УДК 519.21

Ю. В. Боровских

Неравномерная оценка скорости сходимости для L-статистик

1. Линейные функции. Пусть $X_1, X_2, \ldots, X_n$ — независимые случайные величины с одной и той же функцией распределения $F(x)$. Расположим эти величины в порядке возрастания $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$. Тогда $X_{(i)}$ называется i-й порядковой статистикой или i-м членом данного вариационного ряда. Пусть $\{c_{in}\}$ — произвольная последовательность постоянных. Статистики вида $\sum\limits_{i=1}^{n} c_{in} X_{(i)}$ называются линейными функциями порядковых статистик, L-статистиками или L-оценками. Выбор $\{c_{in}\}$ до некоторой степени произволен. Однако в теории оценивания наиболее подходящими оказываются статистики

$$S_n = \frac{1}{n} \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) X_{(i)} \tag{1}$$

и

$$T_n = \sum_{i=1}^{n} X_{(i)} \int\limits_{\frac{i-1}{n}}^{\frac{i}{n}} J(u)\, du \tag{2}$$

с некоторой функцией $J(u)$ на $[0, 1]$. L-статистики можно рассматривать в терминах функционалов, определенных на пространстве функций распределения F. Так, основным функционалом, соответствующим (2), является

$$T(F) = \int\limits_0^1 F^{-1}(t)\, J(t)\, dt, \tag{3}$$

где $F^{-1}(t) = \inf\{x : F(x) \geqslant t\}$, $F(x) \in F$. Для эмпирической функции распределения $F_n(x)$ легко следует

$$T(F_n) = T_n. \tag{4}$$

Поскольку L-статистики играют важную роль в развитии теории оценивания, то естествен вопрос об исследовании их распределений. Однако точные распределения даже при самых простых $F(x)$ и $J(u)$ сложны для анализа. Поэтому основное внимание было сфокусировано на вопросе об асимптотической нормальности L-статистик. Определим

$$\widetilde{F}_n(x) = P\left(\frac{S_n - \mu_n}{\sigma_n} < x\right),$$

где $\mu_n = ES_n$, $\sigma_n^2 = E(S_n - \mu_n)^2$.

Задача исследования асимптотических свойств $\widetilde{F}_n(x)$, конечно, тривиальна, если $F(x)$ — экспоненциальное распределение. Изучение $\widetilde{F}_n(x)$ может быть осуществлено непосредственно и в том случае, когда $F(x)$ — равномерное распределение на $[0, 1]$. В общем случае эта задача, однако, далеко не тривиальна. Оценкой разности $\Delta_n = \sup\limits_x |\widetilde{F}_n(x) - \Phi(x)|$ занимались Розенкранц и О'Рейли [1], де Вет [2]. Впервые соотношение вида $\Delta_n = O\left(\frac{1}{\sqrt{n}}\right)$ было получено Бьерве [3] для трункированных (усеченных) линейных комбинаций порядковых статистик. В этом случае весовая функция $J(u)$ разрывна. Затем Хельмерс [4—5] доказал оценку порядка $O\left(\frac{1}{\sqrt{n}}\right)$ для некоторого класса гладких функций $J(u)$ на $(0,1)$. При этом предполагалось, что исходные случайные величины X_i должны иметь абсолютный третий момент. Вопрос об асимптотическом разложении распределения $\widetilde{F}_n(x)$ в ряд по степеням $\frac{1}{\sqrt{n}}$ исследовали Хельмерс [6] и ван Цвет [7].

Методом Хельмерса будем рассматривать случай гладких весовых функций $J(u)$ на $(0, 1)$.

2. Теорема. *Пусть функция $J(u)$ и распределение $F(x)$ удовлетворяют следующим условиям:*

1) *$J(u)$ имеет непрерывную производную $J'(u)$ на $[0, 1]$, которая удовлетворяет условию Липшица;*

2) $EX_1^2 < \infty$;

3) $\int\limits_0^1 \sqrt{x(1-x)}\, dF^{-1}(x) < \infty$;

4) $\int\limits_0^1 \left(\int\limits_0^1 J'(s)(\delta(s-t)-s)\, dF^{-1}(s)\right)^4 dt < \infty$;

5) $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} J(F(x))\, J(F(y))\, \{F(\min(x, y)) - F(x)F(y)\}\, dxdy > 0$.

Тогда существует постоянная $C(J, F)$, зависящая лишь от J и F, такая, что при всех $-\infty < x < \infty$ для больших n выполняется неравенство

$$\left| P\left(\frac{S_n - ES_n}{\sigma(S_n)} < x\right) - \Phi(x) \right| \leqslant \frac{C(J, F)}{\sqrt{n}\,(1 + x^2)}. \tag{5}$$

Доказательство. Пусть для каждого $n \geqslant 1$ $U_1, U_2, \ldots, U_n$ — последовательность независимых равномерно распределенных на интервале $[0, 1]$ случайных величин, $\Gamma_n(x)$ — эмпирическая функция распределения, отвечающая этой последовательности. Хорошо известно, что совместное распределение $X_1, X_2, \ldots, X_n$ совпадает с совместным распределением $F^{-1}(u_1), F^{-1}(u_2), \ldots, F^{-1}(u_n)$. Положим

$$f(u) = \int\limits_u^1 J(s)\, ds - (1-u)\int\limits_0^1 J(s)\, ds, \tag{6}$$

где $0 < u < 1$.

При (4) и (6) получаем

$$T_n = \int\limits_0^1 f(\Gamma_n(s))\, dF^{-1}(s) + \frac{a}{n} \cdot \sum_{i=1}^{n} F^{-1}(u_i) \tag{7}$$

с вероятностью единица, где $a = \int\limits_0^1 J(s)\, ds$.

Пусть

$$Q_n = \int_0^1 \left((f(s) + (\Gamma_n(s) - s) f'(s) + \left(\frac{\Gamma_n(s) - s)^2}{2!} \cdot f''(s) \right) dF^{-1}(s) + \right.$$

$$+ \frac{a}{n} \sum_{i=1}^{n} F^{-1}(u_i). \tag{8}$$

Условия теоремы гарантируют то, что Q_n определена. Для преобразования подынтегральной функции в (7) применим формулу Тейлора с остаточным членом в форме Лагранжа. Тогда, очевидно

$$T_n = Q_n + R_{n,1}, \tag{9}$$

где

$$R_{n,1} = \frac{1}{2} \int_0^1 (\Gamma_n(s) - s)^2 [f''(s + \theta(\Gamma_n(s) - s)) - f''(s)] \, dF^{-1}(s). \tag{10}$$

Согласно (6) $f''(s) = -J'(s)$. Производная $J'(s)$ по условию теоремы удовлетворяет условию Липшица. Поэтому с вероятностью единица

$$|R_{n,1}| \leqslant C_1 \int_0^1 |\Gamma_n(s) - s|^3 dF^{-1}(s), \tag{11}$$

причем $C_1 > 0$ — постоянная (здесь и всюду в дальнейшем C_i, $i \geqslant 1$, не зависят от n и x). Стало быть, $\sigma^2(T_n - Q_n) \leqslant E(T_n - Q_n)^2 \leqslant C_1^2 \cdot E \times \times \left(\int_0^1 |\Gamma_n(s) - s|^3 \, dF^{-1}(s)\right)^2$. По неравенству Гельдера

$$E\left(\int_0^1 |\Gamma_n(s) - s|^3 dF^{-1}(s) \right)^2 \leqslant \left(\int_0^1 \left(E|\Gamma_n(s) - s|^6\right)^{1/2} dF^{-1}(s) \right)^2.$$

Здесь $\Gamma_n(s) = \frac{1}{n} \sum_{j=1}^{n} \delta(s - u_j)$ при любом s — сумма n независимых одинаково распределенных случайных величин, причем $E\Gamma_n(s) = s$, $\sigma^2(\Gamma_n(s) - s) = s(1-s)$. Тогда по [7] $E|\Gamma_n(s) - s|^6 \leqslant C_2 n^{-3} s(1-s)$, где $C_2 > 0$ — абсолютная постоянная. Таким образом,

$$\sigma^2(T_n - Q_n) \leqslant C_3 n^{-3} \left(\int_0^1 \sqrt{s(1-s)} \, dF^{-1}(s) \right)^2. \tag{12}$$

Вследствие (9) и (12) видим, что L-статистика T_n аппроксимируется величиной Q_n. Распишем Q_n. При помощи (6) преобразуем Q_n:

$$Q_n - EQ_n = -\frac{1}{n} \sum_{i=1}^{n} \int_0^1 J(s) (\delta(s - u_i) - s) \, dF^{-1}(s) -$$

$$-\frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \int_0^1 J'(s) \, \delta(s - u_i) - s) (\delta(s - u_j) - s) \, dF^{-1}(s) +$$

$$+ \frac{1}{2n} \int_0^1 J'(s) \, s(1-s) \, dF^{-1}(s). \tag{13}$$

Далее, при $i,\ j = 1, 2, \ldots, n$ обозначим $g(u_i) = -\frac{1}{2}\int\limits_0^1 J(s)(\delta(s-u_i)-s)\times$ $\times dF^{-1}(s)$, $h(u_i, u_j) = -\frac{1}{2}\int\limits_0^1 J'(s)(\delta(s-u_i)-s)(\delta(s-u_j)-s)\,dF^{-1}(s)$,

$\Phi_n(u_i, u_j) = \frac{n-1}{n} h(u_i, u_j) + g(u_i) + g(u_j)$.

Пусть U_n — U-статистика:

$$U_n = \frac{1}{\binom{n}{2}} \sum_{1 \leqslant i < j \leqslant n} \Phi_n(u_i, u_j). \tag{14}$$

По (13) и (14) можно написать

$$Q_n - U_n - E(Q_n - U_n) = -\frac{1}{2n^2}\sum_{i=1}^{n}\int\limits_0^1 J'(s)((\delta(s-u_i)-s)^2 - s(1-s))\,dF^{-1}(s),$$

и, следовательно,

$$\sigma^2(Q_n - U_n) = \frac{1}{4n^3}\sigma^2\left(\int\limits_0^1 J'(s)(\delta(s-u_1)-s)^2\,dF^{-1}(s)\right).$$

Оценивая выражение справа, получаем

$$\sigma^2(Q_n - U_n) \leqslant \frac{C_4}{4}\cdot\frac{1}{n^3}\left(\int\limits_0^1 \sqrt{s(1-s)}\,dF^{-1}(s)\right)^2. \tag{15}$$

Неравенство (15) показывает, что L-статистика T_n в среднем хорошо аппроксимируется U-статистикой вида (14).

Положим

$$Z_{n1} = \frac{S_n - ES_n}{\sigma(S_n)},\quad Z_{n2} = \frac{T_n - ET_n}{\sigma(T_n)},\quad Z_{n3} = \frac{Q_n - EQ_n}{\sigma(Q_n)},\quad Z_{n4} = \frac{U_n}{\sigma(U_n)}.$$

Нетрудно убедиться в том, что при любом $A > 0$ $P(Z_{n+1} \leqslant x) \geqslant P(Z_{n4} \leqslant x - A) - P(|Z_{n1} - Z_{n4}| \geqslant A)$, $P(Z_{n1} \leqslant x) \leqslant P(Z_{n4} \leqslant x + A) + P(|Z_{n1} - Z_{n4}| \geqslant A)$. Отсюда выводим неравенство

$$|P(Z_{n1} \leqslant x) - \Phi(x)| \leqslant \sum_{i=1}^{4}\Delta_{n,i}(x) + P(|z_{n1} - z_{n4}| \geqslant A), \tag{16}$$

где $\Delta_{n,1}(x) = |P(Z_{n,4} \leqslant x + A) - \Phi(x + A)|$, $\Delta_{n,2}(x) = |P(Z_{n,4} \leqslant x - A) - \Phi(x - A)|$, $\Delta_{n,3}(x) = |\Phi(x + A) - \Phi(x)|$, $\Delta_{n,4}(x) = |\Phi(x - A) - \Phi(x)|$. Пусть в этих выражениях $A = \frac{1 + |x|}{\sqrt{n}}$. При таком выборе A согласно неравенствам для U-статистик в условиях теоремы (см. [11]) при всех $1 \leqslant i \leqslant 4$ и $-\infty < x < \infty$

$$\Delta_{n,i}(x) \leqslant \frac{C_i(J, F)}{\sqrt{n}(1 + x^2)}, \tag{17}$$

где $C_i(J, F)$ не зависит от n и x.

Таким образом, в (16) требуется изучить лишь последнее слагаемое справа. При помощи неравенства Чебышева получаем

$$P\left(|Z_{n1}-Z_{n4}|\geqslant\frac{1+|x|}{\sqrt{n}}\right)\leqslant\sum_{i=1}^{3}P\left(|Z_{n,i}-Z_{n,i+1}|\geqslant\frac{1+|x|}{4\sqrt{n}}\right)\leqslant$$

$$\leqslant\frac{16n}{(1+|x|)^2}\sum_{i=1}^{3}E\,|Z_{n,i}-Z_{n,i+1}|^2. \quad (18)$$

Оценим здесь последовательно каждое из математических ожиданий. Поскольку

$$Z_{n1}-Z_{n2}=\frac{S_n-T_n-E\,(S_n-T_n)}{\sigma(S_n)}+(T_n-ET_n)\,\frac{\sigma(T_n)-\sigma(S_n)}{\sigma(S_n)\cdot\sigma(T_n)}\,, \quad (19)$$

то

$$E\,|Z_{n1}-Z_{n2}|^2\leqslant\frac{2\sigma^2(S_n-T_n)}{\sigma^2(S_n)}+\frac{2\,(\sigma(S_n)-\sigma(T_n))^2}{\sigma^2(S_n)}\,. \quad (20)$$

Дисперсию $\sigma^2(S_n)$ рассмотрим по формуле [10]

$$n\sigma^2(S_n)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}H_n(x,y)\,dxdy, \quad (21)$$

где

$$H_n(x,y)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}J\left(\frac{i}{n+1}\right)J\left(\frac{j}{n+1}\right)[G_{ij}(x,y)-G_i(x)\,G_j(y)],$$

$$G_i(x)=P\,(X_{(i)}\leqslant x),\;\; G_{ij}(x,y)=P\,(X_{(i)}\leqslant x,\,X_{(j)}\leqslant y).$$

По теореме из статьи [9] для любых x, y, i, j, n и $F(x)$ справедливо неравенство

$$G_i(x)\,G_j(y)\leqslant G_{ij}(x,y). \quad (22)$$

Если $M=\sup\limits_{0\leqslant u\leqslant 1}|J(u)|$, то по (22) и (23) с учетом

$$\sigma^2(X_1)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{F(\min(x,y))-F(x)\,F(y)\}\,dxdy$$

имеем

$$n\sigma^2(S_n)\leqslant M^2\sigma^2(X_1). \quad (23)$$

Более точно по теореме Стиглера

$$\lim_{n\to\infty}(n\sigma^2(S_n))=\sigma^2(J,F), \quad (24)$$

где

$$\sigma^2(J,F)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}J(F(x))\,J(F(y))\,\{F(\min(x,y))-F(x)\,F(y)\}\,dxdy. \quad (25)$$

Аналогичным образом, исходя из (1) и (4), выводим формулу

$$n\sigma^2(S_n-T_n)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}H_{n,1}(x,y)\,dxdy, \quad (26)$$

где

$$H_{n,1}(x,y)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}J_1(i)\,J_1(j)\,[G_{ij}(x,y)-G_i(x)\,G_j(y)],$$

$$J_1(i) = J\left(\frac{i}{n+1}\right) - n \int\limits_{\frac{i-1}{n}}^{\frac{i}{n}} J(s)\,ds, \quad i = 1, 2, \ldots, n.$$

Отсюда получаем

$$\sigma^2(S_n - T_n) \leqslant A^2 \frac{\sigma^2(X_1)}{n^3}. \tag{27}$$

Рассматривая правую часть (20), видим прежде всего, что в силу (23) и (27)

$$\frac{\sigma^2(S_n - T_n)}{\sigma^2(S_n)} = O\left(\frac{1}{n^2}\right). \tag{28}$$

Далее справедлива также цепочка соотношений

$$\frac{(\sigma(S_n) - \sigma(T_n))^2}{\sigma^2(S_n)} = \frac{(\sigma^2(S_n) - \sigma^2(T_n))^2}{\sigma^2(S_n)(\sigma(S_n) + \sigma(T_n))^2} \leqslant$$

$$\leqslant \frac{(2\sigma(S_n)\,\sigma(S_n - T_n) + \sigma^2(S_n - T_n))^2}{\sigma^2(S_n)(\sigma(S_n) + \sigma(T_n))^2} = O\left(\frac{1}{n^2}\right).$$

Здесь последнее равенство написано по (23) и (27). В результате

$$E\,|Z_{n1} - Z_{n2}|^2 = O\left(\frac{1}{n^2}\right). \tag{29}$$

Переходя ко второму математическому ожиданию в (18), пишем

$$E\,|Z_{n2} - Z_{n3}|^2 \leqslant \frac{2\sigma^2(T_n - Q_n)}{\sigma^2(Q_n)} + \frac{2(\sigma(T_n) - \sigma(Q_n))^2}{\sigma^2(Q_n)}.$$

Отсюда по (12)

$$E\,|Z_{n2} - Z_{n3}|^2 = O\left(\frac{1}{n^2}\right). \tag{30}$$

Наконец, для последнего математического ожидания в (18) имеем

$$E\,|Z_{n3} - Z_{n4}|^2 \leqslant \frac{2\sigma^2(Q_n - U_n)}{\sigma^2(U_n)} + \frac{2(\sigma(U_n) - \sigma(Q_n))^2}{\sigma^2(Q_n)},$$

что вместе с (15) приводит к оценке

$$E\,|Z_{n3} - Z_{n4}|^2 = O\left(\frac{1}{n^2}\right). \tag{31}$$

Неравенства (16)—(18) и (29)—(31) влекут (5).

Нетрудно показать, что условия 3) и 4) теоремы выполняются, если, например, $EX_1^4 < \infty$. Условия 2) и 3), вообще говоря, неэквивалентны. Это обстоятельство обсуждал Стиглер [10]. Проверка справедливости неравенства (17) и условий теоремы не представляется сложной.

ЛИТЕРАТУРА

1. Rosenkrantz W., O'Reilly N. E. Application of the Skorohod representation theorem to rates of convergence for linear combinations of order statistics.— Ann. Math. Statist., 1972, 43, N 4, p. 1204—1212.
2. T. de Wet. Rates of convergence of linear combinations of order statistics.—South Afr. Statist. G., 1974, 8, p. 35—43.
3. Bjerve S. Error bounds for linear combinations of order statistics.— Ann. Statist., 1977, 5, N 2, p. 357—369.

4. H e l m e r s R. The order of the normal approximation for linear combinations of order statistics with smooth weight functions.— Ann. Probab., 1977, 5, N 6.
5. H e l m e r s R. A Berry-Esseen theorem for linear combinations of order statistics, Mathematical Centre Report SW 54/77, 1977, Amsterdam.
6. H e l m e r s R. Edgeworth expansions for linear combinations of order statistics with smooth weight functions, Mathematical Centre Report SW 44/76, 1976, Amsterdam.
7. Van Z w e t W. R. Asymptotic expansions for the distribution functions of linear combinations of order statistics, Statistical Decision Theory and Related Topics. New York: Academic Press, 1977, 11, p. 421—438.
8. D h a r m a d h i k a r i S. W., J o g d e o K. Bounds on moments of certain random variables, Ann. Math. Statist., 1969, 40, N 4, p. 1506—1508.
9. E s a r y G. D., P r o s h a n F., W a l k u p D. W. Association of random variables with applications, Ann. Math. Statist., 1967, 38, p. 1466—1474.
10. S t i g l e r S. M. Linear functions of order statistics with smooth weight functions, Ann. Statist., 1974, 2, N 4, p. 676—693.
11. Б о р о в с к и х Ю. В. Аппроксимация распределений U-статистик.— Докл. АН УССР, 1979, № 9, с. 695—698.

Институт математики
АН УССР

Поступила в редакцию
25.IX 1979 г.