BANACH CENTER PUBLICATIONS
VOLUME 3

О НЕПАРАМЕТРИЧЕСКОЙ ОЦЕНКЕ КРИВОЙ РЕГРЕССИИ

Ш. А. ХАШИМОВ

Институт Математики им. В. И. Романовского АН УзССР, Ташкент, СССР

Пусть $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ выборка из n независимых наблюдений двумерной случайной величины (ξ, η) с плотностью распределения $f(x, y)$. Обозначим плотность распределения случайной величины ξ через $f(x)$, а кривую регрессии η по ξ через

$$r(x) = \int_{-\infty}^{\infty} y f(y/x)\, dy.$$

Положим

$$\varphi(x) = f(x) \cdot r(x).$$

Будем предполагать, что $f(x)$, $\varphi(x) \in L_2$ и $\{\varphi_j(x)\}_0^\infty$ — полная ортонормированная система функций:

$$|\varphi_j(x)| < C \quad (C \text{ не зависит от } x \text{ и } j).$$

Ряды

$$f(x) = \sum_{j=0}^{\infty} c_j \varphi_j(x), \tag{1}$$

$$\varphi(x) = \sum_{j=0}^{\infty} a_j \varphi_j(x), \tag{1'}$$

где

$$c_j = \int_{-\infty}^{\infty} \varphi_j(x) f(x)\, dx, \qquad a_j = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, \varphi_j(x) f(x, y)\, dx\, dy$$

сходятся в среднем соответственно к $f(x)$ и $\varphi(x)$.

Для оценки неизвестной кривой регрессии $r(x)$ по эмпирическим данным рассмотрим статистику

$$r_n(x) = \frac{\varphi_n(x)}{f_n(x)}, \tag{2}$$

где

$$\varphi_n(x) = \sum_{j=0}^{q(n)} a_{jn}\varphi_j(x), \qquad f_n(x) = \sum_{j=0}^{q(n)} c_{jn}\varphi_j(x),$$

$$a_{jn}^* = \frac{1}{n}\sum_{i=1}^{n} y_i\varphi_j(x_i), \qquad c_{jn} = \frac{1}{n}\sum_{i=1}^{n}\varphi_j(x_i),$$

$$q(n)\to\infty \quad \text{при} \quad n\to\infty.$$

Состоятельность оценки (2) для $r(x)$ была доказана в работе [1].

Обозначим через

$$V_n(x,y) = \sum_{j=0}^{q(n)}\varphi_j(x)\varphi_j(y).$$

Теорема 1. *Если случайная величина* η *ограничена* (*т.е. с вероятностью единица* $-\infty < A \leqslant \eta \leqslant B < \infty$), $V_n(x,y) \geqslant 0$ *и* $q(n) = O(\sqrt{n})$, *то*

$$M r_n(x) = \frac{M\varphi_n(x)}{M f_n(x)} + O\left(\frac{q^2(n)}{n}\right).$$

Теорема 1 показывает, что оценка $r_n(x)$ является асимптотически несмещенной (корректной) оценкой для $r(x)$.

Теорема 2. *Пусть с вероятностью единица* $-\infty < A \leqslant \eta \leqslant B < \infty$,

$$\min_{-\infty < a \leqslant x \leqslant b < \infty} f(x) = \mu > 0.$$

Кроме того, предположим, что ряды (1), (1′) *в отрезке* $[a,b]$ *сходятся соответственно к* $f(x)$, $\varphi(x)$ *равномерно. Если при любом* $\beta > 0$

$$\sum_{n=1}^{\infty} q(n)\exp\{-\beta n/q^2(n)\} < \infty,$$

то с вероятностью единица при $n\to\infty$

$$\sup_{a\leqslant x\leqslant b}|r_n(x)-r(x)| \to 0.$$

Предположим, что имеют место разложения (1), (1′) и $f(x)$, $\varphi(x)$ имеют производные r-порядка ($r = 2,3,\ldots$). Кроме того, пусть коэффициенты разложения (1), (1′) удовлетворяют условию

$$c_j = O\left(\frac{1}{j^r}\right), \quad a_j = O\left(\frac{1}{j^r}\right), \quad r\geqslant 2 \ \text{при} \ j\to\infty. \tag{3}$$

Следующая теорема показывает скорость сходимости оценки $r_n(x)$ к $r(x)$.

Теорема 3. *Пусть выполняется* (3) *и с вероятностью единица* $-\infty < A \leqslant \eta \leqslant B < \infty$,

$$\min_{a\leqslant x\leqslant b} f(x) = \mu > 0.$$

Тогда при выборе

$$q(n) \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2r}} \quad \textit{при} \quad n\to\infty$$

с вероятностью единица

$$\sup_{a\leqslant x\leqslant b}|r_n(x)-r(x)| = O\left(\left(\frac{\ln n}{n}\right)^{\frac{r-1}{2r}}\right).$$

Теоремы 2 и 3 доказываются методом работы [2].

Литература

[1] W. Greblicki, *Identyfikacja statyczna metodą szeregów ortogonalnych*, Pods. Sterow. 4.1 (1974).

[2] M. A. Mirzahmedow, S. A. Hašimov, *On some properties of density estimation*, Colloquia mathematica societatis Janos Bolyai, 1974.

Presented to the Semester
Mathematical Models and Numerical Methods
(February 3–June 14, 1975)