ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 59 2023 Вып. 1

УДК 621.391 : 519.23

 Г.К. Голубев

ПЕРЕПАРАМЕТРИЗОВАННЫЕ ТЕСТЫ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ ДЛЯ ОБНАРУЖЕНИЯ РАЗРЕЖЕННЫХ ВЕКТОРОВ

Рассматривается задача обнаружения разреженного вектора большой размерности на фоне белого гауссовского шума. Предполагается, что неизвестный вектор может иметь только p ненулевых компонент, положение и величина которых неизвестны, а их число, с одной стороны, велико, но с другой – мало по сравнению с его размерностью. Тест максимального правдоподобия (МП) в этой задаче имеет простой вид и, естественно, зависит от p. В статье изучаются статистические свойства перепараметризованных тестов МП, т.е. тестов, построенных на основе предположения, что число ненулевых компонент вектора равно q $(q > p)$, в ситуации, когда на самом деле вектор имеет всего лишь p ненулевых компонент. Показывается, что в некоторых случаях перепараметризованные тесты могут быть лучше стандартных тестов МП.

Ключевые слова: разреженный вектор, белый гауссовский шум, тест максимального правдоподобия.

DOI: 10.31857/S0555292323010047, **EDN:** RMHNYH

§ 1. Введение

Реальный интерес к перепараметризованным статистическим моделям и методам возник сравнительно недавно и связан с задачами глубокого обучения. Оказалось, что во многих случаях обучение перепараметризованной нейронной сети с помощью методов типа градиентного спуска дает хорошие результаты по сравнению с обучением неперепараметризованной сети [1]. Полностью математически удовлетворительного объяснения этого факта, по-видимому, до сих пор не найдено, что связано с очень высокой нелинейностью рассматриваемой задачи [2]. Поэтому внимание привлекли простые статистические модели, в которых наблюдаются похожие эффекты. Одной из таких моделей является модель линейной регрессии со случайными гауссовскими регрессорами, анализ которой в перепараметризованном режиме можно найти, например, в [3]. В обзорной статье [4] приводятся примеры других близких моделей, связанных с обработкой сигналов.

В этой статье будет показано, что в задаче обнаружения разреженных векторов с большим числом ненулевых компонент с помощью метода МП наблюдаются в некотором смысле похожие эффекты, а именно оказывается, что перепараметризованные тесты МП в некоторых случаях могут быть лучше стандартных.

Далее будет рассматриваться задача обнаружения разреженного вектора $\theta^n = (\theta_1, \dots \theta_n)^\top \in \mathbb{R}^n$ по наблюдениям

$$Y_k = \theta_k + \xi_k, \quad k = 1, \dots, n, \tag{1}$$

где ξ_k – независимые стандартные гауссовские случайные величины.

Цель состоит в том, чтобы на основе наблюдений $Y^n = (Y_1, \ldots, Y_n)^\top$ из (1) решить, какая из двух гипотез

$$\mathcal{H}_0 : \theta^n = 0 \quad \text{или} \quad \mathcal{H}_1 : \theta^n \neq 0$$

является справедливой.

Ключевым предположением в этой статье является гипотеза о том, что θ^n – *разреженный* вектор. Существует очень много математических определений разреженности. Для наглядности в этой статье мы ограничимся самым простым. Пусть $I_p^n = \{i_1, \ldots, i_p\}$ – мультииндекс (носитель θ^n), т.е. множество, состоящее из p различных положительных целых чисел, выбранных из множества $\{1, 2, \ldots, n\}$. Тогда θ^n называется разреженным, если существует носитель I_p^n, такой что

$$\theta_k = 0, \quad k \notin I_p^n.$$

Множество всех разреженных векторов в $\mathbb{R}^n$ будем далее обозначать через Θ_p^n. Заметим, что его можно определить эквивалентным образом как

$$\Theta_p^n = \left\{\theta^n \in \mathbb{R}^n : \theta_{(k)}^2 = 0, \ k > p\right\};$$

здесь и далее $\{x_{(1)}, \ldots, x_{(n)}\}$ означает невозрастающую перестановку элементов множества $\{x_1, \ldots, x_n\}$.

Понятие разреженности становится статистически интересным и значимым, когда носитель I_p^n неизвестен, размерность n велика и при этом $p \ll n$. Поэтому в этой статье задача проверки гипотез будет рассматриваться в асимптотической постановке, предполагая, что $n \to \infty$. При этом $p = p(n)$ может быть любой известной функцией n, такой что $p(n) \in [p_\circ(n), p^\circ(n)]$, где $p_\circ(n)$ и $p^\circ(n)$ таковы, что

$$\lim_{n\to\infty} p_\circ(n) = \infty \quad \text{и} \quad \lim_{n\to\infty} \frac{p^\circ(n)\log^{1+\varepsilon}(n)}{n} = 0; \tag{2}$$

здесь и далее ε обозначает любое строго положительное число. Не оговаривая этого особо, будем предполагать, что условия (2) выполнены.

Отметим, что задача проверки гипотез о разреженных векторах в ситуации, когда разреженность фиксирована, например, $p = 1$, а $n \to \infty$, существенно отличается от рассматриваемой далее. Хорошо известно, что при фиксированной разреженности предельные распределения статистик тестов МП и байесовских тестов при $\mathcal{H}_0$ не являются гауссовскими. Для байесовских тестов эти распределения тесно связаны с устойчивыми распределениями. Этот замечательный факт, по-видимому, впервые был установлен в [5] (см. также [6, 7] для задач обнаружения в белом гауссовском шуме). Также хорошо известно, что если $p \to \infty$, то предельные распределения статистик байесовских тестов и тестов МП будут гауссовскими при нулевой гипотезе (см., например, [7]). Но при этом, в отличие от классических (т.е. имеющих низкую размерность) задач проверки гипотез, гауссовость не влечет асимптотической эквивалентности байесовских тестов и тестов МП.

Тест МП для проверки простой гипотезы $\mathcal{H}_0$ против альтернативы $\mathcal{H}_1$ при условии, что $\theta^n \in \Theta_p^n$, легко построить. Он, естественно, основан на максимуме отношения правдоподобия

$$L(Y^n; \Theta_p^n) = \max_{\theta^n \in \Theta_p^n} \prod_{k=1}^n \exp\left\{-\frac{(Y_k - \theta_k)^2}{2} + \frac{Y_k^2}{2}\right\} =$$
$$= \max_{I_p^n} \exp\left\{\frac{1}{2}\sum_{k\in I_p^n} Y_k^2\right\} = \exp\left\{\frac{1}{2}\sum_{k=1}^p Y_{(k)}^2\right\}.$$

Поэтому в качестве статистики этого теста будем использовать

$$M_p(Y^n) \stackrel{\text{def}}{=} \frac{1}{2}\sum_{k=1}^{p} Y_{(k)}^2.$$

Таким образом, тест МП отвергает гипотезу $\mathcal{H}_0$, если

$$M_p(Y^n) \geqslant t_{M_p}(\alpha),$$

где критический уровень $t_{M_p}(\alpha)$ выбирается так, чтобы вероятность ошибки первого рода была равна α, т.е. он определяется как корень уравнения

$$\mathbf{P}_{\mathcal{H}_0}\bigl\{M_p(Y^n) \leqslant t_{M_p}(\alpha)\bigr\} = \mathbf{P}\bigl\{M_p(\xi^n) \leqslant t_{M_p}(\alpha)\bigr\} = \alpha;$$

здесь и далее $\xi^n = (\xi_1, \dots, \xi_n)^\top$ обозначает вектор из независимых стандартных гауссовских случайных величин.

Далее, не оговаривая этого особо, будем считать, что вероятность ошибки первого рода α остается фиксированной при $n \to \infty$.

Мощность (качество) теста МП, как, впрочем, и любого другого теста, принято измерять вероятностью ошибки второго рода

$$\beta_{M_p}(\theta^n) = \mathbf{P}_{\theta^n}\bigl\{M_p(Y^n) \leqslant t_{M_p}(\alpha)\bigr\},$$

где $\mathbf{P}_{\theta^n}$ – вероятностное распределение, порожденное наблюдениями Y^n при альтернативе $\mathcal{H}_1$.

Естественно, нам хотелось бы найти тест, который минимизировал бы ошибку второго рода равномерно при всех θ^n. Однако хорошо известно, что за исключением случая двух простых гипотез эта задача, как правило, не имеет простого решения. Оптимальный тест можно найти, например, для средней (байесовский подход) или максимальной (минимаксный подход) вероятности ошибки второго рода. При этих подходах, очевидно, нужна априорная информация о θ^n. При байесовском подходе она определяется априорным вероятностным распределением на θ^n, а при минимаксном – предположением, что $\theta^n \in \Theta$, где Θ – известное подмножество в $\mathbb{R}^n$. Очевидно, что оптимальные тесты в этом случае будут зависеть от априорной информации.

Если θ^n имеет небольшую размерность, то априорная информация не играет принципиальной роли, поскольку в этом случае практически всю информацию о неизвестном векторе можно извлечь из наблюдений. При больших размерностях θ^n из наблюдений мы получаем недостаточно информации и должны компенсировать ее недостаток априорной информацией, и поэтому тесты будут существенно от нее зависеть.

Чтобы устранить этот принципиальный недостаток байесовского и минимаксного подходов, в статистике используется множественная проверка гипотез. Грубо говоря, этот метод подразумевает, что тест строится не с помощью одной статистики, а с помощью некоторого семейства статистик. При этом предполагается, что это семейство не является очень богатым. В рассматриваемой здесь задаче роль такого семейства может играть, например, $\{M_p(Y^n),\ p = 1, \dots, n\}$. Простейшим хорошо известным, но в то же время достаточно наивным методом множественной проверки гипотез является метод Бонферрони [8]. Обзор некоторых более современных эвристических подходов, включая популярный на практике метод False Discovery Rate [9] и его модификаций, содержится в [10]. Что касается математических результатов множественной проверки гипотез о разреженных векторах, то некоторые из них можно найти, например, в [11].

Отметим, что для обнаружения разреженных векторов из Θ_p^n наряду со стандартным тестом МП можно использовать и любой *перепараметризованный* тест со статистикой $M_q(Y^n)$ при $q > p$. На первый взгляд кажется, что эта идея противоречит здравому смыслу, но мы увидим, что перепараметризованные тесты могут иметь меньшую вероятность ошибки второго рода, чем классический тест МП. В этом можно убедиться, анализируя вероятность ошибки второго рода при всех $\theta^n \in \Theta_p^n$. Этот анализ, по сути дела, и представляет основное содержание настоящей статьи.

§ 2. Предельное распределение статистики теста МП при $\mathcal{H}_0$

Для вычисления критических значений тестов МП нам потребуется следующий результат о распределении суммы квадратов порядковых статистик независимых стандартных гауссовских случайных величин.

Пусть

$$G(x) = \sqrt{\frac{2}{\pi}} \int\limits_{\sqrt{2x}}^{\infty} e^{-u^2/2}\, du, \tag{3}$$

и обозначим через $G^{-1}(x)$ обратную функцию к $G(x)$.

Теорема 1. *При любых $p = p(n) \in [p_\circ(n), p^\circ(n)]$ (см. (2)) имеем*

$$\frac{M_p(\xi^n) - \mu(p;n)}{\sqrt{2p}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad n \to \infty, \tag{4}$$

где

$$\begin{aligned}\mu(p;n) &= \sum_{k=1}^{p} G^{-1}\left(\frac{n+1}{k}\right) = \\ &= p\left\{\frac{n+1}{\sqrt{\pi}p} G^{-1}\left(\frac{p}{n+1}\right) \exp\left[-G^{-1}\left(\frac{p}{n+1}\right)\right] + \frac{1}{2}\right\} + O(1) = \\ &= p\left\{\log\left(\frac{n+1}{\sqrt{\pi}p}\right) - \frac{1}{2}\log\left[1 + \log\left(\frac{n+1}{\sqrt{\pi}p}\right)\right] + 1 + o(1)\right\} + O(1).\end{aligned} \tag{5}$$

Доказательство этой теоремы, как и других вспомогательных результатов, приведено в § 4.

Из (4) и (5) вытекает, что при больших p критическое значение для теста МП можно вычислить как

$$t_{M_p}(\alpha) = p\left\{\frac{n+1}{\sqrt{\pi}p} G^{-1}\left(\frac{p}{n+1}\right) \exp\left[-G^{-1}\left(\frac{p}{n+1}\right)\right] + \frac{1}{2}\right\} + \sqrt{2p}\,t(\alpha), \tag{6}$$

где $t(\alpha)$ – α-квантиль стандартного гауссовского распределения, т.е. корень уравнения

$$\frac{1}{\sqrt{2\pi}} \int\limits_{t(\alpha)}^{\infty} e^{-x^2/2}\, dx = \alpha.$$

Что касается вычисления вероятности ошибки второго рода, то теорема 1 может быть в принципе полезна только в случае, когда эта вероятность не является малой.

Для анализа малых вероятностей ошибки будет использоваться следующее неасимптотическое неравенство. В его формулировке и далее неравенство $\xi \overset{\mathcal{D}}{\geqslant} \eta$ означает, что

$$\mathbf{P}\{\xi \geqslant x\} \leqslant \mathbf{P}\{\eta \geqslant x\} \quad \text{при всех } x \geqslant 0.$$

Л е м м а 1. *Справедливо неравенство*

$$M_p(\xi^n) \overset{\mathcal{D}}{\geqslant} \mu\bigl(p; \Sigma_{n+1}\bigr) + \left\{1 - \frac{1}{2}\left[1 + \log\left(\frac{\Sigma_{n+1}}{\sqrt{\pi} p^\circ(n)}\right)\right]^{-1}\right\} \sum_{s=1}^{p} w(s,p)(1-\chi_s),$$

где:

- *функция* $\mu(\cdot\,,\cdot)$ *определена в* (5);
- χ_s – *независимые стандартные экспоненциально распределенные случайные величины*;
- $\Sigma_{n+1} = \sum_{s=1}^{n+1} \chi_s$;
- $$w(s,p) = \sum_{k=s}^{p} \frac{1}{k}. \tag{7}$$

§ 3. Вероятность ошибки второго рода

К сожалению, простых и одновременно точных оценок для вероятности ошибки второго рода, по-видимому, не существует. Поэтому мы приведем лишь границы сверху для этой величины, предполагая, что $\theta^n \in \Theta_p^n$, но при этом используется тест МП со статистикой $M_q(Y^n)$, $q \geqslant p$.

Чтобы несколько упростить дальнейшее изложение, будем считать, что

$$q \leqslant Kp,$$

где $K \geqslant 1$ – некоторая постоянная.

В силу инвариантности статистики $M_q(Y^n)$ относительно перестановок компонент вектора наблюдений Y^n, без ограничения общности можно считать, что

$$\begin{aligned} Y_k &= \theta_k + \xi_k, && k = 1, \ldots, p, \\ Y_k &= \xi_k, && k = p+1, \ldots, n, \end{aligned}$$

а компоненты вектора θ^n таковы, что $\theta_1^2 \geqslant \theta_2^2 \geqslant \ldots \geqslant \theta_p^2$.

Пусть $p' \leqslant p$ – некоторое целое число. Обозначим для краткости

$$Y'_k = Y_{p'+k}, \quad \xi'_k = \xi_{p'+k}, \quad k = 1, \ldots, n-p',$$

и

$$\|\theta^n\|_{p'}^2 = \sum_{k=1}^{p'} \theta_{(k)}^2.$$

Тогда очевидно, что

$$M_q(Y^n) \geqslant \frac{1}{2} \sum_{k=1}^{p'} Y_k^2 + \frac{1}{2} \sum_{k=1}^{q-p'} Y'^2_{(k)}. \tag{8}$$

Первое слагаемое в правой части этого неравенства имеет вид

$$\frac{1}{2}\sum_{k=1}^{p'} Y_k^2 = \frac{\|\theta^n\|_{p'}^2 + p'}{2} + \sum_{k=1}^{p'} \xi_i\theta_i + \frac{1}{2}\sum_{k=1}^{p'}(\xi_k^2 - 1). \tag{9}$$

Для оценки второго слагаемого в правой части (8) можно воспользоваться леммой Андерсона [12] (см. также [13, гл. II, § 10]), а именно интуитивно понятным неравенством

$$\frac{1}{2}\sum_{k=1}^{q-p'} Y_{(k)}'^2 \overset{\mathcal{D}}{\geqslant} \frac{1}{2}\sum_{k=1}^{q-p'} \xi_{(k)}'^2.$$

Продолжив правую часть этого неравенства с помощью леммы 1, получаем

$$\frac{1}{2}\sum_{k=1}^{q-p'} Y_{(k)}'^2 \overset{\mathcal{D}}{\geqslant} \mu\bigl(q-p'; n-p'\bigr) + (1+o(1))\sum_{k=1}^{q-p'} w(s, q-p')(1-\chi_s),$$

где веса $w(\cdot\,,\cdot)$ определены в (7). Поэтому отсюда и из (8), (9) приходим к следующему неравенству:

$$M_q(Y^n) \overset{\mathcal{D}}{\geqslant} \frac{\|\theta^n\|_{p'}^2 + p'}{2} + \mu(q-p'; n-p') - \zeta(\theta^n, p', q),$$

где

$$\zeta(\theta^n, p', q) = (1+o(1))\sum_{k=1}^{q-p'} w(s, q-p')(\chi_s - 1) + \sum_{k=1}^{p'} \xi_i\theta_i + \frac{1}{2}\sum_{k=1}^{p'}(1-\xi_k^2).$$

Тогда отсюда и из (6) вытекает следующая граница сверху для вероятности ошибки второго рода:

$$\begin{aligned}\beta_{M_q}(\theta^n) &= \mathbf{P}_{\theta^n}\Bigl\{M_q(Y^n) \leqslant \mu(q;n) + \sqrt{2q}t(\alpha)\Bigr\} \leqslant \\ &\leqslant \mathbf{P}\biggl\{\zeta(\theta^n, p', q) \geqslant \frac{\|\theta^n\|_{p'}^2}{2} - \Delta(q, p'; n) + \frac{p'}{2} - \sqrt{2q}t(\alpha)\biggr\},\end{aligned} \tag{10}$$

где

$$\Delta(p', q; n) = \mu(q;n) - \mu(q-p'; n-p').$$

Величина $\Delta(p', q; n)$ играет принципиальную роль в рассматриваемой задаче, и далее потребуется следующая ее аппроксимация, вытекающая из (5) и формулы Тейлора:

$$\begin{aligned}\Delta(p', q; n) &= q\biggl\{\log\biggl(\frac{n}{\sqrt{\pi}q}\biggr) - \frac{1}{2}\log\biggl[1 + \log\biggl(\frac{n}{\sqrt{\pi}q}\biggr)\biggr] + 1 + o(1)\biggr\} - \\ &- (q-p')\biggl\{\log\biggl(\frac{n-p'}{\sqrt{\pi}(q-p')}\biggr) - \frac{1}{2}\log\biggl[1 + \log\biggl(\frac{n-p'}{\sqrt{\pi}(q-p')}\biggr)\biggr] + 1\biggr\} =\end{aligned}$$

$$= p'\left\{\log\left(\frac{n}{\sqrt{\pi}p'}\right) - \frac{1}{2}\log\left[1 + \log\left(\frac{n}{\sqrt{\pi}p'}\right)\right] + 1 + o(1)\right\} =$$

$$= \mu(p'; n) - qh\left(\frac{p'}{q}\right) + o(q), \quad n \to \infty, \tag{11}$$

где

$$h(x) = -x\log(x) - (1-x)h(1-x).$$

Также далее будет нужна аппроксимация распределения случайной величины $\zeta(\theta^n, p', q)$. Обозначим для краткости

$$D(x) = D(x; p', q, \theta^n) = \left[\frac{\|\theta^n\|_{p'}^2}{1+x} + \frac{p'}{2} + \sum_{s=1}^{q-p'} w^2(s,p)\right]. \tag{12}$$

Л е м м а 2. *Для всех $x > 0$, таких что*

$$x \leqslant \frac{(1-\varepsilon)D(0)}{2w^2(1, q-p')}, \tag{13}$$

справедливо неравенство

$$\mathbf{P}\left\{\zeta(\theta^n, p', q) \geqslant \sqrt{2xD(0)}\right\} \leqslant$$

$$\leqslant \exp\left[-r_\circ(x) + O\left(\frac{x(q-p')}{D(0)w(1, q-p')}\right)\right], \quad n \to \infty, \tag{14}$$

где

$$r_\circ(x) = x + \frac{\log[2\pi(1+x)]}{2}.$$

Следующая теорема, представляющая основной результат статьи, является по сути прямым следствием этой леммы и неравенства (10).

Т е о р е м а 2. *Предположим, что для $p' \in [p_\circ(n), p]$, $\theta^n \in \Theta_p^n$ и $A > 0$ выполнены следующие условия*:

$$\|\theta^n\|_{p'}^2 \geqslant 2p'\log\left(\frac{n+1}{\sqrt{\pi}p'}\right) - p'\log\left[1 + \log\left(\frac{n+1}{\sqrt{\pi}p'}\right)\right] + p' - 2qh\left(\frac{p'}{q}\right) +$$

$$+ 4\sqrt{A}\left[p'\log\left(\frac{n+1}{\sqrt{\pi}p'}\right) + A\right]^{1/2} + 4A + \varepsilon q \tag{15}$$

и

$$A \leqslant \frac{(1-\varepsilon)\|\theta^n\|_{p'}^2}{2w^2(1, q-p')}.$$

Тогда при $n \to \infty$

$$\beta_{M_q}(\theta^n) \leqslant \exp\left[-r_\circ(A) + O\left(\frac{q-p'}{w^3(1, q-p')}\right)\right]. \tag{16}$$

Доказательство. Из леммы 2 и неравенства (10) непосредственно вытекает, что если для некоторого $p' \in [p_\circ(n), p]$

$$D(0; p', q, \theta^n) - 2\sqrt{2AD(0; p', q, \theta^n)} \geqslant$$
$$\geqslant 2\Delta(q, p'; n) - \frac{p'}{2} + \sum_{s=1}^{q-p'} w^2(s, p) + 2\sqrt{2q}t(\alpha), \tag{17}$$

то неравенство (16) выполнено.

Условие (15) является упрощенной формой неравенства (17). Чтобы в этом убедиться, заметим, что неравенство

$$y - 2\sqrt{2Ay} \geqslant x, \quad x > 0,$$

эквивалентно неравенству

$$y \geqslant \left[\sqrt{2A} + \sqrt{2A + x}\,\right]^2 = x + 4A + 2\sqrt{2A}\sqrt{2A + x}.$$

Подставив в него

$$x = 2\Delta(q, p'; n) - \frac{p'}{2} + \sum_{s=1}^{q-p'} w^2(s, p) + 2\sqrt{2q}t(\alpha),$$

$$y = D(0; p', q, \theta^n) = \|\theta^n\|^2_{p'} + \frac{p'}{2} + \sum_{s=1}^{q-p'} w^2(s, p)$$

и воспользовавшись (5) и (11), приходим к (15). ▲

Замечание 1. Теорема 2 допускает достаточно простую интерпретацию. Она позволяет оценить разность между минимальными энергиями векторов θ^n, обнаруживаемых стандартным и перепараметризованным тестом МП.

Предположим для простоты, что $A = A(n)$ – медленно растущая функция n, например, $A(n) = \log[p^\circ(n)]$. В этом случае, если выполнено условие

$$\lim_{n\to\infty} \frac{A(n)\log(n)}{p_\circ(n)} = 0,$$

то последние слагаемые в (15) имеют порядок $o(p)$. Поэтому стандартный тест МП сможет с достаточно малой вероятностью ошибки второго рода обнаружить векторы $\theta^n \in \Theta^n_p$ при отношениях сигнал/шум $\|\theta^n\|^2/p$, больших чем

$$2\log\left(\frac{n+1}{\sqrt{\pi}p}\right) - \log\left[1 + \log\left(\frac{n+1}{\sqrt{\pi}p}\right)\right] + 1 + \varepsilon,$$

а перепараметризованный тест сможет сделать это с теми же вероятностями ошибок, но при меньших отношениях сигнал/шум, а именно

$$2\log\left(\frac{n+1}{\sqrt{\pi}p}\right) - \log\left[1 + \log\left(\frac{n+1}{\sqrt{\pi}p}\right)\right] + 1 - \frac{q}{p}h\left(\frac{p}{q}\right) + \varepsilon.$$

Замечание 2. На первый взгляд кажется, что выбирая большое q, можно существенно уменьшить вероятность ошибки второго рода. К сожалению, это слишком оптимистичное предположение с практической точки зрения. Дело в том, что, во-первых, улучшение верхних границ не влечет автоматического улучшения действительной вероятности ошибки. Во-вторых, улучшаются, по сути, только члены третьего порядка в экспоненте вероятности ошибки. И наконец, улучшаются асимптотические границы в ситуации, когда скорость сходимости к ним очень медленная.

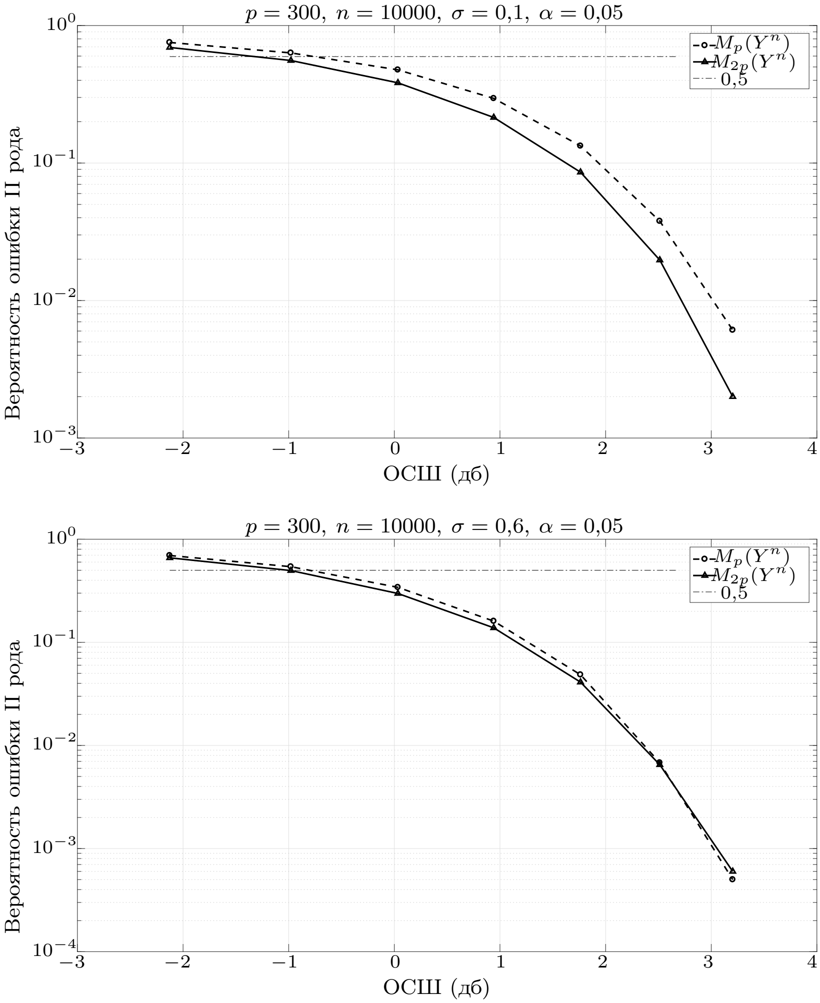

Рис. 1. Ошибки второго рода для стандартного и перепараметризованного теста МП. Графики сверху построены при $\sigma = 0{,}1$, а снизу – при $\sigma = 0{,}6$

К сожалению, медленные скорости сходимости типичны при проверке гипотез о разреженных векторах. На это обращал внимание еще Р.Л. Добрушин в [5].

Замечание 3. Наблюдать на практике небольшие преимущества перепараметризованных тестов возможно только лишь при достаточно больших n и p. Это иллюстрирует рис. 1, на котором показана зависимость вероятности ошибки второго рода от отношения сигнал/шум при $n = 10000$ и $p = 300$ для стандартного теста МП и перепараметризованного теста с $q = 2p = 600$. При этом в качестве ненулевых θ_k использовались независимые одинаково распределенные случайные величины, имеющие следующую структуру:

$$\theta_k = A(b_k + \sigma\xi_k),$$

где A и σ – постоянные, b_k – независимые случайные величины, принимающие значения $\{+1, -1\}$ с равными вероятностями, ξ_k – независимые $\mathcal{N}(0,1)$. В этом случае отношение сигнал/шум равно $A^2(1+\sigma^2)$.

Мы видим, что на втором графике стандартный тест и перепараметризованный оказываются практически эквивалентными. Это объясняется тем, что для "случайных" θ_k^2 неравенство (8) является слишком грубым.

§ 4. Доказательства

Для доказательства теоремы 1 нам потребуется следующий простой факт, который легко проверить с помощью экспоненциального неравенства Чебышева.

Л е м м а 3. *При* $x \geqslant 0$

$$\mathbf{P}\left\{\frac{1}{k}\sum_{s=1}^{k}\chi_s \geqslant (1+x)\right\} \leqslant \exp\{-k[x-\log(1+x)]\}, \quad x>0,$$

$$\mathbf{P}\left\{\frac{1}{k}\sum_{s=1}^{k}\chi_s \leqslant (1-x)\right\} \leqslant \exp\{k[x+\log(1-x)]\}, \quad x\in(0,1).$$

Д о к а з а т е л ь с т в о т е о р е м ы 1 основано на простых и хорошо известных методах и результатах, таких как теорема Пайка [14].

Введем функцию

$$r(x) = -\log[\sqrt{\pi}G(x)].$$

Тогда очевидно, что

$$G^{-1}(x) = r^{-1}\bigl[-\log\bigl(\sqrt{\pi}x\bigr)\bigr], \tag{18}$$

где $r^{-1}(x)$ – функция, обратная к $r(x)$.

Интегрируя по частям правую часть в (3), нетрудно проверить, что

$$r(x) = x + \frac{1}{2}\log(1+x) + O\Bigl(\frac{1}{x^{3/2}}\Bigr), \tag{19}$$

и отсюда с помощью формулы Тейлора находим

$$r^{-1}(x) = x - \frac{\log(1+x)}{2} + \frac{\log(1+x)}{2(1+x)} + O\Bigl(\frac{1}{x^{3/2}}\Bigr). \tag{20}$$

Проверим сначала (5). С помощью (18) и (20) нетрудно убедиться, что

$$\mu(p;n) = \sum_{k=1}^{p} G^{-1}\left(\frac{k}{n+1}\right) = -\int_1^p G^{-1}\left(\frac{x}{n+1}\right)dx + O(1) =$$

$$= -(n+1)\int_{1/(n+1)}^{p/(n+1)} G^{-1}(x)\,dx + O(1) =$$

$$= -(n+1)\int_{G^{-1}[1/(n+1)]}^{G^{-1}[p/(n+1)]} xG'(x)\,dx + O(1) =$$

$$= -\frac{n+1}{\sqrt{\pi}}\int_{G^{-1}[1/(n+1)]}^{G^{-1}[p/(n+1)]} \sqrt{x}e^{-x}\,dx + O(1). \tag{21}$$

Далее, с помощью очевидной замены переменных и интегрирования по частям находим

$$\int\limits_y^\infty \sqrt{x}e^{-x}\,dx = \sqrt{y}e^{-y} + \frac{\sqrt{\pi}}{2}\sqrt{\frac{2}{\pi}}\int\limits_{\sqrt{2y}}^\infty e^{-x^2/2}\,dx = \sqrt{y}e^{-y} + \frac{\sqrt{\pi}}{2}G(y) =$$

$$= \sqrt{\pi}G(y)\biggl(\frac{\sqrt{y}e^{-y}}{\sqrt{\pi}G(y)} + \frac{1}{2}\biggr).$$

Подставив в это равенство (см. (19))

$$e^{-y} = \sqrt{\pi(1+y)}G(y)\biggl[1 + O\biggl(\frac{1}{y^{3/2}}\biggr)\biggr],$$

получаем

$$\frac{1}{\sqrt{\pi}}\int\limits_y^\infty \sqrt{x}e^{-x}\,dx = G(y)\biggl[\sqrt{y(y+1)} + \frac{1}{2} + O\biggl(\frac{1}{\sqrt{y}}\biggr)\biggr] =$$

$$= G(y)\biggl[y + 1 + O\biggl(\frac{1}{\sqrt{y}}\biggr)\biggr].$$

Используя это равенство, (18) и (20), продолжим (21) следующим образом:

$$\mu(p;n) = p\biggl\{\frac{n+1}{\sqrt{\pi}p}G^{-1}\biggl(\frac{p}{n+1}\biggr)\exp\biggl[-G^{-1}\biggl(\frac{p}{n+1}\biggr)\biggr] + \frac{1}{2}\biggr\} -$$

$$-\frac{1}{n+1}\biggl[G^{-1}\biggl(\frac{1}{n+1}\biggr) + 1 + O\biggl(\frac{1}{\sqrt{G^{-1}[1/(n+1)]}}\biggr)\biggr] + O(1) =$$

$$= p\biggl[G^{-1}\biggl(\frac{p}{n+1}\biggr) + 1 + O\biggl(\frac{1}{\sqrt{G^{-1}[p/(n+1)]}}\biggr)\biggr] + O(1) =$$

$$= p\biggl\{\log\biggl(\frac{n+1}{\sqrt{\pi}p}\biggr) - \frac{1}{2}\log\biggl[1 + \log\biggl(\frac{n+1}{\sqrt{\pi}p}\biggr)\biggr] + 1 + o(1)\biggr\} + O(1).$$

Пусть $U_1, \ldots, U_n$ – независимые случайные величины, равномерно распределенные на отрезке $[0,1]$. Тогда (см., например, [14])

$$U_{(k)} \stackrel{\mathcal{D}}{=} 1 - \sum_{s=1}^{k}\chi_s \Bigg/ \sum_{s=1}^{n+1}\chi_s. \tag{22}$$

Далее воспользуемся следующим очевидным тождеством:

$$\frac{\xi_{(k)}^2}{2} \stackrel{\mathcal{D}}{=} F^{-1}(U_{(k)}), \tag{23}$$

где $F^{-1}(\cdot)$ – функция, обратная к

$$F(x) = \mathbf{P}\biggl\{\frac{\xi_i^2}{2} \leqslant x\biggr\} = 1 - G(x).$$

Тогда из (22) и (23) получаем

$$\frac{\xi^2_{(k)}}{2} \overset{\mathcal{D}}{=} G^{-1}(1-U_{(k)}) \overset{\mathcal{D}}{=} G^{-1}\left[\frac{k}{n+1} \times \left(\frac{1}{k}\sum_{s=1}^{k}\chi_s\right) \Big/ \left(\frac{1}{n+1}\sum_{s=1}^{n+1}\chi_s\right)\right]. \tag{24}$$

Для $x^{n+1} \in \mathbb{R}^{n+1}$ обозначим для краткости

$$\mu_k\bigl(x^{n+1}\bigr) = \frac{1}{k}\sum_{s=1}^{k} x_s.$$

Тогда из (24) получаем

$$\frac{\xi^2_{(k)}}{2} \overset{\mathcal{D}}{=} r^{-1}\left[\log\left(\frac{n+1}{\sqrt{\pi}k}\right) - \log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr] + \log\bigl[\mu_{n+1}\bigl(\chi^{n+1}\bigr)\bigr]\right]. \tag{25}$$

Далее разложим правую часть соотношения (25) по формуле Тейлора в точке $\log[(n+1)/(\sqrt{\pi}k)]$. Это можно сделать, если, например, неравенства

$$\Bigl|\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\Bigr| + \Bigl|\log\bigl[\mu_{n+1}\bigl(\chi^{n+1}\bigr)\bigr]\Bigr| \leqslant \frac{1}{1+\varepsilon}\log\left(\frac{n+1}{\sqrt{\pi}p}\right), \quad k = 1,\ldots,p,$$

выполняются с большими вероятностями.

Введем следующее подмножество в $\mathbb{R}^{n+1}$:

$$\Omega^n_{\varepsilon,p} = \bigcap_{k=1}^{n+1}\left\{x \in \mathbb{R}^{n+1} : \Bigl|\log\bigl[\mu_k\bigl(x^{n+1}\bigr)\bigr]\Bigr| + \Bigl|\log\bigl[\mu_{n+1}\bigl(x^{n+1}\bigr)\bigr]\Bigr| \leqslant \frac{1}{1+\varepsilon}\log\left(\frac{n+1}{\sqrt{\pi}p}\right)\right\}.$$

Из леммы 3 имеем

$$\mathbf{P}\left\{\Bigl|\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\Bigr| + \Bigl|\log\bigl[\mu_{n+1}\bigl(\chi^{n+1}\bigr)\bigr]\Bigr| \leqslant \frac{1}{1+\varepsilon}\log\left(\frac{n+1}{\sqrt{\pi}p}\right)\right\} \leqslant$$
$$\leqslant \exp\left[-Ck\left(\frac{n}{\sqrt{\pi}p}\right)^{1/(1+\varepsilon)}\right],$$

где здесь и далее $C > 0$ – некоторая постоянная, и следовательно,

$$\mathbf{P}\bigl\{\chi^{n+1} \notin \Omega^n_{\varepsilon,p}\bigr\} \leqslant \exp\left[-C\left(\frac{n}{\sqrt{\pi}p}\right)^{1/(1+\varepsilon)}\right]. \tag{26}$$

Применяя формулу Тейлора, из (25) получаем, что если $\chi^n \in \Omega^n_{\varepsilon,p}$, то

$$\frac{\xi^2_{(k)}}{2} \overset{\mathcal{D}}{=} G^{-1}\left(\frac{n+1}{k}\right) - \bigl[1+o(1)\bigr]\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr] + \delta_k(\chi^n). \tag{27}$$

где

$$\delta_k(\chi^n) = [1+o(1)]\log\bigl[\mu_{n+1}\bigl(\chi^{n+1}\bigr)\bigr] + o(1)\log^2\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr] +$$
$$+ o(1)\log^2\bigl[\mu_{n+1}\bigl(\chi^{n+1}\bigr)\bigr].$$

Нетрудно проверить, что

$$\mathbf{E}\,\delta^2_k(\chi^n) \leqslant \frac{C}{n} + \frac{C}{k^2}. \tag{28}$$

Из (27) получаем

$$\sum_{k=1}^{p}\left[\frac{\xi_{(k)}^2}{2}-G^{-1}\left(\frac{n+1}{k}\right)\right]=\sum_{k=1}^{p}\left[\frac{\xi_{(k)}^2}{2}-G^{-1}\left(\frac{n+1}{k}\right)\right]\mathbf{1}\bigl\{\chi^{n+1}\in\Omega_{\varepsilon,p}^{n}\bigr\}+$$

$$+\sum_{k=1}^{p}\left[\frac{\xi_{(k)}^2}{2}-G^{-1}\left(\frac{n+1}{k}\right)\right]\mathbf{1}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}=$$

$$=-(1+o(1))\sum_{k=1}^{p}\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]+\sum_{k=1}^{p}\delta_k(\chi^{n+1})\mathbf{1}\bigl\{\chi^{n+1}\in\Omega_{\varepsilon,p}^{n}\bigr\}+$$

$$+(1+o(1))\sum_{k=1}^{p}\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\mathbf{1}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}+$$

$$+\sum_{k=1}^{p}\left[\frac{\xi_{(k)}^2}{2}-G^{-1}\left(\frac{n+1}{k}\right)\right]\mathbf{1}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}. \tag{29}$$

С помощью неравенства Коши – Буняковского и (28) получаем, что

$$\frac{1}{\sqrt{p}}\,\mathbf{E}\sum_{k=1}^{p}\bigl|\delta_k(\chi^{n+1})\bigr|\mathbf{1}\bigl\{\chi^{n+1}\in\Omega_{\varepsilon,p}^{n}\bigr\}\leqslant C\left(\sqrt{\frac{p}{n}}+\frac{\log(p)}{\sqrt{p}}\right)\to 0,\quad n\to\infty.$$

Очевидно также, что

$$\frac{1}{\sqrt{p}}\sum_{k=1}^{3}\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\mathbf{1}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}\xrightarrow{\mathcal{D}}0,\quad n\to\infty, \tag{30}$$

и в силу (26), леммы 3 и неравенства Коши – Буняковского

$$\frac{1}{\sqrt{p}}\,\mathbf{E}\sum_{k=4}^{p}\Bigl|\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\Bigr|\mathbf{1}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}\leqslant$$

$$\leqslant C\sqrt{\mathbf{P}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}}\,\frac{1}{\sqrt{p}}\sum_{k=4}^{p}\frac{1}{\sqrt{k}}\leqslant C\sqrt{\mathbf{P}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}}\to 0,\quad n\to\infty. \tag{31}$$

Наконец, для последнего слагаемого в (29), принимая во внимание второе условие в (2), получаем

$$\mathbf{E}\sum_{k=1}^{p}\left[\frac{\xi_{(k)}^2}{2}-G^{-1}\left(\frac{n+1}{k}\right)\right]\mathbf{1}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}\leqslant$$

$$\leqslant\mathbf{E}\,\mathbf{1}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}\sum_{k=1}^{n}\xi_{(k)}^2+\mathbf{P}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}\sum_{k=1}^{n}G^{-1}\left(\frac{n+1}{k}\right)\leqslant$$

$$\leqslant n\sqrt{\mathbf{P}\bigl\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^{n}\bigr\}}\leqslant n\exp\left[-C\left(\frac{n}{\sqrt{\pi}p}\right)^{1/(1+\varepsilon)}\right]\to 0,\quad n\to\infty. \tag{32}$$

Таким образом, в силу (29)–(32) осталось проверить, что

$$\frac{1}{\sqrt{2p}}\sum_{k=1}^{p}\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\xrightarrow{\mathcal{D}}\mathcal{N}(0,1),\quad n\to\infty. \tag{33}$$

Аналогично (30)–(32) получаем, что

$$\frac{1}{\sqrt{2p}}\sum_{k=1}^{p^{1-\varepsilon}}\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\xrightarrow{\mathcal{D}}0,\quad n\to\infty,\tag{34}$$

и

$$\frac{1}{\sqrt{2p}}\sum_{k=p^{1-\varepsilon}}^{p}\Bigl[\mathbf{E}\log^4\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\Bigr]^{1/4}\leqslant C.\tag{35}$$

Чтобы доказать (33), воспользуемся формулой Тейлора, точнее, аппроксимацией

$$\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]=\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)+O(1)\left[\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)\right]^2,\quad k>p^{1-\varepsilon},\tag{36}$$

которая будет верна, если, например, неравенство

$$\left|\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)\right|\leqslant 0{,}5$$

выполняется при всех $k>p^{1-\varepsilon}$.

Из леммы 3 получаем, что

$$\mathbf{P}\left\{\left|\frac{1}{\sqrt{k}}\sum_{s=1}^{k}(\chi_s-1)\right|\geqslant 0{,}5\right\}\leqslant\exp(-Ck),$$

и поэтому

$$\mathbf{P}\left\{\max_{k>p^{1-\varepsilon}}\left|\frac{1}{\sqrt{k}}\sum_{s=1}^{k}(\chi_s-1)\right|\geqslant 0{,}5\right\}\leqslant\exp(-Cp^{1-\varepsilon}).$$

Отсюда и из (34)–(36) получаем

$$\frac{1}{\sqrt{2p}}\sum_{k=1}^{p}\log\left(\frac{1}{k}\sum_{s=1}^{k}\chi_s\right)-\frac{1}{\sqrt{2p}}\sum_{k=1}^{p}\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)\xrightarrow{\mathcal{D}}0,\quad n\to\infty.$$

Поэтому для завершения доказательства осталось проверить, что

$$\frac{1}{\sqrt{2p}}\sum_{k=1}^{p}\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1),\quad p\to\infty.\tag{37}$$

Очевидно, что

$$\sum_{k=1}^{p}\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)=\sum_{s=1}^{p}(\chi_s-1)w(s,p),\tag{38}$$

где веса $w(s,p)$ определены в (7). Заметим также, что

$$w(s,p)=\int\limits_s^p\frac{1}{x}\,dx+\int\limits_s^p O\Bigl(\frac{1}{x^2}\Bigr)\,dx=\log\Bigl(\frac{p}{s}\Bigr)+O\Bigl(\frac{1}{s}\Bigr)+O\Bigl(\frac{1}{p}\Bigr).$$

Отсюда, интегрируя по частям, получаем

$$\sum_{s=1}^{p} w^2(s,p) = p\int_{1/p}^{1} \log^2(x)\,dx + O(1) = 2p + O\big(\log^2(p)\big),$$

$$\sum_{s=1}^{p} w^3(s,p) = -p\int_{1/p}^{1} \log^3(x)\,dx + O(1) = 6p + O(\log^3(p)), \tag{39}$$

$$\sum_{s=1}^{p} w^4(s,p) = p\int_{1/p}^{1} \log^4(x)\,dx + O(1) = 24p + O(\log^4(p)).$$

Поэтому с помощью формулы Тейлора находим, что для всех $|\lambda| \leqslant (1-\varepsilon)/\log(p)$

$$\begin{aligned}&\mathbf{E}\exp\left[\lambda\sum_{s=1}^{p}(\chi_s-1)w(s,p)\right] = \exp\left[-\sum_{s=1}^{p}\log[1-\lambda w(s;p)] - \lambda\sum_{s=1}^{p}w(s,p)\right] = \\ &= \exp\left[\frac{\lambda^2}{2}\sum_{s=1}^{p}w^2(s,p) + \frac{\lambda^3}{3}\sum_{s=1}^{p}w^3(s,p) + \frac{\lambda^4}{4}\sum_{s=1}^{p}w^4(s,p) + O\Big(\frac{\lambda^5 p}{\varepsilon^5}\Big)\right] = \\ &= \exp\big\{(1+o(1))p\lambda^2\big\}.\end{aligned} \tag{40}$$

Это соотношение очевидным образом доказывает (37) и, следовательно, теорему 1. ▲

Доказательство леммы 1 практически непосредственно вытекает из (25) и выпуклости функции

$$f(x) = r^{-1}\left[\log\left(\frac{1}{\sqrt{\pi}k}\sum_{s=1}^{n+1}\chi_s\right) - \log(1+x)\right]. \quad ▲$$

Доказательство леммы 2 основано на методе Лапласа [15], который используется для вычисления интеграла в правой части следующего тождества:

$$\begin{aligned}&\mathbf{P}\big\{\zeta(\theta^n,p',q) \geqslant y\big\} = \\ &= \frac{1}{2\pi i}\int_{-\infty}^{\infty}\exp\big\{-ity - \log(t) + \log\big[\mathbf{E}\exp\big(it\zeta(\theta^n,p',q)\big)\big]\big\}\,dt = \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\Big\{-it - \log(i\sqrt{2\pi}t) + \log\big[\mathbf{E}\exp\big(ity^{-1}\zeta(\theta^n,p',q)\big)\big]\Big\}\,dt.\end{aligned} \tag{41}$$

Мы опустим технические детали этого метода, чтобы не загромождать изложение, и сосредоточимся на его принципиальных элементах.

С помощью формулы Тейлора находим (см. (40)), что при $|\lambda| \leqslant (1-\varepsilon)/w(1,q-p')$

$$\begin{aligned}&\log\big\{\mathbf{E}\exp\big[\lambda\zeta(\theta^n,p',q)\big]\big\} = \frac{\lambda p'}{2} - \frac{p'}{2}\log(1+\lambda) + \frac{\lambda^2\|\theta^n\|_{p'}^2}{2(1+\lambda)} - \\ &- \sum_{s=1}^{q-p'}\big\{\log[1-\lambda w(s,q-p')] + \lambda w(s,q-p')\big\} =\end{aligned}$$

$$
\begin{aligned}
&= \frac{\lambda^2}{2}\left[\frac{\|\theta^n\|_{p'}^2}{1+\lambda} + \frac{p'}{2} + \sum_{s=1}^{q-p'} w^2(s,q-p')\right] + \frac{\lambda^3}{3}\sum_{s=1}^{q-p'} w^3(s,q-p') + \lambda^4 O\left(\frac{q-p'}{\varepsilon^4}\right) = \\
&= \frac{\lambda^2}{2}D(\lambda) + \lambda^3 M + \lambda^4 O\left(\frac{q-p'}{\varepsilon^4}\right), \qquad (42)
\end{aligned}
$$

где функция $D(\cdot)$ определена в (12), а (см. (39))

$$
M = \frac{1}{3}\sum_{s=1}^{q-p'} w^3(s,q-p') = 2(q-p')(1+o(1)), \quad q-p' \to \infty.
$$

Полагая в (42) $\lambda = it/y$, получаем

$$
\log\Big\{\mathbf{E}\exp\big[ity^{-1}\zeta(\theta^n,p',q)\big]\Big\} = -\frac{t^2}{2y^2}D\left(\frac{it}{y}\right) + \frac{it^3}{y^3}M + O\left(\frac{t^4(q-p')}{y^4\varepsilon^4}\right). \qquad (43)
$$

Метод Лапласа основан на квадратичной аппроксимации (см. (41) и (43)) функции

$$
\begin{aligned}
&F(t) = -it - \log(i\sqrt{2\pi}t) + \log\big[\mathbf{E}\exp\big(ity^{-1}\zeta(\theta^n,p',q)\big)\big] = \\
&= -it - \frac{t^2}{2y^2}D\left(\frac{it}{y}\right) + i\frac{t^3}{y^3}M - \log(i\sqrt{2\pi}t) + O\left(\frac{t^4(q-p')}{y^4\varepsilon^4}\right) \qquad (44)
\end{aligned}
$$

в окрестности некоторой точки t_y, находящейся вблизи точки экстремума $F(\cdot)$. В качестве такой точки можно выбрать, например,

$$
t_y = -\frac{iy^2}{D(0)}.
$$

Заметим, что поскольку в рассматриваемом случае $y = \sqrt{2xD(0)}$, то

$$
t_y = -2ix \quad \text{и} \quad \frac{t_y}{y} = -i\sqrt{\frac{2x}{D(0)}}.
$$

Тогда из (13) и (44) получаем

$$
\begin{aligned}
&F(t_y) = -x + x\left[\frac{1}{D(0)}D\left(\sqrt{\frac{2x}{D(0)}}\right) - 1\right] - \log\big(2\sqrt{2\pi}x\big) + \\
&+ O\left(\frac{x(q-p')}{D(0)w(1;q-p')}\right), \\
&F'(t_y) = -i + \frac{i}{D(0)}\left[D\left(\sqrt{\frac{2x}{D(0)}}\right) + \frac{1}{2}\sqrt{\frac{2x}{D(0)}}D'\left(\sqrt{\frac{2x}{D(0)}}\right) - D(0)\right] - \frac{1}{x} + \\
&+ O\left(\frac{q-p'}{D(0)\sqrt{w(1;q-p')}}\right), \qquad (45) \\
&F''(t_y) = -\frac{1}{2xD(0)}\left[D\left(\sqrt{\frac{2x}{D(0)}}\right) + \frac{3}{2}\sqrt{\frac{2x}{D(0)}}D'\left(\sqrt{\frac{2x}{D(0)}}\right) + \right. \\
&\left. + \frac{2x}{D(0)}D''\left(\sqrt{\frac{2x}{D(0)}}\right)\right] - \frac{1}{x^2} + O\left(\frac{q-p'}{xD(0)\sqrt{w(1;q-p')}}\right).
\end{aligned}
$$

Нетрудно проверить, что

$$\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}\exp\left[\frac{F''(t_y)}{2}(t-t_y)^2+F(t_y)(t-t_y)+F(t_y)\right]dt=$$
$$=\exp\left[F(t_y)-\frac{[F'(t_y)]^2}{2F''(t_y)}-\frac{1}{2}\log[-F''(t_y)]\right]. \tag{46}$$

Поэтому заметив, что

$$\frac{1}{D(0)}D\left(\sqrt{\frac{2x}{D(0)}}\right)-1\leqslant 0,$$

из (45) находим

$$F(t_y)-\frac{[F'(t_y)]^2}{2F''(t_y)}-\frac{1}{2}\log[-F''(t_y)]\leqslant -x-\frac{1}{2}\log(x)+O\left(\frac{x(q-p')}{w(1,q-p')}\right).$$

Это неравенство вместе с (41) и (46) доказывает (14). ▲

В заключение автор выражает благодарность анонимному рецензенту за замечания, способствовавшие улучшению статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Zhang C., Bengio S., Hardt M., Recht B., Vinyals O.* Understanding Deep Learning (Still) Requires Rethinking Generalization // Commun. ACM. 2021. V. 64. № 3. P. 107–115. https://doi.org/10.1145/3446776
2. *Belkin M.* Fit without Fear: Remarkable Mathematical Phenomena of Deep Learning through the Prism of Interpolation // Acta Numer. 2021. V. 30. P. 203–248. https://doi.org/10.1017/S0962492921000039
3. *Belkin M., Hsu D., Xu J.* Two Models of Double Descent for Weak Features // SIAM J. Math. Data Sci. 2020. V. 2. № 4. P. 1167–1180. https://doi.org/10.1137/20M1336072
4. *Dar Y., Muthukumar V., Baraniuk R.G.* A Farewell to the Bias-Variance Tradeoff? An Overview of the Theory of Overparameterized Machine Learning, https://arxiv.org/abs/2109.02355 [stat.ML], 2021.
5. *Добрушин Р.Л.* Одна статистическая задача теории обнаружения сигнала на фоне шума в многоканальной системе, приводящая к устойчивым законам распределения // Теория вероятн. и ее примен. 1958. Т. 3. № 2. С. 173–185. https://www.mathnet.ru/rus/tvp4928
6. *Бурнашев М.В., Бегматов И.А.* Об одной задаче обнаружения сигнала, приводящей к устойчивым распределениям // Теория вероятн. и ее примен. 1990. Т. 35. № 3. С. 557–560. https://www.mathnet.ru/rus/tvp1261
7. *Ingster Yu.I., Suslina I.A.* Nonparametric Goodness-of-Fit Testing Under Gaussian Models // Lect. Notes Statist. V. 169. New York: Springer-Verlag, 2003. https://doi.org/10.1007/978-0-387-21580-8
8. *Bonferroni C.E.* Teoria statistica delle classi e calcolo delle probabilità // Pubbl. del R. Ist. Super. di Sci. Econ. e Commer. di Firenze. V. 8. Firenze: Seeber, 1936.
9. *Benjamini Y., Hochberg Y.* Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing // J. Roy. Statist. Soc. Ser. B. 1995. V. 57. № 1. P. 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
10. *Benjamini Y.* Simultaneous and Selective Inference: Current Successes and Future Challenges // Biom. J. 2010. V. 52. № 6. P. 708–721. https://doi.org/10.1002/bimj.200900299

11. *Donoho D., Jin J.* Higher Criticism Thresholding: Optimal Feature Selection When Useful Features are Rare and Weak // Proc. Natl. Acad. Sci. U.S.A. 2008. V. 105. № 39. P. 14790–14795. https://doi.org/10.1073/pnas.0807471105
12. *Anderson T.W.* The Integral of a Symmetric Unimodal Function over a Symmetric Convex Set and Some Probability Inequalities // Proc. Amer. Math. Soc. 1955. V. 6. № 2. P. 170–176. https://doi.org/10.1090/S0002-9939-1955-0069229-1
13. *Ибрагимов И.А., Хасьминский Р.З.* Асимптотическая теория оценивания. М.: Наука, 1979.
14. *Pyke R.* Spacings // J. Roy. Statist. Soc. Ser. B. 1965. V. 27. № 3. P. 395–436; 437–449 (discussion). https://doi.org/10.1111/j.2517-6161.1965.tb00602.x; https://doi.org/10.1111/j.2517-6161.1965.tb00603.x
15. *Федорюк М.В.* Асимптотика: интегралы и ряды. М.: Наука, 1987.

Голубев Георгий Ксенофонтович
Институт проблем передачи информации
им. А.А. Харкевича РАН, Москва
golubev.yuri@gmail.com

Поступила в редакцию
16.05.2022
После доработки
06.12.2022
Принята к публикации
03.01.2023