Записки научных
семинаров ПОМИ
Том 408, 2012 г.

,

$\boxed{\text{Ю.И.Ингстер}}$, **Н. А. Степанова**

АДАПТИВНАЯ СЕЛЕКЦИЯ КОМПОНЕНТ РАЗРЕЖЕННОЙ ФУНКЦИИ РЕГРЕССИИ

§1. Введение

В этой статье мы изучаем проблему селекции компонент разреженного сигнала в непрерывной регрессионной модели. Мы предполагаем, что функция многих переменных $f \in \mathcal{F}_d \subset L_2([0,1]^d) = L_2^d$ наблюдается в гауссовском белом шуме интенсивности ε. Модель описывается уравнением

$$X_\varepsilon = f + \varepsilon W, \tag{1}$$

где W – гауссовский белый шум, заданный на $[0,1]^d$, $\varepsilon > 0$ – интенсивность шума и $\mathcal{F}_d$ – некоторое подмножество квадратично-интегрируемых на $[0,1]^d$ гладких функций. В этой модели "наблюдение" преставляет собой функцию $X_\varepsilon\colon L_2^d \to \mathcal{G}$, принимающую значения на множестве $\mathcal{G}$ нормальных случайных величин, такую, что если $\xi = X_\varepsilon(\phi)$, $\eta = X_\varepsilon(\psi)$, где $\phi, \psi \in L_2^d$, то $\mathbf{E}(\xi) = (f,\phi)$, $\mathbf{E}(\eta) = (f,\psi)$ и $\mathbf{Cov}(\xi,\eta) = \varepsilon^2(\phi,\psi)$. Для каждой функции $f \in L_2^d$ наблюдение X_ε определяет гауссовскую меру $\mathbf{P}_{\varepsilon,f}$ со средним f и ковариационным оператором $\varepsilon^2 I$, где I – тождественный оператор (см. [8,16]). В настоящей работе изучается случай растущей размерности, то есть $d = d_\varepsilon \to \infty$ при $\varepsilon \to 0$. Предполагается, что f имеет аддитивную разреженную структуру, и задача состоит в точной селекции компонент функции f.

Проблема селекции (отбора факторов) при построении множественной регрессии – классическая задача математической статистики.

Ключевые слова: аддитивная регрессия, разреженный сигнал, точное восстановление, адаптивная селекция.

Работа первого автора поддержана грантом РФФИ 11-01-00577 и грантом Канадского совета в области естественнонаучных и прикладных исследований (NSERC) во время визита первого автора в Карлтон Университет зимой 2012 года.

Работа второго автора поддержана грантом NSERC.

В применении к непараметрическим моделям она изучалась различными авторами в самых разнообразных предположениях (см., например, [2,7,15] и ссылки там же). В отличие от предыдущих исследований в этой области, основанных большей частью на статистической теории оценивания, наш подход к данной проблеме тесно связан с теорией проверки гипотез и ведет к построению асимптотически минимаксного селектора, являющегося адаптивным по отношению к индексу разреженности.

Предположение аддитивности и разреженности функции f связано с наличием эффекта "проклятия размерности". Данный эффект возникает при оценивании и обнаружении сигнала, зависящего от многих переменных, и проявляется в существенном ухудшении качества оценок и тестов с увеличением размерности. Одной из известных моделей, позволяющих эффективно бороться с проклятием размерности, является функциональная модель дисперсионного анализа. Данная модель предполагает, что функция регрессии многих переменных $f \in L_2^d$ представима в виде (см., например, [14])

$$f(\mathbf{x}) = \sum_{u\subset\{1,\ldots,d\}} f_u(\mathbf{x}_u), \quad \int_0^1 f_u(\mathbf{x}_u)dx_j = 0 \ \text{ п.в.} \quad \text{для всех } \ j \in u,$$

где $\mathbf{x}_u = \{x_j,\, j \in u\}$ – $|u|$-мерный подвектор вектора $\mathbf{x} \in [0,1]^d$, и суммирование ведется по всем 2^d подмножествам множества $\{1,\ldots,d\}$. Важным частным случаем этой модели является аддитивная модель с функцией регрессии вида

$$f(\mathbf{x}) = \mu + \sum_{j=1}^d f_j(x_j), \quad \int_0^1 f_j(x_j)dx_j = 0,$$
$$\mu \in \mathbb{R}, \quad \mathbf{x} = (x_1,\ldots,x_d) \in [0,1]^d,$$

ставшая популярной после работы Стоуна [18].

Другая идея, эффективно используемая при работе с непараметрическими регрессионными моделями, состоит в предположении, что функция регрессии имеет разреженную структуру. Условие разреженности обеспечивает идентифицируемость статистических моделей и дает возможность для их более простой интерпретации. Кроме того, оно ведет к уменьшению вычислительных затрат и поэтому весьма популярно в современной статистической литературе (см. [15] и ссылки там же).

В настоящей работе изучается непараметрическая аддитивная модель с разреженной функцией регрессии. Кроме условий регулярности, накладываемых на функцию регрессии f и необходимых для получения содержательной непараметрической задачи, мы предполагаем, что f имеет *разреженную структуру* вида

$$f(\mathbf{x}) = \sum_{j=1}^{d} \eta_j f_j(x_j), \quad \mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d, \tag{2}$$

где $\eta_j \in \{0,1\}$ – параметры, соответствующие наличию j-й компоненты сигнала ($\eta_j = 1$) или ее отсутствию ($\eta_j = 0$). Предполагается, что η_j удовлетворяют условию $\sum\limits_{j=1}^{d} \eta_j = d^{1-\beta}$ для некоторого $\beta \in (0,1)$. Величина $d^{1-\beta}$ есть число ненулевых сигналов f_j в разложении (2). Параметр β будем называть *индексом разреженности.* Чем больше значение β, тем меньше активных компонент содержит сигнал f. Аддитивные функции с разреженной структурой вида (2) были введены в работе [5] в связи с задачей обнаружения сигнала. Мы же рассматриваем эти функции в контексте задачи селекции.

Для более точной постановки задачи введем множество

$$\mathcal{H}_{d,\beta} = \left\{ \eta = (\eta_1, \ldots, \eta_d) : \eta_j \in \{0,1\}, 1 \leqslant j \leqslant d, \sum_{j=1}^{d} \eta_j = d^{1-\beta} \right\}.$$

Требуется по наблюдению X_ε в модели (1) с функцией регрессии f вида (2) оценить вектор $\eta = (\eta_1, \ldots, \eta_d) \in \mathcal{H}_{d,\beta}$. Другими словами, требуется определить наличие или отсутствие компонент f_j, составляющих сигнал f. Эту задачу можно рассматривать как задачу *селекции.* Чтобы задача селекции была содержательной, сигналы f_j не должны быть слишком "слабыми". Более точно, для каждого $j = 1, \ldots, d$, сигнал f_j должен быть *обнаружим* в задаче проверки гипотезы $H_{0j} : f_j \equiv 0$ против альтернативы $H_{1j} : f_j \in \mathcal{F}(r_\varepsilon)$, где $\mathcal{F}(r_\varepsilon) = \{f \in \mathcal{F} : \|f\|_2 \geqslant r_\varepsilon\}$, $\mathcal{F}$ – некоторое подмножество L_2^1 и $\|f\|_2$ – норма в пространстве L_2^1. Обнаружение f_j представляется возможным в случае, когда параметр

$r_\varepsilon > 0$, характеризующий альтернативу, превосходит некоторое положительное число r_ε^*, называемое *критическим радиусом*. Цель настоящей работы состоит в определении критического радиуса r_ε^* и в построении асимптотически минимаксного *селектора* для вектора $\eta = (\eta_1, \ldots, \eta_d) \in \mathcal{H}_{d,\beta}$, гарантирующего *точное воспроизведение* сигнала $f(\mathbf{x}) = \sum_{j=1}^{d} \eta_j f_j(x_j)$. Наше исследование дополняет результаты, полученные в работе [6] относительно точного воспроизведения разреженного нормального вектора.

Точная постановка задачи состоит в следующем. Для каждой компоненты f_j сигнала f будем рассматривать задачу проверки гипотезы $H_{0j} : f_j \equiv 0$ против альтернативы $H_{1j} : f_j \in \mathcal{F}(r_\varepsilon)$. Гипотезы H_{0j} и H_{1j} являются асимтотически различимыми (неразличимыми), если $r_\varepsilon / r_\varepsilon^* \to \infty$ $(r_\varepsilon / r_\varepsilon^* \to 0)$ при $\varepsilon \to 0$. Асимптотическая различимость (неразличимость) означает, что минимаксная ошибка обнаружения сигнала для альтернативы H_{1j} стремится к нулю (единице) при $\varepsilon \to 0$. В случае, когда H_{0j} и H_{1j} асимптотически различимы, мы говорим, что сигнал f_j *обнаружим*. Если гипотезы H_{0j} and H_{1j} асимптотически различимы (неразличимы) в случае, когда $\liminf r_\varepsilon / r_\varepsilon^* > 1$ $(\limsup r_\varepsilon / r_\varepsilon^* < 1)$, то критический радиус r_ε^* называется *точным*. Знание (точного) критического радиуса позволяет сделать содержательной задачу проверки гипотезы $H_{0j} : f_j \equiv 0$ против альтернативы $H_{1j} : f_j \in \mathcal{F}(r_\varepsilon)$ путем выбора параметра r_ε, превосходящего критический радиус r_ε^*. В противном случае, сигнал f_j слишком слаб для того, чтобы быть обнаружимым.

Пусть далее $\mathcal{F}_d(r_\varepsilon)$ есть множество функций вида $f = \sum_{j=1}^{d} \eta_j f_j \in \mathcal{F}_d$, где $f_j \in \mathcal{F}(r_\varepsilon)$ для каждого $j = 1, \ldots, d$. Вернемся к модели (1) и предположим, что каждая компонента f_j сигнала $f \in \mathcal{F}_d(r_\varepsilon)$ обнаружима. Как и в работе [6], мы определяем качество селектора $\eta^* = \eta^*(X_\varepsilon)$, воспроизводяшего вектор η, с помощью *расстояния Хэммингa*:

$$|\eta^* - \eta| = \sum_{j=1}^{d} |\eta_j^* - \eta_j|.$$

Обозначим через $\mathbf{E}_{f,\eta}$ математическое ожидание по мере $\mathbf{P}_{\varepsilon,f}$ с функцией f вида (2). Требуется найти *селектор* $\eta^* = \eta^*(X_\varepsilon)$, удовлетворяющий условию

$$\limsup_{\varepsilon\to 0} \sup_{\eta\in\mathcal{H}_{d,\beta}} \sup_{f\in\mathcal{F}_d(r_\varepsilon)} \mathbf{E}_{f,\eta}|\eta^*-\eta| = 0, \tag{3}$$

и показать, что при нарушении условия обнаружимости,

$$\liminf_{\varepsilon\to 0} \inf_{\widetilde{\eta}} \sup_{\eta\in\mathcal{H}_{d,\beta}} \sup_{f\in\mathcal{F}_d(r_\varepsilon)} \mathbf{E}_{f,\eta}|\widetilde{\eta}-\eta| > 0, \tag{4}$$

то есть, что в этом случае точная селекция невозможна. При справедливости соотношения (3) будем говорить, что селектор η^* позволяет *точно воспроизводить* сигнал $f(\mathbf{x}) = \sum_{j=1}^{d} \eta_j f_j(x_j)$, наблюдаемый в гауссовском белом шуме. Если при этом справедливо неравенство (4), то процедура селекции, основанная на η^*, является *наилучшей* (в минимаксном смысле).

Наряду с задачей точной селекции, можно рассматривать задачу *почти точной* селекции, когда требуется найти селектор η^*, для которого (ср. с (3)) $\sup_{\eta\in\mathcal{H}_{d,\beta}} \sup_{f\in\mathcal{F}_d(r_\varepsilon)} \mathbf{E}_{f,\eta}|\eta^*-\eta| = o(d^{1-\beta})$, что соответствует правильному выбору *почти всех* значимых компонент. В адаптивной постановке данная задача требует отдельного изучения, и в настоящей работе не рассматривается.

§2. Предварительные сведения

2.1. Соболевские шары. Рассмотрим класс функций, определяемый в терминах соболевской полунормы порядка $\sigma > 0$. Для начала предположим, что $\sigma > 0$ – целое число, и что функция $f(x)$, $x \in [0,1]$, имеет производные до порядка σ включительно, причем производная $f^{(\sigma-1)}$ абсолютно непрерывна, и $\int_0^1 \left(f^{(\sigma)}(x)\right)^2 dx < \infty$. Для такой функции зададим полунорму $\|f\|_{\sigma,2}$ равенством

$$\|f\|_{\sigma,2}^2 = \int_0^1 \left(f^{(\sigma)}(x)\right)^2 dx, \tag{5}$$

и обозначим через $\mathcal{F}_\sigma$ соболевский шар единичного радиуса, то есть,

$$\mathcal{F}_\sigma = \{f \in L_2^1 : \|f\|_{\sigma,2} \leqslant 1\}.$$

В случае нецелочисленного параметра $\sigma > 0$, все производные $f^{(m)}$ порядка $0 \leqslant m \leqslant [\sigma]$ предполагаются 1-периодическими, то есть,

$f^{(m)}(0) = f^{(m)}(1)$, и полунорма $\|f\|^2_{\sigma,2}$ задается с помощью разложения f в ряд Фурье. Именно, пусть $\{\phi_k(x)\}_{k\in\mathbb{Z}}$ – стандартный базис Фурье в пространстве $L_2^1 = L_2[0,1]$:

$$\phi_0(x) = 1, \quad \phi_l(x) = \sqrt{2}\cos(2\pi l x), \quad \phi_{-l}(x) = \sqrt{2}\sin(2\pi l x), \quad l > 0.$$

Для функции $f(x) = \sum\limits_{l\in\mathbb{Z}} \theta_l \phi_l(x) \in L_2^1$ с коэффициентами Фурье $\theta_l = (f, \phi_l)$ положим

$$\|f\|^2_{\sigma,2} = \sum_{l\in\mathbb{Z}} c_l^2 \theta_l^2, \tag{6}$$

где

$$c_l^2 = c_l^2(\sigma) = (2\pi|l|)^{2\sigma}, \quad \sigma > 0, \quad l \in \mathbb{Z}.$$

При целочисленном параметре $\sigma > 0$, в предположении периодичности, данное определение совпадает с определением (5). Соболевский класс $\mathcal{F}_\sigma$ произвольного порядка $\sigma > 0$ определяется как

$$\mathcal{F}_\sigma = \{f \in L_2^1 : \|f\|_{\sigma,2} \leqslant 1\},$$

где $\|f\|_{\sigma,2}$ задано в (6). Мы также будем иметь дело с подмножеством множества $\mathcal{F}_\sigma$, из которого удален шар в пространстве L_2^1 радиуса $r_\varepsilon > 0$:

$$\mathcal{F}_\sigma(r_\varepsilon) = \{f \in \mathcal{F}_\sigma : \|f\|_2 \geqslant r_\varepsilon\}.$$

Здесь и далее параметр гладкости $\sigma > 0$ предполагается известным.

2.2. Формулировка задачи в терминах последовательностей. Рассматриваемую задачу можно представить в несколько иной, более удобной с технической точки зрения, форме. Следуя работе [5], рассмотрим класс аддитивных функций d переменных с разреженной структурой, определяемый следующим образом:

$$\mathcal{F}_{d,\beta,\sigma} = \left\{ f : f(\mathbf{x}) = \sum_{j=1}^{d} \eta_j f_j(x_j), \int_0^1 f_j(x_j)dx_j = 0, \right.$$
$$\left. f_j \in \mathcal{F}_\sigma, \eta = (\eta_j) \in \mathcal{H}_{d,\beta} \right\}.$$

Регрессионную модель (1), в которой сигнал f принадлежит функциональному классу $\mathcal{F}_{d,\beta,\sigma}$, можно переписать в терминах квадратично-суммируемых последовательностей. Для этого в пространстве L_2^d рассмотрим базис $\{\phi_l(\mathbf{x})\}_{l\in\mathbb{Z}^d}$ вида

$$\phi_l(\mathbf{x}) = \prod_{k=1}^{d}\phi_{l_k}(x_k), \quad \mathbf{x} = (x_1,\dots,x_d)\in[0,1]^d, \quad l = (l_1,\dots,l_d)\in\mathbb{Z}^d.$$

Для функции $f(\mathbf{x}) = \sum\limits_{l\in\mathbb{Z}^d}\theta_l\phi_l(\mathbf{x}) \in L_2^d$ с коэффициентами Фурье $\theta_l = (f,\phi_l)$, $l\in\mathbb{Z}^d$, и числа $\sigma>0$, положим

$$\|f\|^2_{d,\sigma,2} = \sum_{l\in\mathbb{Z}^d} c_l^2\theta_l^2,$$

где

$$c_l^2 = c_l^2(\sigma) = \sum_{j=1}^{d}(2\pi|l_j|)^{2\sigma}, \quad l=(l_1,\dots,l_d)\in\mathbb{Z}^d.$$

Далее, для вектора $l\in\mathbb{Z}^d$, у которого j-ая компонента равняется k, а остальные принимают нулевое значение, определим функцию

$$\phi_{j,k}(\mathbf{x}) = \phi_l(\mathbf{x}) = \phi_k(x_j)$$

и обозначим $\theta_{j,k} = (f,\phi_{j,k})$. Тогда соболевская полунорма функции $f = \sum\limits_{j=1}^{d}\eta_j f_j \in \mathcal{F}_{d,\beta,\sigma}$ переписывается в виде

$$\|f\|^2_{d,\sigma,2} = \sum_{j=1}^{d}\eta_j\|f_j\|^2_{\sigma,2} = \sum_{j=1}^{d}\eta_j\sum_{k\in\mathbb{Z}}(2\pi|k|)^{2\sigma}\theta^2_{j,k}.$$

Имея ортонормированную систему $\{\phi_{j,k}(\mathbf{x}) : 1\leqslant j\leqslant d,\ k\in\mathbb{Z}\}$, рассмотрим статистики

$$X_{j,k} = \eta_j\theta_{j,k} + \varepsilon\xi_{j,k}, \quad \xi_{j,k} \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \quad 1\leqslant j\leqslant d, \quad k\in\mathbb{Z}, \qquad (7)$$

где $X_{j,k} = X_\varepsilon(\phi_{j,k})$ – эмпирические коэффициенты Фурье, $\theta_{j,k} = (f,\phi_{j,k}) = \int\limits_0^1\phi_k(x_j)f_j(x_j)dx_j$ – k-й коэффициент Фурье j-й компоненты

f_j сигнала f, а "вектор" $(\eta_1\theta_1, \dots, \eta_d\theta_d)$ состоит из последовательностей $\eta_j\theta_j = (\eta_j\theta_{j,k})_{k\in\mathbb{Z}}$ и принадлежит множеству

$$\Theta_{d,\beta,\sigma} = \Big\{(\eta_1\theta_1, \dots, \eta_d\theta_d) : (\eta_j) \in \mathcal{H}_{d,\beta},\, \theta_j = (\theta_{j,k}) \in l_2(\mathbb{Z}),$$
$$\sum_{k\in\mathbb{Z}} (2\pi|k|)^{2\sigma}\theta_{j,k}^2 \leqslant 1,\, 1 \leqslant j \leqslant d\Big\}.$$

Легко видеть, что это – достаточные статистики в рассматриваемой задаче. Они имеют нормальное распределение $N(\eta_j\theta_{j,k}, \varepsilon^2)$, которое зависит от коэффициентов Фурье $\theta_{j,k}$ относительно системы $\{\phi_{j,k}\}$ и не зависит от выбора самой системы. Поэтому вместо задачи точного воспроизведения сигнала $f \in \mathcal{F}_{d,\beta,\sigma}$ в модели (1) мы можем изучать эквивалентную ей задачу точной селекции координат вектора $(\eta_1\theta_1, \dots, \eta_d\theta_d) \in \Theta_{d,\beta,\sigma}$ в модели (7).

Для нахождения условий, при которых точное воспроизведение сигнала $f = \sum_j \eta_j f_j \in \mathcal{F}_{d,\beta,\sigma}$ возможно, нам понадобится рассмотреть одномерную задачу (одну и ту же для каждого $j = 1, \dots, d$) проверки гипотезы $H_{0j} : f_j \equiv 0$ против альтернативы $H_{1j} : f_j \in \mathcal{F}_\sigma(r_\varepsilon)$, где

$$\mathcal{F}_\sigma(r_\varepsilon) = \mathcal{F}_\sigma(r_\varepsilon) = \{f : f \in \mathcal{F}_\sigma, \|f\|_{\sigma,2} \leqslant 1, \|f\|_2 \geqslant r_\varepsilon\}.$$

В терминах последовательностей данные гипотезы переписываются в виде

$$H_{0j} : \theta_j = 0, \quad H_{1j} : \theta_j \in \Theta_\sigma(r_\varepsilon),$$

где $\theta_j = (\theta_{j,k})_{k\in\mathbb{Z}}$ и

$$\Theta_\sigma(r_\varepsilon) = \Big\{\theta = (\theta_k)_{k\in\mathbb{Z}} \in l_2(\mathbb{Z}) : \sum_{k\in\mathbb{Z}} (2\pi|k|)^{2\sigma}\theta_k^2 \leqslant 1,\, \sum_{k\in\mathbb{Z}} \theta_k^2 \geqslant r_\varepsilon^2\Big\}.$$

Прежде чем перейти к изучению проблемы точного воспроизведения сигнала в функциональной модели регрессии, мы приведем аналогичный результат для более простого векторного случая.

2.3. Селекция компонент разреженного нормального вектора. Пусть $\eta = (\eta_1, \dots, \eta_d) \in \mathcal{H}_{d,\beta}$ и пусть $a = a(d, \beta)$ есть известный положительный параметр. Предположим, что имеется d-мерный нормальный вектор наблюдений $X = (X_1, \dots, X_d)$ возрастающей размерности (то есть $d \to \infty$) вида

$$X = \mu + \xi, \quad \xi \sim N(0, I_d), \tag{8}$$

с разреженным вектором средних значений $\mu = a\eta = (a\eta_1, \dots, a\eta_d)$ и единичной ковариационной матрицей I_d. В данных предположениях задача селекции состоит в оценивании вектора η при наличии "мешающего" параметра a. Очевидно, эта задача является содержательной только тогда, когда число a не очень "мало".

Пусть $\widehat{\eta} = \widehat{\eta}(X) = (\widehat{\eta}_1(X), \dots, \widehat{\eta}_d(X))$ есть селектор, построенный по вектору наблюдений X. Обозначим через $|\widehat{\eta} - \eta|$ расстояние Хэмминга между селектором $\widehat{\eta}$ и вектором η, то есть, $|\widehat{\eta} - \eta| = \sum_{j=1}^{d} |\widehat{\eta}_j - \eta_j|$. Предположим, что качество селекции измеряется с помощью максимального риска

$$R(\widehat{\eta}) = \sup_{\eta \in \mathcal{H}_{d,\beta}} \mathbf{E}_\eta |\widehat{\eta} - \eta|.$$

В этих и чуть более общих предположениях проблема точной селекции изучалась в работе [6], где показано, что точная селекция осуществима при условии $\liminf_{d\to\infty} \frac{a}{\sqrt{\log d}} > \sqrt{2}(1 + \sqrt{1-\beta})$, и недостижима, когда $\limsup_{d\to\infty} \frac{a}{\sqrt{\log d}} < \sqrt{2}(1 + \sqrt{1-\beta})$.

Можно показать, что при выполнении условия $\liminf_{d\to\infty} \frac{a}{\sqrt{\log d}} > \sqrt{2}(1 + \sqrt{1-\beta})$, точная селекция в векторной модели (8) обеспечивается оценкой $\widehat{\eta}$ вектора η вида

$$\begin{gathered} \widehat{\eta} = \widehat{\eta}(X) = (\widehat{\eta}_1(X), \dots, \widehat{\eta}_d(X)), \\ \widehat{\eta}_j(X) = \mathbb{I}(X_j > \sqrt{(2+\delta)\log d}), \quad j = 1, \dots, d, \end{gathered} \tag{9}$$

где δ – некоторое (малое) положительное число. Если же параметр a удовлетворяет условию $\limsup_{d\to\infty} \frac{a}{\sqrt{\log d}} < \sqrt{2}(1 + \sqrt{1-\beta})$, то справедливо неравенство

$$\liminf_{\varepsilon\to 0} \inf_{\widetilde{\eta}} \sup_{\eta \in \mathcal{H}_{d,\beta}} \mathbf{E}_\eta |\widetilde{\eta} - \eta| > 0,$$

то есть, в этом случае точная селекция невозможна. Данный факт устанавливается путем оценивания снизу минимаксного риска $\inf_{\widetilde{\eta}} \sup_{\eta \in \mathcal{H}_{d,\beta}} \mathbf{E}_\eta |\widetilde{\eta} - \eta|$ байесовским риском, который, при подходящем выборе априорного распределения вектора $\eta \in \mathcal{H}_{d,\beta}$, строго положителен (см. [6, п. 4]).

§3. Основные результаты

В этом параграфе мы сформулируем условия, при которых точная селекция в функциональной задаче регрессии становится возможной, и построим селектор, обеспечивающий точную селекцию (см. теорему 1). Построенный нами селектор *адаптивен* по отношению к индексу разреженности β и неулучшаем в смысле минимаксного подхода (см. теорему 2). Вместо непрерывной регрессионной модели мы будем рассматривать более удобную с технической точки зрения эквивалентную ей модель в пространстве последовательностей. Напомним, что число переменных функции $f \in \mathcal{F}_{d,\beta,\sigma}$ предполагается растущим и удовлетворяющим условию $d = d_\varepsilon \to \infty$ при $\varepsilon \to \infty$.

3.1. Точное воспроизведение разреженного сигнала. Рассмотрим эллипсоид в пространстве $l_2(\mathbb{Z})$ с удаленной из него малой окрестностью точки $\theta = 0$:

$$\Theta_\sigma(r_\varepsilon) = \left\{\theta = (\theta_k)_{k\in\mathbb{Z}} \in l_2(\mathbb{Z}) : \sum_{k\in\mathbb{Z}} (2\pi|k|)^{2\sigma}\theta_k^2 \leqslant 1,\ \sum_{k\in\mathbb{Z}} \theta_k^2 \geqslant r_\varepsilon^2\right\}.$$

Обозначим через $\theta^*(r_\varepsilon) = (\theta_k^*(r_\varepsilon))_{k\in\mathbb{Z}}$ решение экстремальной задачи

$$\frac{1}{2\varepsilon^4}\sum_{k\in\mathbb{Z}} \theta_k^4 \to \inf_{\theta\in\Theta_\sigma(r_\varepsilon)}, \tag{10}$$

и через $u_\varepsilon^2(r_\varepsilon) = u_\varepsilon^2(\Theta_\sigma(r_\varepsilon))$ ее значение, то есть

$$u_\varepsilon^2(r_\varepsilon) = \frac{1}{2\varepsilon^4}\inf_{\theta\in\Theta_\sigma(r_\varepsilon)}\sum_{k\in\mathbb{Z}} \theta_k^4 = \frac{1}{2\varepsilon^4}\sum_{k\in\mathbb{Z}} \left(\theta_k^*(r_\varepsilon)\right)^4.$$

Функция $u_\varepsilon^2(r_\varepsilon)$ играет важную роль в минимаксной теории проверки гипотез, поэтому ее появление в задаче селекции вполне естественно. Отметим, что $u_\varepsilon(r_\varepsilon)$ является неубывающей функцией аргумента r_ε и обладает своего рода свойством “непрерывности” (см., например, [9, п. 3.2], [12, §5.2.3]). Именно, для любого $\epsilon > 0$ существуют $\Delta > 0$ и $\varepsilon_0 > 0$ такие, что для всех $\delta \in (0, \Delta)$ и $\varepsilon \in (0, \varepsilon_0)$,

$$u_\varepsilon(r_\varepsilon) \leqslant u_\varepsilon((1+\delta)r_\varepsilon) \leqslant (1+\epsilon)u_\varepsilon(r_\varepsilon).$$

Кроме того, для рассматриваемого класса σ-гладких функций известны вид экстремальной последовательности $(\theta_k^*(r_\varepsilon)_{k\in\mathbb{Z}}$ (см., например, теорему 2 в [13]) и точная асимптотика вида (см. [12, §4.3.2], а также теорему 4 в [13])

$$u_\varepsilon(r_\varepsilon) \sim C(\sigma) r_\varepsilon^{2+1/(2\sigma)} \varepsilon^{-2}, \quad \varepsilon \to 0, \tag{11}$$

где $C(\sigma) > 0$ – известная константа, зависящая от σ и выражаемая в терминах гамма-функции.

По аналогии с векторным случаем (см. п. 2.3) естественно ожидать, что точное воспроизведение разреженного сигнала $f(\mathbf{x}) = \sum_{j=1}^{d} \eta_j f_j(x_j)$ в функциональной регрессионной модели возможно при $\|f_j\|_2 \geqslant r_\varepsilon$, $j = 1, \dots, d$, где для параметра $r_\varepsilon > 0$ справедливо

$$\liminf_{\varepsilon \to 0} \frac{u_\varepsilon(r_\varepsilon)}{\sqrt{\log d}} > \sqrt{2}(1 + \sqrt{1-\beta}). \tag{12}$$

Вернемся к модели (7) в пространстве последовательностей и рассмотрим случайные величины

$$t_j = t_j(\beta) = \sum_{k \in \mathbb{Z}} \omega_k(r_\varepsilon^*(\beta)) \left[\left(\frac{X_{j,k}}{\varepsilon} \right)^2 - 1 \right], \quad j = 1, \dots, d, \tag{13}$$

где для любого $r_\varepsilon > 0$ весовые функции $\omega_k(r_\varepsilon)$ определяются формулой

$$\omega_k(r_\varepsilon) = \frac{1}{2\varepsilon^2} \frac{(\theta_k^*(r_\varepsilon))^2}{u_\varepsilon(r_\varepsilon)}, \quad k \in \mathbb{Z},$$

и удовлетворяют условию $\sum_{k \in \mathbb{Z}} \omega_k^2(r_\varepsilon) = 1/2$, а число $r_\varepsilon^*(\beta) > 0$ есть решение уравнения (ср. с (12))

$$u_\varepsilon(r_\varepsilon^*(\beta)) = \sqrt{2}(1 + \sqrt{1-\beta}) \sqrt{\log d}. \tag{14}$$

Хорошо известно, что в случае *известного* параметра β асимптотически минимаксный тест для проверки гипотезы $H_{0j} : \theta_j \equiv 0$ против альтернативы $H_{1j} : \theta_j \in \Theta_\sigma(r_\varepsilon)$ имеет вид $\psi_\varepsilon = \mathbb{I}(t_j > u_\varepsilon(r_\varepsilon)/2)$. Причем, данная задача проверки гипотез имеет смысл, когда $r_\varepsilon / r_\varepsilon^* \to \infty$, где *критический радиус* $r_\varepsilon^* \to 0$ определяется из соотношения $u_\varepsilon(r_\varepsilon^*) \asymp 1$ (эти и другие общие факты минимаксной теории проверки гипотез можно найти в монографии [12], работах [4] и [17], а также в цикле обзорных статей [9–11]). Если теперь предположить, что размерность d и радиус $r_\varepsilon > 0$ удовлетворяют соотношениям

$$\log d = o(\varepsilon^{-2/(2\sigma+1)}), \quad \liminf_{\varepsilon \to 0} \frac{u_\varepsilon(r_\varepsilon)}{\sqrt{\log d}} > \sqrt{2}(1 + \sqrt{1-\beta}),$$

то можно показать, что для некоторого $\delta > 0$ оценка (ср. с (9))

$$\widehat{\eta} = (\widehat{\eta}_1, \ldots, \widehat{\eta}_d), \quad \widehat{\eta}_j = \mathbb{I}\left(t_j > \sqrt{(2+\delta)\log d}\right), \quad j = 1, \ldots, d,$$

обеспечивает точную селекцию компонент вектора $\eta \in \mathcal{H}_{d,\beta}$, и как следствие – точное воспроизведение сигнала $f \in \mathcal{F}_{d,\beta,\sigma}$ в модели (1).

Рассмотрим теперь более реалистичную постановку задачи, когда значение индекса разреженности $\beta \in (0,1)$ неизвестно. В этом случае требуется адаптировать процедуру селекции ко всевозможным значениям β, то есть построить *адаптивную* оценку η^* вектора $\eta \in \mathcal{H}_{d,\beta}$. Для этого рассмотрим "решетку" на интервале $(0,1)$ вида

$$\beta_m = m\Delta, \quad m = 1, \ldots, M = [1/\Delta],$$

с числом узлов $M = M_d$, уловлетворяющим соотношению

$$M \to \infty, \quad \log M = o(\log d), \quad d \to \infty.$$

Пусть для каждого $m = 1, \ldots, M$ параметр $r^*_{\varepsilon,m} > 0$ определяется уравнением (ср. с (14))

$$u_\varepsilon(r^*_{\varepsilon,m}) = \sqrt{2}(1 + \sqrt{1-\beta_m})\sqrt{\log d}. \tag{15}$$

Как и в случае известного β, рассмотрим статистики типа взвешенной статистики хи-квадрат (ср. с (13))

$$t_{j,m} = \sum_{k\in\mathbb{Z}} \omega_k(r^*_{\varepsilon,m})\left[\left(\frac{X_{j,k}}{\varepsilon}\right)^2 - 1\right], \quad j = 1, \ldots, d, \quad m = 1, \ldots, M, \tag{16}$$

с весовыми функциями

$$\omega_k(r^*_{\varepsilon,m}) = \frac{1}{2\varepsilon^2}\frac{(\theta^*_k(r^*_{\varepsilon,m}))^2}{u_\varepsilon(r^*_{\varepsilon,m})}, \quad \sum_{k\in\mathbb{Z}} \omega^2_k(r^*_{\varepsilon,m}) = 1/2. \tag{17}$$

Далее, для каждого $j = 1, \ldots, d$ и $m = 1, \ldots, M$, положим

$$\eta_{j,m} = \mathbb{I}\left(t_{j,m} > \sqrt{(2+\delta)(\log d + \log M)}\right),$$

и зададим адаптивную оценку η^* вектора $\eta \in \mathcal{H}_{d,\beta}$ равенством

$$\eta^* = (\eta^*_1, \ldots, \eta^*_d), \quad \eta^*_j = \max_{1\leqslant m\leqslant M} \eta_{j,m}, \quad j = 1, \ldots, d. \tag{18}$$

Идея построенной нами оценки проста и состоит в следующем. Мы полагаем, что в j-м месте есть сигнал, если он обнаруживается хотя бы при одном β_m. Тогда вероятность пропустить сигнал будет меньше, чем при β_m близком к "правильному" значению β (она приблизительно такая же, как у статистик с неизменным порогом), а вероятность обнаружения в неправильном месте будет меньше суммы соответствующих вероятностей по всем $m = 1, \dots, M$, и мала за счет выбора порогового значения.

Прежде, чем формулировать основные результаты, введем в рассмотрение множество

$$\Theta_{\sigma,d}(r_\varepsilon) = \Big\{\theta = (\theta_1, \dots, \theta_d) : \theta_j = (\theta_{j,k}) \in l_2(\mathbb{Z}),$$
$$\sum_{k\in\mathbb{Z}} (2\pi|k|)^{2\sigma}\theta_{j,k}^2 \leqslant 1,\ \sum_{k\in\mathbb{Z}} \theta_{j,k}^2 \geqslant r_\varepsilon^2,\ 1 \leqslant j \leqslant d\Big\}.$$

Справедлива следующая теорема.

Теорема 1. *Предположим, что* $\log d = o(\varepsilon^{-2/(2\sigma+1)})$, *и что величина* $r_\varepsilon > 0$ *удовлетворяет условию*

$$\liminf_{\varepsilon\to 0} \frac{u_\varepsilon(r_\varepsilon)}{\sqrt{\log d}} > \sqrt{2}(1 + \sqrt{1-\beta}).$$

Тогда при $\varepsilon \to 0$ *справедливо соотношение*

$$\sup_{\eta\in\mathcal{H}_{d,\beta}} \sup_{\theta\in\Theta_{\sigma,d}(r_\varepsilon)} \mathbf{E}_{\eta,\theta}|\eta - \eta^*| \to 0,$$

где η^* *– оценка вектора* η, *определяемая формулой* (18).

Теорема 1 утверждает, что если одновременно для всех $j = 1, \dots, d$, гипотезы $H_{0j} : \theta_j \equiv 0$ и $H_{1j} : \theta_j \in \Theta_\sigma(r_\varepsilon)$ асимптотически различимы, то процедура селекции, основанная на η^*, выбирает в точности “правильные” компоненты вектора $\eta \in \mathcal{H}_{d,\beta}$ и таким образом обеспечивает точное воспроизведение $(\eta_1\theta_1, \dots, \eta_d\theta_d)$, равномерно по $\mathcal{H}_{d,\beta}$ и $\Theta_{\sigma,d}(r_\varepsilon)$.

Следующий результат показывает, что при нарушении условия обнаружимости точная селекция невозможна.

Теорема 2. *Предположим, что* $\log d = o(\varepsilon^{-2/(2\sigma+1)})$, *и что величина* $r_\varepsilon > 0$ *удовлетворяет условию*

$$\limsup_{\varepsilon\to 0} \frac{u_\varepsilon(r_\varepsilon)}{\sqrt{\log d}} < \sqrt{2}(1 + \sqrt{1-\beta}).$$

Тогда при $\varepsilon \to 0$ *справедливо неравенство*

$$\liminf_{\varepsilon\to 0} \inf_{\widetilde{\eta}} \sup_{\eta\in\mathcal{H}_{d,\beta}} \sup_{\theta\in\Theta_{\sigma,d}(r_\varepsilon)} \mathbf{E}_{\eta,\theta}|\eta - \widetilde{\eta}| > 0,$$

где нижняя грань берется по всевозможным оценкам $\widetilde{\eta}$ вектора η в (7).

Замечание 1. Как следует из теорем 1 и 2, точный критический радиус $r_\varepsilon^* = r_\varepsilon^*(\beta)$, наличие которого делает задачу селекции содержательной, определяется из уравнения (14), в котором функция $u_\varepsilon(r_\varepsilon)$ удовлетворяет соотношению (11).

Замечание 2. Сравнение теорем 1 и 2 с результатами работы [6], полученными для векторного случая и кратко изложенными в п. 2.3, показывает, что в функциональной задаче селекции роль параметра a принимает на себя величина $u_\varepsilon(r_\varepsilon)$, где $u_\varepsilon^2(r_\varepsilon)$ – значение экстремальной задачи (10).

§4. Доказательство теорем

4.1. Доказательство теоремы 1. Прежде чем переходить к доказательству теоремы 1, мы установим ряд свойств статистик $t_{j,m}$ (см. определение (16)) и поясним необходимость условия

$$\log d = o\left(\varepsilon^{-2/(2\sigma+1)}\right) \tag{19}$$

в утверждении теоремы 1.

Из леммы 6.1 в [5] следует, что при нулевой гипотезе $H_{0j} : \theta_j = 0$

$$\mathbf{E}_0(t_{j,m}) = 0, \quad \mathbf{Var}_0(t_{j,m}) = 1,$$

в то время как при альтернативе $H_{1j} : \theta_j \in \Theta_\sigma(r_\varepsilon)$, где при достаточно малых ε малый параметр $r_\varepsilon > 0$ удовлетворяет условию $r_\varepsilon / r_{\varepsilon,m}^* > 1$, мы имеем

$$\begin{gathered}\mathbf{E}_{\theta_j}(t_{j,m}) = \varepsilon^{-2} \sum_{k\in\mathbb{Z}} \omega_k(r_{\varepsilon,m}^*)\theta_{j,k}^2 \geqslant u_\varepsilon(r_{\varepsilon,m}^*),\\ \mathbf{Var}_{\theta_j}(t_{j,m}) = 1 + O(\mathbf{E}_{\theta_j}(t_{j,m}) \max_{k\in\mathbb{Z}} \omega_k(r_{\varepsilon,m}^*)).\end{gathered} \tag{20}$$

Более того, при указанных ограничениях на r_ε and $r_{\varepsilon,m}^*$ справедлив следующий результат.

Лемма 3. *Пусть величина $T = T_\varepsilon \to -\infty$ и весовые функции $\omega_k(r_{\varepsilon,m}^*)$, определеляемые равенством* (17), *таковы, что при $\varepsilon \to 0$*

$$T \max_{k\in\mathbb{Z}} \omega_k(r_{\varepsilon,m}^*) \to 0 \quad \text{и} \quad \mathbf{E}_{\theta_j}(t_{j,m}) \max_{k\in\mathbb{Z}} \omega_k(r_{\varepsilon,m}^*) \to 0. \tag{21}$$

Тогда при $\varepsilon \to 0$

$$\mathbf{P}_0(t_{j,m} \leqslant T) \leqslant \exp\left(-\frac{T^2}{2}(1+o(1))\right),$$

и для любого $j = 1, \dots, d$, *равномерно по* $\theta_j \in \Theta_\sigma(r_\varepsilon)$,

$$\mathbf{P}_{\theta_j}(t_{j,m} - \mathbf{E}_{\theta_j}(t_{j,m}) \leqslant T) \leqslant \exp\left(-\frac{T^2}{2}(1+o(1))\right).$$

Доказательство этой важной леммы повторяет доказательство предложения 6.1 в работе [5] и потому опускается. При использовании леммы 1 в доказательстве теоремы 1 величина T оказывается порядка $O(\sqrt{\log d})$. Поэтому и в силу (20), оба условия в (21) выполняются, когда

$$\sqrt{\log d}\, \max_{k \in \mathbb{Z}} \omega_k(r^*_{\varepsilon,m}) \to 0, \quad \varepsilon \to 0. \tag{22}$$

При нашем определении $\omega_k(r^*_{\varepsilon,m})$, которое зависит от рассматриваемого в данной работе класса гладких функций $\mathcal{F}_\sigma$, условие (22) выполняется, когда размерность $d = d_\varepsilon$ растет согласно (19).

Действительно, нам нужно удовлетворить требованию (22), в котором

$$\omega_k(r^*_{\varepsilon,m}) = \frac{1}{2\varepsilon^2}\frac{(\theta^*_{j,k}(r^*_{\varepsilon,m}))^2}{u_\varepsilon(r^*_{\varepsilon,m})} =: \frac{(v^*_{j,k}(r^*_{\varepsilon,m}))^2}{2u_\varepsilon(r^*_{\varepsilon,m})},$$

где $(\theta^*_{j,k}(r_\varepsilon))_{k\in\mathbb{Z}}$ – экстремальная последовательность в задаче (одной и той же для всех $j = 1, \dots, d$)

$$\frac{1}{2\varepsilon^4}\sum_{k\in\mathbb{Z}} \theta^4_{j,k} \to \inf_{\theta_j \in \Theta_\sigma(r_\varepsilon)},$$

а $u^2_\varepsilon(r_\varepsilon)$ – значение этой задачи. Чтобы иметь возможность осуществить точную селекцию, необходимо, чтобы гипотезы $H_{0j} : \theta_j \equiv 0$ и $H_{1j} : \theta_j \in \Theta_\sigma(r_\varepsilon)$ были асимптотически различимы одновременно для всех $j = 1, \dots, d$. Из общей минимаксной теории проверки гипотез известно, что $u^2_\varepsilon(r_\varepsilon)$ определяет пороговое значение оптимального теста в данной задаче. Известно также (см., например, [13, п. 2]), что

$$(v^*_{j,k}(r_\varepsilon))^2 \asymp r_\varepsilon^{2+1/\sigma}\varepsilon^{-2}, \quad u_\varepsilon(r_\varepsilon) \asymp r_\varepsilon^{2+1/(2\sigma)}\varepsilon^{-2}, \tag{23}$$

и потому для всех $k \in \mathbb{Z}$

$$\omega_k(r_\varepsilon) = \frac{(v^*_{j,k}(r_\varepsilon))^2}{2u_\varepsilon(r_\varepsilon)} \asymp \frac{r_\varepsilon^{2+1/\sigma}\varepsilon^{-2}}{r_\varepsilon^{2+1/(2\sigma)}\varepsilon^{-2}} = r_\varepsilon^{1/(2\sigma)}. \tag{24}$$

Далее, гипотезы H_{0j} и H_{1j} асимптотически различимы, когда $r_\varepsilon/r_\varepsilon^* \to \infty$ (или, при наличии точного критического радиуса r_ε^*, когда $\liminf r_\varepsilon/r_\varepsilon^* > 1$), где критический радиус r_ε^* удовлетворяет соотношению $u_\varepsilon(r_\varepsilon^*) \asymp (r_\varepsilon^*)^{2+1/(2\sigma)}\varepsilon^{-2}$ с одной стороны (см. (23)) и соотношению $u_\varepsilon(r_\varepsilon^*) \asymp u_\varepsilon(r_{\varepsilon,m}^*) \asymp \log^{1/2} d$ с другой стороны (см. (15) и замечание 1). Поэтому, используя упомянутое выше свойство непрерывности функции $u_\varepsilon(r_\varepsilon)$, находим

$$r_\varepsilon^* \asymp r_{\varepsilon,m}^* \asymp \left(\varepsilon^4 \log d\right)^{\sigma/(4\sigma+1)}. \tag{25}$$

Осталось заметить, что (22) и (24) позволяют получить соотношение $(r_{\varepsilon,m}^*)^{1/(2\sigma)} \log^{1/2} d = o(1)$, которое, ввиду (25), эквивалентно $\varepsilon^{2/(4\sigma+1)}(\log d)^{(2\sigma+1)/(4\sigma+1)} = o(1)$, что и приводит к (19).

Доказательство теоремы 1. По определению селектора η^* имеем

$$\begin{aligned}
&\sup_{\eta\in\mathcal{H}_{d,\beta}} \sup_{\eta\in\Theta_{\sigma,d}(r_\varepsilon)} \mathbf{E}_{\eta,\theta}|\eta-\eta^*| = \sup_{\eta\in\mathcal{H}_{d,\beta}} \sup_{\eta\in\Theta_{\sigma,d}(r_\varepsilon)} \mathbf{E}_{\eta,\theta} \sum_{j=1}^{d} |\eta_j-\eta_j^*| \\
&= \sup_{\eta\in\mathcal{H}_{d,\beta}} \sup_{\eta\in\Theta_{\sigma,d}(r_\varepsilon)} \mathbf{E}_{\eta,\theta}\left(\sum_{j:\eta_j=0}\eta_j^* + \sum_{j:\eta_j=1}(1-\eta_j^*)\right) \\
&= \sup_{\eta\in\mathcal{H}_{d,\beta}}\left(\sum_{j:\eta_j=0}\mathbf{E}_0(\eta_j^*) + \sum_{j:\eta_j=1}\sup_{\theta_j\in\Theta_\sigma(r_\varepsilon)}\mathbf{E}_{\theta_j}(1-\eta_j^*)\right) \\
&= \sup_{\eta\in\mathcal{H}_{d,\beta}}\left(\sum_{j:\eta_j=0}\mathbf{E}_0\left[\max_{1\leqslant m\leqslant M}\mathbb{I}\left(t_{j,m} > \sqrt{(2+\delta)(\log d+\log M)}\right)\right]\right. \\
&\left.+\sum_{j:\eta_j=1}\sup_{\theta_j\in\Theta_\sigma(r_\varepsilon)}\mathbf{E}_{\theta_j}\left[1-\max_{1\leqslant m\leqslant M}\mathbb{I}\left(t_{j,m} > \sqrt{(2+\delta)(\log d+\log M)}\right)\right]\right) \\
&= (d-d^{1-\beta})\mathbf{E}_0\left[\max_{1\leqslant m\leqslant M}\mathbb{I}\left(t_{1,m} > \sqrt{(2+\delta)(\log d+\log M)}\right)\right] \\
&+ d^{1-\beta}\sup_{\theta_1\in\Theta_\sigma(r_\varepsilon)}\mathbf{E}_{\theta_1}\left[1-\max_{1\leqslant m\leqslant M}\mathbb{I}\left(t_{1,m} > \sqrt{(2+\delta)(\log d+\log M)}\right)\right] \\
&=: I_1 + I_2.
\end{aligned} \tag{26}$$

Для первого слагаемого, используя вариант леммы 1 с $T = T_\varepsilon \to \infty$, при $d \to \infty$ получаем следующую оценку сверху:

$$
\begin{aligned}
I_1 &\leqslant d \sum_{m=1}^{M} \mathbf{E}_0 \left[\mathbb{I} \left(t_{1,m} > \sqrt{(2+\delta)(\log d + \log M)} \right) \right] \\
&\leqslant Md \max_{1 \leqslant m \leqslant M} \mathbf{P}_0 \left(t_{1,m} > \sqrt{(2+\delta)(\log d + \log M)} \right) \\
&= O(Md) \exp\left(-\frac{2+\delta}{2} (\log d + \log M) \right) = O((Md)^{-\delta/2}) = o(1).
\end{aligned} \tag{27}
$$

Рассмотрим второе слагаемое и заметим, что

$$
1 - \max_{1 \leqslant m \leqslant M} \mathbb{I} \left(t_{1,m} > \sqrt{(2+\delta)(\log d + \log M)} \right)
$$

равняется единице, когда $t_{1,m} \leqslant \sqrt{(2+\delta)(\log d + \log M)}$ для всех $m = 1, \dots, M$, и нулю в противном случае. Поэтому

$$
\begin{aligned}
I_2 &= d^{1-\beta} \sup_{\theta_1 \in \Theta_\sigma(r_\varepsilon)} \mathbf{P}_{\theta_1} \left(\bigcap_{m=1}^{M} \left\{ t_{1,m} \leqslant \sqrt{(2+\delta)(\log d + \log M)} \right\} \right) \\
&\leqslant d^{1-\beta} \sup_{\theta_1 \in \Theta_\sigma(r_\varepsilon)} \min_{1 \leqslant m \leqslant M} \mathbf{P}_{\theta_1} \left(t_{1,m} \leqslant \sqrt{(2+\delta)(\log d + \log M)} \right) \\
&= d^{1-\beta} \sup_{\theta_1 \in \Theta_\sigma(r_\varepsilon)} \min_{1 \leqslant m \leqslant M} \mathbf{P}_{\theta_1} \Big(t_{1,m} - \mathbf{E}_{\theta_1}(t_{1,m}) \\
&\qquad\qquad \leqslant \sqrt{(2+\delta)(\log d + \log M)} - \mathbf{E}_{\theta_1}(t_{1,m}) \Big).
\end{aligned} \tag{28}
$$

Далее, чтобы применить лемму 1 к правой части (28), нужно показать, что для некоторого индекса m, такого что при всех достаточно малых ε, выполняется неравенство $r_\varepsilon / r^*_{\varepsilon,m} > 1$, справедливо соотношение

$$
\sqrt{(2+\delta)(\log d + \log M)} - \inf_{\theta_1 \in \Theta_\sigma(r_\varepsilon)} \mathbf{E}_{\theta_1}(t_{1,m}) \to -\infty, \quad \varepsilon \to 0. \tag{29}
$$

По условию, для некоторого $\delta > 0$ (можно считать, что $\delta = o(1)$)

$$
u_\varepsilon(r_\varepsilon) \geqslant \sqrt{2}(1 + \sqrt{1-\beta}) \sqrt{\log d}\, (1+\delta).
$$

Следовательно, в силу “непрерывности” $u_\varepsilon(r_\varepsilon)$, найдется близкое к β значение β_m, определяющее $r^*_{\varepsilon,m}$, такое что (см. (15))

$$
\begin{aligned}
u_\varepsilon(r^*_{\varepsilon,m}) &= \sqrt{2}(1 + \sqrt{1-\beta}) \sqrt{\log d}\, (1 + o(1)) \\
&\geqslant \sqrt{2}(1 + \sqrt{1-\beta}) \sqrt{\log d}\, (1 + \delta + o(1)).
\end{aligned}
$$

Таким образом, для некоторого числа $\delta_1 > 0$,

$$r_\varepsilon \geqslant (1+\delta_1) r^*_{\varepsilon,m} =: B r^*_{\varepsilon,m}, \quad B > 1.$$

Далее, в силу предложения 3.1 из [5], при некотором $\delta_2 > 0$, таком что $B^2(1+o(1)) > 1+\delta_2$, получаем, используя (20),

$$\inf_{\theta_1\in\Theta_\sigma(r_\varepsilon)} \mathbf{E}_{\theta_1}(t_{1,m}) \geqslant \inf_{\theta_1\in\Theta_\sigma(Br^*_{\varepsilon,m})} \varepsilon^{-2} \sum_{k\in\mathbb{Z}} \omega_k(r^*_{\varepsilon,m})\theta^2_{1,k}$$
$$\geqslant B^2 \inf_{\theta_1\in\Theta_\sigma(r^*_{\varepsilon,m})} \varepsilon^{-2} \sum_{k\in\mathbb{Z}} \omega_k(r^*_{\varepsilon,m})\theta^2_{1,k} = B^2 \inf_{\theta_1\in\Theta_\sigma(r^*_{\varepsilon,m})} \mathbf{E}_{\theta_1}(t_{1,m}) = B^2 u_\varepsilon(r^*_{\varepsilon,m})$$
$$= B^2\sqrt{2}(1+\sqrt{1-\beta})\sqrt{\log d}\,(1+o(1)) > \sqrt{2}(1+\sqrt{1-\beta})\sqrt{\log d}\,(1+\delta_2).$$

Отсюда, для некоторого малого числа $\delta > 0$ при достаточно малых ε (и следовательно при достаточно больших d),

$$\inf_{\theta_1\in\Theta_\sigma(r_\varepsilon)} \mathbf{E}_{\theta_1}(t_{1,m}) > \sqrt{2}(1+\sqrt{1-\beta})\sqrt{\log d}(1+\delta_2)$$
$$> \sqrt{(2+\delta)(\log d + \log M)},$$

что и доказывает (29).

Поэтому, применяя лемму 1 в правой части неравенства (28), мы получаем при $d\to\infty$ (и малом $\delta > 0$)

$$I_2 \leqslant d^{1-\beta} \exp\left(-\frac{\left(\sqrt{(2+\delta)} - \sqrt{2}(1+\sqrt{1-\beta})(1+\delta_2)\right)^2}{2} \log d\,(1+o(1))\right)$$
$$= O\left(d^{-\beta-\delta/2-(1+\sqrt{1-\beta})^2(1+\delta_2)^2+\sqrt{2(2+\delta)}(1+\sqrt{1-\beta})(1+\delta_2)}\right) = o(1). \qquad (30)$$

Подставляя (27) и (30) в (26), выводим требуемый результат:

$$\sup_{\eta\in\mathcal{H}_{d,\beta}} \sup_{\theta\in\Theta_{\sigma,d}(r_\varepsilon)} \mathbf{E}_{\eta,\theta}|\eta-\eta^*| \leqslant I_1 + I_2 = o(1), \quad \varepsilon\to 0.$$

Теорема 1 доказана. □

4.2. Доказательство теоремы 2. При доказательстве теоремы 2 будем предполагать, что η_j и θ_j, $j = 1,\ldots,d$, есть реализации независимых случайных величин и последовательностей, соответственно. При таком байесовском подходе и подходящем выборе априорных распределений, минимаксный риск

$$R_\varepsilon = \inf_{\widetilde{\eta}} \sup_{\eta\in\mathcal{H}_{d,\beta}} \sup_{\theta\in\Theta_{\sigma,d}(r_\varepsilon)} \mathbf{E}_{\eta,\theta}|\eta-\widetilde{\eta}|$$

будет не меньше байесовского риска, для которого, в свою очередь, существует положительная нижняя грань. Как и в теореме 1, ограничение на рост размерности d играет здесь ключевую роль. Оно гарантирует, что нормированная экстремальная последовательность $(v^*_{j,k})_{k\in\mathbb{Z}}$, определяемая формулой (32), удовлетворяет важному для нашего доказательства условию $v^*_{j,k} = o(1)$ при $\varepsilon \to 0$.

Доказательство теоремы 2. Как и прежде, обозначим через $\theta^*_j = (\theta^*_{j,k})_{k\in\mathbb{Z}}$ экстремальную последовательность в задаче (одной и той же для всех $j = 1, \dots, d$)

$$\frac{1}{2\varepsilon^4}\sum_{k\in\mathbb{Z}}\theta^4_{j,k} \to \inf_{\theta_j\in\Theta_\sigma(r_\varepsilon)}.$$

Пусть априорное распределение "вектора" $\theta = (\theta_1, \dots, \theta_d) \in \Theta_{\sigma,d}(r_\varepsilon)$ имеет вид

$$\pi_\theta(d\theta) = \prod_{j=1}^{d}\pi_{\theta_j}(d\theta_j), \quad \pi_{\theta_j}(d\theta_j) = \prod_{k\in\mathbb{Z}}\left(\frac{\delta_{-\theta^*_{j,k}} + \delta_{\theta^*_{j,k}}}{2}\right)(d\theta_{j,k}),$$

где δ_x есть δ-мера с единичной массой в точке x. Обозначим теперь через

$$p = d^{1-\beta}/d = d^{-\beta}$$

долю ненулевых компонент вектора $\eta = (\eta_1, \dots, \eta_d) \in \mathcal{H}_{d,\beta}$. Априорное распределение η естественно положить равным

$$\pi_\eta(d\eta) = \prod_{j=1}^{d}\pi_{\eta_j}(d\eta_j), \quad \pi_{\eta_j}(d\eta_j) = ((1-p)\delta_0 + p\delta_1)(d\eta_j).$$

Тогда, в предположении независимости $\theta = (\theta_j)$ и $\eta = (\eta_j)$, справедливо неравенство

$$\begin{aligned}
R_\varepsilon &= \inf_{\widetilde{\eta}}\sup_{\eta\in\mathcal{H}_{d,\beta}}\sup_{\theta\in\Theta_{\sigma,d}(r_\varepsilon)}\mathbf{E}_{\eta,\theta}|\eta-\widetilde{\eta}| \geqslant \inf_{\widetilde{\eta}}\mathbf{E}_{\pi_\eta}\mathbf{E}_{\pi_\theta}\mathbf{E}_{\eta,\theta}|\eta-\widetilde{\eta}| \\
&= \inf_{\widetilde{\eta}}\mathbf{E}_{\pi_\eta}\mathbf{E}_{\pi_\theta}\mathbf{E}_{\eta,\theta}\sum_{j=1}^{d}|\eta_j-\widetilde{\eta}_j| \\
&= \inf_{\widetilde{\eta}}\sum_{j=1}^{d}\mathbf{E}_{\pi_{\eta_j}}\mathbf{E}_{\pi_{\theta_j}}\mathbf{E}_{\eta_j\theta_j}|\eta_j-\widetilde{\eta}_j| =: R_{B,\varepsilon},
\end{aligned}$$

где $\mathbf{E}_{\eta_j\theta_j}$ – математическое ожидание по мере $\mathbf{P}_{\eta_j\theta_j}$, индуцированной наблюдением $X_j = (X_{j,k})_{k\in\mathbb{Z}}$, состоящим из независимых случайных величин $X_{j,k}$, $k \in \mathbb{Z}$, распределенных по нормальному закону $N(\eta_j\theta_{j,k}, \varepsilon^2)$, а $R_{B,\varepsilon}$ – байесовский риск по отношению к введенным априорным распределениям.

Рассмотрим смесь по распределению $\mathbf{P}_{\eta_j\theta_j}$, определяемую формулой

$$\begin{aligned}\mathbf{P}_{\pi,\eta_j}(dX_j) &= \mathbf{E}_{\pi_{\theta_j}}\mathbf{P}_{\eta_j\theta_j}(dX_{j,k})\\ &= \prod_{k\in\mathbb{Z}}\left(\frac{N(-\eta_j\theta^*_{j,k},\varepsilon^2)+N(\eta_j\theta^*_{j,k},\varepsilon^2)}{2}\right)(dX_{j,k}).\end{aligned}\tag{31}$$

В частности, при $\eta_j = 0$, $\mathbf{P}_{\pi,0}(dX_j) = \prod_{k\in\mathbb{Z}} N(0,\varepsilon^2)(dX_{j,k})$. Используя обозначение

$$v^*_{j,k} = \frac{\theta^*_{j,k}}{\varepsilon},\tag{32}$$

мы получаем относительно вероятностной меры $\mathbf{P}_{\pi,\eta_j}$

$$Y_{j,k} := \frac{X_{j,k}}{\varepsilon} = \eta_j v^*_{j,k} + \xi_{j,k} \overset{\text{ind.}}{\sim} \mathcal{N}(\eta_j v^*_{j,k}, 1), \quad 1 \leqslant j \leqslant d, \quad k \in \mathbb{Z}.$$

Далее, обозначая $Y_j = (Y_{j,k})_{k\in\mathbb{Z}}$, переписываем отношение правдоподобия в виде

$$\frac{d\mathbf{P}_{\pi,\eta_j}}{d\mathbf{P}_{\pi,0}}(Y_j) = \prod_{k\in\mathbb{Z}}\exp\left(-\frac{\eta_j(v^*_{j,k})^2}{2}\right)\cosh\left(\eta_j v^*_{j,k}Y_{j,k}\right).\tag{33}$$

Отсюда, принимая во внимание тот факт, что η_j принимает лишь два значения, ноль и 1 с вероятностями $(1-p)$ и p, соответственно, продолжим нашу оценку снизу:

$$\begin{aligned}R_\varepsilon \geqslant R^*_\varepsilon &= \sum_{j=1}^{d}\inf_{\widetilde{\eta}_j}\mathbf{E}_{\pi_{\eta_j}}\mathbf{E}_{\pi,\eta_j}|\eta_j-\widetilde{\eta}_j|\\ &= \sum_{j=1}^{d}\inf_{\widetilde{\eta}_j}\left[(1-p)\mathbf{E}_{\pi,0}(\widetilde{\eta}_j)+p\mathbf{E}_{\pi,1}(1-\widetilde{\eta}_j)\right],\end{aligned}$$

где $\inf_{\widetilde{\eta}_j}(1-p)\mathbf{E}_{\pi,0}(\widetilde{\eta}_j) + p\mathbf{E}_{\pi,1}(1-\widetilde{\eta}_j)$ – байесовский риск в задаче проверки двух простых гипотез

$$H_0 : \mathbf{P} = \mathbf{P}_{\pi,0}, \quad H_1 : \mathbf{P} = \mathbf{P}_{\pi,1},$$

с вероятностными мерами $\mathbf{P}_{\pi,0}$ and $\mathbf{P}_{\pi,1}$, определенными в (31). В частности, при нулевой гипотезе, вектор $Y_j = (Y_{j,k})_{k\in\mathbb{Z}}$ имеет нормальное распределение с плотностью $p_{\pi,0}(t) = \prod_{k\in\mathbb{Z}} (2\pi)^{-1/2} \exp(-t_k^2/2\varepsilon^2)$, $t = (t_k)_{k\in\mathbb{Z}}$. Согласно формуле (33), отношением правдоподобия в данной задаче является статистика

$$\Lambda_\pi = \Lambda_\pi(Y_j) = \frac{d\mathbf{P}_{\pi,1}}{d\mathbf{P}_{\pi,0}}(Y_j) = \prod_{k\in\mathbb{Z}} \exp\left(-\frac{(v_{j,k}^*)^2}{2}\right) \cosh\left(v_{j,k}^* Y_{j,k}\right),$$

и оптимальный (байесовский) тест η_B имеет вид (см., например, [3, § 8.11])

$$\eta_B(Y_j) = \mathbb{I}\left(\Lambda_\pi(Y_j) \geqslant \frac{1-p}{p}\right).$$

Используя полученный результат, мы выводим

$$R_\varepsilon \geqslant R_{B,\varepsilon} = d\left[(1-p)\mathbf{P}_{\pi,0}\left(\Lambda_\pi(Y_j) \geqslant \frac{1-p}{p}\right) + p\mathbf{P}_{\pi,1}\left(\Lambda_\pi(Y_j) < \frac{1-p}{p}\right)\right].$$

Таким образом, для доказательства теоремы 2 нужно показать, что для некоторого числа $c > 0$ при достаточно малых ε байесовский риск $R_{B,\varepsilon}$ удовлетворяет условию $R_{B,\varepsilon} \geqslant c$. Последнее неравенство заведомо выполняется, когда для некоторой константы $C > 0$ при достаточно больших d справедливо неравенство

$$d(1-p)\mathbf{P}_{\pi,0}\left(\Lambda_\pi(Y_j) \geqslant \frac{1-p}{p}\right) \geqslant C, \tag{34}$$

к проверке которого мы и переходим.

Введем случайные величины

$$\lambda_\pi = \lambda_\pi(Y_j) = \log \Lambda_\pi(Y_j) = \sum_{k\in\mathbb{Z}} \lambda_{\pi,k}(Y_{j,k}), \tag{35}$$

$$\lambda_{\pi,k} = \lambda_{\pi,k}(Y_{j,k}) = -\frac{(v_{j,k}^*)^2}{2} + \log\cosh(v_{j,k}^* Y_{j,k}), \tag{36}$$

и, обозначая

$$H := \log \frac{1-p}{p} \sim \beta \log d,$$

перепишем интересующую нас вероятность в виде

$$\begin{aligned}\mathbf{P}_{\pi,0}\left(\Lambda_\pi > \frac{1-p}{p}\right) &= \mathbf{P}_{\pi,0}(\lambda_\pi > H) = \mathbf{E}_{\pi,0}\mathbb{I}(\lambda_\pi > H) \\ &= \mathbf{E}_{\pi,0}\left(e^{h\lambda_\pi} e^{-h\lambda_\pi}\mathbb{I}(\lambda_\pi - H > 0)\right).\end{aligned}$$

Далее, используя для краткости вместо $\mathbf{P}_{\pi,0}$ обозначение $\mathbf{P}_0$, введем вероятностную меру $\mathbf{P}_h$, зависщую от параметра $h > 0$, следующим образом:

$$\frac{d\mathbf{P}_h}{d\mathbf{P}_0}(Y_j) := \frac{e^{h\lambda_\pi(Y_j)}}{\Psi(h)}, \quad \Psi(h) = \mathbf{E}_{\mathbf{P}_0} e^{h\lambda_\pi(Y_j)}.$$

Мы имеем

$$\begin{aligned}\mathbf{P}_0\left(\Lambda_\pi > \frac{1-p}{p}\right) &= \mathbf{E}_{\mathbf{P}_h}\left(\frac{d\mathbf{P}_0}{d\mathbf{P}_h}(Y_j) e^{h\lambda_\pi} e^{-h\lambda_\pi}\mathbb{I}(\lambda_\pi - H > 0)\right) \\ &= \Psi(h)\mathbf{E}_{\mathbf{P}_h}\left[e^{-h\lambda_\pi}\mathbb{I}(\lambda_\pi - H > 0)\right] \\ &= \Psi(h)e^{-hH}\mathbf{E}_{\mathbf{P}_h}\left[e^{-h(\lambda_\pi - H)}\mathbb{I}(\lambda_\pi - H > 0)\right] \\ &=: \Psi(h)e^{-hH}J(h).\end{aligned}$$

Отсюда получаем, что

$$\log \mathbf{P}_0\left(\Lambda_\pi \geqslant \frac{1-p}{p}\right) = \log\Psi(h) - hH + \log J(h). \tag{37}$$

Мы покажем, что при надлежащем выборе параметра $h = h_\varepsilon > 0$ и при достаточно малых ε слагаемое $\log J(h)$ мало по сравнению с $\log\Psi(h)$ и hH.

Именно, предположим, что h удовлетворяет равенству

$$\mathbf{E}_{\mathbf{P}_h}\lambda_\pi = H. \tag{38}$$

Тогда для случайной величины

$$Z_h := \frac{\lambda_\pi - H}{\sigma_h}, \quad \sigma_h^2 = \mathbf{Var}_{\mathbf{P}_h}(\lambda_\pi), \tag{39}$$

справедливо

$$\mathbf{E}_{\mathbf{P}_h}(Z_h) = 0, \quad \mathbf{Var}_{\mathbf{P}_h}(Z_h) = 1.$$

Далее рассмотрим вероятностные меры $\mathbf{P}_{h,k}$, $k \in \mathbb{Z}$, определяемые соотношением

$$\frac{d\mathbf{P}_h}{d\mathbf{P}_0}(Y_j) = \frac{e^{h\sum_k \lambda_{\pi,k}(Y_{j,k})}}{\mathbf{E}_{\mathbf{P}_0} e^{h\sum_k \lambda_{\pi,k}(Y_{j,k})}} = \prod_{k\in\mathbb{Z}} \frac{e^{h\lambda_{\pi,k}(Y_{j,k})}}{\mathbf{E}_{\mathbf{P}_0} e^{h\lambda_{\pi,k}(Y_{j,k})}} =: \prod_{k\in\mathbb{Z}} \frac{d\mathbf{P}_{h,k}}{d\mathbf{P}_0}(Y_{j,k}),$$

где $\lambda_{\pi,k} = \lambda_{\pi,k}(Y_{j,k}) = -(v_{j,k}^*)^2/2 + \log\cosh(v_{j,k}^* Y_{j,k})$, а величины $v_{j,k}^*$ заданы в (32). Имеет место следующий результат, справедливый при достаточно малых ε.

Лемма 4. *Для* $\mathbf{P}_{h,k}$*-среднего и дисперсии статистик* $\lambda_{\pi,k}$ *имеем*

$$\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}) = \frac{(v_{j,k}^*)^4}{2}(h-1/2) + o\left((v_{j,k}^*)^4\right),$$
$$\mathbf{Var}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}) = \frac{(v_{j,k}^*)^4}{2} + o\left((v_{j,k}^*)^4\right),$$

и, следовательно,

$$\mathbf{E}_{\mathbf{P}_h}(\lambda_\pi) = u_\varepsilon^2(r_\varepsilon)(h-1/2)(1+o(1)),$$
$$\sigma_h^2 = \mathbf{Var}_{\mathbf{P}_h}(\lambda_\pi) = u_\varepsilon^2(r_\varepsilon)(1+o(1)).$$

Используя методы доказательства леммы 2, можно показать, что случайная величина Z_h в определении (39) асимптотически нормальна (см. лемму 3). Доказательство обоих лемм дано в Приложении.

Лемма 5. *При* $\varepsilon \to 0$ *выполняется условие Ляпунова, т.е.*

$$L_h^{(4)} = \frac{\sum_k \mathbf{E}_{\mathbf{P}_{h,k}}\left(\lambda_{\pi,k} - \mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^4}{\left(\sum_k \mathbf{Var}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^2} \to 0,$$

и, следовательно, в пределе Z_h *имеет стандартное нормальне распределение.*

Так как параметр h выбран таким образом, что справедливо (см. (38) и лемму 2)

$$\mathbf{E}_{\mathbf{P}_h}(\lambda_\pi) = (h-1/2)\, u_\varepsilon^2(r_\varepsilon)(1+o(1)) = H \sim \beta \log d,$$

то при $\varepsilon \to 0$

$$h \sim \frac{1}{2} + \frac{\beta \log d}{u_\varepsilon^2(r_\varepsilon)}. \tag{40}$$

А поскольку по условию теоремы справедливо соотношение $u_\varepsilon^2(r_\varepsilon) \asymp \log d$, то из (40) находим, что

$$h = O(1). \tag{41}$$

Вернемся к (37) и рассмотрим функцию

$$J(h) = \mathbf{E}_{\mathbf{P}_h}\left[e^{-h\sigma_h Z_h}\mathbb{I}(Z_h > 0)\right].$$

Очевидно, что

$$\mathbf{P}_h(Z_h \in [0,\delta])e^{-h\sigma_h\delta} \leqslant J(h) \leqslant 1,$$

где по лемме 3 при достаточно малых δ

$$\mathbf{P}_h(Z_h \in [0,\delta]) \sim \delta/\sqrt{2\pi} \geqslant C > 0.$$

Отсюда получаем

$$-h\sigma_h\delta + \log\delta - \frac{1}{2}\log(2\pi) \leqslant \log J(h) \leqslant 0, \tag{42}$$

где в силу (41) и леммы 2,

$$h\sigma_h \asymp u_\varepsilon(r_\varepsilon) \asymp \sqrt{H} \asymp \sqrt{\log d}, \tag{43}$$

и потому $h\sigma_h \to \infty$ при $\varepsilon \to 0$. Следовательно, при подходящем выборе малого δ в левой части (42), мы приходим к соотношению $\log J(h) = o(h\sigma_h) = o(\sqrt{H})$. Откуда, учитывая (37), получаем

$$\log \mathbf{P}_0\left(\Lambda_\pi \geqslant \frac{1-p}{p}\right) = \log\Psi(h) - hH + o(\sqrt{H}). \tag{44}$$

Обратимся теперь к функции

$$\Psi(h) = \mathbf{E}_{\mathbf{P}_0} e^{h\lambda_\pi} = \prod_{k\in\mathbb{Z}} \mathbf{E}_{\mathbf{P}_0} e^{h\lambda_{\pi,k}} =: \prod_{k\in\mathbb{Z}} \Psi_k(h),$$

где λ_π и $\lambda_{\pi,k}$ определяются соответственно выражениями (35) и (36). Заметим, что при $\varepsilon \to 0$ имеет место $h = O(1)$ (см. (41)) и, кроме того,

$$v^*_{j,k} = o(1).$$

Действительно, мы знаем, что гипотезы H_{0j} и H_{1j} асимптотически неразличимы, когда $r_\varepsilon/r^*_\varepsilon \to 0$ (или, при наличие точного критического радиуса r^*_ε, когда $\limsup r_\varepsilon/r^*_\varepsilon < 1$). Поэтому равенство $v^*_{j,k} = o(1)$ при $\varepsilon \to 0$ есть прямое следствие условий теоремы 2, "непрерывности" функции $u_\varepsilon(r_\varepsilon)$ и соотношений (23) и (25).

Покажем, что при $h = O(1)$ и $v^*_{j,k} = o(1)$,

$$\Psi(h) = \prod_{k\in\mathbb{Z}}\left[\frac{h^2-h}{4}(v^*_{j,k})^4 + o\left((v^*_{j,k})^4\right)\right]$$
$$= \exp\left(\frac{h^2-h}{2}\, u^2_\varepsilon(r_\varepsilon)(1+o(1))\right). \tag{45}$$

Для этого рассмотрим функцию

$$\widetilde{p}(v,y) = p(v,y) + \frac{v^4}{4}, \quad p(v,y) = -\frac{v^2}{2} + \log\cosh(vy),$$

относительно которой известно (см. формулу (6.6) в [1]), что если $h = O(1)$ и $v \to 0$, то

$$\mathbf{E}_{\mathbf{P}_0} e^{h\widetilde{p}(v, Y_{j,k})} = \exp\left(\frac{h^2 v^4}{2} + o(v^4)\right).$$

Поэтому мы имеем

$$\Psi_k(h) = \mathbf{E}_{\mathbf{P}_0}\left(e^{h\widetilde{p}(v^*_{j,k}, Y_{j,k})} e^{-h(v^*_{j,k})^4/4}\right)$$
$$= \exp\left(\frac{h^2 - h}{4}(v^*_{j,k})^4 + o\left((v^*_{j,k})^4\right)\right), \quad (46)$$

и, как следствие, равенство (45).

Завершая доказательство, отметим (см. (43)), что $u_\varepsilon^2 = u_\varepsilon^2(r_\varepsilon) \asymp H \asymp \log d$. Поэтому, подставляя (40) и (45) в (44), находим, что при всех достаточно малых ε

$$\log \mathbf{P}_0\left(\Lambda_\pi \geqslant \frac{1-p}{p}\right) = \frac{h^2 - h}{2}\, u_\varepsilon^2 + o(H) - hH + o(\sqrt{H})$$
$$= \left(-\frac{1}{8} + \frac{H^2}{2u_\varepsilon^4}\right) u_\varepsilon^2 - \frac{H}{2} - \frac{H^2}{u_\varepsilon^2} + o(H),$$

откуда

$$\log \mathbf{P}_0\left(\Lambda_\pi \geqslant \frac{1-p}{p}\right) = -\frac{u_\varepsilon^2}{8} - \frac{H^2}{2u_\varepsilon^2} - \frac{H}{2} + o(H),$$

и, следовательно, при достаточно малых ε

$$\mathbf{P}_0\left(\Lambda_\pi \geqslant \frac{1-p}{p}\right) \sim \exp\left(-\frac{u_\varepsilon^2}{8} - \frac{H^2}{2u_\varepsilon^2} - \frac{H}{2}\right) = \exp\left(-\frac{(H + u_\varepsilon^2/2)^2}{2u_\varepsilon^2}\right).$$

А так как условие $\limsup_{\varepsilon \to 0} \dfrac{u_\varepsilon(r_\varepsilon)}{\sqrt{\log d}} < \sqrt{2}(1 + \sqrt{1-\beta})$, влечет при достаточно малых ε неравенство

$$d \exp\left(-\frac{(H + u_\varepsilon^2/2)^2}{2u_\varepsilon^2}\right) \geqslant C > 0,$$

то справедливость (34) установлена, и теорема 2 доказана. $\square$

Приложение

Доказательство леммы 2. Рассмотрим математическое ожидание

$$\begin{aligned}\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}) &= \frac{1}{\Psi_k(h)}\mathbf{E}_{\mathbf{P}_0}(\lambda_{\pi,k}) \\ &= \frac{1}{\Psi_k(h)}\mathbf{E}_{\mathbf{P}_0}\left(\exp\left\{-\frac{h(v_{j,k}^*)^2}{2} + h\log\cosh(v_{j,k}^* Y_{j,k})\right\}\right. \\ &\qquad \left.\times\left[-\frac{(v_{j,k}^*)^2}{2} + \log\cosh(v_{j,k}^* Y_{j,k})\right]\right)\end{aligned}$$

и напомним, что $\mathbf{P}_{\pi,\eta_j}$-распределение компонент вектора $Y_j = (Y_{j,k})_{k\in\mathbb{Z}}$ есть нормальное распределение со средним $\eta_j v_{j,k}^*$, где $\eta_j = 0$ или 1, и единичной дисперсией. (Распределение $\mathbf{P}_{\pi,\eta_j}$ определяется формулой (31), и мы используем более короткое обозначение $\mathbf{P}_0$ для $\mathbf{P}_{\pi,0}$.) Для краткости положим

$$v = v_{j,k}^*, \quad Y = Y_{j,k}$$

и напомним, что $v = o(1)$. Тогда, используя равенство (46), мы можем продолжить:

$$\begin{aligned}\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}) &= \left[\exp\left(\frac{h^2-h}{4}v^4 + o\left(v^4\right)\right)\right]^{-1} \\ &\times\left[-\frac{v^2}{2}\exp\left(-\frac{hv^2}{2}\right)\mathbf{E}_{\mathbf{P}_0}\left\{\exp\left(h\log\cosh(vY)\right)\right\}\right. \qquad (47) \\ &+ \left.\exp\left(-\frac{hv^2}{2}\right)\mathbf{E}_{\mathbf{P}_0}\left\{\log\cosh(vY)\exp\left(h\log\cosh(vY)\right)\right\}\right].\end{aligned}$$

Первое математическое ожидание в правой части (47) можно представить в виде

$$\begin{aligned}\mathbf{E}_{\mathbf{P}_0}\left\{\exp\left(h\log\cosh(vY)\right)\right\} &= \mathbf{E}_{\mathbf{P}_0}\left\{\exp\left(h\log\cosh(vY)\mathbb{I}(|Y|\leqslant 1/\sqrt{v})\right)\right\} \\ &+ \mathbf{E}_{\mathbf{P}_0}\left\{\exp\left(h\log\cosh(vY)\mathbb{I}(|Y|> 1/\sqrt{v})\right)\right\} =: L_1 + L_2.\end{aligned}$$

Заметим, что если $|Y| < 1/\sqrt{v}$, то $|vY| < \sqrt{v} \to 0$. Поэтому, с использованием неравенства $\cosh(x) \leqslant \exp(x^2/2)$, $x\in\mathbb{R}$, остаточный член L_2

оценивается следующим образом:

$$L_2 = \int\limits_{|y|>1/\sqrt{v}} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \exp\left(h \log\cosh(vy)\right)\, dy$$
$$\leqslant \int\limits_{|y|>1/\sqrt{v}} \frac{e^{-(y^2/2)(1-hv^2)}}{\sqrt{2\pi}}\, dy = O\left(v^{1/2} e^{-1/(2v)}\right).$$

Далее, принимая во внимание разложения

$$\log\cosh x = \frac{x^2}{2} - \frac{x^4}{12} + o(x^4), \quad e^x = 1 + x + \frac{x^2}{2} + o(x^2), \quad x \to 0, \tag{48}$$

мы находим, что

$$L_1 = \mathbf{E}_{\mathbf{P}_0}\left[1 + \frac{hv^2Y^2}{2} + \left(\frac{h^2v^4}{8} - \frac{hv^4}{12}\right) Y^4 \left(1 + o(1)\right)\right]$$
$$- \mathbf{E}_{\mathbf{P}_0}\left[\left(1 + \frac{hv^2Y^2}{2} + \left(\frac{h^2v^4}{8} - \frac{hv^4}{12}\right) Y^4 \left(1 + o(1)\right)\right) \mathbb{I}\left(|Y| > 1/\sqrt{v}\right)\right]$$
$$= 1 + \frac{hv^2}{2} + \left(\frac{3h^2v^4}{8} - \frac{hv^4}{4}\right)(1 + o(1)) + O(v^{1/2}e^{-1/(2v)}) = 1 + \frac{hv^2}{2} + O(v^4),$$

и потому

$$\mathbf{E}_{\mathbf{P}_0}\left\{\exp\left(h\log\cosh(vY)\right)\right\} = L_1 + L_2 = 1 + \frac{hv^2}{2} + O(v^4). \tag{49}$$

Аналогичные рассуждения, основанные на (48) и том факте, что если $|y| < 1/\sqrt{v}$ и $v = o(1)$, то

$$\exp\left(h\log\cosh(vy)\right) = 1 + \frac{hv^2y^2}{2} + O(v^4y^4), \tag{50}$$

ведут к представлению второго математического ожидания в правой части (47) в виде

$$\mathbf{E}_{\mathbf{P}_0}\left\{\log\cosh(vY)\exp\left(\log\cosh(vY)\right)\right\}$$
$$= \mathbf{E}_{\mathbf{P}_0}\left(\frac{v^2Y^2}{2} + \frac{hv^4Y^4}{4} - \frac{v^4Y^4}{12}\right) + o(v^4)$$
$$= \frac{v^2}{2} + \frac{3hv^4}{4} - \frac{v^4}{4} + o(v^4). \tag{51}$$

Подставляя (49) и (51) в (47), выводим

$$\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}) = \left[\exp\left(\frac{h^2-h}{4}v^4 + o\left(v^4\right)\right)\right]^{-1} \exp\left(-\frac{hv^2}{2}\right)$$
$$\times\left[-\frac{v^2}{2}\left(1+\frac{hv^2}{2}+O(v^4)\right)+\left(\frac{v^2}{2}+\frac{3hv^4}{4}-\frac{v^4}{4}+o(v^4)\right)\right]$$
$$= (1+O(v^2))\left[\frac{v^4}{2}\left(h-\frac{1}{2}\right)+o\left(v^4\right)\right] = \frac{v^4}{2}\left(h-\frac{1}{2}\right)+o\left(v^4\right).$$

Это дает искомое выражение для $\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})$.

Далее, рассмотрим дисперсию

$$\mathbf{Var}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}) = \mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}^2) - \left(\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^2.$$

Требуется вычислить

$$\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}^2) = \Psi_k^{-1}(h)\mathbf{E}_{\mathbf{P}_0}\left(\lambda_{\pi,k}^2\frac{d\mathbf{P}_{h,k}}{d\mathbf{P}_0}(Y)\right) = \Psi_k^{-1}(h)\mathbf{E}_{\mathbf{P}_0}\left(e^{h\lambda_{\pi,k}}\lambda_{\pi,k}^2\right)$$
$$= \Psi_k^{-1}(h)\mathbf{E}_{\mathbf{P}_0}(e^{h\lambda_{\pi,k}}\lambda_{\pi,k}^2\mathbb{I}(|Y|\leqslant 1/\sqrt{v}))$$
$$+\Psi_k^{-1}(h)\mathbf{E}_{\mathbf{P}_0}(e^{h\lambda_{\pi,k}}\lambda_{\pi,k}^2\mathbb{I}(|Y|> 1/\sqrt{v})) =: M_1+M_2,$$

где $\lambda_{\pi,k} = -\frac{v^2}{2}+\log\cosh(vY)$, $h=O(1)$, $v=o(1)$ и, согласно (46),

$$\Psi_k(h) = 1+O(v^4). \tag{52}$$

Для первого слагаемого M_1, используя (48) и (50), мы получаем выражение

$$M_1 = \Psi_k^{-1}(h)\mathbf{E}_{\mathbf{P}_0}\left[e^{-\frac{hv^2}{2}+h\log\cosh(vY)}\left(-\frac{v^2}{2}+\log\cosh(vY)\right)^2\mathbb{I}(|Y|\leqslant 1/\sqrt{v})\right]$$
$$= \mathbf{E}_0\left[e^{-\frac{hv^2}{2}}\left(1+\frac{hv^2Y^2}{2}+O(v^4Y^4)\right)\right.$$
$$\left.\times\left(-\frac{v^2}{2}+\frac{v^2Y^2}{2}-\frac{v^4Y^4}{12}(1+o(1))\right)^2\mathbb{I}(|Y|\leqslant 1/\sqrt{v})\right]\times(1+O(v^4)),$$

откуда после простых вычислений с использованием формулы

$$e^{-\frac{hv^2}{2}} = 1-\frac{hv^2}{2}+\frac{h^2v^4}{8}+o(v^4)$$

выводим

$$M_1 = \mathbf{E}_{\mathbf{P}_0}\left(\frac{v^4}{4} - \frac{v^4Y^2}{2} + \frac{v^4Y^4}{4}\right) + o(v^4) = \frac{v^4}{2} + o(v^4).$$

Далее, замечая, что $M_2 = o(v^4)$, получаем

$$\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^2_{\pi,k}) = M_1 + M_2 = \frac{v^4}{2} + o(v^4).$$

Возвращаясь к полным обозначениям и используя тот факт, что

$$\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}) = \frac{(v^*_{j,k})^4}{2}(h - 1/2) + o\left((v^*_{j,k})^4\right),$$

находим

$$\begin{aligned}\mathbf{Var}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}) &= \mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^2_{\pi,k}) - \left(\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^2 = \frac{(v^*_{j,k})^4}{2} + o\left((v^*_{j,k})^4\right),\\ \mathbf{Var}_{\mathbf{P}_h}(\lambda_\pi) &= \sum_{k\in\mathbb{Z}} \mathbf{Var}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k}) = u^2_\varepsilon(r_\varepsilon)(1+o(1)).\end{aligned}$$

Лемма 2 доказана. □

Доказательство леммы 3. Среднее значение в числителе дроби Ляпунова равняется

$$\begin{aligned}\mathbf{E}_{\mathbf{P}_{h,k}}\left(\lambda_{\pi,k} - \mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^4 &= \mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^4_{\pi,k}) - 4\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^3_{\pi,k})\\ &+ 6\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^2_{\pi,k})\left(\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^2 - 3\left(\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^4, \quad (53)\end{aligned}$$

и потому, согласно лемме 2,

$$\begin{aligned}&\mathbf{E}_{\mathbf{P}_{h,k}}\left(\lambda_{\pi,k} - \mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^4\\ &\qquad = \mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^4_{\pi,k}) + O\left((v^*_{j,k})^4\right)\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^3_{\pi,k}) + O\left((v^*_{j,k})^{12}\right).\end{aligned}$$

Вычисление моментов $\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^3_{\pi,k})$ и $\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^4_{\pi,k})$ проводится по той же схеме, что и вычисление среднего и дисперсии в лемме 2. Именно, использование соотношений (48) и (52) приводит к выражениям

$$\begin{aligned}\mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^3_{\pi,k}) &= \Psi_k^{-1}(h)\mathbf{E}_{\mathbf{P}_0}\left(e^{h\lambda_{\pi,k}}\lambda^3_{\pi,k}\right) = O((v^*_{j,k})^6),\\ \mathbf{E}_{\mathbf{P}_{h,k}}(\lambda^4_{\pi,k}) &= \Psi_k^{-1}(h)\mathbf{E}_{\mathbf{P}_0}\left(e^{h\lambda_{\pi,k}}\lambda^4_{\pi,k}\right) = O((v^*_{j,k})^8).\end{aligned}$$

Поэтому, принимая во внимание (53), мы находим, что

$$\mathbf{E}_{\mathbf{P}_{h,k}}\left(\lambda_{\pi,k} - \mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^4 \asymp (v^*_{j,k})^8 \leqslant \max_{k\in\mathbb{Z}}(v^*_{j,k})^4 \sum_{k\in\mathbb{Z}}(v^*_{j,k})^4.$$

Отсюда, из леммы 2 и формулы (43) получаем требуемый результат:

$$L_h^{(4)} = \frac{\sum_{k\in\mathbb{Z}} \mathbf{E}_{\mathbf{P}_{h,k}} \left(\lambda_{\pi,k} - \mathbf{E}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^4}{\left(\sum_{k\in\mathbb{Z}} \mathbf{Var}_{\mathbf{P}_{h,k}}(\lambda_{\pi,k})\right)^2} \asymp \frac{\sum_{k\in\mathbb{Z}}(v_{j,k}^*)^8}{\left(\sum_{k\in\mathbb{Z}}(v_{j,k}^*)^4\right)^2}$$

$$\leqslant \frac{\max\limits_{k\in\mathbb{Z}}(v_{j,k}^*)^4}{\sum_{k\in\mathbb{Z}}(v_{j,k}^*)^4} = \frac{o(1)}{u_\varepsilon^2(r_\varepsilon)} = \frac{o(1)}{O(\log d)} = o(\log^{-1} d).$$

Лемма 3 доказана. □

Литература

1. C. Butucea, Yu. I. Ingster, *Detection of a sparse submatrix of a high-dimensional noisy matrix.* — http://arxiv.org/abs/1109.0898 (2012).
2. L. Comminges, A. S. Dalalyan, *Tight conditions for consistency of variable selection in the context of high dimensionality.* — http://arxiv.org/abs/1106.4293 (2012).
3. M. De Groot, *Optimal Statistical Decisions.* McGraw-Hill Book Company, New York, 1970.
4. М. С. Ермаков, *Минимаксное обнаружение сигнала в гауссовском белом шуме.* — Теория вероятн. и ее примен. **35**, No. 4 (1991), 667–679.
5. G. Gayraud, Yu. I. Ingster, *Detection of sparse variable functions.* — http://arxiv.org/abs/1011.6369 (2012).
6. C. R. Genovese, J. Jin, L. Wasserman, *Revisiting marginal regression.* — http://arxiv.org/abs/0911.4080 (2009).
7. J. Huang, J. L. Howoritz, F. Wei, *Variable selection in nonparametric additive models.* — Ann. Statist. **38** (2010), 2282–2313.
8. I. A. Ibragimov, R. Z. Khasminskii, *Some estimation problems on infinite dimensional Gaussian white noise.* — In: Festschrift for Lucien Le Cam. Research Papers in Probability and Statistics. Springer-Verlag, New York (1997), pp. 275–296.
9. Yu. I. Ingster, *Asymptotically minimax hypothesis testing for nonparametric alternatives. I.* — Math. Methods Statist. **2**, No. 2 (1993), 85–114.
10. Yu. I. Ingster, *Asymptotically minimax hypothesis testing for nonparametric alternatives. II.* — Math. Methods Statist. **2**, No. 3 (1993), 171–189.
11. Yu. I. Ingster, *Asymptotically minimax hypothesis testing for nonparametric alternatives. III.* — Math. Methods Statist. **2**, No. 4 (1993), 249–268.
12. Yu. I. Ingster, I. A. Suslina, *Nonparametric Goodness-of-Fit Testing Under Gaussian Models.* Lect. Notes Statist., **169**, Springer-Verlag, New York, 2003.
13. Yu. I. Ingster, I. A. Suslina, *On estimation and detection of smooth function of many variables.* — Math. Methods Statist. **14** (2005), 299–331.
14. Y. Lin, *Tensor product space ANOVA models.* — Ann. Statist. **28** (2000), 734–755.
15. G. Raskutti, M. J. Wainwright, B. Yu, *Minimax-optimal rates for sparse additive models over kernel classes via convex programming.* — http://arxiv.org/abs/1008.3654 (2011).
16. A. V. Skorohod, *Integration in Hilbert Spaces*, Springer-Verlag, Berlin–New York, 1974.

17. V. G. Spokoiny, *Adaptive hypothesis testing using wavelets.* — Ann. Statist. **24**, No. 6 (1996), 2477–2498.
18. C. J. Stone, *Additive regression and other nonparametric models.* — Ann. Statist. **13**, No. 2 (1985), 689–705.

Ingster Yu.I., Stepanova N. A. Adaptive variable selection in nonparametric sparse regression.

We study the problem of exact recovery of an unknown multivariate function f observed in the continuous regression model. It is assumed that, in addition to some smoothness constraints, f possesses an additive sparse structure determined by the sparsity index $\beta \in (0, 1)$. As a consequence of the additive sparsity assumption, the recovery problem transforms to a variable selection problem. Conditions for exact variable selection are provided, and a family of asymptotically minimax variable selection procedures is constructed. The procedures are adaptive in the sparsity index β.

С.-Петербургский Государственный
Электротехнический Университет,
ул. Профессора Попова 5,
Санкт-Петербург 197376, Россия

Поступило 15 октября 2012 г.

Карлтон Университет,
Колонел Бай Драйв 1125,
Оттава, Онтарио K1S 5B6, Канада
E-mail: nstep@math.carleton.ca