gg(\log n)^{-1}. \tag{3.13}$$

Рассмотрим последовательности числовых наборов вида

$$j_{n,l}=[C+\frac{2\log_2 n-\log_2\log k_n)}{1+4\sigma_{n,l}}],$$

$$b_{n,l}=\frac{1}{2}\left(2^{-j_{n,l}+1}\log k_n\right)^{1/4};\ l=1,...,k_n, \tag{3.14}$$

где $C>0$ – некоторая постоянная, $[t]$ – целая часть t. Из (3.13) следует, что $j_{n,1}>...>j_{n,k_n}\gg1$, $\max_l b_{n,l}=o(1)$. Выберем меры $\pi_{n,l,\eta}$ следующим образом:

$$\pi_{n,l,i,j}=\begin{cases}\delta_0, & j\ne j_{n,l};\\ (\delta_{-b_{n,l}}+\delta_{b_{n,l}})/2, & j=\ j_{n,l},\ i=1,...,2^{j_{n,l}}.\end{cases} \tag{3.15}$$

Соотношение (3.15) означает, что относительно мер $\pi_{n,l}$ случайные функции $f(x)$ имеют вид:

$$f(x) = 1 + n^{-1/2} b_{n,l} \sum_{i=1}^{2^{j_l}} \epsilon_i \phi_{i,j_l}(x); \;\; P(\epsilon_i = 1) = P(\epsilon_i = -1) = 1/2, \tag{3.16}$$

где $\epsilon_1, ..., \epsilon_{2^{j_l}}$ – независимые бернуллиевские случайные величины $j_l = j_{n,l}$. В силу (3.9) для нормы L_∞, функции (3.16) являются плотностями распределения при

$$\max_l n^{-1/2} 2^{j_l/2} b_{n,l} = o(1) \tag{3.17}$$

и условие $f \in F(K, \rho_n)$ вытекает из следующих:

$$\max_l n^{-1/2} 2^{j_l(1/2+\sigma_{n,l})} b_{n,l} \le C_1, n^{-1/2} 2^{j_l/2} b_{n,l} \ge C_2 \rho_n(\tau_{n,l}). \tag{3.18}$$

Из (3.17), (3.18) вытекает требуемое соотношение $\pi_n(F_n(K, \rho_n)) = 1$.

Наконец, для построенных наборов последовательностей $\bar\pi_{n,l}$ выполнено условие ортогональности:

$$(\bar\pi_{n,l_1}, \bar\pi_{n,l_2}) = \begin{cases} 0\ , & l_1 \ne l_2; \\ 2^{j_l+1} \sinh^2(b_{n,l}^2/2) \sim 2^{j_l-1} b_{n,l}^4, & l_1 = l_2 = l, \end{cases} \tag{3.19}$$

так что в силу Леммы 1 для выполнения условия неразличимости (3.12) достаточно, чтобы $\max_l k_n^{\delta-1} \exp(2^{j_l-1} b_{n,l}^4) < 1$ для некоторого $\delta \in (0,1)$, что, в свою очередь, вытекает из соотношения:

$$\max_l 2^{j_l-1} b_{n,l}^4 < (1-\delta) \log k_n. \tag{3.20}$$

Справедливость соотношений (3.17)–(3.20) легко проверяется с учетом соотношений (3.14) при некоторых постоянных C, L_0 и выбора ρ_n^{ad} в соответствии с (1.6). Теорема доказана.

4. Доказательство Теоремы 2

Для доказательства Теоремы 2 и Замечания 2 нужно проверить соотношение $\alpha(\psi_n) \to 0$, которое вытекает из следующего:

$$\sum_{l=r_n}^{m_n} P_{f_0}^n (L_{n,k_l} > T_n) \to 0; k_l = 2^l, \tag{4.21}$$

а также, с учетом известных теорем вложения, показать, что для некоторых постоянных $B > 0$, $L_1 > 0$ и для любых $\sigma_{\max} > 0$, $R_{\max} > R_{\min} > 0$, любых последовательностей $\tau_n = (\sigma_n, p_n, q_n, h_n, R_n)$: $1 \le p_n = q_n \le 2$, $h_n = \infty$, $(\log\log\log n)^{1+\delta}/\log n \le \sigma_n \le \sigma_{\max}$, $R_{\min} \le R_n \le R_{\max}$, а также $f_n \in F_n(\tau_n, \rho_n)$ при $\rho_n(\tau_n) > L_1\rho_n^{ad}(\sigma_n)$ найдется такая последовательность l_n, $r_n \le l_n \le m_n$, что

$$P^n_{f_n}(L_{n,k_{l_n}} \le T_n) \to 0. \tag{4.22}$$

Докажем (4.21). Для функции распределения стандартного нормального закона имеем: $\Phi(-\sqrt{2\log m_n}) \asymp 1/m_n\sqrt{\log m_n} = o(1/m_n)$. Поэтому соотношение (4.21) вытекает из оценки вероятностей умеренных уклонений статистик L_{n,k_l} при $m_n \ge l_n > \log\log n$:

$$P^n_{f_0}(L_{n,k_l} > T_n) \le \Phi(-T_n) + \xi_n,\ \xi_n = o(m_n^{-1}). \tag{4.23}$$

Для доказательства (4.23) используем представление L_{n,k_l} как вырожденной U-статистики второго порядка с неотрицательно определенным симметричным ядром $K_{n,l}(x,y)$, порожденным центрированными индикаторными функциями разбиения $\Delta_{k_l} = \{\delta_{1,k_l}, ..., \delta_{k_l,k_l}\}$:

$$L_{n,k_l} = C_n \sum_{1\le s<t\le n} K_{n,l}(x_s, x_t);$$

$$K_{n,l}(x,y) = \sum_{i=1}^{k_l} (\mathbf{1}_{\delta_{i,k_l}}(x) - 1/k_l)(\mathbf{1}_{\delta_{i,k_l}}(y) - 1/k_l),$$

где $C_n = (2K_{n,2}/n(n-1))^{-1/2} \sim (2k_l)^{1/2}/n$,

$$K_{n,2} = \int\int K^2_{n,l}(x,y)dxdy = \|K_{n,l}\|_2^2 = (1-1/k_l)/k_l \sim 1/k_l$$

(здесь и ниже область интегрирования есть интервал $[0,1]$, $K_{n,l}$ есть интегральный оператор в $L_2(0,1)$ с ядром $K_{n,l}(x,y)$). Используем также мартингальное представление вырожденной U-статистики второго порядка:

$$L_{n,k_l} = C_n \sum_{j=2}^{n} \xi_{n,j};\ \xi_{n,j} = \sum_{s<j} K_{n,l}(x_s, x_j),$$

где $\xi_{n,j}$ образуют серию мартингал-разностей относительно σ-алгебр $\Im_j$, порожденных наблюдениями $\{x_1, ..., x_j\}$ (ср. с работой Холла [15]), а также оценки скорости сходимости в центральной предельной теореме (см. Липцер и Ширяев [16], неравенство (9)), из которых вытекает неравенство:

$$\sup_t |P^n_{f_0}(L_{n,k_l} < t) - \Phi(t)| \le$$

$$\leqslant C_1 \left(D \left(\sum_{j=2}^n E(\xi_{n,j}|\Im_{j-1}) \right) \right)^{1/6} + C_2 \left(\sum_{j=2}^n E\xi^4_{n,j} \right)^{1/8}$$

(здесь и ниже средние и дисперсии рассматриваются относительно меры $P^n_{f_0}$; C – абсолютные постоянные). Нетрудно получить оценки (см., например, [15])

$$\sum_{j=2}^n E\xi^4_{n,j} \le C \left(\frac{K_{n,4}}{n^2 K^2_{n,2}} + \frac{G_n}{nK^2_{n,2}} \right),$$

$$D \left(\sum_{j=2}^n E(\xi_{n,j}|\Im_{j-1}) \right) \le C \left(\frac{G_{n,2}}{K^2_{n,2}} + \frac{G_n}{nK^2_{n,2}} \right),$$

где $K_{n,4} = \int\int K^4_{n,l}(x,y)dxdy$,

$$G_n(x,y) = \int K_{n,l}(x,z)K_{n,l}(y,z)dz, \ G_{n,2} = \int\int G^2_n(x,y)dxdy.$$

При этом

$$G_{n,2} \le \|K_{n,l}\|^2_2 \|K_{n,l}\|^2_\infty \le K_{n,2} k_l^{-2},$$

так как для равномерной нормы оператора $K_{n,l}$ имеем: $\|K_{n,l}\|_\infty \le 1/k_l$, а также $K_{n,4} \le \sup_{x,y} |K_{n,l}(x,y)| K_{n,2} \le K_{n,2}$, поскольку $\sup_{x,y} |K_{n,l}(x,y)| \le 1$.

Кроме того для неотрицательно определенного симметричного ядра $K_{n,l}(x,y)$ с $K_{n,2} < \infty$ справедливо неравенство

$$G_n \le K_{n,2} K^{1/2}_{n,4}. \tag{4.24}$$

Действительно, разлагая ядро по ортонормальному собственному базису $K_{n,l}(x,y) = \sum_i \lambda_i \psi_i(x)\psi_i(y)$, $\lambda_i \ge 0$, имеем:

$$K_{n,2} = \sum_i \lambda_i^2, \ K_{n,4} = \sum_{i,j,k,l} \lambda_i\lambda_j\lambda_k\lambda_l r^2_{ijkl}, \ G_n = \sum_{i,j} \lambda_i^2 \lambda_j^2 r^2_{iijj},$$

где $r_{ijkl} = \int \psi_i(x)\psi_j(x)\psi_k(x)\psi_k(x)dx$, так что

$$\sum_{i,j} \lambda_i^2 \lambda_j^2 r_{iijj}^2 \le \left(\sum_{i,j} \lambda_i^2 \lambda_j^2\right)^{1/2} \left(\sum_{i,j} \lambda_i^2 \lambda_j^2 r_{iijj}^2\right)^{1/2},$$

откуда следует (4.24).

Из приведенных выше оценок следует, что

$$K_{n,4}/n^2 K_{n,2}^2 = O(k_l/n^2),$$

$$G_{n,2}/K_{n,2}^2 = O(k_l^{-1}),\ G_n/nK_{n,2}^2 = O(k_l^{1/2}/n),$$

откуда получаем (4.23), поскольку для любого $B > 0$ при достаточно больших $B_1,\ B_2$

$$k_l^{-1} \le 2^{-r_n} \ll m_n^{-B},\ k_l/n^2 \le 2^{m_n}/n^2 \asymp (\log n)^{-B_2} \ll m_n^{-B}.$$

Для доказательства (4.22) воспользуемся оценками P_f^n-средних (обозначим их $h_{n,l}(f)$) и дисперсий для $P_{f_0}^n$-нормированных U-статистик L_{n,k_l} (см., например, [2], п. 5.5):

$$E_{n,f}(L_{n,k_l}) = h_{n,l}(f) = \sqrt{\frac{n(n-1)}{2}} \frac{(f, K_{n,l}f)}{\|K_{n,l}\|_2} \sim n(k_l/2)^{1/2}(f, K_{n,l}f),$$

$$D_{n,f}(L_{n,k_l}) = \|K_{n,l}\|_2^{-2}((\|K_{n,l}\|_{2,f}^2 - (f, K_{n,l}f)^2) +$$

$$+ 2(n-1)(\|K_{n,l}f\|_{2,f}^2 - (f, K_{n,l}f)^2)).$$

Здесь для рассматриваемого ядра при $p_{n,j} = P_f(\delta_{j,k_l})$ имеем: $(f, K_{n,l}f) = \sum_{j=1}^{k_l}(p_{n,j} - i/k_l)^2$,

$$\|K_{n,l}\|_{2,f}^2 = \int\int K_{n,l}^2(x,y)f(x)f(y)dxdy =$$

$$= \|K_{n,l}\|_2^2 + (1 - 2/k_l)(f, K_{n,l}f),$$

$$\|K_{n,l}f\|_{2,f}^2 = \int (K_{n,l}f)^2(x)f(x)dx = \sum_{j=1}^{k_l} p_{n,j}(p_{n,j} - i/k_l)^2 \le$$

$$\le (f, K_{n,l}f)/k_l + (f, K_{n,l}f)^{3/2},$$

откуда при $t_{n,l} = k_l^{1/4}/n^{1/2}$ легко получить оценки:

$$D_{n,f}(L_{n,k_l}) \le C\left(1 + h_{n,l}(f) + t_{n,l}h_{n,l}^{3/2}(f) + t_{n,l}^2 h_{n,l}^2(f)\right). \tag{4.25}$$

Используя неравенство Чебышева при $h_{n,l_n}(f_n) > T_n$, видим, что

$$P_{f_n}^n(L_{n,k_{l_n}} \le T_n) = P_{f_n}^n(h_{n,l_n}(f_n) - L_{n,k_{l_n}} \ge h_{n,l_n}(f_n) - T_n) \le$$
$$\le D_{n,f_n}(L_{n,k_{l_n}})/(h_{n,l_n}(f_n) - T_n)^2,$$

и так как $t_{n,l_n} \to 0$, в силу оценки (4.25) для выполнения (4.22) достаточно, чтобы

$$\liminf h_{n,l_n}(f_n)/T_n > 1. \tag{4.26}$$

Заметим теперь, что при $p \le 2$

$$h_{n,l_n}(f_n) \sim n(2k_l)^{-1/2}\|Pr_{k_l}(f_n - f_0)\|_2^2 \geqslant n(2k_l)^{-1/2}\|Pr_{k_l}(f_n - f_0)\|_p^2, \tag{4.27}$$

где Pr_k есть оператор ортонормального проектирования в $L_2(0,1)$ на пространство функций, постоянных на интервалах разбиения Δ_k. Воспользуемся следующим неравенством

$$\|Pr_k(f - f_0)\|_p \ge B_1\|f - f_0\|_p - B_2 k^{-\sigma}, \ f - f_0 \in B_{p,\infty}^\sigma. \tag{4.28}$$

Неравенство (4.28) доказано в [2] (см. неравенство (5.16)) для шаров Никольского H_p^σ, совпадающих с шарами Бесова при $h = \infty$. Доказательство основано на возможности кусочной аппроксимации функций из шаров Никольского H_p^σ полиномами степени $m = m[\sigma]$ с точностью $k^{-\sigma}$ в L_p-норме, где k – число интервалов аппроксимации. Из доказательства легко увидеть, что положительные постоянные B_1, B_2 можно выбрать не зависящими от σ, p, R при $0 < \sigma \le \sigma_{\max}$, $1 \le p \le 2$, $0 < R_{\min} \le R \le R_{\max}$.

Выберем целочисленную последовательность l_n так, чтобы

$$n_*^{2/(1+4\sigma_n)} \le k_{l_n} = 2^{l_n} \le 2n_*^{2/(1+4\sigma_n)}; \quad n_* = n/\sqrt{\log\log n}.$$

Поскольку $(\log\log n)^{1+\delta}/\log n \le \sigma_n \le \sigma_{\max}$, при этом $r_n \ll l_n \le m_n$. Используя (4.27), (4.28), имеем:

$$\begin{aligned} h_{n,k_{l_n}}(f_n) &\ge (n/\sqrt{2k_{l_n}})\left(B_1\|f - f_0\|_p - B_2 k_{l_n}^{-\sigma_n}\right)^2 \\ &\ge n/(\sqrt{2k_{l_n}})\left(B_1 L_1 \rho_n^{ad}(\sigma_n) - B_2 k_{l_n}^{-\sigma_n}\right)^2 \\ &\ge (n/2)n_*^{1/(1+4\sigma_n)}\left(B_1 L_1 n_*^{-2\sigma_n/(1+4\sigma_n)} - B_2 n_*^{-2\sigma_n/(1+4\sigma_n)}\right)^2 \\ &\sim ((B_1 L_1 - B_2)^2/2)n/n_* \end{aligned}$$

откуда при достаточно большом L_1 следует (4.26). Теорема доказана.

Литература

1. Ю. И. Ингстер, *Минимаксная проверка непараметрических гипотез о плотности распределения в метриках* L_p. Теория вероятностей и ее примен. **31**, No. 2 (1986), 384–389.
2. Yu. I. Ingster, *Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III.* Mathematical Methods of Statistics, **2**, Nos. 2–4 (1993), 85–114, 171–189, 249–268.
3. Ю. И. Ингстер, *Минимаксная проверка гипотез о плотности распределения для эллипсоидов в* l_p. Теория вероятностей и ее примен. **39**, No. 3 (1994), 530–553..
4. М. С. Ермаков, *Асимптотически минимаксные критерии проверки непараметрических гипотез о плотности распределения.* Зап. Научн. Семин. ЛОМИ **166** (1988), 44–53.
5. М. С. Ермаков, *Минимаксная непараметрическая проверка гипотез о плотности распределения.* Теория вероятностей и ее примен. **39**, No. 3 (1994), 488–512.
6. Lepski and V. G. Spokoiny, *Minimaxnonparametric hypothesis testing: the case of an inhomogeneous alternative.* Technical Report 44, Humboldt-Univ. (1995), Berlin.
7. V. G. Spokoiny, *Adaptive hypothesis testing using wavelets.* Technical Report 176, Weierstrass Institute, (1995), Berlin.
8. V. G. Spokoiny, *Adaptive and spatially adaptive testing of nonparametric hypothesis.* Technical Report 234, Weierstrass Institute, (1996), Berlin.
9. O. V. Lepski and A. B. Tsybakov, *Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point.* Discussion Paper 91, Humboldt-Univ., (1996), Berlin.
10. H. Triebel, *Theory of function spaces. II.* Basel, Birkhauser, 1992.
11. A. Cohen, I. Daubechies, and P. Vial, *Wavelets on an interval and fast wavelet transforms.* Applied and Computational Harmonic Analysis (A) **1** (1993), 54–81.
12. О. В. Лепский, *О задаче адаптивного оценивания в гауссовском белом шуме.* Теория вероятностей и ее примен. **35**, No. 4 (1990), 454–466.
13. Ю. И. Ингстер, *Асимптотически минимаксная проверка гипотезы независимости.* Зап. Научн. Семин. ЛОМИ **153** (1986), 60–72.
14. М. В. Бурнашев, *О минимаксном обнаружении неточно известного сигнала на фоне белого гауссовского шума.* Теория вероятностей и ее примен. **24**, No. 1 (1979), 107–119.
15. P. Hall, *Central limit theorem for integrated square error of multivariate nonparametric density estimators.* J. Multivar. Anal. **14** (1984), 1–16.
16. Р. Ш. Липцер, А. Н. Ширяев, *О скорости сходимости в центральной предельной теореме для семимартингалов.* Теория вероятностей и ее примен. **27**, No. 1 (1982), 3–14.

Ingster Yu.I. Adaptive chi-square tests.

We consider minimax hypothesis testing problem $H_0 : f = f_0,\ f_0(x) \equiv 1$ on a distribution density f of i.i.d. observations $X_1, ..., X_n,\ X_i \in [0, 1],\ n \to \infty$ versus alternative corresponding to smooth densities f which are distant enough from f_0. A distance between f_0 and f is measured in L_p-norm and a smoothness σ of f is measured in L_q-norm. A priory the values σ, p, q are not fixed but satisfy to constraints $1 \le p \le 2,\ p \le q,\ \sigma > 0$. We show that optimal minimax rate is provided by test procedures which are based on the union of chi-square tests with increasing number of cells.

С.-Петербургский Государственный
Университет путей сообщений

Поступило 5 ноября 1997 г.