

И. ПЕТЕРСЕН

СОПОСТАВЛЕНИЕ МЕТОДА ВОСПРОИЗВОДЯЩИХ ЯДЕР С МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

Рассматривается метод идентификации функций из класса с воспроизводящим ядром в случае конечномерного класса и исследуется связь этого метода с методом наименьших квадратов.

1. Пусть H — линейное пространство непрерывных функций $h = h(x)$, определенных на множестве Q со скалярным произведением, порожденным вероятностной мерой $\xi = \xi(dx)$, $\int_Q \xi(dx) = 1$. Пусть $F \subset H$ — конечномерный линейный класс функций с базисом $\varphi_0, \varphi_1, \dots, \varphi_m$, $\varphi_0(x) \equiv 1$. Обозначая $\varphi^T = (\varphi_0, \varphi_1, \dots, \varphi_m)$, имеем $f = \varphi^T \beta$ для любого $f \in F$, где β суть $(m+1)$ -мерный вектор числовых коэффициентов. В этих предположениях матрица $\Gamma = \int_Q \varphi(x) \varphi^T(x) \xi(dx)$ положительно определена и ядро $K(x, y) = \varphi^T(x) \Gamma^{-1} \varphi(y)$ является воспроизводящим ядром [1] для класса F в том смысле, что

$$\int_Q K(x, y) f(y) \xi(dy) = f(x) \quad (1)$$

при любом $f \in F$ и все линейные комбинации $\sum a_i K(x, x_i) \in F$ при $x_i \in Q$. Наряду с классом F рассмотрим p -мерный, $p \leq (m+1)(m+2)/2$, линейный класс функций G , порожденный всеми различными парными произведениями $\varphi_i \varphi_j$ ($i, j = 0, 1, \dots, m$). Пусть для функции H известна некоторая квадратурная формула по мере ξ , точная на классе G , т. е. определены точки $x_i \in Q$, $i = 1, \dots, N$ и соответствующие веса c_i , так что для любой $g \in G$

$$\int_Q g(x) \xi(dx) = \sum_{i=1}^N c_i g(x_i). \quad (2)$$

Условие (2) равносильно условию

$$\Gamma = \sum_{i=1}^N c_i \varphi(x_i) \varphi^T(x_i). \quad (3)$$

По определению воспроизводящего ядра имеем

$$\int_Q K^2(x, y) \xi(dy) = K(x, x). \quad (4)$$

Так как $K^2(x, y) \in G$ при фиксированном x , то из (4) и (2) следует

$$\sum_{i=1}^N c_i K^2(x, x_i) = K(x, x). \quad (5)$$

2. Рассмотрим теперь проблему идентификации неизвестной функции $f \in F$ по результатам измерений $f(x)$ с аддитивными некоррелиро-

ванными случайными ошибками в узлах x_1, \dots, x_N квадратурной формулы (2):

$$\begin{aligned} z_i &= \hat{f}(x_i) + n_i, \\ \mathbf{M}n_i &= 0, \quad \mathbf{M}n_i n_k = 0 \quad \text{при } i \neq k. \end{aligned} \quad (6)$$

Предполагаем, что точность измерения в отдельных точках x_i распределена пропорционально величинам $d_i > 0$, так что

$$\mathbf{M}n_i^2 = \frac{\sigma^2}{Nd_i}, \quad \sum_{i=1}^N d_i = 1. \quad (7)$$

В частности, при равнооточных измерениях $d_i = \frac{1}{N}$ и $\mathbf{M}n_i^2 = \sigma^2$.

В [2] нами был предложен метод идентификации (назовем его методом воспроизводящих ядер) неизвестных функций из (в общем бесконечномерных) классов функций с известным воспроизводящим ядром. Соответствующая этому методу оценка $\hat{f}(x)$ имеет вид

$$\hat{f}(x) = \sum_{i=1}^N c_i K(x, x_i) z_i. \quad (8)$$

Оценка (8) построена применением квадратурной формулы (2) к (1) и заменой $\hat{f}(x_i)$ на z_i . В силу условия (3) оценка (8) в данном случае несмещенная:

$$\begin{aligned} \mathbf{M}\hat{f}(x) &= \sum_{i=1}^N c_i K(x, x_i) \mathbf{M}z_i = \sum_{i=1}^N c_i \varphi^T(x) \Gamma^{-1} \varphi(x_i) \varphi^T(x_i) \beta = \\ &= \varphi^T(x) \Gamma^{-1} \sum_{i=1}^N c_i \varphi(x_i) \varphi^T(x_i) \beta = \varphi^T(x) \beta = \hat{f}(x). \end{aligned}$$

Дисперсия оценки $\hat{f}(x)$ выражается в виде

$$\mathbf{D}\hat{f}(x) = \frac{\sigma^2}{N} \sum_{i=1}^N \frac{c_i^2}{d_i} K^2(x, x_i). \quad (9)$$

Наряду с оценкой (8) рассмотрим оценку наименьших квадратов $\tilde{f}(x)$, соответствующую предположениям (6) и (7). Имеем [3]

$$\tilde{f}(x) = \varphi^T(x) \left[\sum_{i=1}^N d_i \varphi(x_i) \varphi^T(x_i) \right]^{-1} \sum_{i=1}^N d_i \varphi(x_i) z_i \quad (10)$$

и

$$\mathbf{D}\tilde{f}(x) = \frac{\sigma^2}{N} \varphi^T(x) \left[\sum_{i=1}^N d_i \varphi(x_i) \varphi^T(x_i) \right]^{-1} \varphi(x). \quad (11)$$

Если в (10) положить $d_i = c_i$ ($i = 1, \dots, N$), то с учетом (3) имеем

$$\tilde{f}(x) = \sum_{i=1}^N c_i \varphi^T(x) \Gamma^{-1} \varphi(x_i) z_i = \sum_{i=1}^N c_i K(x, x_i) z_i = \hat{f}(x) \quad (12)$$

и

$$\mathbf{D}\tilde{f}(x) = \frac{\sigma^2}{N} K(x, x). \quad (13)$$

Таким образом, оценка $\hat{f}(x)$ метода воспроизводящих ядер совпадает с оценкой $\tilde{f}(x)$ метода наименьших квадратов, если коэффициенты квад-

ратурной формулы (2) положительны и точность измерений в узлах пропорциональна соответствующим весам. В частности, при равноточных измерениях $\hat{f}(x) = \tilde{f}(x)$, если коэффициенты квадратурной формулы равны: $c_i = 1/N$ ($i = 1, \dots, N$).

3. Качество оценки $\hat{f}(x)$ метода воспроизводящих ядер естественно охарактеризовать величиной $\hat{\delta} = \frac{N}{\sigma^2} \max_{x \in Q} \mathbf{D}\hat{f}(x)$ и сравнить с соответствующей

величиной для оценки метода наименьших квадратов. При заданных Q и F величина $\hat{\delta}$ зависит, с одной стороны, от меры ξ и, с другой стороны, от квадратурной формулы (2), т. е. от N , x_i , c_i ($i = 1, \dots, N$) и от распределения d_1, \dots, d_N точности измерений по узлам x_1, \dots, x_N . Для разделения этих зависимостей введем для узлов

x_1, \dots, x_N и коэффициентов c_1, \dots, c_N оценку $\tilde{f}(x)$, считая ее оценкой наименьших квадратов $\hat{f}(x)$ при условии, что точность измерения в точках x_1, \dots, x_N распределена как c_1, \dots, c_N . Отметим, что бесконечным увеличением числа измерений можно, повторяя измерения в точках x_1, \dots, x_N с частотами $c_1/d_1, \dots, c_N/d_N$, как угодно точно преобразовать исходное распределение точности d_1, \dots, d_N в распределение c_1, \dots, c_N . При соблюдении условия несмещенности (3) введенная

оценка $\tilde{f}(x)$ зависит не от выбранной квадратурной формулы и от распределения точности в узлах этой формулы, а лишь от матрицы Γ , т. е.

от меры ξ . Следовательно, и $\tilde{\delta} = \frac{N}{\sigma^2} \max_{x \in Q} \mathbf{D}\tilde{f}(x) = \max_{x \in Q} K(x, x)$ зависит

только от ξ . Кифером установлено [4], что минимальное значение $\tilde{\delta} = \tilde{\delta}(\xi)$ равно размерности класса функций F : $\min_{\xi} \max_{x \in Q} K(x, x) = m + 1$.

Отношение

$$\mu = (m + 1) / \max_{x \in Q} K(x, x) \quad (14)$$

является поэтому естественным показателем качества меры ξ для заданных Q и F . Назовем μ оптимальностью меры ξ . С другой стороны, отношение

$$\nu = \tilde{\delta} / \hat{\delta} \quad (15)$$

мерит эффективность оценки $\hat{f}(x)$ метода воспроизводящих ядер при заданной мере ξ по сравнению с оценкой наименьших квадратов $\tilde{f}(x)$. Назовем ν эффективностью оценки $\hat{f}(x)$. Мерой качества оценки $\hat{f}(x)$ считаем величину

$$\eta = (m + 1) / \hat{\delta} = \mu \nu, \quad (16)$$

которую назовем оптимальностью оценки $\hat{f}(x)$. Очевидно, $0 < \mu, \nu, \eta \leq 1$. Надо иметь в виду, что оценки с оптимальностью 1 в общем не существуют при конечном числе измерений даже среди оценок наименьших квадратов.

Укажем на следующую оценку снизу эффективности оценки $\hat{f}(x)$. По (9) и (5)

$$\begin{aligned} \hat{\delta} &= \max_{x \in Q} \sum_{i=1}^N \frac{c_i^2}{d_i} K^2(x, x_i) \leq \max_i \frac{c_i}{d_i} \cdot \max_{x \in Q} \sum_{i=1}^N c_i K^2(x, x_i) = \\ &= \max_i \frac{c_i}{d_i} \cdot \max_{x \in Q} K(x, x) \end{aligned}$$

и, следовательно,

$$v \geq 1 / \max_i \frac{c_i}{d_i}. \quad (17)$$

На основании (17) для максимальной дисперсии $\hat{f}(x)$ получаем также оценку

$$\max_{x \in Q} D\hat{f}(x) \leq \frac{\sigma^2}{N} \max_{x \in Q} K(x, x) \cdot \max_i \frac{c_i}{d_i}. \quad (18)$$

На практике обычно желательно, чтобы количество измерений N , необходимое для построения оценки $\hat{f}(x)$, было по возможности меньше. Так как класс F имеет размерность $m+1$, то минимальным количеством измерений для построения несмещенной оценки является $m+1$. Отношение $N/(m+1)$ мерит избыточность количества измерений и является еще одним показателем качества конкретной оценки $\hat{f}(x)$.

4. Для примера рассмотрим проблему идентификации полинома двух переменных третьей степени на единичном круге $x_1^2 + x_2^2 \leq 1$ с применением равномерной меры $\xi(dx) = \frac{1}{\pi} dx_1 dx_2$. В этом случае вектор $\varphi^T(x) = (1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_2^3)$ имеет $m+1 = 10$ координат. Матрицы Γ и Γ^{-1} получаются следующие:

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 & 1/4 & 0 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 & 0 & 0 & 1/8 & 0 & 1/24 & 0 \\ 0 & 0 & 1/4 & 0 & 0 & 0 & 0 & 1/24 & 0 & 1/8 \\ 1/4 & 0 & 0 & 1/8 & 0 & 1/24 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/24 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 1/24 & 0 & 1/8 & 0 & 0 & 0 & 0 \\ 0 & 1/8 & 0 & 0 & 0 & 0 & 5/64 & 0 & 1/64 & 0 \\ 0 & 0 & 1/24 & 0 & 0 & 0 & 0 & 1/64 & 0 & 1/64 \\ 0 & 1/24 & 0 & 0 & 0 & 0 & 1/64 & 0 & 1/64 & 0 \\ 0 & 0 & 1/8 & 0 & 0 & 0 & 0 & 1/64 & 0 & 5/64 \end{pmatrix}$$

$$\Gamma^{-1} = \begin{pmatrix} 4 & 0 & 0 & -6 & 0 & -6 & 0 & 0 & 0 & 0 \\ 0 & 36 & 0 & 0 & 0 & 0 & -48 & 0 & -48 & 0 \\ 0 & 0 & 36 & 0 & 0 & 0 & 0 & -48 & 0 & -48 \\ -6 & 0 & 0 & 18 & 0 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 24 & 0 & 0 & 0 & 0 & 0 \\ -6 & 0 & 0 & 6 & 0 & 18 & 0 & 0 & 0 & 0 \\ 0 & -48 & 0 & 0 & 0 & 0 & 80 & 0 & 48 & 0 \\ 0 & 0 & -48 & 0 & 0 & 0 & 0 & 144 & 0 & 48 \\ 0 & -48 & 0 & 0 & 0 & 0 & 48 & 0 & 144 & 0 \\ 0 & 0 & -48 & 0 & 0 & 0 & 0 & 48 & 0 & 80 \end{pmatrix}$$

Соответствующее воспроизводящее ядро можно выписать в виде

$$\begin{aligned}
 K(x, y) = & 4 - 6(y_1^2 + y_2^2) + 12y_1[3 - 4(y_1^2 + y_2^2)]x_1 + \\
 & + 12y_2[3 - 4(y_1^2 + y_2^2)]x_2 + 6(3y_1^2 + y_2^2 - 1)x_1^2 + \\
 & + 24y_1y_2x_1x_2 + 6(y_1^2 + 3y_2^2 - 1)x_2^2 + \\
 & + 16y_1(5y_1^2 + 3y_2^2 - 3)x_1^3 + 48y_2(3y_1^2 + y_2^2 - 1)x_1^2x_2 + \\
 & + 48y_1(y_1^2 + 3y_2^2 - 1)x_1x_2^2 + 16y_2(3y_1^2 + 5y_2^2 - 3)x_2^3. \quad (19)
 \end{aligned}$$

Среди известных квадратурных формул для круга с постоянным весом и точных для полиномов шестой степени наименьшее число узлов $N=12$ (избыточность 1,2) имеет формула Гаммера [5]. Узлы этой формулы расположены по четыре на трех окружностях вокруг начала с радиусами соответственно $r_1=0,457$, $r_2=0,866$ и $r_3=0,911$. Узлы первой и третьей окружностей находятся на двух перпендикулярных диаметрах, а второй окружности сдвинуты на 45° . Коэффициенты для узлов одной окружности равны. При первой окружности $a_1=0,1231$, при второй $a_2=0,0741$ и при третьей $a_3=0,0527$. В случае равноточных измерений получаем по (17) для эффективности v оценки $\hat{f}(x)$ границу $v \geq \frac{1}{12a_1} = 0,677$. Оценка $\hat{f}(x)$ совпадает с оценкой наименьших квадратов ($v=1$), если точности измерений на указанных трех окружностях соотносятся как $a_1 : a_2 : a_3$.

При $x=y$ получаем из (19)

$$K(x, x) = K(r) = 4 + 36r^2 - 78r^4 + 80r^6, \quad r^2 = x^2 + y^2. \quad (20)$$

Функция $K(r)$ имеет локальные минимумы при $r=0$ и $r=\sqrt{2/5}$ с $K(0)=4$, $K(\sqrt{2/5})=6,24$ и локальный максимум при $r=1/2$ с $K(1/2)=6,375$. Абсолютного максимума на $[0, 1]$ $K(r)$ достигает при $r=1$ с $K(1)=30$. Для единичного круга, таким образом, оптимальность равномерной меры $\mu = \frac{1}{3}$. Но, учитывая, что все узлы расположены в замкнутом круге с радиусом r_3 , естественно при определении $\max K(r)$ ограничиваться этим кругом. Соответствующий абсолютный максимум $K(r_3)=15,924$, что уже соответствует оптимальности 0,628. Отметим еще, что в окружности вокруг начала с радиусом $r_4=0,8366$ имеем $K(r) \leq 10$ и с радиусом $r_5=0,6892$ даже $K(r) \leq 6,375$. Следовательно, для равноточных измерений в указанных 12 узлах имеем на основании (18)

$$\begin{aligned}
 \max_{x_1^2 + x_2^2 \leq 1} D\hat{f}(x) & \leq 3,7\sigma^2, & \max_{x_1^2 + x_2^2 \leq r_3^2} D\hat{f}(x) & \leq 2\sigma^2, \\
 \max_{x_1^2 + x_2^2 \leq r_4^2} D\hat{f}(x) & \leq 1,3\sigma^2, & \max_{x_1^2 + x_2^2 \leq r_5^2} D\hat{f}(x) & \leq 0,8\sigma^2.
 \end{aligned}$$

Если точность измерений при удалении от начала уменьшается так, что на расстояниях r_1 , r_2 и r_3 дисперсии ошибок измерений соответственно

равны: $\frac{\sigma^2}{12a_1} = 0,68\sigma^2$, $\frac{\sigma^2}{12a_2} = 1,12\sigma^2$ и $\frac{\sigma^2}{12a_3} = 1,58\sigma^2$, то $\hat{j}(x)$ является оценкой наименьших квадратов и

$$\max_{x_1^2 + x_2^2 \leq 1} D\hat{j}(x) = \frac{5}{2}\sigma^2, \quad \max_{x_1^2 + x_2^2 \leq r_1^2} D\hat{j}(x) = 1,327\sigma^2,$$

$$\max_{x_1^2 + x_2^2 \leq r_4^2} D\hat{j}(x) = \frac{5}{6}\sigma^2, \quad \max_{x_1^2 + x_2^2 \leq r_5^2} D\hat{j}(x) = \frac{17}{32}\sigma^2.$$

Приближенная теория оптимального планирования регрессионных экспериментов Кифера [6] предлагает для единичного круга и полиномов третьей степени в качестве оптимальной такую меру, которая сосредоточена на двух окружностях с радиусами 0,5155 и 1, с соответствующими весами 0,3077 и 0,6923 и равномерна на обеих окружностях. Для такой меры $\max K(x, x) = 10$ ($\mu = 1$). Однако для дискретной

реализации этой меры, с одной стороны, надо иметь по меньшей мере 14 точек (по семь на каждой окружности) [7] и, с другой стороны, для эффективности соответствующей оценки точность на внешней окружности должна быть в 2,26 раза больше, чем на внутренней. На практике бывает обычно наоборот — точность измерений при удалении от центра эксперимента уменьшается. Для получения приближенно эффективной оценки при равноточных измерениях на окружности с $r = 1$ надо иметь

$\frac{0,3077}{0,6923} \cdot 7 \approx 16$ точек, т. е. вообще 23 точки. Отсюда видно, что в данном случае равномерная мера по всему кругу позволяет построить более приемлемый план эксперимента, чем оптимальная в смысле приближенной теории Кифера, — за счет некоторого увеличения дисперсии оценки вблизи границы области дисперсия меньше в большей части области и меньше также необходимое количество точек измерений.

ЛИТЕРАТУРА

1. Ароншайн Н., Теория воспроизводящих ядер. Математика, период. сб. переводов иностранных статей, 7, № 2 (1963).
2. Петерсен И., Автоматика и телемеханика, № 6, 95 (1969).
3. Рао С. Р., Линейные статистические методы и их применение, М., 1968.
4. Kiefer J., Wolfowitz J., Ann. Math. Stat., 30, 271 (1959).
5. Hammer P. C., Stroud A. H., Math. Tables and Other Aids to Comp., 12, 272 (1958).
6. Farrell R. H., Kiefer J., Walbran A., Proceedings of the Fifth Berkeley Symposium of Mathematical Statistics and Probability, 1, 113 (1967).
7. Gardiner D. A., Grandage A. H. E., Hader R. J., Ann. Math. Stat., 30 (1959).

Институт кибернетики
Академии наук Эстонской ССР

Поступила в редакцию
14/IV 1969

I. PETERSEN

TAASTAVATE TUUMADE NING VÄHIMRUUTUDE MEETODITE VÕRDLUS

Artiklis käsitletakse identifitseerimiseks kasutatavat taastavate tuumade meetodit [2] lõplikdimensionaalse funktsioonide klassi puhul. Esitatakse tingimused, mille korral taastavate tuumade meetodi hinnang on vähimruutude hinnang. Tuletatakse mõned võrra-

tused selle hinnangu dispersiooni jaoks. Näitena vaadeldakse kahemuutuja kuuppolünoomi identifitseerimist 12 mõõtmise järgi ühikringis. Selgub, et sel juhul annab ühtlane mõõt vastuvõetavama katsete plaani kui optimaalne mõõt J. Kieferi [6] ligikaudse teooria mõttes.

I. PETERSEN

COMPARISON OF THE METHOD OF REPRODUCING KERNELS AND THE METHOD OF THE LEAST SQUARES

The method of reproducing kernels for identification proposed by the author earlier [2] is studied in the case of identification from a known finite dimensional class. The conditions are established for this estimator to be the least squares estimator, and some bounds for the dispersion of the estimator are given. As an example the two-dimensional case of estimation of a cubic based on twelve measurements in the unit circle is considered. It is shown that in this case the uniform measure leads to better design of experiments than the optimal measure in the sense of the approximate theory of J. Kiefer [6].