УДК 519.2

Ю.Г. Дмитриев, Ж.Н. Зенкова

ЯДЕРНАЯ ОЦЕНКА ПЛОТНОСТИ НЕРАВНОПЛЕЧНО СИММЕТРИЧНОГО РАСПРЕДЕЛЕНИЯ

Предлагается ядерная оценка неизвестной плотности распределения, модифицированная с учетом знания ее неравноплечной симметрии относительно известного центра. Показана сходимость оценки к истинной плотности в среднеквадратическом.

При обработке статистических данных нередко возникает необходимость построения оценки неизвестной функции плотности распределения исследуемой случайной величины. На практике очень часто возникают ситуации, когда существует дополнительная информация о случайной величине, например ее положительности, ограниченности, непрерывности, симметричности, моментах и пр. Источником этой информации могут служить условия эксперимента, теоретические выводы, физический смысл анализируемой случайной величины и т.д. Следовательно, возникают задачи учета дополнительной информации при построении оценок плотности, а также исследовании качества модифицированных статистик.

S_p^c -неравноплечная симметрия функции распределения

Пусть случайная величина ξ имеет функцию распределения $F(x)$ с плотностью $f(x)$ относительно меры Лебега на прямой $x \in \mathbf{R}$.

Определение. Будем говорить, что функция распределения (ф.р.) $F(x)$, $x \in \mathbf{R}$, обладает *свойством* S_p^c*-неравноплечной симметрии* относительно центра c с известным весом

$$p = F(c), \tag{1}$$

если она удовлетворяет условию

$$F(\max\{x, S(x)\}) = 1 - \frac{1-p}{p} F(\min\{x, S(x)\}), \tag{2}$$

где функция $S(x)$ – непрерывна, монотонно убывает и удовлетворяет требованиям

$$(S)^{-1}(x) = S(x)\,,\ S(c_1) = c_0\,,\ S(c) = c,$$

где

$$c_0 = \arg\min_{x \in R}(F(x)),\ \ c_1 = \arg\max_{x \in R}(F(x)).$$

Заметим, что при

$$p = F(c) = 0{,}5\,,\ S(x) = 2c - x$$

получим обычную симметрию

$$F(x) = 1 - F(2c_1 - x + 0) \tag{3}$$

относительно центра c.

Оценка S_p^c -неравноплечно симметричной функции плотности

Пусть $X = \{X_1, \ ..., \ X_N\}$ – последовательность независимых одинаково распределенных наблюдений над случайной величиной ξ (выборка объема N) с неизвестной ф.р. $F(x)$, $x \in \mathbf{R}$, обладающей свойством (2).

Оценка плотности $f(x)$ по выборке X с учетом свойства S_p^c -неравноплечной симметрии строится по формуле

$$f_N^{Sp}(x) = g(x)\big(f_N(x) - S'(x) f_N(S(x))\big), \tag{4}$$

где

$$g(x) = \begin{cases} p, & x \le c, \\ -(1-p), & x > c, \end{cases} \tag{5}$$

$$f_N(x) = \frac{1}{N h_N} \sum_{i=1}^{N} K\left(\frac{x - X_i}{h_n}\right), \tag{6}$$

где $f_N(x)$ – ядерная оценка плотности Розенблатта–Парзена, $K(t)$ – ядро оценки (обычно это некоторая плотность распределения вероятностей), h_N – параметр размытости [1].

Формулу (4) можно также представить в виде

$$f_N^{Sp}(x) = \frac{g(x)}{N h_N} \sum_{i=1}^{N} \left(K\left(\frac{x - X_i}{h_n}\right) - S'(x) K\left(\frac{S(x) - X_i}{h_n}\right)\right). \tag{7}$$

Теорема 1 [2]**.** Если ядро

$$K(t) \ge 0,\ t \in \mathbf{R}, \tag{8}$$

$$\int_{-\infty}^{+\infty} K(t)\,dt = 1, \tag{9}$$

$$\sup_{t \in \mathbf{R}} K(t) < \infty, \tag{10}$$

$$\lim_{|t| \to \infty} t K(t) = 0, \tag{11}$$

h_N – последовательность положительных величин, такая, что

$$\lim_{N \to \infty} h_N = 0, \tag{12}$$

то в любой точке непрерывности функции $f(x)$

$$\lim_{N \to \infty} \mathbf{M} f_N(x) = f(x). \tag{13}$$

Теорема приводится без доказательства.

Теорема 2. Пусть плотность $f(x)$ непрерывна и обладает свойством (2). Тогда при выполнении условий (8)–(11)

$$\lim_{N\to\infty} \mathbf{M} f_N^{Sp}(x) = f(x). \tag{14}$$

Доказательство. Пользуясь (13), получим

$$\begin{aligned} &\lim_{N\to\infty} \mathbf{M} f_N^{Sp}(x) = \\ &= g(x) \lim_{N\to\infty} \mathbf{M}\big(f_N(x) - S'(x) f_N(S(x))\big) = \\ &= g(x) \lim_{N\to\infty} \mathbf{M} f_N(x) - g(x) S'(x) \lim_{N\to\infty} \mathbf{M} f_N(S(x)) = \\ &= g(x)\big(f(x) - S'(x) f(S(x))\big). \end{aligned}$$

Пусть $x \le c$. Применим свойство (2), которое для плотностей имеет вид

$$f(x) = p(x) S'(x) f(S(x)), \tag{15}$$

где

$$p(x) = \begin{cases} -\dfrac{p}{1-p}, \ x \le c; \\ -\dfrac{1-p}{p}, \ x > c. \end{cases}$$

Будем иметь

$$S'(x) f(S(x)) = -\frac{1-p}{p} f(x), \quad g(x) = p,$$

поэтому

$$\lim_{N\to\infty} \mathbf{M} f_N^{Sp}(x) = p\left(f(x) + \frac{1-p}{p} f(x)\right) = f(x).$$

Для $x > c$ доказательство строится аналогично.

Теорема доказана.

Таким образом, модифицированная оценка (4) является асимптотически несмещенной.

В [2, 3] показано, что если выполняются условия (8)–(12), плотность $K(t)$ имеет все моменты,

$$\int_{-\infty}^{+\infty} t K(t) dt = 0, \tag{16}$$

$$\int_{-\infty}^{+\infty} K^2(t) dt < \infty, \tag{17}$$

$$\lim_{N\to\infty} N h_N = \infty, \tag{18}$$

то

$$\lim_{N\to\infty} \mathbf{M}\big(f_N(x) - f(x)\big)^2 = 0. \tag{19}$$

Теорема 3. Пусть плотность $f(x)$ непрерывна и обладает свойством (2), ядро $K(t)$ имеет все моменты. Тогда при выполнении условий (6), (7), (11)

$$\lim_{N\to\infty} \mathbf{M}\big(f_N^{Sp}(x) - f(x)\big)^2 = 0. \tag{20}$$

Доказательство. Пусть $x \le c$, тогда, воспользовавшись (15), получим

$$-pS'(x) f(S(x)) = (1-p) f(x),$$

следовательно,

$$f(x) = p f(x) - p S'(x) f(S(x)),$$

поэтому

$$\begin{aligned} &\lim_{N\to\infty} \mathbf{M}\big(f_N^{Sp}(x) - f(x)\big)^2 = \lim_{N\to\infty} \mathbf{M}\big(p f_N(x) - \\ &- pS'(x) f_N(S(x)) - p f(x) + pS'(x) f(S(x))\big)^2 = \\ &= p^2 \lim_{N\to\infty} \mathbf{M}\big(f_N(x) - f(x)\big)^2 - 2p^2 S'(x) \times \\ &\times \lim_{N\to\infty} \mathbf{M}\{\big(f_N(x) - f(x)\big)\big(f_N(S(x)) - f(S(x))\big)\} + \\ &+ p^2 \lim_{N\to\infty} \mathbf{M}\big(f_N(S(x)) - f(S(x))\big)^2. \end{aligned}$$

Применив формулу (19), получим

$$\begin{aligned} &\lim_{N\to\infty} \mathbf{M}\big(f_N^{Sp}(x) - f(x)\big)^2 = -2p^2 S'(x) \times \\ &\times \lim_{N\to\infty} \mathbf{M}\{\big(f_N(x) - f(x)\big)\big(f_N(S(x)) - f(S(x))\big)\} = \\ &= -2p^2 S'(x) \lim_{N\to\infty} \operatorname{cov}\big(f_N(x), f_N(S(x))\big). \end{aligned}$$

В [4. С. 145] показано, что для непрерывной на **R** плотности $f(x)$ при выполнении условий теоремы 1 для достаточно больших N

$$\begin{aligned} &\operatorname{cov}\big(f_N(x), f_N(y)\big) \approx \\ &\approx \frac{\delta(x,y)}{2Nh_N}\big(f(x) + f(y)\big) \int_{-\infty}^{+\infty} K^2(t) dt + o\left(\frac{1}{Nh_N}\right), \end{aligned} \tag{21}$$

где $\delta(x,y) = \{1 : x = y, 0 : x \ne y\}$ – функция Кронекера.

В данном случае $\delta(x,y) = 1$ лишь при $x = S(x)$, что возможно только тогда, когда $x = c = S(c)$. Таким образом, выражение (21) принимает следующий вид:

$$\operatorname{cov}\big(f_N(x), f_N(S(x))\big) \approx \frac{f(c)}{Nh_N} \int_{-\infty}^{+\infty} K^2(t) dt + o\left(\frac{1}{Nh_N}\right).$$

По условиям теоремы $\lim_{N\to\infty} Nh_N = \infty$, $\int_{-\infty}^{+\infty} K^2(t) dt < \infty$. Значение $f(c) < \infty$ в силу интегрируемости $f(x)$, тогда

$$\lim_{N\to\infty} \operatorname{cov}\big(f_N(x), f_N(S(x))\big) = 0,$$

из чего следует (20).

Для случая $x > c$ доказательство строится аналогично.

Теорема доказана.

Таким образом, использование дополнительной информации об S_p^c-неравноплечной симметрии случайной величины позволяет получить асимптотически несмещенную модифицированную ядерную оценку плотности с нулевым в асимптотике среднеквадратическим отклонением.

ЛИТЕРАТУРА

1. *Rosenblatt M.* Remarks on some nonparametric estimates of a density functions // Ann. Math. Statist. 1956. Vol. 27, № 3. P. 832–837.
2. *Parzen E.* On estimation of a probability density function and mode // Ann. Math. Statist. 1962. Vol. 33. P. 1065–1076.
3. *Тарасенко Ф.П.* Непараметрическая статистика. Томск: Изд-во ТГУ, 1976. 294 с.
4. *Непараметрическое* оценивание функционалов от распределений стационарных последовательностей / В.А. Васильев, А.В. Добровидов, Г.М. Кошкин / Отв. ред. Н.А. Кузнецов. М.: Наука, 2004. 508 с.

Статья представлена кафедрой теоретической кибернетики факультета прикладной математики и кибернетики Томского государственного университета, поступила в научную редакцию «Кибернетика» 30 мая 2006 г.