УДК 519.21

В. С. Королюк, О. Меляева

Мартингальные оценки для распределения уклонения оценки плотности

Пусть $X_1, \ldots, X_n$ — независимые вещественные случайные величины с плотностью распределения $f(x)$. В качестве статистической оценки плотности $f(x)$ применяются ядерные оценки вида (см. [2—4])

$$f_n(x) = (na_n)^{-1} \sum_{i=1}^{n} W((x - X_i)/a_n) \tag{1}$$

с условиями для положительных чисел a_n:

$$a_n \to 0, \quad na_n \to \infty, \ n \to \infty. \tag{2}$$

Ядро $W(x)$ — заданная симметричная относительно нуля плотность распределения.

Мерой качества этой оценки служит среднеквадратическое уклонение *

$$B_n^2 = \int_{-\infty}^{+\infty} [f_n(x) - Ef_n(x)]^2 h(x)\,dx \tag{3}$$

с неотрицательной весовой функцией $h(x)$.

Известно [1, 3], что при некоторых дополнительных условиях уклонение B_n^2 асимптотически нормально, что следует из центральной предельной теоремы (ЦПТ) для мартингалов [5]. Однако дополнительные условия существования зависят от метода сведения задачи к ЦПТ для мартингалов (см. [1, 4, 6]).

В настоящей работе используется метод представления B_n^2 в виде U-статистики с переменным (зависящим от n) ядром [3]. При этом удается получить вполне естественные условия выполнимости ЦПТ и, более того, оценить скорость сходимости распределения B_n^2 к предельному нормальному закону.

Теорема 1. *Пусть функции $f(x)$ и $h(x)$ непрерывны, ограничены и таковы, что конечны интегралы*

$$\int |h(x) - \int f(y) h(y)\,dy|^2 f(x)\,dx, \quad \int [f(x) h(x)]^2\,dx \tag{4}$$

и плотность вероятностей $W(x)$ интегрируема с квадратом. Тогда при выполнении условия (2) выполняется соотношение при $n \to \infty$

$$\Delta_n = \sup_x |P\{\sigma_n^{-1}(B_n^2 - EB_n^2) < x\} - \Phi(x)| \to 0. \tag{5}$$

Здесь $\sigma_n^2 = E(B_n^2 - EB_n^2)^2$, $\Phi(x)$ — стандартное нормальное распределение.

Для доказательства теоремы 1 представим B_n^2 в виде U-статистики (см. [1 — 3]). Сначала представим B_n^2 в виде

$$B_n^2 = \sum_{i,j=1}^{n} \psi(X_i, X_j) \tag{6}$$

с ядром

$$\psi(x, y) = \int H_n(x, u) H_n(y, u) h(u)\,du, \tag{7}$$

где

$$H_n(x, u) = (na_n)^{-1} [W((x - u)/a_n) - EW((X_1 - u)/a_n)]. \tag{8}$$

Из (6) — (8) следует

$$B_n^2 - EB_n^2 = \sum_{i,j=1} \Phi(X_i, X_j) \tag{9}$$

* Интегрирование здесь и далее ведется по всей вещественной прямой.

с ядром

$$\Phi(X_i, X_j) = 2\psi(X_i, X_j) + \frac{1}{n-1}[\psi(X_i, X_i) - E\psi(X_i, X_i)] +$$

$$+ \frac{1}{n-1}[\psi(X_j, X_j) - E\psi(X_j, X_j)]. \tag{10}$$

Для ядра (10) вычислим формулы (см. [1, с 37]):

$$g_1(x) = E[\Phi(X_1, X_2)/X_1 = x] = \frac{1}{n-1}[\psi(x, x) - E\psi(X_1, X_1)],$$

$$g_2(x, y) = \Phi(x, y) - g_1(x) - g_1(y) = 2\psi(x, y). \tag{11}$$

Заметим, что ядро (10) существенно зависит от n. Поэтому ЦПТ не следует непосредственно из результатов работы [2]. В соответствии с ЦПТ для U-статистик, сформулированной в работе [1, с. 33], будем проверять выполнение следующих условий при $n \to \infty$:

$$l_1(\Phi) \to 0, \quad l_2(\Phi) \to 0, \tag{12}$$

где при некотором фиксированном $\delta > 0$

$$l_1(\Phi) = n^{-\delta/2} \sum_{k=1}^{2} (Eg_k^2)^{-1-\delta/2} E|g_k|^{2+\delta}, \tag{13}$$

$$l_2(\Phi) = n^{-4} E|G_2(X_1, X_2)|^2 / \Big[\sum_{k=1}^{2} n^{-k} Eg_k^2\Big]. \tag{14}$$

Здесь

$$G_2(x, y) = E[g_2(x, X_1) g_2(y, X_1)]. \tag{15}$$

Оценим математические ожидания, входящие в (13) и (14) с ядром (10). Из (7) и (8) после замены переменной интегрирования имеем

$$\psi(x, x) = a_n^{-1} n^{-2} \int \Big(W(u) - a_n \int W(y) f(x + a_n(u + y))\, dy\Big)^2 h(x + a_n u)\, du.$$

Так что при $n \to \infty$

$$a_n n^2 \psi(x, x) \to h(x) \int W^2(u)\, du. \tag{16}$$

Аналогично при $n \to \infty$ получаем

$$a_n n^2 E\psi(X_1, X_1) \to \int h(x) f(x)\, dx \int W^2(u)\, du. \tag{17}$$

Теперь при некотором $p > 0$ имеем (см. (11))

$$E|g_1|^p = (n-1)^p \int |\psi(x, x) - E\psi(X_1, X_1)|^p f(x)\, dx.$$

Следовательно, при $n \to \infty$

$$(n-1)^p n^{2p} a_n^p E|g_1|^p \to \Big(\int W^2(u)\, du\Big)^p \int \Big|h(x) - \int h(y) f(y)\, dy\Big|^p f(x)\, dx. \tag{18}$$

Рассмотрим теперь $E|g_2|^p$ в соответствии с (11) и (7). Имеем представление

$$E|g_2|^p = 2^p \iint \Big|\int H_n(x, u) H_n(y, u) h(u)\, du\Big|^p f(x) f(y)\, dx dy. \tag{19}$$

После замены переменных $u = x + a_n u'$, $y = x + a_n y'$ с учетом (8) при $n \to \infty$ получим

$$2^{-p} n^{2p} a_n^{p-1} E|g_2|^p \to \int \Big(\int W(u) W(u - y)\, du\Big)^p dy \int (f(x) h(x))^2\, dx. \tag{20}$$

Осталось оценить $E|G_2(X_1, X_2)|^p$. С учетом (11) и (15) имеем представление

$$E|G_2(X_1, X_2)|^p = 4^p \int\int \Big| \int\int\int H_n(x, u) H_n(y, u) H_n(y, v) H_n(y_1, v) \times$$
$$\times h(u) h(v) \, du dv f(y_1) \, dy_1 \Big|^p f(x) f(y) \, dx dy. \quad (21)$$

После соответствующих замен переменных при $n \to \infty$ получим

$$4^{-p} n^{4p} a_n^{p-1} E|G_2(X_1, X_2)|^p \to \int \Big(\int W_2(u) W_2(u+v) \, du \Big)^p \times$$
$$\times dv \int h^{2p}(x) f^{p+2}(x) \, dx. \quad (22)$$

Теперь применим результаты (18), (20) и (22) при $p = 2$ и $p = 2 + \delta$ для оценки $l_k(\Phi)$, $k = 1, 2$ (см. (13) и (14)). После несложных преобразований с учетом условий теоремы будем иметь

$$l_1(\Phi) \leqslant C_1 (na_n)^{-\delta/2}, \quad l_2(\Phi) \leqslant C_2 a_n \quad (23)$$

с некоторыми константами C_1 и C_2. Из условия (3) и (23) следует утверждение теоремы.

З а м е ч а н и е 1. Условие (2) является оптимальным в том смысле, что при его выполнении оценка плотности (1) состоятельна в среднем квадратическом, если $f(x)$ непрерывна и ограничена.

З а м е ч а н и е 2. Условия теоремы 1 на функции $f(x)$, $h(x)$ и $W(x)$ отличны от условий работ [4, 6], где преполагается дифференцируемость $f(x)$ и конечность вторых моментов $W(x)$.

Применение теоремы 5.1 работы [1] позволяет оценить скорость сходимости Δ_n к нулю.

Т е о р е м а 2. *При выполнении условий теоремы 1 и дополнительном предположении конечности функционалов*

$$\int \Big| h(x) - \int h(y) f(y) \, dy \Big|^2 (h(x) f(x))^2 \, dx,$$
$$\int \Big| h^2(x) f(x) - \int h^2(y) f^2(y) \, dy \Big|^2 f(x) \, dx \quad (24)$$

справедливо неравенство

$$\Delta_n \leqslant C a_n^{1/8}. \quad (25)$$

1. *Королюк В. С., Боровских Ю. В.* Мартингальная теория U-статистик.— Киев, 1984.— 56 с.— (Препринт / АН УССР, Ин-т математики; 84.63).
2. *Королюк В. С., Боровских Ю. В.* Асимптотический анализ распределений статистик.— Киев : Наук. думка, 1984.— 304 с.
3. *Боровских Ю. В.* Проблема аппроксимации распределений U-статистик и функционалов Мизеса, II.— Киев, 1980.— 56 с.— (Препринт / АН УССР, Ин-т математики; 80.7).
4. *Надарая Э. А.* Применение центральной предельной теоремы для мартингалов к исследованию предельного распределения квадратического уклонения оценки плотности типа ядра.— Сообщ. АН ТССР, 1984, **113**, № 2, с. 253—256.
5. *Липцер Р. Ш., Ширяев А. Н.* Функциональная центральная предельная теорема для семимартингалов.— Теория вероятностей и ее применения, 1980, **25**, № 4, с. 683—703.
6. *Hall P.* Central limit theorem for integrated square error of multivariate nonparametric density etstimators.— J. Multivar. Anal., 1984, **14**, N 1, p. 1—16.

Ин-т математики АН УССР, Киев

Получено 19.09.85