УДК 519.21

Ю. В. Боровских

Двумерная аппроксимация U-статистик

1. Пусть имеется c независимых последовательностей $\{X_{ji},\ i=1,2,\ldots\}$, $j=1,\ldots,c$, независимых случайных величин, причем все величины из j-й последовательности принимают значения в измеримом пространстве $(\mathfrak{X}_j, \mathcal{A}_j)$ и имеют на нем одно и то же вероятностное распределение P_j. Пусть $p=1,2$; $\Phi^{(p)}(x_{ji},\ i=1,\ldots,m_{jp},\ j=1,\ldots,c)$—две измеримые функции, которые являются симметрическими относительно m_{jp} аргументов j-го множества. Рассмотрим двумерную обобщенную U-статистику $U=(U^{(1)},U^{(2)})$, координаты которой $U^{(p)}$, $p=1,2$, определяются как (см. [1, 2])

$$U^{(p)}=\prod_{j=1}^{c}\binom{n_j}{m_{jp}}^{-1}\sum\Phi^{(p)}(x_{j\alpha},\ \alpha=i_{j1},\ldots,i_{jm_{jp}},\ 1\leqslant j\leqslant c), \qquad (1)$$

где суммирование осуществляется по всем индексам $j=1,\ldots,c$, $1\leqslant i_{j1}<\ldots$ $\ldots<i_{jm_{jp}}\leqslant n_j$, $n_j\geqslant m_j$, причем предполагается, что $E\Phi^{(p)}=0$, $E(\Phi^{(p)})^2<\infty$, $p=1,2$. Функция $\Phi^{(p)}$ называется ядром $U^{(p)}$-статистики, целочисленный вектор $(m_{1p},\ldots,m_{cp})$ — степенью ядра, число c — порядком $U^{(p)}$-статистики (1). Обозначим

$$g_j^{(p)}(x)=E(\Phi^{(p)}(X_{11},\ldots,X_{1m_{1p}};\ldots,X_{c1},\ldots,X_{cm_{cp}})\,|\,X_{j1}=x),$$

$$(\sigma_j^{(p)})^2=E(g_j^{(p)}(X_{j1}))^2,\quad \sigma_j^{(1,2)}=E(g_j^{(1)}(X_1)\,g_j^{(2)}(X_{j1})),\quad j=1,\ldots,c,$$

$$(\hat{\sigma}_p)^2 = \sum_{j=1}^{c} m_{jp}^2 n_j^{-1} (\sigma_j^{(p)})^2, \quad p = 1, 2, \quad \rho_c = (\hat{\sigma}_1 \hat{\sigma}_2)^{-1} \sum_{j=1}^{c} m_{j1} m_{j2} n_j^{-1} \sigma_j^{(1,2)}.$$

Далее, пусть $\Phi_c(x_1, x_2) = P(\sigma_{11}^{-1} U^{(1)} < x_1; \sigma_{22}^{-1} U^{(2)} < x_2)$, где $\sigma_{ij} = E(U^{(i)} U^{(j)})$, $\Phi(x_1, x_2)$ — двумерное нормальное распределение с характеристической функцией

$$\varphi(t_1, t_2) = \exp\left(-\frac{1}{2}(t_1^2 + 2\rho_c t_1 t_2 + t_2^2)\right),$$

где $x_i, t_i \in R$.

Двумерная центральная предельная теорема для обобщенных U-статистик утверждает (см. [2]), что при $n = \min(n_1, \ldots, n_c) \to \infty$ величина

$$\Delta = \sup_{x_1, x_2} |\Phi_c(x_1, x_2) - \Phi(x_1, x_2)| \to 0, \tag{2}$$

если $\max(\sigma_1^{(p)}, \ldots, \sigma_c^{(p)}) > 0$, $p = 1, 2$. В работе [1] рассмотрена оценка скорости сходимости в (2) и доказано, что если $E|\Phi^{(p)}|^3 < \infty$, $p = 1, 2$, то $\Delta = O(n^{-1/2})$. При более специальных условиях на ядро $\Phi^{(p)}$ такая же оценка получена в [3] с использованием общих результатов и методов из [4].

В данной работе оценка $\Delta = O(n^{-1/2)})$ вычислена при условиях, которые слабее, чем $E|\Phi^{(p)}|^3 < \infty$, $p = 1, 2$. Всюду в дальнейшем для простоты считаем, что $\sigma_j^{(p)} > 0$ при всех $j = 1, \ldots, c$ и $p = 1, 2$. Обозначим $\Gamma_c^{(p)} = \max_{1 \leqslant j \leqslant c} E|g_j^{(p)}(X_{j1})|^3$, $p = 1, 2$.

2. Теорема. *Пусть выполнены условия*

$$\max(\Gamma_c^{(p)}, E(\Phi^{(p)})^2) < \infty, \quad p = 1, 2. \tag{3}$$

Тогда при $n \to \infty$

$$\sup_{x_1, x_2} |\Phi_c(x_1, x_2) - \Phi(x_1, x_2)| = O\ n^{-1/2}). \tag{4}$$

Доказательство. Будем следовать методу, изложенному в [1]. Обозначим

$$\hat{U}^{(p)} = \sum_{j=1}^{c} \sum_{i=1}^{n_j} m_{jp} n_j^{-1} g_j^{(p)}(X_{ji}), \quad Y_{i_{11}, \ldots, i_{cm_{cp}}}^{(p)} = \Phi^{(p)}(X_{1i_{11}}, \ldots, X_{ci_{cm_{cp}}}) -$$

$$- \sum_{j=1}^{c} \sum_{k=1}^{m_{jp}} g_j^{(p)}(X_{ji_{jk}}).$$

Случайные величины $Y^{(p)}$ обладают свойством вырожденности:

$$EY_{i_{11}, \ldots, i_{cm_{cp}}}^{(p)} = 0, \quad E(Y_{i_{11}, \ldots, i_{cm_{cp}}} | X_{ji_{jk}}) = 0. \tag{5}$$

Далее, пусть

$$S^{(p)} = (\hat{\sigma}_p)^{-1} \hat{U}^{(p)}, \quad \delta^{(p)} = (\hat{\sigma}_p)^{-1}(U^{(p)} - \hat{U}^{(p)}), \quad \delta_1^{(p)} = (\hat{\sigma}_p)^{-1} \prod_{=1}^{c} \binom{n_j}{m_{jp}}^{-1} \times$$

$$\times \sum Y_{i_{11}, \ldots, i_{cm_{cp}}}^{(p)},$$

где суммирование осуществляется по всем индексам $1 \leqslant i_{j1} < \ldots < i_{jm_{jp}} \leqslant k_j$, $k_j = n_j - [\sqrt{n_j}]$, $j = 1, \ldots, c$, $[\sqrt{n_j}]$ — целая часть числа n_j.

Согласно [1]

$$\sup_{x_1, x_2} |\Phi_c(x_1, x_2) - \Phi(x_1, x_2)| \leqslant \sum_{=1}^{4} \Delta_j. \tag{6}$$

Здесь

$$\Delta_1 = \sup_{x_1, x_2} \left| \Phi(x_1, x_2) - \Phi\left(x_1 \frac{\sigma_{11}}{\hat{\sigma}_1}, x_2 \frac{\sigma_{22}}{\hat{\sigma}_2}\right) \right|, \quad \Delta_2 = \sup_{x_1, x_2} | \Phi(x_1, x_2) -$$

$$- \Phi(x_1 + \hat{\sigma}_1, \; x_2 + \hat{\sigma}_2)|,$$

$$\Delta_3 = P(|\delta^{(1)} - \delta_1^{(1)}| > \hat{\sigma}_1) + P(|\delta^{(2)} - \delta_1^{(2)}| > \hat{\sigma}_2), \quad \Delta_4 = \sup_{x_1, x_2} | P(S^{(1)} +$$

$$+ \delta_1^{(1)} < x_1, S^{(2)} + \delta_1^{(2)} < x_2) - \Phi(x_1, x_2)|.$$

Оценка Δ_1. По свойствам нормальной функции распределения (см. [1])

$$\Delta_1 \leqslant \sup_{x_1} \left| \Phi(x_1) - \Phi\left(x_1 \frac{\sigma_{11}}{\hat{\sigma}_1}\right) \right| + \sup_{x_2} \left| \Phi(x_2) - \Phi\left(x_2 \frac{\sigma_{22}}{\hat{\sigma}_2}\right) \right|, \tag{7}$$

где $\Phi(x)$ — стандартная нормальная функция распределения. Известно (см. [2]), что

$$\sigma_{ij} = \prod_{\gamma=1}^{c} \binom{n_\gamma}{m_{\gamma j}}^{-1} \sum_{d_1=0}^{m_1(i,j)} \cdots \sum_{d_c=0}^{m_c(i,j)} \prod_{\gamma=1}^{c} \binom{m_{\gamma i}}{d_\gamma} \binom{n_\gamma - m_{\gamma i}}{m_{\gamma j} - d_\gamma} \eta_{d_1 \ldots d_c}^{(i,j)}. \tag{8}$$

Здесь $i, j = 1, 2$, $m_\gamma^{(i,j)} = \min(m_{\gamma i}, m_{\gamma j})$, $\eta_{d_1 \ldots d_c}^{(i,j)}$ обозначает ковариацию случайных величин $\Phi^{(i)}(X_{11}, \ldots, X_{1m_{1i}}; \ldots; X_{c1}, \ldots, X_{cm_{ci}})$ и $\Phi^{(j)}(X_{11}, \ldots, X_{1d_1}, X'_{1d_1+1}, \ldots \ldots, X'_{1m_{1j}}; \ldots; X_{c1}, \ldots X_{cd_c}, X'_{cd_c+1}, \ldots, X'_{cm_{cj}})$, причем случайные величины X_{ji} и X'_{ji} предполагаются независимыми и имеющими одно и то же распределение P_j при всех i.

Из свойств функции распределения $\Phi(x)$ имеем

$$\sup_{x_1} \left| \Phi(x_1) - \Phi\left(x_1 \frac{\sigma_{11}}{\hat{\sigma}_1}\right) \right| \leqslant \frac{1}{\sqrt{\pi e}} (\sigma_{11} - \hat{\sigma}_1) \hat{\sigma}_1^{-1},$$

откуда с помощью (8) получаем

$$(\sigma_{11} - \hat{\sigma}_1) \hat{\sigma}_1^{-1} = O(1/n), \quad n \to \infty.$$

Аналогичная оценка справедлива и для второго слагаемого справа в (7). Поэтому при $n \to \infty$

$$\Delta_1 = O(n^{-1}). \tag{9}$$

Оценка Δ_2. Сначала имеем

$$\Delta_2 \leqslant \sup_{x_1} | \Phi(x_1) - \Phi(x_1 + \hat{\sigma}_1)| + \sup_{x_2} | \Phi(x_2) - \Phi(x_2 + \hat{\sigma}_2)|. \tag{10}$$

Так как

$$\sup_{x_p} | \Phi(x_p) - \Phi(x_p + \hat{\sigma}_p)| \leqslant \frac{1}{\sqrt{2\pi}} \hat{\sigma}_p, \quad p = 1, 2,$$

то отсюда и из (10) следует, что при $n \to \infty$

$$\Delta_2 = O(n^{-1/2}). \tag{11}$$

Оценка Δ_3. По неравенству Чебышева

$$P(| \delta^{(p)} - \delta_1^{(p)}| > \hat{\sigma}_p) \leqslant \hat{\sigma}_p^{-2} E(\delta^{(p)} - \delta_1^{(p)})^2, \quad p = 1, 2.$$

В [1] с учетом (5) показано, что $E(\delta^{(p)} - \delta_1^{(p)})^2 = O(n^{-3/2})$. Стало быть, при $n \to \infty$

$$\Delta_3 = O(n^{-1/2}). \tag{12}$$

Оценка Δ_4. По неравенству Садиковой [5]

$$\Delta_4 \leqslant \frac{1}{2\pi^2} \int\limits_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} \int\limits_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} |f(t_1, t_2) - h(t_1, t_2)| \, |t_1 t_2|^{-1} dt_1 dt_2 + 2 \sup_{x_1} |P(S^{(1)} +$$

$$+ \delta_1^{(1)} < x_1) - \Phi(x_1)| + 2 \sup_{x_2} |P(S^{(2)} + \delta_1^{(2)} < x_2) - \Phi(x_2)| + \varepsilon_1 n^{-1/2}, \quad (13)$$

где ε — достаточно малое число, ε_1 — некоторая положительная постоянная,

$$f(t_1, t_2) = Ee^{it_1(S^{(1)} + \delta_1^{(1)}) + it_2(S^{(2)} + \delta_1^{(2)})} - Ee^{it_1(S^{(1)} + \delta_1^{(1)})} Ee^{it_2(S^{(2)} + \delta_1^{(2)})},$$

$$h(t_1, t_2) = \varphi(t_1, t_2) - \varphi(0, t_2)\varphi(t_1, 0).$$

В [1] доказано, что при условии (3)

$$\sup_{x_p} |P(S^{(p)} + \delta_1^{(p)} < x_p) - \Phi(x_p)| = O(n^{-1/2}) \quad (14)$$

для $p = 1, 2$.

Рассмотрим интегральное выражение в (13). Очевидно, что

$$|f(t_1, t_2) - h(t_1, t_2)| \leqslant |p(t_1, t_2) - h(t_1, t_2)| + |p(t_1, t_2) - f(t_1, t_2)|.$$

Здесь $p(t_1, t_2) = Ee^{it_1 S^{(1)} + it_2 S^{(2)}} - Ee^{it_1 S^{(1)}} Ee^{it_2 S^{(2)}}$.

Интеграл от $|p(t_1, t_2) - h(t_1, t_2)|$ не превышает величину порядка $O(n^{-1/2})$ (см. [5]). Таким образом,

$$\Delta_4 = \frac{1}{2\pi^2} \int\limits_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} \int\limits_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} |p(t_1, t_2) - f(t_1, t_2)| \, |t_1 t_2|^{-1} dt_1 dt_2 + O(n^{-1/2}).$$

Из соображений независимости (см. [1]) функции $p(t_1, t_2)$ и $f(t_1, t_2)$ допускают следующие представления:

$$p(t_1, t_2) = \prod_{j=1}^{c} \left(\varphi_j \left(\frac{t_1 m_{j1}}{\hat{\sigma}_1 n_j}, \frac{t_2 m_{j2}}{\hat{\sigma}_2 n_j} \right) \right)^{n_j} - \prod_{j=1}^{c} \left(\varphi_j^{(1)} \left(\frac{t_1 m_{j1}}{\hat{\sigma}_1 n_j} \right) \varphi_j^{(2)} \left(\frac{t_2 m_{j2}}{\hat{\sigma}_2 n_j} \right) \right)^{n_j},$$

$$f(t_1, t_2) = \prod_{j=1}^{c} \left(\varphi_j \left(\frac{t_1 m_{j1}}{\hat{\sigma}_1 n_j}, \frac{t_2 m_{j2}}{\hat{\sigma}_2 n_j} \right) \right)^{n_j - k_j} E \exp(it_1 (S_1^{(1)} + \delta_1^{(1)}) + it_2 (S_1^{(2)} +$$

$$+ \delta_1^{(2)})) - \prod_{j=1}^{c} \left(\varphi_j^{(1)} \left(\frac{t_1 m_{j1}}{\hat{\sigma}_1 n_j} \right) \varphi_j^{(2)} \left(\frac{t_2 m_{j2}}{\hat{\sigma}_2 n_j} \right) \right)^{n_j - k_j} E \exp(it_1 (S_1^{(1)} + \delta_1^{(1)})) \times$$

$$\times E \exp(it_2 (S_1^{(2)} + \delta_1^{(2)})),$$

$$\varphi_j(t_1, t_2) = E \exp(it_1 g_j^{(1)}(X_{j1}) + it_2 g_j^{(2)}(X_{j1})), \quad \varphi_j^{(1)}(t_1) = \varphi_j(t_1, 0),$$

$$\varphi_j^{(2)}(t_2) = \varphi_j(0, t_2).$$

Анализ этих представлений с помощью метода, изложенного в [1], приводит к оценке

$$\Delta_4 = O(n^{-1/2}). \quad (15)$$

Из (6), (9), (11), (12) и (15) вытекает (4). Теорема доказана.

1. *Королюк В. С., Боровских Ю. В.* Асимптотический анализ распределений статистик.— Киев : Наук. думка, 1984.— 304 с.
2. *Sugiura N.* Multisample and multivariate nonparametric tests Based on *U*-statistics and their asymptotic efficiencies.— Osaka J. Math., 1965, 2, N 2, p. 385—426.
3. *Helmers R., Janssen P.* On the Berry—Esseen theorem for multivariate *U*-statistics.— Report SW 90/82/Math. Centr. Amsterdam, 1982, p. 1—22.
4. *Бхаттачария Р. Н., Ранга Р. Рао.* Аппроксимация нормальным распределением и асимптотические разложения.— М.: Наука, 1982.— 286 с.
5. *Садикова С. М.* Двумерные аналоги неравенства Эссеена с применением к центральной предельной теореме.— Теория вероятностей и ее применения, 1966, 11, вып. 3, с. 369—380.

Ленингр. ин-т текстильной и легкой пром-сти Получено 26.09.84