

К вопросу о скорости сходимости вырожденной U-статистики

Ольга ЯНУШКЯВИЧЕНЕ (МП)

e-mail: olgjan@takas.lt

1. Введение и формулировка результата

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

$$T = T(X_1, \dots, X_n) = n^{-1} \sum_{1 \leq i < k \leq n} h(X_i, X_k), \quad (1)$$

предполагая, что

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

Обозначим

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

и предположим, что

$$\beta_3 < \infty, \quad \beta_1 > 0. \quad (3)$$

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

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

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

Введем следующие обозначения:

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

$$\Delta_n = \rho(T, T_0) = \sup_x |F(x) - F_0(x)|. \quad (6)$$

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

$$\Delta_n \leq c\beta_3 q_1^{-1} n^{-1/7} + c q_1^{-1} \sum_{i \geq 1} |q_i| n^{-1/4}. \quad (7)$$

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 $G, G_i, 1 \leq i \leq n$ – независимые одинаково распределенные случайные вектора $G_i = (G_{1,i}, G_{2,i}, \dots)$ со значениями в \mathbf{R}^∞ , где $G_{1,i}, G_{2,i}, \dots$ – независимые одинаково нормально распределенные случайные величины. Пусть

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

Используя неравенство треугольника, получаем

$$\rho(T, T_0) \leq \rho(T(X_1, \dots, X_n), T(G_1, \dots, G_n)) + \rho(T(G_1, \dots, G_n), T_0). \quad (9)$$

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

$$\rho(T(X_1, \dots, X_n), T(G_1, \dots, G_n)) \leq c q_1^{-1} \beta_3 n^{-1/7}, \quad (10)$$

и

$$\rho(T(G_1, \dots, G_n), T_0) \leq c q_1^{-1} n^{-1/4}. \quad (11)$$

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

$$\rho(T(X_1, \dots, X_n), T(G_1, \dots, G_n)) \leq \frac{c}{q_1} \sqrt{\varepsilon} + \Delta, \quad (12)$$

где

$$\Delta = \max_{\varphi} \left| \mathbf{E}\varphi(T(X_1, \dots, X_n)) - \mathbf{E}\varphi(T(G_1, \dots, G_n)) \right|,$$

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

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

либо

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

Оценим

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

Представим $T(X_1, \dots, X_n)$ в виде

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

где

$$h(X_1, X_2) + h(X_1, X_3) + \dots + h(X_1, X_n) = l_1,$$

$$h(X_2, X_3) + \dots + h(X_2, X_n) = l_2,$$

.....

$$h(X_{n-2}, X_{n-1}) + h(X_{n-2}, X_n) = l_{n-2},$$

$$h(X_{n-1}, X_n) = l_{n-1}.$$

Аналогично

$$h(G_1, G_2) + h(G_1, G_3) + \dots + h(G_1, G_{n-1}) + h(G_1, G_n) = l_1^*,$$

$$h(G_2, G_3) + \dots + h(G_2, G_n) = l_2^*,$$

.....

$$h(G_{n-2}, G_{n-1}) + h(G_{n-2}, G_n) = l_{n-2}^*,$$

$$h(G_{n-1}, G_n) = l_{n-1}^*,$$

и

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

Нетрудно видеть, что

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

$$\begin{aligned}
& + \mathbf{E}\varphi\left(\frac{1}{n}(l_1^* + l_2^* + l_3 + \dots + l_{n-1})\right) - \mathbf{E}\varphi\left(\frac{1}{n}(l_1^* + l_2^* + l_3^* + l_4 + \dots + l_{n-1})\right) \\
& + \dots - \mathbf{E}\varphi\left(\frac{1}{n}(l_1^* + l_2^* + l_3^* + \dots + l_{n-1}^*)\right)\Big|.
\end{aligned}$$

Используя неравенство треугольника, получаем

$$\begin{aligned}
\Delta = & \left| \mathbf{E}\varphi\left(\frac{1}{n}(l_1 + l_2 + \dots + l_{n-1})\right) - \mathbf{E}\varphi\left(\frac{1}{n}(l_1^* + l_2 + \dots + l_{n-1})\right) \right| \\
& + \dots + \left| \mathbf{E}\varphi\left(\frac{1}{n}(l_1^* + l_2^* + \dots + l_{n-2}^* + l_{n-1})\right) - \mathbf{E}\varphi\left(\frac{1}{n}(l_1^* + l_2^* + \dots + l_{n-1}^*)\right) \right|.
\end{aligned}$$

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

$$\begin{aligned}
& \left| \mathbf{E}\varphi\left(\frac{1}{n}(l_1 + R)\right) - \mathbf{E}\varphi\left(\frac{1}{n}(l_1^* + R)\right) \right| \\
& = \left| \mathbf{E}\varphi\left(\frac{1}{n}(R)\right) - \mathbf{E}\varphi\left(\frac{1}{n}(R)\right) + \mathbf{E}\varphi'\left(\frac{1}{n}(R)\right)\frac{l_1}{n} - \mathbf{E}\varphi'\left(\frac{1}{n}(R)\right)\frac{l_1^*}{n} \right. \\
& \quad + \frac{1}{2}\mathbf{E}\varphi''\left(\frac{1}{n}(R)\right)\frac{l_1^2}{n^2} - \frac{1}{2}\mathbf{E}\varphi''\left(\frac{1}{n}(R)\right)\frac{l_1^{*2}}{n^2} \\
& \quad \left. + \frac{1}{2}\mathbf{E}(1-\tau)^2\varphi'''\left(\frac{1}{n}(R+l_1\tau)\right) \cdot \frac{l_1^3}{n^3} - \frac{1}{2}\mathbf{E}(1-\tau)^2\varphi'''\left(\frac{1}{n}(R+l_1^*\tau)\right)\frac{l_1^{*3}}{n^3} \right|.
\end{aligned}$$

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

Используя условия (8), получаем:

$$\begin{aligned}
\Delta = & \left| \frac{1}{n}\mathbf{E}\varphi'\left(\frac{1}{n}(R)\right)\mathbf{E}_1 l_1 - \frac{1}{n}\mathbf{E}\varphi'\left(\frac{1}{n}(R)\right)\mathbf{E}_1 l_1^* \right. \\
& + \frac{1}{2n^2}\left(\mathbf{E}\varphi''\left(\frac{1}{n}(R)\right)\mathbf{E}_1 l_1 - \mathbf{E}\varphi''\left(\frac{1}{n}(R)\right)\mathbf{E}_1 l_1^{*2}\right) \\
& \left. + \frac{1}{2n^3}\left(\mathbf{E}(1-\tau)^2\varphi'''\left(\frac{1}{n}(R+l_1\tau)\right)l_1^3 - \mathbf{E}(1-\tau)^2\varphi'''\left(\frac{1}{n}(R+l_1^*\tau)\right)l_1^{*3}\right) \right| \\
\leq & \frac{c}{n^3 \cdot \varepsilon^3}(\mathbf{E}|l_1^*|^3 + \mathbf{E}|l_1|^3).
\end{aligned}$$

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

Оценим $\mathbf{E}|l_1|^3$.

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

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,

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

Следовательно,

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

Оставшиеся слагаемые оцениваются аналогично. Таким образом, получаем

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

Возвращаясь к формуле (12), получаем:

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

Выбирая ε оптимальным образом, получаем:

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

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

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

(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. Yanushkeviciene, Optimal rates of convergence of second degree polynomials in several metrics, *Journal of Mathematical Sciences*, to appear (2005).

REZIUOMĖ

O. Januškevičienė. Apie išsigimusios U-statistikos konvergavimo greičį

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, \dots be independent identically distributed random variables taking values in a measurable space (Θ, \mathfrak{R}) . Let $h(x, y)$ be real valued measurable symmetric function of the arguments $x, y \in \mathfrak{R}$. Assume that $Eh(x, X) = 0$, for all x . We consider U-statistics of type

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

Let $q_i, i \geq 1$ 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

$$\Delta_n = \rho(T, T_0) \leq c\beta_3 q_1^{-1} n^{-1/7} + cq_1^{-1} \sum_{i \geq 1} q_i n^{-1/4},$$

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.