ЛИТЕРАТУРА

1. **Мишев Д.** Дистанционное исследование Земли из космоса: Пер. с болг.— М.: Мир, 1985.
2. **Аэрокосмические** исследования Земли.— М.: Наука, 1979.
3. **Исследование** природной среды с пилотируемых орбитальных станций.— Л.: Гидрометеоиздат, 1972.
4. **Прэтт У.** Цифровая обработка изображений.— М.: Мир, 1982.
5. **Ahuja N., Schacter B. J.** Image models // ACM Comput. Surv.— 1981.— V. 13, N 4.— P. 373.
6. **Kashyap R. L.** Analysis and synthesis of image patterns by spatial interaction models // Progress in Pattern Recognition/Ed. L. M. Kanal, A. Rosenfeld.— Amsterdam; New York; Oxford: North-Holland Publishing Company, 1981.
7. **Кендалл М., Моран П.** Геометрические вероятности.— М.: Наука, 1972.
8. **Getis A., Boots B.** Models of Spatial Processes.— L., Cambridge, 1978.
9. **Иванов В. А., Иванченко Г. А.** Математическое обеспечение статистического анализа аэрофотоснимков леса // Автометрия.— 1982.— № 4.
10. **Косых В. П., Пустовских А. И., Тарасов Е. В., Яковенко Н. С.** Морфологический процессор // Там же.— 1984.— № 4.
11. **Mandelbrot B. B., Van Ness J. W.** Fractional brownian motions, fractional noises and applications // SIAM Rev.— 1968.— V. 10, N 4.— P. 422.
12. **Mandelbrot B. B.** Fractials: Form, Chance and Dimension.— San Francisco: Freeman, 1977.
13. **Fournier A., Fussel D., Carpenter L.** Computer rendering of stochastic models // Communs ACM.— 1982.— V. 25, N 6.— P. 371.
14. **Woods J. W.** Two-dimensional discrete marcovian fields // IEEE Trans. of Inf. Theory.— 1972.— V. IT-18, N 2.— P. 232.

Поступила в редакцию 5 октября 1987 г.

УДК 519.24

В. Г. АЛЕКСЕЕВ

(Москва)

О ВЫБОРЕ ПАРАМЕТРОВ ОЦЕНКИ КРИВОЙ РЕГРЕССИИ С ПОМОЩЬЮ ПЕРЕКРЕСТНОЙ ПРОВЕРКИ

Статистическое оценивание кривой регрессии является одной из тех немногих задач, где сама «природа» идет навстречу исследователю, помогая ему в выборе параметров оценки таким образом, чтобы ошибка оценивания в условиях данного эксперимента была близка к минимальной. Решающая роль в корректировке параметров оценки искомой функции регрессии с помощью самой выборки принадлежит методу перекрестной проверки, описанному, например, в [1, 2]. В настоящей работе метод перекрестной проверки применен к непараметрической оценке, зависящей от двух функциональных и двух числовых параметров.

Итак, пусть (ξ, η) — двумерная случайная величина с плотностью вероятности $f(x, y)$ и

$$\lambda(x) = \langle(\eta \mid \xi = x)\rangle = \int y f(x, y)\, dy \Big/ \int f(x, y)\, dy. \tag{1}$$

Здесь (и всюду в дальнейшем) интеграл без указания пределов обозначает интегрирование в пределах от $-\infty$ до $+\infty$, а угловые скобки $\langle\ \rangle$ являются символом математического ожидания.

Оценку величины $\lambda(x)$ по выборке $\{(\xi_i, \eta_i),\ i = 1, \ldots, n\}$ из n независимых наблюдений случайной величины (ξ, η) будем искать в виде

$$\lambda_n(x) = \sum_{i=1}^{n} \left[\eta_i v\left(\frac{\xi_i - x}{b}\right) \Big/ nb\right] \Big/ \sum_{i=1}^{n} \left[u\left(\frac{\xi_i - x}{a}\right) \Big/ na\right], \tag{2}$$

где ядра (весовые функции) $u(x)$ и $v(x)$, $x \in R = (-\infty, \infty)$, четны, ограничены, тождественно обращаются в нуль вне интервала $(-1, 1)$ и нормируются условием

$$\int_{-1}^{1} u(x)\,dx = \int_{-1}^{1} v(x)\,dx = 1,$$

а масштабные множители (параметры размытости) $a = a(n)$ и $b = b(n)$ положительны.

Наиболее существенное отличие предлагаемой нами оценки (2) от оценки, рассмотренной в работах [1, 2] и многих других, состоит в том, что параметры $v(\cdot)$ и b, стоящие в числителе (2), могут не совпадать с параметрами $u(\cdot)$ и соответственно a, стоящими в знаменателе. Пусть $p_n(x)$ и $\gamma_n(x)$ — знаменатель и числитель правой части формулы (2). Величины $p_n(x)$ и $\gamma_n(x)$ могут рассматриваться как оценки знаменателя $p(x) = \int f(x, y)\,dy$ и числителя $\gamma(x) = \int yf(x, y)\,dy$ правой части формулы (1). При этом параметры обеих статистических оценок желательно было бы выбирать независимо в соответствии с теми или иными предположениями относительно степени гладкости функций $p(x)$ и $\gamma(x) = \lambda(x)p(x)$. Здесь, однако, нельзя забывать о том, что величина $p_n(x)$, оценивающая плотность вероятности $p(x)$ случайной величины ξ, стоит в знаменателе оценки (2). Это накладывает на выбор параметров оценки (2) определенные ограничения. Легко видеть, что оценка (2) будет иметь смысл при любых значениях аргумента x, если

$$b \leqslant a \text{ и } u(x) > 0 \text{ для всех } x \in (-1,\ 1). \qquad (3)$$

Что же касается ядра $v(x)$, то оно может быть выбрано и знакопеременным, т. е. принимающим на интервале $(-1, 1)$ как положительные, так и отрицательные значения. Введем в рассмотрение порядок r, $r = 2,\ 4,\ \ldots$, ядра $v(x) = v_r(x)$, определяемый как наименьшее четное число k, для которого

$$\int_{-1}^{1} x^k v(x)\,dx \neq 0.$$

Применение ядер $v_r(x)$ высших порядков (т. е. порядков $r > 2$) может привести к многократному уменьшению ошибки оценивания, если: а) оцениваемая функция $\gamma(x)$ имеет непрерывные производные достаточно высоких порядков и б) объем выборки n не слишком мал. При этом с ростом объема выборки n растет как порядок r оптимального ядра $v(x)$, так и достигаемый с его помощью выигрыш в точности оценивания (по этому поводу см., например, работы [3, 4], посвященные родственным задачам непараметрического оценивания). Различные наборы весовых функций $v_r(x)$, $r = 2,\ 4,\ \ldots$, могут быть найдены в [5, 6]. Исследователю, впервые вступающему на путь применения знакопеременных весовых функций, может быть рекомендован набор $v_r(x)$, $r = 2,\ 4,\ \ldots$, начинающийся с функций

$$v_2(x) = (3/4)\,(1 - x^2)\,g(x);$$
$$v_4(x) = (15/32)\,(3 - 10x^2 + 7x^4)\,g(x);$$
$$v_6(x) = (105/256)\,(5 - 35x^2 + 63x^4 - 33x^6)\,g(x);$$
$$v_8(x) = (315/4096)\,(35 - 420x^2 + 1386x^4 - 1716x^6 + 715x^8)\,g(x);$$
$$v_{10}(x) = (3465/65536)\,(63 - 1155x^2 + 6006x^4 - 12870x^6 + 12155x^8 - 4199x^{10})\,g(x),$$

где $g(x) = I_{(-1,1)}(x)$ — индикатор интервала $(-1, 1)$.

Метод перекрестной проверки, описанный ниже, позволяет улучшить выбор параметров $u(x)$, $v(x)$, a и b оценки (2), а заодно и уточнить наши априорные предположения относительно степени гладкости функций $p(x)$ и $\gamma(x)$. Пусть $\lambda_{n-1}^{(k)}(x)$ — оценка вида (2), построенная по выборке $\{(\xi_i, \eta_i),\ i=\overline{1, n},\ i\neq k\}$, и

$$\Delta_n = \Delta_n(u, v, a, b) = \sum_{k=1}^{n} \left[\lambda_{n-1}^{(k)}(\xi_k) - \eta_k\right]^2 \alpha(\xi_k),$$

где $\alpha(x)$ — некоторая неотрицательная весовая функция, выбираемая исследователем в соответствии с целями и задачами данного эксперимента.

В отличие от работ [1, 2], в которых для случая

$$v(x) = u(x) \quad \text{и} \quad b = a = h \tag{4}$$

предлагается искать минимум величины Δ_n лишь по параметру h и объявить соответствующее этому минимуму значение h оптимальным, здесь рекомендуем искать минимум величины Δ_n в гораздо более широкой области значений функциональных параметров $u(x)$ и $v(x)$ и масштабных множителей a и b. А именно: вместо соотношений (4) предполагаем лишь, что выполняются условия (3). Наилучшим для данного эксперимента будем считать тот выбор параметров $u(x)$, $v(x)$, a и b, удовлетворяющих условиям (3), при котором величина Δ_n обращается в минимум.

В ряде случаев принятые нами зависимости (3) относительно параметров $u(x)$, $v(x)$, a и b могут быть еще более ослаблены. Предположим, например, что: а) весовая функция $\alpha(x)$ положительна лишь на некотором множестве A, на котором $p(x) > c > 0$, и б) объем выборки n не слишком мал. В этом случае условия (3) могут быть опущены. В силу предположения «а» и равномерной (с вероятностью 1) сходимости оценки $p_n(x)$ к плотности вероятности $p(x)$ (см., например, [7—9]) при достаточно больших n оценка $p_n(x)$ будет с вероятностью 1 положительной для всех $x \in A$ даже в случае применения знакопеременного ядра $u(x)$.

Разумеется, расширение области допустимых значений параметров $u(x)$, $v(x)$, a и b не является самоцелью. Оно позволяет нам, используя метод перекрестной проверки, улучшить качество статистического оценивания кривой регрессии $\lambda(x)$.

ЛИТЕРАТУРА

1. **Hall P.** Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function // Zeitschr. für Wahrscheinlichkeitstheorie und verw. Gebiete.— 1984.— Bd 67, N 2.— S. 175.
2. **Collomb G., Sarda P., Vieu P.** Weak pointwise consistency of the cross validatory window estimate in nonparametric regression estimation // Commentationes Mathematicae Universitatis Carolinae.— 1985.— V. 26, N 4.— P. 789.
3. **Алексеев В. Г.** О непараметрических оценках плотности вероятности и ее производных // ППИ.— 1982.— Т. 18, № 2.
4. **Alekseev V. G.** On the use of alternating kernels in nonparametric statistical estimation // Lecture Notes in Mathematics.— Berlin; Heidelberg; New York; Tokyo: Springer-Verlag, 1983.— V. 1021.
5. **Алексеев В. Г.** Некоторые практические рекомендации по спектральному анализу гауссовых стационарных случайных процессов // ППИ.— 1973.— Т. 9, № 4.
6. **Алексеев В. Г.** О вычислении спектров стационарных случайных процессов по выборкам большого объема // ППИ.— 1980.— Т. 16, № 1.
7. **Алексеев В. Г.** Об оценке плотности вероятности и ее производных // Мат. заметки.— 1972.— Т. 12, № 5.
8. **Конаков В. Д.** Теорема об уклонении эмпирической меры и ее приложения // Теория вероятностей и ее применения.— 1984.— Т. 29, № 1.
9. **Singh R. S.** Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density // Ann. Statist.— 1977.— V. 5, N 2.— P. 394.

Поступила в редакцию 2 апреля 1987 г.