Записки научных
семинаров ПОМИ
Том 368, 2009 г.

С. В. Решетов

МИНИМАКСНЫЙ РИСК ДЛЯ КВАДРАТИЧНО ВЫПУКЛЫХ МНОЖЕСТВ

Введение

Пусть на отрезке [0, 1] наблюдается процесс

$$y(t) = s(t) + x(t),$$

где $s(t) \in L^2_{[0,\,1]}$ – неизвестная подлежащая оцениванию функция, $x(t)$ – стационарный процесс с известной спектральной плотностью f. Априорная информация заключается в том, что s лежит в компактном подмножестве пространства $L^2_{[0,\,1]}$. Пусть $s = \sum\limits_{i=1}^{\infty} \theta_i \varphi_i$ – разложение функции s по ортонормированному базису $\{\varphi_j\}$ в $L^2_{[0,\,1]}$. Задача оценивания функции s в этой модели в некотором смысле эквивалентна задаче оценивания вектора $(\theta_1, \theta_2, \dots)$, заведомо лежащего в подходящим образом выбранном компактном множестве $\Theta \subset l_2$, по наблюдениям

$$y_i = \theta_i + \sigma_i \mathbf{x}_i, \quad i = 1, 2, \dots,$$

где

$$\sigma_i = \int\limits_0^1 |\widehat{\varphi}_i(t)|^2 \, f(t)\, dt, \quad \mathbf{x}_i \in \mathcal{N}(0,1), \quad i = 1, 2, \dots,$$

$\widehat{\varphi}$ – преобразование Фурье функции φ. Подобный переход от непрерывной к дискретной модели был впервые предложен Н. Н. Ченцовым [1]. В настоящей работе мы рассмотрим случай, когда случайные величины $\mathbf{x}_i$, $i = 1, 2, \dots$, независимы, параметрическое множество Θ ортосимметрично, компактно, выпукло и квадратично выпукло.

Прежде, чем перейти к основной задаче этой работы, рассмотрим простейший одномерный случай. Пусть имеется единственное наблюдение

$$y = \theta + \sigma \mathbf{x},$$

Ключевые слова : минимаксный риск, линейный минимаксный риск, квадратично выпуклые множества, бесконечномерные прямоугольники.

где $\mathbf{x}$ – стандартная нормальная случайная величина. Требуется оценить θ, если известно, что $|\theta| \leq \tau$. Мы обозначим $\nabla_N(\tau, \sigma)$ минимаксный риск в этой задаче:

$$\nabla_N(\tau, \sigma) = \inf_{\widehat{\theta}(y)} \sup_{|\theta| \leq \tau} \mathbf{E}\left(\widehat{\theta} - \theta\right)^2, \tag{1}$$

где inf берётся по всем оценкам $\widehat{\theta}(y)$. Буква “N” ставится для того, чтобы отличать минимаксный (нелинейный – nonlinear) риск от минимаксного линейного риска ∇_L, в определении которого в (1) inf берется лишь по линейным оценкам. Оценка, на которой достигается inf в (1), называется минимаксной, а в случае, когда inf берется лишь по линейным оценкам, – линейной минимаксной. Хорошо известно [2], что в такой модели минимаксная линейная оценка имеет вид

$$\widehat{\theta}^*(y) = \frac{\tau^2}{\tau^2 + \sigma^2} y,$$

а минимаксный линейный риск

$$\nabla_L(\tau, \sigma) = \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2}. \tag{2}$$

Обозначим

$$\mu^* = \sup_{\tau, \sigma} \frac{\nabla_L(\tau, \sigma)}{\nabla_N(\tau, \sigma)}.$$

Ибрагимов и Хасьминский [3] показали, что $\mu^* < \infty$, а в работе Donoho, Liu, MacGibbon [4] доказано, что $\mu^* \leq 1.25$. Таким образом, в одномерной модели минимаксный нелинейный риск не может сильно отличаться от минимаксного линейного.

Вернемся к общей модели. Рассмотрим задачу оценивания неизвестного вектора $\theta = (\theta_1, \theta_2, \dots) \in l_2$ по имеющимся наблюдениям

$$y_i = \theta_i + \sigma_i \mathbf{x}_i, \quad i = 1, 2, \dots \tag{3}$$

где $\mathbf{x}_i$, $i = 1, 2, \dots,$ – независимые стандартные нормальные случайные величины, а числа $\sigma_i \neq 0$, $i = 1, 2, \dots$ известны. Априорная информация состоит в том, что вектор θ лежит в компактном множестве $\Theta \subset l_2$.

Качество оценки $\widehat{\theta} = \widehat{\theta}(y) = \left(\widehat{\theta}_1(y), \widehat{\theta}_2(y), \ldots\right)$, $y = (y_1, y_2, \ldots)$, построенной по наблюдениям (3), измеряется величиной квадратичного риска

$$\mathcal{R}\left(\widehat{\theta}, \Theta\right) = \sup_{\theta\in\Theta} \mathbf{E}\left\|\widehat{\theta} - \theta\right\|_{l_2}^2 .$$

Минимаксным риском называется величина

$$\mathfrak{R}_N(\sigma, \Theta) = \inf_{\widehat{\theta}} \mathcal{R}\left(\widehat{\theta}, \Theta\right), \quad \sigma = (\sigma_1, \sigma_2, \ldots),$$

где inf берётся по всем измеримым функциям $\widehat{\theta} = \widehat{\theta}(y)$ со значениями в l_2. Буква “N” снова ставится для того, чтобы отличать минимаксный нелинейный риск от минимаксного линейного риска $\mathfrak{R}_L$, в определении которого в (4) inf берется лишь по линейным оценкам. Оценка, на которой достигается inf в (4), называется минимаксной, а в случае, когда inf берется лишь по линейным оценкам, – линейной минимаксной.

Как мы сказали выше, в этой работе рассматривается случай, когда Θ компактно, ортосимметрично, выпукло и квадратично выпукло.

Множество Θ называется ортосимметричным, если из того, что $\theta = (\theta_i)_{i=1}^\infty \in \Theta$, следует, что $(\pm\theta_i)_{i=1}^\infty \in \Theta$ для любого набора знаков $\pm$.

Множество Θ квадратично выпукло, если выпукло множество

$$\Theta^2 = \left\{\left(\theta_i^2\right)_{i=1}^\infty : (\theta_i)_{i=1}^\infty \in \Theta\right\}.$$

Мы будем использовать обозначение $\Pi(\tau)$, $\tau = (\tau_1, \tau_2, \ldots)$ для бесконечномерного прямоугольника:

$$\Pi(\tau) = \{\theta = (\theta_1, \theta_2, \ldots) : |\theta_i| \le \tau_i, \, i = 1, 2, \ldots\}.$$

Основные результаты этой работы содержатся в следующих двух теоремах.

Теорема 1. *Пусть Θ компактно, ортосимметрично, выпукло и квадратично выпукло. Тогда для минимаксного линейного риска в задаче оценивания вектора $\theta \in \Theta$ по наблюдениям* (3) *справедливо следующее равенство*

$$\mathfrak{R}_L(\sigma, \Theta) = \sup\left\{\mathfrak{R}_L(\sigma, \Pi(\tau)) : \Pi(\tau) \subset \Theta\right\}. \tag{5}$$

Теорема 2. *Пусть* Θ *компактно, ортосимметрично, выпукло и квадратично выпукло. Тогда для минимаксного риска в задаче оценивания вектора* $\theta \in \Theta$ *по наблюдениям* (3) *справедливы следующие неравенства*

$$\begin{aligned} &\sup \{\mathfrak{R}_L(\sigma, \Pi(\tau)) : \Pi(\tau) \subset \Theta\} \geq \mathfrak{R}_N(\sigma, \Theta) \\ &\geq (\mu^*)^{-1} \sup \{\mathfrak{R}_L(\sigma, \Pi(\tau)) : \Pi(\tau) \subset \Theta\}. \end{aligned} \tag{6}$$

Эти две теоремы позволяют свести задачу, в которой $\Theta \subset l_2$ – произвольное компактное, ортосимметричное, выпуклое и квадратично выпуклое множество к хорошо изученному случаю бесконечномерного прямоугольника. Отметим, что теорема 1 доказана в работе Donoho, Liu, MacGibbon [4] при условии, что величины σ_i, $i = 1, 2, \ldots$, равны между собой и, стало быть, отделены от нуля. Отказ от этого условия существенно меняет доказательство.

Доказательство теорем

Доказательство первой теоремы состоит из двух частей:

1. Рассмотрим множество $\Theta_+ = \{\theta = (\theta_i) : \theta_i \geq 0,\ \theta \in \Theta\}$. Ясно, что для любого $\theta \in \Theta_+$ множество $\Pi(\theta) \subset \Theta$. Мы покажем, что существует такое $v \in \Theta_+$, что

$$\sup \{\mathfrak{R}_L(\sigma, \Pi(\theta)) : \Pi(\Theta) \subset \Theta\} = \mathfrak{R}_L(\sigma, \Pi(v)). \tag{7}$$

Эта часть доказана в [4] для одинаковых σ_i, $i = 1, 2, \ldots$, и идея доказательства не меняется, если отказаться от их равенства, поэтому эта часть будет изложена менее подробно, чем вторая.

2. Пусть теперь $\widehat{\theta}^*$ – минимаксная линейная оценка в задаче, когда параметрическое множество – $\Pi(v)$. Другими словами, $\widehat{\theta}^*$ такова, что

$$\mathfrak{R}_L(\sigma, \Pi(v)) = \mathcal{R}\left(\widehat{\theta}^*, \Pi(v)\right).$$

Хорошо известно [4], что $\widehat{\theta}^*$ существует и имеет вид $\widehat{\theta}^* = (b_1 y_1, b_2 y_2, \ldots)$, где $b_i = b_i(v_i, \sigma_i) = \frac{v_i^2}{v_i^2 + \sigma_i^2}$, $i = 1, 2, \ldots$. Мы покажем, что для всех $\theta \in \Theta$ выполняется неравенство:

$$R\left(\widehat{\theta}^*, v\right) \geq R\left(\widehat{\theta}^*, \theta\right), \tag{8}$$

где $R\left(\widehat{\theta}, \theta\right) = \mathbf{E}\left\|\widetilde{\theta} - \theta\right\|_{l_2}^2$.

Отсюда,

$$\sup\left\{\mathfrak{R}_L\left(\sigma, \Pi\left(\theta\right)\right) : \Pi\left(\theta\right) \subset \Theta\right\} = \mathfrak{R}_L\left(\sigma, \Pi\left(v\right)\right) = R\left(\widehat{\theta}^*, v\right)$$
$$\geq \sup_{\theta \in \Theta} R\left(\widehat{\theta}^*, \theta\right) \geq \mathfrak{R}_L\left(\sigma, \Theta\right),$$

и, так как, с другой стороны, очевидно,

$$\mathfrak{R}_L\left(\sigma, \Theta\right) \geq \mathfrak{R}_L\left(\sigma, \Pi\left(v\right)\right) = \sup\left\{\mathfrak{R}_L\left(\sigma, \Pi\left(\theta\right)\right) : \Pi\left(\theta\right) \subset \Theta\right\},$$

получаем справедливость (5).

Как было сказано выше, $\Pi\left(\tau\right) \subset \Theta$ для любого $\tau \in \Theta_+$. Введем в рассмотрение следующий функционал J:

$$J\left(\tau\right) = \mathfrak{R}_L\left(\sigma, \Pi\left(\tau\right)\right), \quad \tau \in \Theta_+.$$

Имеем

$$\sup\left\{\mathfrak{R}_L\left(\sigma, \Pi\left(\tau\right)\right) : \Pi\left(\tau\right) \subset \Theta\right\} = \sup_{\Theta_+} J\left(\tau\right).$$

Хорошо известно [4], что

$$\mathfrak{R}_L\left(\sigma, \Pi\left(\tau\right)\right) = \sum_{i=1}^{\infty} \nabla_L\left(\sigma_i, \tau_i\right),$$

откуда, с помощью (2), получаем

$$\nabla_L\left(\sigma_i, \tau_i\right) = \frac{\tau_i^2 \sigma_i^2}{\tau_i^2 + \sigma_i^2}, \quad i = 1, 2, \ldots.$$

Так что,

$$J\left(\tau\right) = \sum_{i=1}^{\infty} \frac{\tau_i^2 \sigma_i^2}{\tau_i^2 + \sigma_i^2},$$

а значит,

$$\left|J\left(\theta\right) - J\left(\tau\right)\right| = \left|\sum_{i=1}^{\infty}\left(\frac{\theta_i^2 \sigma_i^2}{\theta_i^2 + \sigma_i^2} - \frac{\tau_i^2 \sigma_i^2}{\tau_i^2 + \sigma_i^2}\right)\right| \leq \sum_{i=1}^{\infty}\left|\theta_i^2 - \tau_i^2\right|.$$

С помощью леммы Шеффе и последнего неравенства легко показать [4], что J непрерывен в l_2, а значит, так как Θ – компакт, J достигает максимум на Θ_+, что и доказывает (7).

Прежде, чем перейти ко второй части доказательства, сделаем небольшое отступление. Зададим на множестве

$$\Theta^2 = \left\{ t\,(\theta) = \left(\theta_1^2,\, \theta_2^2,\, \dots\right) \,:\, \theta \in \Theta \right\}$$

функционал G по правилу

$$G\,(t) = G\,(t\,(\theta)) \equiv J\,(\theta)\,, \quad \theta \in \Theta_+.$$

Понятно, что $\Theta_+^2 = \Theta^2$ вследствие ортосимметричности Θ, и Θ^2 выпукло, так как Θ квадратично выпукло. Обозначив $t^v = t(v)$, имеем в силу того, что v – точка максимума функционала J на Θ_+:

$$G\,(t) \le G\,(t^v) \quad \text{для всех} \quad t \in \Theta^2.$$

Теперь для $\varepsilon \in [0,1]$, $\theta \in \Theta$ обозначим

$$t^{\varepsilon,\,v} = t^{\varepsilon,\,v}\,(\theta) = (1-\varepsilon)\,t(v) + \varepsilon t\,(\theta) = (1-\varepsilon)\,t^v + \varepsilon t\,(\theta)\,.$$

Ясно, что в силу выпуклости Θ^2 этот вектор лежит в Θ^2 при любых $\varepsilon \in [0,1]$, $\theta \in \Theta$. Несложные вычисления, вместе с (9), показывают, что при любых $\varepsilon \in [0,1]$, $t \in \Theta^2$ справедливо неравенство

$$G\,(t^v) - G\,(t^{\varepsilon,\,v}) = \varepsilon \sum_{i=1}^{\infty} \frac{\sigma_i^4\,(t_i^v - t_i)}{(t_i^{\varepsilon,\,v} + \sigma_i^2)\,(t_i^v + \sigma_i^2)} \ge 0. \tag{10}$$

Вернемся к доказательству. Напомним, что линейная минимаксная оценка в случае, когда параметрическое множество – прямоугольник $\Pi\,(v)$, имеет вид $\widehat{\theta}^* = (b_1\,y_1,\, b_2\,y_2,\, \dots)$, где

$$b_i = b_i\,(v_i,\, \sigma_i) = \frac{{v_i}^2}{{v_i}^2 + \sigma_i^2} = \frac{t_i^v}{t_i^v + \sigma_i^2}, \quad i = 1,\, 2,\, \dots\,.$$

Отсюда, используя (3), получаем

$$R\left(\widehat{\theta}^*, \theta\right) = \sum_{i=1}^{\infty} (1-b_i)^2\,\theta_i^2 + \sum_{i=1}^{\infty} b_i^2 \sigma_i^2.$$

Оба ряда, участвующие в этом равенстве, сходятся, поскольку

$$\sum_{i=1}^{\infty}(1-b_i)^2\,\theta_i^2 \le \sum_{i=1}^{\infty}\theta_i^2, \quad \sum_{i=1}^{\infty} b_i^2\sigma_i^2 \le \sum_{i=1}^{\infty} v_i^2,$$

а векторы θ и v лежат в l_2.

Как было сказано выше, (5) следует из (8). Следовательно, мы должны показать, что $R\left(\widehat{\theta}^{*},\theta\right)$ достигает свой максимум при $\theta = v$. Так как $\sum\limits_{i=1}^{\infty} b_i^2\sigma_i^2$ не зависит от θ_i, а $\sum\limits_{i=1}^{\infty}(1-b_i)^2\,\theta_i^2$ не зависит от знаков θ_i, то, очевидно, достаточно показать, что

$$\sum_{i=1}^{\infty}(1-b_i)^2\left(v_i^2-\theta_i^2\right) \ge 0 \quad \text{для всех} \quad \theta \in \Theta_+,$$

или

$$\sum_{i=1}^{\infty}(1-b_i)^2\left(t_i^v - t_i\right) \ge 0 \quad \text{для всех} \quad t \in \Theta^2. \tag{11}$$

Будем доказывать от противного. Предположим, что нашлось такое $t \in \Theta^2$, что (11) не выполнено. Тогда, найдётся такое положительное $\delta = \delta(\theta)$, что

$$\sum_{i=1}^{\infty}(1-b_i)^2\left(t_i^v - t_i\right) = \sum_{i=1}^{\infty}\frac{\sigma_i^4\left(t_i^v - t_i\right)}{\left(t_i^v+\sigma_i^2\right)^2} \le -\delta. \tag{12}$$

Из сходимости ряда $\sum\limits_{i=1}^{\infty}\frac{\sigma_i^4|t_i^v-t_i|}{\left(t_i^v+\sigma_i^2\right)^2}$ заключаем, что для любого положительного δ_1 найдётся такое целое положительное $N_1 = N_1(\delta_1)$, что

$$\sum_{i=N_1+1}^{\infty}\frac{\sigma_i^4\left|t_i^v - t_i\right|}{\left(t_i^v+\sigma_i^2\right)^2} < \delta_1. \tag{13}$$

Далее, так как

$$\sum_{i=1}^{\infty}\frac{\sigma_i^4\left|t_i^v - t_i\right|}{\left(t_i^{\varepsilon,\,v}+\sigma_i^2\right)\left(t_i^v+\sigma_i^2\right)} \le \sum_{i=1}^{\infty}\left|t_i^v - t_i\right|,$$

то для любого положительного δ_2 можно найти такое $N_2 = N_2(\delta_2)$, не зависящее от ε, что

$$\sum_{i=N_2+1}^{\infty} \frac{\sigma_i^4\,(t_i^v - t_i)}{(t_i^{\varepsilon,\,v} + \sigma_i^2)\,(t_i^v + \sigma_i^2)} < \delta_2. \tag{14}$$

Взяв $N = \max(N_1, N_2)$, получаем из (13)

$$\sum_{i=1}^{N} \frac{\sigma_i^4\,(t_i^v - t_i)}{(t_i^v + \sigma_i^2)^2} < -\delta + \delta_1, \tag{15}$$

а из (14) –

$$\sum_{i=N+1}^{\infty} \frac{\sigma_i^4\,(t_i^v - t_i)}{(t_i^{\varepsilon,\,v} + \sigma_i^2)\,(t_i^v + \sigma_i^2)} < \delta_2. \tag{16}$$

Так как $g(\varepsilon) = \sum\limits_{i=1}^{N} \frac{\sigma_i^4 (t_i^v - t_i)}{(t_i^{\varepsilon,\,v} + \sigma_i^2)(t_i^v + \sigma_i^2)}$ непрерывна на $(0,\,1)$, и $g(0) = \sum\limits_{i=1}^{N} \frac{\sigma_i^4 (t_i^v - t_i)}{(t_i^v + \sigma_i^2)^2}$, то в силу (15) для любого положительного δ_3 мы можем выбрать такое $\varepsilon \in (0, 1)$, что

$$\sum_{i=1}^{N} \frac{\sigma_i^4\,(t_i^v - t_i)}{(t_i^{\varepsilon,\,v} + \sigma_i^2)\,(t_i^v + \sigma_i^2)} < -\delta + \delta_1 + \delta_3. \tag{17}$$

Выбрав теперь $\delta_1 = \delta_2 = \delta_3 = \frac{\delta}{4}$, с учётом (16) и (17), получаем

$$\varepsilon \sum_{i=1}^{\infty} \frac{\sigma_i^4\,(t_i^v - t_i)}{(t_i^{\varepsilon} + \sigma_i^2)\,(t_i^v + \sigma_i^2)} \le -\frac{\varepsilon\delta}{4} < 0,$$

что противоречит (10), а значит, (11) верно, и теорема 1 доказана. $\square$

Вторая теорема является несложным следствием первой. Как было сказано выше,

$$\mathfrak{R}_L(\sigma, \Pi(\tau)) = \sum_{i=1}^{\infty} \nabla_L(\sigma_i,\, \tau_i)\,.$$

Аналогичное равенство справедливо и для нелинейного минимаксного риска [4]:

$$\mathfrak{R}_N(\sigma, \Pi(\tau)) = \sum_{i=1}^{\infty} \nabla_N(\sigma_i,\, \tau_i)\,.$$

Здесь ∇_N – минимаксный риск в одномерной задаче. Поэтому, используя определение μ^*, получаем

$$\mathfrak{R}_N\left(\sigma, \Pi\left(\tau\right)\right) \leq \mathfrak{R}_L\left(\sigma, \Pi\left(\tau\right)\right) \leq \mu^* \mathfrak{R}_N\left(\sigma, \Pi\left(\tau\right)\right).$$

И, окончательно, с помощью теоремы 1,

$$\begin{aligned} &\sup\left\{\mathfrak{R}_L\left(\sigma, \Pi\left(\tau\right)\right) : \Pi\left(\tau\right) \subset \Theta\right\} = \mathfrak{R}_L\left(\sigma, \Theta\right) \geq \mathfrak{R}_N\left(\sigma, \Theta\right) \\ &\geq \sup\left\{\mathfrak{R}_N\left(\sigma, \Pi\left(\tau\right)\right) : \Pi\left(\tau\right) \subset \Theta\right\} \\ &\geq \left(\mu^*\right)^{-1} \sup\left\{\mathfrak{R}_L\left(\sigma, \Pi\left(\tau\right)\right) : \Pi\left(\tau\right) \subset \Theta\right\}. \end{aligned}$$

Теорема доказана. □

Литература

1. Н. Н. Ченцов, *Оценка неизвестной плотности распределения по наблюдениям.* — ДАН СССР **147** (1962), 45–48.
2. М. С. Пинскер, *Оптимальная фильтрация квадратично-интегрируемых сигналов на фоне гауссовского шума.* — Проблемы передачи информации **16(2)** (1980), 52–68.
3. I. A. Ibragimov, R. Z. Has'minskii, *Nonparametric estimation of the value of a linear functional in Gaussian white noise.* — Theory Probab. Appl. **29** (1984), 1–32.
4. D. L. Donoho, R. C. Liu, B. MacGibbon, *Minimax risk over hyperrectangles, and implications.* — Ann. Statist. **18**, No. 3 (1990), 1416–1437.

Reshetov S. V. Minimax risk for quadratically convex sets.

We consider the problem of estimating the vector $\theta = (\theta_1, \theta_2, \dots) \in \Theta \subset l_2$ on the observations $y_i = \theta_i + \sigma_i \mathbf{x}_i$, $i = 1, 2, \dots$, where $\mathbf{x}_i$ are i.i.d. $N(0, 1)$, the parametric set Θ is compact, orthosymmetric, convex and quadratically convex. We show that in that case the minimax risk is not very different from $\sup \mathfrak{R}_L(\Pi)$, where $\mathfrak{R}_L(\Pi)$ is the minimax linear risk in the same problem with the parametric set Π and sup is taken over all the hyperrectangles $\Pi \subset \Theta$. Donoho, Liu, and McGibbon (1990) have obtained this result for the case of equal σ_i, $i = 1, 2, \dots$.

С.-Петербургский
государственный университет,
Университетеский пр. 28, Старый Петергоф,
198504 Санкт-Петербург, Россия
E-mail: screzh@mail.ru

Поступило 18 октября 2009 г.