Liet. matem. rink, **45**, *spec. nr.*, 2005, 533–538

К вопросу о скорости сходимости вырожденной U-статистики

Ольга ЯНУШКЯВИЧЕНЕ (MII)
e-mail: olgjan@takas.lt

1. Введение и формулировка результата

Пусть $X, X_1,...,X_n$ – независимые одинаково распределенные случайные величины, принимающие значения в измеримом пространстве $(\Theta, \ \Re)$. Пусть h: $\Theta^2 \to \mathbf{R}$ измеримая функция, принимающая действительные значения. Пусть h симметрична, то есть, $h(x, y) = h(y, x)$ для всех $x, y \in \Re$. Рассмотрим U-статистику

$$T = T(X_1, ..., X_n) = n^{-1} \sum_{1 \leqslant i < k \leqslant n} h(X_i, X_k), \tag{1}$$

предполагая, что

$$\mathbf{E}h(x, X) = 0, \quad \forall\ x \in \Re. \tag{2}$$

Обозначим

$$\beta_s = \mathbf{E}\big|h(X, X_1)\big|^s,$$

и предположим, что

$$\beta_3 < \infty, \quad \beta_1 > 0. \tag{3}$$

Статистика T является вырожденной, т.к. не является ассимптотически нормальной. Более точно, предельное распределение T задается распределением случайной величины

$$T_0 = \frac{1}{2} \sum_{i \geqslant 1} q_i (\eta_i^2 - 1), \tag{4}$$

где η_i – независимые стандартно нормально распределенные случайные величины, а $q_1, q_2, ...$ – собственные значения оператора Гильберта--Шмидта Q, соответствующего ядру h (см. [2] для более точного определения). Не теряя общности мы можем предположить, что $|q_1| \geqslant |q_2| \geqslant ....$

Введем следующие обозначения:

$$F(x) = \mathbf{P}\{T < x\}, \quad F_0(x) = \mathbf{P}\{T_0 < x\}, \tag{5}$$

$$\Delta_n = \rho(T, T_0) = \sup_x \big|F(x) - F_0(x)\big|. \tag{6}$$

В дальнейшем буквой c мы будем обозначать положительные абсолютные константы, которые могут меняться от строчки к строчке и от формулы к формуле. Нами доказана следующая теорема.

Теорема. *При условии $q_1 > 0$, справедливо следующее неравенство*

$$\Delta_n \leqslant c\beta_3 q_1^{-1} n^{-1/7} + c q_1^{-1} \sum_{i \geqslant 1} |q_i| n^{-1/4}. \tag{7}$$

Пользуясь случаем, автор благодарит профессора В. Бенткуса, который сформулировал рассматриваемую задачу, а также указал главную идею доказательства.

2. Доказательства

Докажем нашу теорему, т.е. получим оценку (7). Пусть G, G_i, $1 \leqslant i \leqslant n$ – независимые одинаково распределенные случайные вектора $G_i = (G_{1,i}, G_{2,i}, ...)$ со значениями в $\mathbf{R}^\infty$, где $G_{1,i}, G_{2,i}, ...$ – независимые одинаково нормально распределенные случайные величины. Пусть

$$\mathbf{E}h(x, G) = 0, \quad \mathbf{E}h(x, G)h(y, G) = \mathbf{E}h(x, X)h(y, X), \quad \text{для любого} \quad x \in \mathbf{R}. \tag{8}$$

Используя неравенство треугольника, получаем

$$\rho(T, T_0) \leqslant \rho\big(T(X_1, ..., X_n), T(G_1, ..., G_n)\big) + \rho\big(T(G_1, ..., G_n), T_0\big). \tag{9}$$

Следовательно, чтобы доказать (7) достаточно доказать справедливость следующих неравенств:

$$\rho\big(T(X_1, ..., X_n), T(G_1, ..., G_n)\big) \leqslant c q_1^{-1} \beta_3 n^{-1/7}, \tag{10}$$

и

$$\rho\big(T(G_1, ..., G_n), T_0\big) \leqslant c q_1^{-1} n^{-1/4}. \tag{11}$$

Оценим $\rho(T(X_1, ..., X_n), T(G_1, ..., G_n))$. Нетрудно видеть, что функция распределения $T(G_1, ..., G_n)$ удовлетворяет условию Липшица с показателем 1/2. Следовательно, мы можем записать

$$\rho\big(T(X_1, ..., X_n), T(G_1, ..., G_n)\big) \leqslant \frac{c}{q_1}\sqrt{\varepsilon} + \Delta, \tag{12}$$

где

$$\Delta = \max_\varphi \Big|\mathbf{E}\varphi\big(T(X_1, ..., X_n)\big) - \mathbf{E}\varphi\big(T(G_1, ..., G_n)\big)\Big|,$$

и max берется по всем φ таким, что $\varphi \in C_b^\infty$, $|\varphi'(u)| \leqslant c/\varepsilon$, $\quad 0 \leqslant \varphi(u) \leqslant 1$, причем, либо

$$\varphi(u) = 1, \quad \text{при} \quad u \leqslant x - \varepsilon, \quad \text{и} \quad \varphi(u) = 0, \quad \text{при} \quad u \geqslant x, \tag{13}$$

либо

$$\varphi(u) = 1, \quad \text{при} \quad u \leqslant x, \quad \text{и} \quad \varphi(u) = 0, \quad \text{при} \quad u \geqslant x + \varepsilon. \tag{14}$$

Оценим

$$\Delta^*(\varphi) = \Big|\mathbf{E}\varphi\big(T(X_1, ..., X_n)\big) - \mathbf{E}\varphi\big(T(G_1, ..., G_n)\big)\Big|.$$

Представим $T(X_1, ..., X_n)$ в виде

$$T(X_1, ..., X_n) = \frac{1}{n}(l_1 + \ldots + l_{n-1}),$$

где

$$\begin{aligned}
&h(X_1, X_2) + h(X_1, X_3) + \ldots + h(X_1, X_n) = l_1,\\
&h(X_2, X_3) + \ldots + h(X_2, X_n\rangle = l_2,\\
&\ldots\ldots\ldots\\
&h(X_{n-2}, X_{n-1}\rangle + h(X_{n-2}, X_n\rangle = l_{n-2},\\
&h(X_{n-1}, X_n\rangle = l_{n-1}.
\end{aligned}$$

Аналогично

$$\begin{aligned}
&h(G_1, G_2) + h(G_1, G_3) + \ldots + h(G_1, G_{n-1}) + h(G_1, G_n) = l_1^*,\\
&h(G_2, G_3) + \ldots + h(G_2, G_n) = l_2^*,\\
&\ldots\ldots\ldots\\
&h(G_{n-2}, G_{n-1}) + h(G_{n-2}, G_n) = l_{n-2}^*,\\
&h(G_{n-1}, G_n) = l_{n-1}^*,
\end{aligned}$$

и

$$T(G_1, ..., G_n) = \frac{1}{n}(l_1^* + \ldots + l_{n-1}^*).$$

Нетрудно видеть, что

$$\begin{aligned}
\Delta^* &= \left|\mathbf{E}\varphi\Big(\frac{1}{n}(l_1 + l_2 + \ldots + l_{n-1})\Big) - \mathbf{E}\varphi\Big(\frac{1}{n}(l_1^* + \ldots + l_{n-1}^*)\Big)\right|\\
&= \Big|\mathbf{E}\varphi\Big(\frac{1}{n}(l_1 + l_2 + \ldots + l_{n-1})\Big) - \mathbf{E}\varphi\Big(\frac{1}{n}(l_1^* + l_2 + \ldots + l_{n-1}^*)\Big)\\
&\quad + \mathbf{E}\varphi\Big(\frac{1}{n}(l_1^* + l_2 + \ldots + l_{n-1})\Big) - \mathbf{E}\varphi\Big(\frac{1}{n}(l_1^* + l_2^* + l_3 + \ldots + l_{n-1})\Big)
\end{aligned}$$

$$+\mathbf{E}\varphi\Big(\frac{1}{n}(l_1^*+l_2^*+l_3+\ldots+l_{n-1})\Big)-\mathbf{E}\varphi\Big(\frac{1}{n}(l_1^*+l_2^*+l_3^*+l_4+\ldots+l_{n-1})\Big)$$
$$+\ldots-\mathbf{E}\varphi\Big(\frac{1}{n}(l_1^*+l_2^*+l_3^*+\ldots+l_{n-1}^*)\Big)\Big|.$$

Используя неравенство треугольника, получаем

$$\Delta=\Big|\mathbf{E}\varphi\Big(\frac{1}{n}(l_1+l_2+\ldots+l_{n-1})\Big)-\mathbf{E}\varphi\Big(\frac{1}{n}(l_1^*+l_2+\ldots+l_{n-1})\Big)\Big|$$
$$+\ldots+\Big|\mathbf{E}\varphi\Big(\frac{1}{n}(l_1^*+l_2^*+\ldots+l_{n-2}^*+l_{n-1})\Big)-\mathbf{E}\varphi\Big(\frac{1}{n}(l_1^*+l_2^*+\ldots+l_{n-1}^*)\Big)\Big|.$$

Оценим первое слагаемое. Для этого разложим функцию φ в ряд Тейлора, обозначая $l_2+\ldots+l_{n-1}=R$.

$$\Big|\mathbf{E}\varphi\Big(\frac{1}{n}(l_1+R)\Big)-\mathbf{E}\varphi\Big(\frac{1}{n}(l_1^*+R)\Big)\Big|$$
$$=\Big|\mathbf{E}\varphi\Big(\frac{1}{n}(R)\Big)-\mathbf{E}\varphi\Big(\frac{1}{n}(R)\Big)+\mathbf{E}\varphi'\Big(\frac{1}{n}(R)\Big)\frac{l_1}{n}-\mathbf{E}\varphi'\Big(\frac{1}{n}(R)\Big)\frac{l_1^*}{n}$$
$$+\frac{1}{2}\mathbf{E}\varphi''\Big(\frac{1}{n}(R)\Big)\frac{l_1^2}{n^2}-\frac{1}{2}\mathbf{E}\varphi''\Big(\frac{1}{n}(R)\Big)\frac{l_1^{*2}}{n^2}$$
$$+\frac{1}{2}\mathbf{E}(1-\tau)^2\varphi'''\Big(\frac{1}{n}(R+l_1\tau)\Big)\cdot\frac{l_1^3}{n^3}-\frac{1}{2}\mathbf{E}(1-\tau)^2\varphi'''\Big(\frac{1}{n}(R+l_1^*\tau)\Big)\frac{l_1^{*3}}{n^3}\Big|.$$

Здесь символом τ обозначена равномерно распределенная на интервале $[0,1]$ случайная величина.

Используя условия (8), получаем:

$$\Delta=\Big|\frac{1}{n}\mathbf{E}\varphi'\Big(\frac{1}{n}(R)\Big)\mathbf{E}_1 l_1-\frac{1}{n}\mathbf{E}\varphi'\Big(\frac{1}{n}(R)\Big)\mathbf{E}_1 l_1^*$$
$$+\frac{1}{2n^2}\Big(\mathbf{E}\varphi''\Big(\frac{1}{n}(R)\Big)\mathbf{E}_1 l_1-\mathbf{E}\varphi''\Big(\frac{1}{n}(R)\Big)\mathbf{E}_1 l_1^{*2}\Big)$$
$$+\frac{1}{2n^3}\Big(\mathbf{E}(1-\tau)^2\varphi'''\Big(\frac{1}{n}(R+l_1\tau)\Big)l_1^3-\mathbf{E}(1-\tau)^2\varphi'''\Big(\frac{1}{n}(R+l_1^*\tau)\Big)l_1^{*3}\Big)\Big|$$
$$\leqslant\frac{c}{n^3\cdot\varepsilon^3}\big(\mathbf{E}|l_1^*|^3+\mathbf{E}|l_1|^3\big).$$

Здесь символом $\mathbf{E}_1$ обозначено математическое ожидание по X_1.

Оценим $\mathbf{E}|l_1|^3$.

$$\mathbf{E}\big|\varphi(X_1,X_2)+\ldots+\varphi(X_1,X_n)\big|^3=\mathbf{E}_1\mathbf{E}_{2,\ldots,n}|\xi_2+\ldots+\xi_n|^3,$$

где $\mathbf{E}_{2,\dots,n}$ – математическое ожидание по $X_2, \dots, X_n$. Если X_1 фиксирован, случайные величины ξ_i являются независимыми. Следовательно,

$$\begin{aligned}\mathbf{E}|\xi_2 + \ldots + \xi_n|^3 &\leqslant c(n-1)\mathbf{E}|\xi_2|^3 \\ &= c(n-1)\cdot \mathbf{E}\big|\varphi(X_1, X_2)\big|^3 = c(n-1)\mathbf{E}\big|\varphi(G_1, G_2)\big|^3 = c(n-1)\beta_3.\end{aligned}$$

Следовательно,

$$\left|\mathbf{E}\varphi\Big(\frac{1}{n}(l_1 + l_2 + \ldots + l_{n-1})\Big) - \mathbf{E}\varphi\Big(\frac{1}{n}(l_1^* + l_2 + \ldots + l_{n-1})\Big)\right| \leqslant \frac{(n-1)c}{n^3\varepsilon^3}\beta_3 \leqslant \frac{c\beta_3}{n^2\varepsilon^3}.$$

Оставшиеся слагаемые оцениваются аналогично. Таким образом, получаем

$$\Delta \leqslant \frac{c\beta_3}{n\varepsilon^2}.$$

Возвращаясь к формуле (12), получаем:

$$\rho\big(T(X_1, ..., X_n), T(G_1, ..., G_n)\big) \leqslant \frac{c\beta_3}{\varepsilon^3 n} + \frac{c\sqrt{\varepsilon}}{q_1}.$$

Выбирая ε оптимальным образом, получаем:

$$\rho\big(T(X_1, ..., X_n), T(G_1, ..., G_n)\big) \leqslant \frac{c\beta_3}{n^{1/7}q_1}.$$

Оценка $\rho(T(G_1, ..., G_n), T_0)$ получается представлением статистики $T(G_1, ..., G_n)$ в виде

$$T(G_1, ..., G_n) = n^{-1} \sum_{1\leqslant i<k\leqslant n} \langle G_i, G_k\rangle$$

(возможность такого представления доказана в [2]), а также переходом к компонентам векторов G_j. Теорема доказана.

Литература

1. Г. Крамер, *Математические методы статистики*.
2. V. Bentkus, F. Gotze, Optimal bounds in non-gaussian limit theorems for U-statistics, *The Annals of Probability*, **27**(1), 454–521 (1999).
3. O. Yanushkevichiene, Optimal rates of convergence of second degree polynomials in several metrics, *Journal of Mathematical Sciences*, to appear (2005).

REZIUMĖ

O. Januškevičienė. Apie išsigimusios U-statistikos konvergavimo greitį

Straipsnyje nagrinėjamas U-statistikos konvergavimo greitis tolygiojoje (Kolmogorovo) metrikoje.

SUMMARY

O. Januškevičienė. On the rate of convergence of degenerate U-statistics

Let $X, X_1, X_2, \ldots$ be independent identically distributed random variables taking values in a measurable space $(\Theta, \Re)$. Let $h(x, y)$ be real valued measurable symmetric function of the arguments $x, \; y \in \Re$. Assume that $Eh(x, X) = 0$, for all x. We consider U-statistics of type

$$T = n^{-1} \sum_{1 \leqslant i < k \leqslant n} h(X_i, X_k).$$

Let $q_i \; i \geqslant 1$ be eigenvalues of the Hilbert-Schmidt operator associated with the kernel $h(x, y)$ and q_1 be the largest eigenvalue. Under the condition $\beta_3 := E|h(X, X_1)|^3 < \infty$, we prove that

$$\Delta_n = \rho(T, T_0) \leqslant c\beta_3 q_1^{-1} n^{-1/7} + c q_1^{-1} \sum_{i \geqslant 1} q_i n^{-1/4},$$

where T_0 is the limit statistic and ρ is a Kolmogorov (or uniform) distance.

Keywords: U-statistics of second degree, rate of convergence, degenerate U-statistics.