ПОСЛЕДОВАТЕЛЬНОЕ НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНОВАНИЕ ИНТЕРВАЛАМИ ФИКСИРОВАННОЙ ШИРИНЫ ФУНКЦИИ РЕГРЕССИИ

Г. Г. Рахимова, Г. Т. Турсунов

Национальный университет Узбекистана имени Мирзо Улугбека
Ташкент, Узбекистан
E-mail: rakhimova_gulnoza@mail.ru

В данной работе рассмотрены непараметрическая оценка интервалами фиксированной ширины функции регрессии. Получена асимптотическая нормальность этой оценки. Для момента остановки показаны асимптотическая состоятельность и эффективность.

Ключевые слова: случайный вектор, регрессия, стохастическая аппроксимация, сильная состоятельность, асимптотическая нормальность, фиксированная ширина.

Рассмотрим случайный вектор $(\xi; \eta)$, $\xi \in [0, 1]$, $\eta \in R_1$. Обозначим через $f(x)$ плотность вероятности случайной величины ξ, через $r(x) = M(\eta/\xi = x)$, $x \in [0; 1]$, функцию регрессии.

Пусть $(\xi_1; \eta_1)$, $(\xi_2; \eta_2)$, …, $(\xi_n; \eta_n)$, … – независимые одинаково распределенные с $(\xi; \eta)$ случайные векторы. Непараметрические оценки функции регрессии $r(x)$ и её свойства изучены многими авторами (см., например [1–3]).

В работе [1] П. Ревеша, используя метод стохастической аппроксимации, построена следующая реккурентная оценка для $r(x)$:

$$r_{n+1}(x) = r_n(x) + \frac{1}{(n+1)a_{n+1}} K\left(\frac{x-\xi_{n+1}}{a_{n+1}}\right)\cdot(\eta_{n+1} - r_n(x)), \tag{1}$$

где $r_0(x) = 0$, $0 \le x \le 1$, $a_n = n^{-\alpha}$, $0 < \alpha < 1$, и $K(x) = 1$, если $|x| \le 0{,}5$, $K(x) = 0$, в противном случае. Он получил условия асимптотической сильной состоятельности и асимптотической нормальности оценки (1). Обозначим:

$$\sigma^2(x) = \frac{2f(x) - 1 + \alpha}{b(x)f(x)}, \quad b(x) = M((\eta - r(x))^2/\xi = x).$$

Введем условия:

(A): $\dfrac{1-\alpha}{2} \le f(x) \le c$ для всех $x \in [0; 1]$, $f(x)$ дифференцируема и $\left|\dfrac{df(x)}{dx}\right| \le c$, $r(x)$ дифференцируема и $\left|\dfrac{dr(x)}{dx}\right| \le c$.

(B): $M(e^{|\eta - r(x)|}/\xi = x) \le c$.

Здесь и в дальнейшем c означает конечную постоянную, не всегда одну и ту же. Символ $\Rightarrow$ означает слабую сходимость функций распределений случайных величин и векторов, а также всех конечномерных распределений случайных процессов и полей.

Теорема 1 [1]. Если выполнены условия (A), (B) и 1/2 < α < 1, то для всех $x \in [0;\,1]$, при $n \to \infty$

$$\sigma(x)\sqrt{na_n}\,(r_n(x) - r(x)) \Rightarrow N(0{,}1),$$

здесь $N(0,1)$ – стандартная нормальная случайная величина со средним 0 и дисперсией 1.

Пусть $N(\varepsilon)$ – целочисленная неотрицательная случайная величина, определенная на том же вероятностном пространстве $(\Omega,\ \Im,\ P)$, где были заданы случайные вектора $(\xi_k,\ \eta_k)$, $k \ge 1$, такая, что

(C): $\dfrac{N(\varepsilon)}{n(\varepsilon)} \xrightarrow{P} \nu$ при $\varepsilon \to 0$,

здесь ν – с вероятностью 1 положительная случайная величина и $n(\varepsilon)$ – неслучайная неотрицательная функция такая, что $n(\varepsilon) \to \infty$ при $\varepsilon \to 0$.

Теорема 2. Если выполнены условия (A), (B), (C) и 1/2 < α < 1, то для всех $x \in [0;\,1]$, при $\varepsilon \to 0$

$$\sigma(x) N(\varepsilon)^{\frac{1-\alpha}{2}} (r_{N(\varepsilon)}(x) - r(x)) \Rightarrow N(0{,}1).$$

Используя эту предельную теорему, построим доверительный интервал фиксированной ширины для $r_n(x)$.

Пусть $0 < \gamma < 1$, $a = \Phi^{-1}\left(\dfrac{1+\gamma}{2}\right)$, $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} \exp\left(-\dfrac{y^2}{2}\right) dy$.

Если для некоторого $\varepsilon > 0$ такого, что $\dfrac{\varepsilon\sigma(x)}{a} \cdot n^{\frac{1-\alpha}{2}} \ge 1$, то

$$P\{r(x) \in (r_n(x) - \varepsilon,\ r_n(x) + \varepsilon)\} \ge P\left\{\sigma(x) \cdot n^{\frac{1-\alpha}{2}} \,|\, r_n(x) - r(x) \,| \le a\right\}.$$

Следовательно, в силу теоремы 1

$$\lim_{\varepsilon \to 0} P\{r(x) \in (r_{n(\varepsilon)} - \varepsilon,\ r_{n(\varepsilon)} + \varepsilon)\} \ge 2\Phi(a) - 1 = \gamma,$$

где $n(\varepsilon) = \min(n \ge 1 : n \ge n_0(\varepsilon))$, $n_0(\varepsilon) = \left(\dfrac{a^2}{\varepsilon^2 \sigma^2(x)}\right)^{\frac{1}{1-\alpha}}$.

Момент остановки $n(\varepsilon)$ обладает тем недостатком, что в его определении через $\sigma(x)$ участвуют неизвестные $f(x)$ и $b(x)$. Поэтому вместо $n(\varepsilon)$ рассмотрим случайный момент остановки

$$N(\varepsilon) = \min\left(n \ge 1 : n \ge \left(\frac{a^2}{\varepsilon^2 \sigma_n^2}\right)^{\frac{1}{1-\alpha}}\right), \tag{2}$$

здесь

$$\sigma_n^2(x) = \frac{2 f_n(x) - 1 + \alpha}{b_n(x) f_n(x)}, \quad f_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} K\left(\frac{x - \xi_i}{a_n}\right),$$

$$b_n(x) = \frac{\varphi_n(x)}{f_n(x)} - r_n^2(x), \quad \varphi_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} \eta_i^2 K\left(\frac{x - \xi_i}{a_n}\right).$$

Обозначим через $\Re$ множество функций регрессий, удовлетворяющих определенным условиям.

Следуя [4], введём следующие определения.

Определение 1. Доверительный интервал

$$I_\varepsilon(r_{N(\varepsilon)}(x)) = (r_{N(\varepsilon)}(x) - \varepsilon,\ r_{N(\varepsilon)}(x) + \varepsilon)$$

фиксированной ширины 2ε называется асимптотически состоятельным, если

$$\lim_{\varepsilon\to 0} P\{r(x) \in I_\varepsilon(N(\varepsilon))\} \geq \gamma$$

для всех $r(x) \in \Re$ и некоторой $0 < \gamma < 1$.

Определение 2. Момент остановки $N(\varepsilon)$, определенный в (2), называется асимптотически эффективным, если доверительный интервал $I_\varepsilon(N(\varepsilon))$ асимптотически состоятелен и

$$\lim_{\varepsilon\to 0} \frac{M(N(\varepsilon))}{n(\varepsilon)} = 1.$$

Заметим, что $n(\varepsilon)$ равен минимальному количеству наблюдений, необходимому для того, чтобы асимптотическая при $\varepsilon \to 0$ вероятность накрытия интервалом $I_\varepsilon(n(\varepsilon))$ «известной» функции регрессии была не меньше чем γ.

Теорема 3. Если выполнены условия (A), (B), (C) и $1/2 < \alpha < 1$, то доверительный интервал $I_\varepsilon(N(\varepsilon))$ фиксированной ширины с моментом остановки (2) асимптотически состоятелен.

Теорема 4. Если выполнены условия (A), (B), (C), $1/2 < \alpha < 2/3$ и $c_1 \leq \eta \leq c_2$, то момент остановки (2) является асимптотически эффективным.

Теорема 5. Если выполнены условия (A), (B), (C) и $1/2 < \alpha < 1$, то

$$P\left(\left|\frac{N(\varepsilon)}{n(\varepsilon)} - 1\right| \geq \varepsilon^{2\alpha}\right) = O(\varepsilon^{\alpha}) \quad \text{при } \varepsilon \to 0.$$

Теорема 6. Если выполнены условия (A), (B), (C), $1/2 < \alpha < 2/3$ и $c_1 \leq \eta \leq c_2$, то

$$P\{r(x) \in I_\varepsilon(N(\varepsilon))\} = \gamma + O\left(\varepsilon^{\frac{2\alpha}{3}}\right) \quad \text{при } \varepsilon \to 0.$$

ЛИТЕРАТУРА

1. *Revesz, P.* How to apply the method of stochastic approximation in the non-parametric estimation of a regression function / P. Revesz // Math. Oper. Statist. ser. Statistics. 1977. V. 8. No. 1, P. 119–126.

2. *Надарая, Э. А.* Об оценке регрессии / Э. А. Надарая // Теория вер. и ее прим. 1964. № 9. С.157–159.

3. *Samanta, M.* Nonparametric estimation of a multivariate multiple regression function / M. Samanta, R. Mugisha // Statistische Hefte. 1984. 25. P. 297–317.

4. *Chow, J. S.* On the asymptotic theory of fixed-width sequential confidence intervals for the mean / J. S. Chow, H. Robbins // Ann. Math.Statist. 1965. V. 36, No 2, P. 457–462.