Записки научных
семинаров ПОМИ
Том 368, 2009 г.

И. А. Ибрагимов

ОБ АДАПТИВНОМ ОЦЕНИВАНИИ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ

1. Прежде чем переходить к задачам оценивания плотности распределения рассмотрим более простую задачу оценивания сигнала в гауссовском белом шуме. Именно, допустим, что на интервале $[0,1]$ наблюдается случайная функция

$$X_\varepsilon(t) = \int_0^t f(s)\,ds + \varepsilon w(t),$$

$w(t)$ – Винеровский процесс, f – подлежащая оценке неизвестная функция. Параметр ε предполагается известным. Априори известно, что неизвестная функция f принадлежит заданному (известному) классу функций $F \subset L_2(0,1)$. Функция потерь оценки $\widehat{f}$ задается L_2-нормой уклонения $\widehat{f}$ от f и задача заключается в построении оценок $\widehat{f}$ с достаточно малой функцией риска $E_f\{\|\widehat{f} - f\|^2\}$, $\|\cdot\|$ обозначает норму в $L_2(0,1)$.

Один естественный способ оценивания, проекционные оценки, был предложен в 1961 г. Н. Н. Ченцовым ([1, 2]). Он заключается в следующем. Пусть $e_1, e_2, \ldots$ – какой-нибудь ортонормированный базис в $L_2(0,1)$, так что

$$f(t) = \sum_1^\infty a_j \cdot e_j(t), \quad a_j = (f, e_j) = \int_0^1 f(t) e_j(t)\,dt.$$

Естественной оценкой для a_j является

$$X_j = X_j(e_j) = \int_0^1 e_j(t) dX_\varepsilon(t).$$

Ключевые слова : плотность распределения, непараметрическое оценивание, адаптивная оценка.

Настоящая работа частично поддержана грантами РФФИ 08-01-00692, РФФИ-ННИО 09-01-9131, НШ 638.2008.1 и Программой фундаментальных исследований ОМ РАН “Современные проблемы теоретической математики” .

Если теперь положить $\widehat{f}_N(t) = \sum\limits_1^N X_j \cdot e_j(t)$, то риск этой оценки

$$E_f \|\widehat{f}_N - f\|^2 = \varepsilon^2 N + \sum_{N+1}^{\infty} |a_j|^2.$$

Естественно далее выбрать такой базис e_j, для которого $\sup\limits_{f\in\mathbb{F}} \sum\limits_{N+1}^{\infty} |a_j|^2$ минимален, а затем взять в качестве оценки f_ε^* для f такую оценку $\widehat{f}_N$, для которой $\varepsilon^2 N$ приблизительно совпадает с $\sum\limits_{N+1}^{\infty} |a_j|^2$.

Например, пусть $\mathbb{F}$ есть эллипсоид

$$\mathbb{F}(\beta; \{e_i\}; L) = \Big\{ f = \sum_1^\infty a_j e_j : \sum_1 |a_j|^2 |j|^{2\beta} \leqslant L^2 \Big\}.$$

Ясно, что в этом случае

$$E_f \|\widehat{f}_N - f\|^2 \leq \varepsilon^2 N + L^2 N^{-2\beta}$$

и, полагая $N(\varepsilon)$ равным целой части $\left(\frac{L}{\varepsilon}\right)^{\frac{2}{2\beta+1}}$, мы найдем, что для $f_\varepsilon^* = \widehat{f}_{N(\varepsilon)}$

$$\sup_{f\in\mathbb{F}} E_f \|f_\varepsilon^* - f\|^2 \leq 2L^{\frac{2}{2\beta+1}} \varepsilon^{\frac{4\beta}{2\beta+1}}. \tag{1}$$

С другой стороны, можно доказать, что для любой оценки $\widehat{f}$

$$\sup_{f\in\mathbb{F}} E_f \|\widehat{f} - f\|^2 \geq c L^{\frac{2}{2\beta+1}} \varepsilon^{\frac{4\beta}{2\beta+1}}, \tag{2}$$

c – некоторая константа (см., например, [3, Теорема 7.5.3]).

Таким образом, рассмотренный метод приводит к достаточно хорошим оценкам. Его недостаток заключается в том, что для конструкции оценки необходимо знать априори не только базис $\{e_i\}$, что еще достаточно приемлемо (например, если $\mathbb{F}$ состоит из гладких 1-периодических функций, то естественный базис составляют экспоненты $e^{2\pi i j t}$, $j = 0, \pm 1, \dots$), но и параметры β и L.

Позднее был предложен естественный и весьма эффективный метод, позволяющий избавиться от последнего недостатка. Этот метод

заключается в следующем. Выберем некоторый базис $\{e_i\}$ в L_2, положим $N = [\varepsilon^{-2}]$ и определим оценку

$$\widetilde{f}_\varepsilon(t) = \sum_1^N \overline{X}_j e_j, \tag{3}$$

где на этот раз

$$\overline{X}_j = \begin{cases} X_j = \int\limits_0^1 e_j(t)\, dX_\varepsilon(t), & \text{если } |X_j| \geq \lambda\varepsilon; \\ 0, & \text{если } |X_j| < \lambda\varepsilon, \end{cases} \tag{4}$$

а порог $\lambda = \lambda(\varepsilon) = 8\sqrt{\ln \frac{1}{\varepsilon}} + 1$.

Конструкция оценки $\widetilde{f}_\varepsilon$ уже зависит только от базиса. В то же время риск оценки $\widetilde{f}_\varepsilon$, вообще говоря, отличается от риска f_ε^* лишь на $\ln 1/\varepsilon$. Приведем два примера вычисления риска $\widetilde{f}_\varepsilon$.

Пусть $\{e_j\}$ – ортонормальный базис в L_2. Пусть $\varphi(x) \geq 1$ – положительная монотонно возрастающая до бесконечности функция, определенная на $[0, \infty)$. Обозначим $\mathcal{E}(\varphi, L)$ эллипсоид

$$\mathcal{E}(\varphi, L) = \Big\{ f \in L_2 : f = \sum_1^\infty a_j e_j, \quad \sum_1^\infty a_j^2 \varphi^2(j) \leq L^2 \Big\}.$$

Положим $H_0(x) = x\varphi^2(x)$. Легко видеть, что можно построить проективную оценку $f_\varepsilon^* = \widehat{f}_{N(\varepsilon)}$ такую, что для всех $f \in \mathcal{E}(\varphi, L)$

$$E_f \|\widehat{f}_\varepsilon - f\|^2 \leq 2\varepsilon^2 H_0^{-1}\left(\frac{L^2}{\varepsilon^2}\right),$$

где вообще $G^{-1}(x)$ означает функцию обратную к G. Очевидно, $H_0(x) \geq H(x) = \int_0^x \varphi^2(y)\, dy$, так что $H_0^{-1}(x) \leq H^{-1}(x)$,

$$E_f \|\widehat{f}_\varepsilon - f\|^2 \leq 2\varepsilon^2 H_0^{-1}\left(\frac{L^2}{\varepsilon^2}\right) \leq 2L^2\varepsilon^2 H^{-1}\left(\frac{1}{\varepsilon^2}\right). \tag{5}$$

Теорема 1. *Пусть* $\mathbb{F} \subseteq \mathcal{E}(\varphi, L)$. *Пусть* $\widetilde{f}_\varepsilon$ *адаптивная оценка* (3). *Тогда для всех* $f \in \mathbb{F}$

$$E_f \|\widetilde{f}_\varepsilon - f\|^2 \le C(L^2+1)\varepsilon^2 |\ln \varepsilon| H^{-1}(\varepsilon^{-2}), \tag{6}$$

где C *– абсолютная постоянная.*

Рассмотрим еще вариант теоремы 1, когда элементы ортонормальной системы заиндексированы мультииндексом $j = (j_1, \dots, j_d)$, пробегающим множество всех целых векторов $(j_1, \dots, j_d)$. Такая задача возникает при наблюдении функции $f \in L_2(G)$, $G \subseteq R^d$, в гауссовском белом шуме интенсивности ε. Схема наблюдений выглядит следующим образом. Для любой функции $\varphi \in L_2(G)$ наблюдается

$$X_\varepsilon(\varphi) = \int\limits_G \varphi(t) f(t)\, dt + \xi_\varepsilon(\varphi),$$

где $\xi_\varepsilon(\varphi)$ гауссовские с $E\xi_\varepsilon(\varphi) = 0$ и $E\xi_\varepsilon(\varphi)\xi_\varepsilon(\psi) = \int\limits_G \varphi(t)\psi(t)\, dt$. Если, например, $G = [0,1]^d$, а параметрическое множество состоит из достаточно гладких периодических функций на G, то естественным образом возникает базис $\{e^{2\pi i(j,t)}\}$ и классы функций вида

$$f(x) = \sum a_j e^{2\pi i(x)}, \quad \sum |a_j|^2 (|j_1|^{2\beta_1} + \dots + |j_d|^{2\beta_d}) \le L^2.$$

Пусть $\{e_n\}$ – какой-нибудь ортонормальный базис в $L_2(G)$, индексированный мультииндексом n. Пусть функция $\psi(y_1, \dots, y_d) = \psi(y)$ определена для $y_j \ge 0$ и удовлетворяет следующим условиям.

1. $\psi(y) \ge 1$; 2. $\psi(y)$ монотонно возрастает по всем аргументам y_j, $j = 1, \dots, d$; 3. $\psi(y) \to \infty$, $y \to \infty$. Положим

$$\varphi(x) = \varphi(x_1, \dots, x_d) = \psi(|x_1|, \dots, |x_d|), \quad x \in R^d.$$

Обозначим $\mathcal{E}(\varphi, L)$ следующее подмножество $L_2(G)$

$$\mathcal{E}(\varphi, L) = \Big\{ f \in L_2(G), \quad f = \sum_j a_j e_j \ : \ \sum |a_j|^2 \varphi^2(j) \le L^2 \Big\}.$$

Для заданной функции $\varphi(x)$ определим функции $a(t)$, $H(t)$ равенствами

$$a(t) = \#\{n : \varphi(n) < t\}, \quad H(t) = \int\limits_0^t [a^{-1}(s)]^2\, ds.$$

Определим оценку $\widetilde{f}_\varepsilon$ равенством

$$\widetilde{f}_\varepsilon(x) = \sum_{j=(j_1,\dots,jd),|j_k|\le N} \overline{X}_j e_j(x), \quad N = [\varepsilon^{-2}],$$

где, как и выше, коэффициенты

$$\overline{X}_j = \begin{cases} X_\varepsilon(e_j), & \text{если } |X_\varepsilon(e_j)| \ge \lambda\varepsilon; \\ 0, & \text{если } |X_\varepsilon(e_j)| < \lambda\varepsilon, \end{cases}$$

а порог $\lambda = 8\sqrt{(\ln\varepsilon)} + 1$.

Теорема 2. *Если* $\mathbb{F} \subseteq \mathcal{E}(\varphi, L)$, *то для всех* $f \in \mathbb{F}$

$$E_f\|\widetilde{f}_\varepsilon - f\|^2 \le Cd\varepsilon^2|\ln(\varepsilon)|H^{-1}(L^2\varepsilon^{-2}). \tag{7}$$

Возможно, в теореме 2 и присутствует некоторый элемент новизны, но мы не будем останавливаться на доказательстве теорем 1, 2. Тем более, что дальнейшее усовершенствование конструкции позволило построить адаптивные оценки, качество которых не хуже качества f^*_ε. О построении таких оценок см., например, [4], [5] и литературу там.

Основное содержание настоящей статьи – следующий пункт, где мы рассмотрим континуальный аналог указанного выше метода на примере оценивания плотности распределения.

2. Ниже мы рассматриваем задачу оценивания неизвестной плотности распределения $f(x)$, $f \in L_2(R^2)$, по выборке

$$X_1, \dots, X_n \tag{8}$$

объема n, так что $X_1, \dots, X_n$ – независимые d-мерные случайные векторы. Относительно неизвестной плотности $f(x)$ будет предполагаться априори известным, что она принадлежит континуальным аналогам классов $\mathcal{E}(\varphi, L)$ предыдущего пункта, к определению которых мы сейчас перейдем.

Пусть $\psi(t) = \psi(t_1, \dots, t_d)$, $t_i \ge 0$ – монотонно возрастающая по каждому аргументу функция такая, что $\psi(t) \ge 1$, $\psi(t) \to \infty$, $t \to \infty$. Положим

$$\varphi(t) = \varphi(t_1, \dots, t_d) = \psi(|t_1|, \dots, |t|_d), \quad t \in R^d.$$

Если функция $g \in R^d$, обозначим $\widetilde{g}$ ее преобразование Фурье,

$$\widetilde{g}(t) = \int\limits_{R^d} d^{i(t,x)} g(x)\, dx,$$

так что

$$g(x) = \frac{1}{(2\pi)^d} \int\limits_{R^d} e^{-1(t,x)} \widetilde{g}(t)\, dt.$$

Положим

$$\mathcal{E}(\varphi, L) = \Big\{ g \in L_2(R^d) : \int\limits_{R^d} |\widetilde{g}(t)|^2 |\varphi(t)|^2\, dt \le L^2 \Big\}. \tag{9}$$

Примерами таких классов служат классы гладких функций $g(x)$, для которых, например

$$\int\limits_{R^d} |\widetilde{g}(t)|^2 (1 + |t_1|^{2\beta_1} + \ldots + |t_d|^{2\beta d}) \leqslant L^2$$

или

$$\int\limits_{R^d} |\widetilde{g}(t)|^2 (1 + |t_1|^{2\beta_1}) \ldots (1 + |t_d|^{2\beta d}) \le L^2.$$

Еще один пример доставляют аналитические функции, для которых

$$\int\limits_{R^d} |\widetilde{g}(t)|^2 \exp\{\lambda_1, |t_1| + \ldots + |\lambda_d| t|^d\}\, dt \leqslant L^2.$$

Обозначим $h(t)$ характеристическую функцию величин X_j, соответствующую плотности распределения $f(x)$. Тогда условие $f \in \mathcal{E}(\varphi, L)$ означает, что

$$\int\limits_{R^d} |h(t)|^2 |\varphi(t)|^2\, dt \le L^2. \tag{10}$$

Пусть $\mathcal{P}_n$ – эмпирическое распределение выборки (8). Через $h_n(t)$ обозначим соответствующую ему эмпирическую характеристическую функцию

$$h_n(t) = \int\limits_{R^d} e^{i(t,x)} d\mathcal{P}_n = \frac{1}{n} \sum_{j=1}^{n} e^{i(t,X_j)}.$$

Континуальным аналогом проекционных оценок п. 1 будут, например, оценки

$$\widehat{f}_T(x) = \frac{1}{(2\pi)^d} \int\limits_{-T_1}^{T_1} \dots \int\limits_{-T_d}^{T_d} e^{-i(t,x)} h_n(t)\, dt, \quad T = (T_1, \dots, T_d).$$

Для таких оценок (здесь и ниже $\|\cdot\|$ означает норму в $L^2(R^d)$)

$$E_f \|\widehat{f}_T - f\|_2^2 = \frac{1}{(2\pi)^d} \left\{ \int\limits_{\Gamma(T)} E_f |h_n(t) - h(t)|^2\, dt + \int\limits_{\Gamma(T)} |h(t)|^2\, dt \right\}, \quad (11)$$

где $\Gamma(T)$ – параллелепипед $[-T_1, T_1] \times \ldots \times [-T_d, T_d]$. Легко видеть, что

$$E_f |h_n(t) - h(t)|^2 = \frac{1}{n}(1 - |f(t)|^2),$$

а

$$\int\limits_{\Gamma(T)} |h(t)|^2\, dt \le \frac{L^2}{m_T^2}, \quad m_T = \inf_{t \in \Gamma(T)} |\varphi(t)|.$$

Таким образом,

$$E_f \|\widehat{f}_T - f\|_2^2 \le \frac{\mathrm{vol}(\Gamma(T))}{n} + \frac{L^2}{m^2}. \quad (12)$$

Выбирая $T(n)$ так, чтобы $\frac{\mathrm{vol}(\Gamma(T(n)))}{n} = \frac{L^2}{m_{T(n)}^2}$, и полагая $\widehat{f}_{T(n)} = f_n^*$, мы придем к оценке f_n^*, являющейся континуальным аналогом оценки f_ε^* п. 1. Для нее

$$E_f \|f_n^* - f\|^2 \le 2 \frac{L^2}{m_{T(n)}^2}.$$

Если, например, $\varphi(x_1, \dots, x_d) = 1 + |x_1|^{\beta_1} + \ldots + |x_d|^{\beta_d}$, то

$$E_f \|f_n^* - f\|^2 \le 2^{\frac{2d\beta+1}{2\beta+1}} \left(\frac{L^2}{d}\right)^{\frac{1}{2\beta+1}} \cdot n^{-\frac{2\beta}{2\beta+1}}, \quad (13)$$

где число β определяется из соотношения $\frac{1}{\beta} = \frac{1}{\beta_1} + \ldots + \frac{1}{\beta_d}$.

Укажем теперь аналог адаптивной оценки $\widetilde{f}_\varepsilon$. Обозначая эту оценку $\widetilde{f}_n(x)$, определим ее следующим образом:

$$\widetilde{f}_n(x) = \frac{1}{(2\pi)^d} \int\limits_{\Gamma_n} e^{-i(t,x)} \widetilde{h}_n(t)\, dt.$$

Здесь Γ_n – куб $[-n,n]^d$, а статистика

$$\widetilde{h}_n(t) = \begin{cases} h_n(t), & \text{если } |h_n(t)| > 8\sqrt{\frac{\ln n}{n}}; \\ 0, & \text{если } |h_n(t)| \le 8\sqrt{\frac{\ln n}{n}} \end{cases}.$$

Чтобы указать границы для функции риска оценки $\widetilde{f}_n$, введем функцию распределения $a(z)$ функции $\varphi(t)$,

$$a(z) = \int\limits_{\{t:\varphi(t)<x\}} dt$$

и положим

$$H(y) = \int\limits_0^y (a^{-1}(z))^2\, dz.$$

Положим также

$$A_\varphi \sup_i \sup_{t_i} \frac{\varphi(1,1,\dots,1,t_i,1,\dots,1)}{\varphi(0,0,\dots,0,t_i,0,\dots,0)}.$$

Последняя величина есть характеристика однородности функции $\varphi(t)$.

Теорема 3. *Если $f \in \mathcal{E}(\varphi, L)$, то*

$$E_f \|\widetilde{f}_n - f\|_2^2 \le C(L^2+1)(A_\varphi+1)\frac{\ln n}{n} H^{-1}(n),$$

C – абсолютная постоянная.

Доказательство. Аналогично (11) и (12),

$$E_f \|\widetilde{f}_n - f\|^2 \le \frac{1}{(2\pi)^d} \left\{ \int\limits_{\Gamma_n} E_f |\widetilde{h}_n(t) - h(t)|^2 + \frac{L^2}{m_n^2} \right\}, \tag{14}$$

где $m_n^2 = \inf\limits_{t\in\overline{\Gamma}_n} |\varphi(t)|$. Для оценки математического ожидания в правой части последнего неравенства нам понадобится следующее неравенство.

Лемма 1 (неравенство Хёффдинга). *Пусть $\xi_1, \dots, \xi_n$ – такие независимые случайные величины, что $a_k \le \xi_k \le b_k$. Тогда для любого $\alpha > 0$*

$$P\left\{\sum_1^n \xi_j - \sum_1^n E\xi_j \ge \alpha\right\} \le \exp\left\{-\frac{2\alpha^2}{\sum\limits_1^n (b_k - a_k)^2}\right\}. \tag{15}$$

Доказательство леммы, кроме оригинальной работы [6], можно найти в монографии А. Б. Цыбакова [5]; схема доказательства неравенства указана также в [7, стр. 25].

Полагая $\lambda = 8d\sqrt{\ln n}$, запишем интеграл в правой части (14) в виде

$$\int\limits_{\Gamma_n} E_f\left\{|h_n(t) - h(t)|^2 \mathbb{I}_{\{|h_n| > \frac{\lambda}{\sqrt{n}}\}}\right\} dt + \int\limits_{\Gamma_n} |h(t)|^2 P\left\{|h_n(t)| \le \frac{\lambda}{\sqrt{n}}\right\} dt. \tag{16}$$

Заметим теперь, что если $|h(t)| \le \frac{\lambda}{2\sqrt{n}}$, то, в силу (15),

$$\begin{aligned} E_f\left\{|h_n(t) - h(t)|^2 \mathbb{I}_{\{|h_n| > \frac{\lambda}{\sqrt{n}}\}}\right\} &\le 4P_f\left\{|h_n(t)| > \frac{\lambda}{\sqrt{n}}\right\} \\ &\le 4P_f\left\{|h_n(t) - h(t)| > \frac{\lambda}{2\sqrt{n}}\right\} \le 8e - \frac{\lambda^2}{32} \end{aligned} \tag{17}$$

(мы пользуемся тем, что либо $|\operatorname{Re} h_n - \operatorname{Re} h|$, либо $|\operatorname{Im} h_n - \operatorname{Im} h|$ больше $\frac{\lambda}{4\sqrt{n}}$).

Очевидно,

$$E_f\left\{|h_n(t) - h(t)|^2 \mathbb{I}\{|h_n| > \frac{\lambda}{\sqrt{n}}\}\right\} \le E_f\{|h_n - h|^2\} = \frac{1 - |h(t)|^2}{n} \le \frac{1}{n}. \tag{18}$$

Далее, если $|f(t)| \ge \frac{3\lambda}{2\sqrt{n}}$, то, подобно (17),

$$P_f\left\{|h_n(t)| \leqslant \frac{\lambda}{\sqrt{n}}\right\} \le 8e^{-\frac{\lambda^2}{32}}. \tag{19}$$

Наконец, тривиально

$$|h(t)|^2 P\left\{|h_n(t)| \le \frac{\lambda}{\sqrt{n}}\right\} \le |h(t)|^2. \tag{20}$$

Неравенства (17)–(20) показывают, что сумма интегралов в (16) не превосходит

$$\frac{c_1}{n} \int\limits_{\{|h(t)|>\frac{\lambda}{2\sqrt{n}}\}} dt + c_2 \int\limits_{\{|h(t)|\le\frac{3\lambda}{2\sqrt{n}}\}} |h(t)|^2\, dt + \frac{c_3}{n}, \tag{21}$$

где c_i – абсолютные постоянные.

Чтобы оценить интегралы в (21), нам понадобятся некоторые сведения о монотонных перестановках функций (см. [8, гл. 10]). Пусть $g(x)$ – вещественная измеримая функция, определенная на измеримом пространстве $(\mathfrak{X}, \mathfrak{A})$ с мерой μ на $\mathfrak{A}$. Пусть

$$M_g(y) = M(y) = \mu\{x : g(x) \ge y\},$$
$$m_g(y) = m(y) = \mu\{x : g(x) < y\}$$

– функции распределения g.

Функция $\bar{g}(z)$, обратная к $M_g(y)$, т.е.

$$\bar{g}(z) = \inf\big\{y \,:\, M_g(y) \le z\big\},$$

называется монотонно убывающей перестановкой функции $g(x)$.

Функция $\underline{g}(z)$, обратная к $m_f(y)$, называется монотонно возрастающей перестановкой $g(x)$.

Заметим, что функции $g, \bar{g}, \underline{g}$ имеют одинаковые функции распределения, так что для любой F

$$\int\limits_{\mathfrak{X}} F(f(x))\, d\mu = -\int\limits_{-\infty}^{\infty} F(u) dM(u) = \int\limits_{-\infty}^{\infty} F(u) dm(u)$$
$$= \int\limits_{-\infty}^{\infty} F(\bar{f}(z))\, dz = \int\limits_{-\infty}^{\infty} F(\underline{f}(z))\, dz. \tag{22}$$

Лемма 3. *Пусть $g_1(x)$, $g_2(x)$ – две неотрицательные измеримые функции на $\{\mathfrak{X}, \mathfrak{A}, \mu\}$. Тогда*

$$\int\limits_{\mathfrak{X}} g_1(x) g_2(x)\, dx \ge \int\limits_{0}^{\infty} \overline{g}_1(z) \underline{g}_2(z)\, dz. \tag{23}$$

Положим $\xi(s) = \int\limits_{\{|h(t)|>s\}} dt$. В силу (22) и (23)

$$L^2 \geq \int\limits_{R^d} |h(t)|^2 |\varphi(t)|^2\, dt \geq \int\limits_0^\infty |\bar{h}(z)|^2 |\underline{\varphi}(z)|^2\, dz$$

$$\geq s^2 \int\limits_{\{\bar{h}(z)>s\}} |\underline{\varphi}(z)|^2\, dz = s^2 \int\limits_0^{\xi(s)} (a^{-1}(\tau))^2\, d\tau = s^2 H(\xi(s)).$$

Следовательно,

$$\xi(s) \leq H^{-1}\left(\frac{b^2}{s^2}\right) \tag{24}$$

и потому первый интеграл в (21)

$$\int\limits_{\{|h(t)|>\frac{\lambda}{2\sqrt{n}}\}} = \xi\left(\frac{\lambda}{2\sqrt{n}}\right) \leq H^{-1}\left(\frac{4nL^2}{\lambda^2}\right). \tag{25}$$

Второй интеграл в (21) (интегрирование идет лишь по $t \in \Gamma_n$)

$$\int\limits_{\{|h(t)|\leq\frac{3\lambda}{2\sqrt{n}}\}} |h(t)|^2 dt \leq (2n)^d n^{-d-1} + \int\limits_{\{n^{\frac{-d-1}{2}}\leq|h(t)|\leq\frac{3\lambda}{2\sqrt{n}}\}}$$

$$\leq \frac{2^d}{n} + \sum_{p=1}^{d\ln n} \int\limits_{G_p} |h(t)|^2\, dt,$$

интегрирование ведется по областям

$$G_p = \left\{t : \frac{3\lambda}{2\sqrt{n}e^{p+1}} \leq |h(t)| \leq \frac{3\lambda}{2\sqrt{n}e^p}\right\}.$$

Используя введенную выше функцию $\xi(s)$, мы найдем, что

$$\int\limits_{\{|h(t)|\leqslant\frac{3\lambda}{2\sqrt{n}}\}} |h(t)|^2\, dt \leq \frac{2^d}{n} + \frac{9\lambda^2}{4n} \sum_{p\leqslant d\ln n} e^{-2p}\xi\left(\frac{3\lambda}{ne^{p+1}}\right)$$

$$\leq \frac{2^d}{n} + \frac{9\lambda^2}{4n} \sum_{p\leq d\ln n} e^{-2p} H^{-1}\left(\frac{4}{9}\frac{L^2e^{2p}n}{\lambda^2}\right). \tag{26}$$

Чтобы оценить последнюю сумму, отметим следующие простые свойства функции H.

Лемма 4. *Пусть $b(z) \geq 1$ – определенная на $[0, \infty)$ монотонно возрастающая до бесконечности функция. Тогда*

$$B(x) = \int_0^x b(z)\,dz$$

непрерывная монотонно возрастающая выпуклая функция. Обратная функция $B^{-1}(x)$ – монотонно возрастающая вогнутая функция. Кроме того,

1. *Для всех $A \geq 1$*

$$B^{-1}(Ax) \leq AB^{-1}(x); \tag{27}$$

2.

$$\frac{1}{x}B^{-1}(x) \geq \frac{1}{b(x)} \tag{28}$$

Доказательство. Производная $c(x)$ функции $B^{-1}(x)$ равна $\frac{1}{b(B^{-1}(x)}$, так что $c(x) \downarrow 0$, когда $x \to \infty$. Следовательно,

$$B^{-1}(Ax) = \int_0^{Ax} b(z)\,dz = A\int_0^x b(Az)\,dz \leq A\int_0^x b(z)\,dz = AB^{-1}(x).$$

Соотношение (27) доказано.

Далее, функция $c(x) \downarrow$, так что

$$H^{-1}(x) = \int_0^x c(z)\,dz \geq xc(x) = \frac{x}{b(H^{-1}(x))}.$$

Неравенство $c(x) \leq 1$ влечет неравенство $H^{-1}(x) \leq x$. Следовательно, $b(H^{-1}(x)) \leq b(x)$ и неравенство (28) доказано.

Возвращаясь к (26), найдем с учетом леммы, что

$$\int\limits_{\{|h(t)| \leq \frac{3\lambda}{2\sqrt{n}}\}} |h(t)|^2\,dt \leq \frac{2^d}{n} + \frac{9\lambda^2}{4n}H^{-1}\left(\frac{4}{9}L^2 n\right) \cdot \sum_1^\infty e^{-2p} + \frac{9\lambda^2}{4n}H^{-1}\left(\frac{4}{9}L^2 n\right) \cdot d\ln n. \tag{29}$$

Используя неравенства (25) и (29), мы найдем из (14), что

$$E_f\{\|\widetilde{f}_n - f\|^2\} \le c\frac{\ln n}{n}H^{-1}(4L^2 n) + \frac{1}{(2\pi)^d}\frac{L^2}{m_n^2}. \tag{30}$$

Чтобы закончить доказательство, нам осталось оценить m_n. В силу монотонности функции φ

$$m_n = \min(\varphi(n,0,\dots,0),\dots,\varphi(0,\dots,0,n)).$$

Допустим для определенности, что $m_n = \varphi(n,0,\dots,0)$. Если точка $t = (t_1,\dots,t_d) \in \{s : |s_1| \le n, |s_2| \le 1,\dots,|s_d| \le 1\}$, то в силу монотонности φ выполняется неравенство $\varphi(t) \le \varphi(n,1,\dots,1)$. Следовательно, $a(\varphi(n,1,\dots,1)) \ge 2n > n$. Из последнего неравенства следует, что

$$m_n = \varphi(n,0,\dots,0) \ge A_\varphi^{-1}\varphi(n,1,\dots,1) \ge A_\varphi^{-1}a^{-1}(n).$$

В силу неравенства (28)

$$(a^{-1}(n))^2 \le \frac{1}{n}H^{-1}(n)$$

и потому

$$\frac{1}{m_n^2} \le \frac{A_\varphi^2}{n}H^{-1}(n). \tag{31}$$

Отсюда и из (30) следует неравенство теоремы. Теорема доказана.

Замечание 1. Простые примеры показывают, что условие однородности, вводящее постоянную A_φ, в рамках данной конструкции адаптивных оценок не может быть отброшено. Чтобы показать это, положим $d = 2$ и $\varphi(t,s) = \varphi_\varepsilon(t,s) = (1+\varepsilon|t|)(1+\varepsilon^{-1}|s|)$, $\varepsilon > 0$. Рассмотрим наблюдения с характеристической функцией $f_\varepsilon(t,s) = f(t,s) = (1+\varepsilon^2 t^2)^{-1}\ (1+\varepsilon^{-2}s^2)$ (т.е. компоненты вектора X_1 независимы и имеют распределения Коши). Интеграл

$$\int\limits_{R^2} |f(t,s)|^2|\varphi(t,s)|^2\,dt\,ds = \left(\int\limits_{-\infty}^{\infty}\frac{(1+|t|)^2}{(1+t^2)^2}\,dt\right)^2 = L^2$$

не зависит от ε. Точно так же функция

$$a(x) = \mathrm{mes}\{(t,s) : \varphi_\varepsilon(t,s) < x\} = \iint\limits_{\{(1+|t|)(1+|s|)<x\}} dt\,ds \sim 4x\ln x, x \to \infty,$$

не зависит от ε. Следовательно, и функция

$$H(x) = \int\limits_0^x (a^{-1}(y))^2\,dy \sim \frac{1}{12}\frac{x^3}{\ln^2 x}$$

не зависит от ε. Функция

$$H^{-1}(x) \sim (12)^{1/3} x^{1/3} (\ln x)^{2/3}$$

тоже не зависит от ε.

В то же время интегралы по $\overline{\Gamma}_n$, $\overline{\Gamma}_n = \{(t,s) : |t| \le n,\ |s| \le n\}$, таковы что

$$\int\limits_{\overline{\Gamma}_n} |f(t,s)|^2\,dt\,ds \ge \int\limits_n^\infty dt \int\limits_{-\infty}^\infty |f(t,s)|^2\,ds$$
$$= \pi(\frac{\pi}{2} - \operatorname{arctg} n\varepsilon) \ge \frac{\pi}{n\varepsilon}(1 + \circ(1)), \quad n\varepsilon \to \infty.$$

Следовательно, неравенство вида

$$\int\limits_{\overline{\Gamma}_n} |f(t,s)|^2\,dt\,ds \le \frac{c}{n} H^{-1}(n) \ln n$$

с абсолютной константой c невозможно.

Замечание 2. Выносить множитель L^2 из под знака функции H^{-1}, т.е. использовать неравенство $H^{-1}(b^2 n) \le L^2 H^{-1}(n)$, конечно, необязательно. В правой части теоремы тогда будет стоять вообще говоря более точный множитель $H^{-1}((L^2+1)n)$.

Замечание 3. Разумеется, можно рассматривать разложение плотности $f(x)$ по какому-нибудь счетному базису в $L_2(R^d)$ или более общо в $L_2(R^d, d\mu)$, μ – какая-нибудь мера на борелевских подмножествах R^d, и строить затем проективные оценки по отношению к этим разложениям. Собственно именно так в работе Н. Н. Ченцова [1] и были введены проективные оценки. Затем, подобно тому, как это указано в п. 1, можно строить адаптивные варианты проекционных оценок. Такие построения проделаны, например, в [9] и [10]. В качестве базиса в этих работах выбирается система вэйвлетов (всплесков). Как

уже отмечалось в п. 1, риск таких адаптивных оценок не уступает риску оптимальных оценок. Подобные построения можно провести и в случае наших континуальных базисов.

Замечание 4. Некоторые идеи доказательства заимствованы мною из доклада D. Picard и G. Kerkyacharian'a о простроении адаптивных оценок в задаче п. 1 на одной из конференций (к сожалению, я не помню где и когда происходила эта конференция).

Литература

1. Н. Н. Ченцов, *Оценка неизвестной плотности распределения по наблюдениям.* — Доклады АН СССР **147**, No. 1 (1962), 45–48.
2. Н. Н. Ченцов, *Избранные труды. Математика.* Физматгиз, М. (2001).
3. И. А. Ибрагимов, Р. З. Хасьминский, *Асимптотическая теория оценивания.* Наука, М. (1979).
4. A. Tsybakov, *Introduction to Nonparametric Estimation.* Springer Series in Statistics, Springer (2009).
5. A. Tsybakov, *Introduction a l'estimation non-parametric.* Springer (2004).
6. W. Hoeffding, *Probability inequalities for sums of bounded random variables.* — J. Amer. Statist. Assoc. **58**, No. 301 (1963), 13–30.
7. А. Н. Ширяев, *Задачи по теории вероятностей.* МНЦ МО, М. (2006).
8. Г. Харди, Дж. Литтлвуд, Дж. Полиа, *Неравенства.* ИЛ, М. (1948).
9. D. Donoho, I. Johnstone, G. Kerkyacharian, D. Picard, *Denstity estimation by wavelet thresholding.* — Ann. Statist. **24** (1996), 508–539.
10. A. Juditsky, *Wavelet estimators: Adapting to unknown smoothness.* — Math. Methods Statist **6** (1997), 1–25.

Ibragimov I. A. On adaptive estimation of probability density functions.

The paper presents a method of adaptive estimation for a class of probability density functions. This method is a continual analogue of some known methods.

Санкт-Петербургское отделение
Математического института
им. В. А. Стеклова РАН,
Фонтанка 27, 191023 С.-Петербург, Россия

E-mail: ibr32@pdmi.ras.ru

Поступило 19 октября 2009 г.