Math-Net.Ru

Общероссийский математический портал

Л. В. Гадасина, Неравенства Берри–Эссеена для U -статистик, *Теория вероятн. и ее примен.*, 2003, том 48, выпуск 1, 151–155

DOI: <https://doi.org/10.4213/tvp305>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 3.85.7.115

26 апреля 2023 г., 13:33:32

КРАТКИЕ СООБЩЕНИЯ

© 2003 г.

ГАДАСИНА Л. В.*

НЕРАВЕНСТВА БЕРРИ-ЭССЕЕНА ДЛЯ U -СТАТИСТИК

В статье получена граница Берри-Эссеена для невырожденных U -статистик степени 2, построенных по независимым случайным величинам, не обязательно имеющим одинаковое распределение, при условии существования абсолютных моментов третьего порядка для первого семейства канонических функций и абсолютных моментов порядка $\frac{5}{3}$ для второго семейства канонических функций разложения Хёфдинга.

Ключевые слова и фразы: U -статистика, неравенство Берри-Эссеена, скорость сходимости.

1. Введение. Пусть $X_1, \dots, X_n$ — независимые случайные величины со значениями в измеримом пространстве $(X, \mathcal{B})$. Предположим, что $n \geq 2$, и рассмотрим U -статистику

$$U_n = \sum_{1 \leq i < j \leq n} \Phi_{ij}(X_i, X_j),$$

где $\Phi_{ij}: X^2 \rightarrow \mathbf{R}$ — симметрические функции такие, что $\mathbf{E}|\Phi_{ij}(X_i, X_j)| < \infty$ для всех $1 \leq i < j \leq n$. Определим канонические функции

$$\begin{aligned} g_j(X_j) &= \mathbf{E}[U_n | X_j] - \mathbf{E}U_n = \sum_{i=1}^{j-1} \left\{ \mathbf{E}[\Phi_{ij}(X_i, X_j) | X_j] - \mathbf{E}\Phi_{ij} \right\} \\ &+ \sum_{i=j+1}^n \left\{ \mathbf{E}[\Phi_{ji}(X_j, X_i) | X_j] - \mathbf{E}\Phi_{ji} \right\}, \end{aligned}$$

где $j = 1, \dots, n$, и

$$\begin{aligned} g_{ij}(X_i, X_j) &= \mathbf{E}[U_n | X_i, X_j] - \mathbf{E}[U_n | X_i] - \mathbf{E}[U_n | X_j] + \mathbf{E}U_n \\ &= \Phi_{ij}(X_i, X_j) - \mathbf{E}[\Phi_{ij}(X_i, X_j) | X_i] - \mathbf{E}[\Phi_{ij}(X_i, X_j) | X_j] + \mathbf{E}\Phi_{ij}, \end{aligned}$$

где $1 \leq i < j \leq n$. Будем обозначать для удобства $g_j(X_j)$ и $g_{ij}(X_i, X_j)$ через g_j и g_{ij} соответственно. Тогда в соответствии с представлением Хёфдинга $U_n = \sum_{j=1}^n g_j + \sum_{1 \leq i < j \leq n} g_{ij} + \mathbf{E}U_n$. Положим $\sigma_n^2 = \sum_{j=1}^n \mathbf{E}g_j^2 > 0$, т.е. будем рассматривать только невырожденные U -статистики.

Обозначим $\Delta_n = \sup_x |\mathbf{P}\{\sigma_n^{-1}(U_n - \mathbf{E}U_n) < x\} - \Phi(x)|$, где $\Phi(x)$ — функция распределения стандартной нормальной случайной величины.

В 1980 г. была получена оценка скорости сходимости к нормальному распределению порядка $O(n^{-1/2})$ в предположении $\max_{i < j} \mathbf{E}|\Phi(X_i, X_j)|^3 < \infty$ (см. [9]). Далее, в 1989 г. в [8] это условие было ослаблено и было показано, что достаточно выполнения условий $\max_j \mathbf{E}|g_j|^3 < \infty$, $\max_{i < j} \mathbf{E}|g_{ij}|^3 < \infty$. Примеры, приведенные в работе [7],

* Петербургский государственный университет путей сообщения, кафедра прикладной математики, Московский пр., 9, 190031 С.-Петербург, Россия.

показывают, что показатель $\frac{5}{3}$ является наилучшим возможным. В 1998 г. в [6] максимальные моменты канонических функций были заменены суммами моментов и была получена скорость сходимости Берри-Эссеена в предположении $\sum_{j=1}^n \mathbf{E}|g_j|^3 < \infty$ и $\sum_{i < j} \mathbf{E}g_{ij}^2 < \infty$.

2. Результаты. Рассмотрим усеченные случайные величины

$$g_j^n = g_j I(|g_j| \leq \sigma_n), \quad j = 1, \dots, n, \quad g_{ij}^n = g_{ij} I(|g_{ij}| \leq \sigma_n), \quad j = 1 \leq i < j \leq n,$$

где $I(A)$ — индикаторная функция множества $A \in \mathcal{B}$.

Теорема 1. Если $\sigma_n^2 > 0$, то при всех $\frac{4}{3} \leq p < 2$ и $n \geq 2$

$$\begin{aligned} \Delta_n \leq & 2.1 \sigma_n^{-2} \sum_{j=1}^n \mathbf{E}g_j^2 I(|g_j| > \sigma_n) + 152.5 \sigma_n^{-3} \sum_{j=1}^n \mathbf{E}|g_j|^3 I(|g_j| \leq \sigma_n) \\ & + 1417.5 \sigma_n^{-1} \sum_{i < j} \mathbf{E}|g_{ij}| I(|g_{ij}| > \sigma_n) + 78.5 \sigma_n^{-3} \sum_{i < j} \mathbf{E}|g_i^n g_j^n g_{ij}^n| \\ & + \left(5174.9 + \frac{197.6}{(2-p)^{2/p+1}} + \frac{733.4}{(2-p)^2} \right) \sigma_n^{-p} \sum_{i < j} \mathbf{E}|g_{ij}|^p I(|g_{ij}| \leq \sigma_n). \end{aligned} \quad (1)$$

Следствие 1. Пусть $p > \frac{4}{3}$ и, кроме того, выполнены следующие условия:

$$\begin{aligned} \Lambda_{1,n}(\epsilon) &= \sigma_n^{-2} \sum_{j=1}^n \mathbf{E}g_j^2 I(|g_j| > \epsilon \sigma_n) \rightarrow 0, \\ \Lambda_{2,n}(\epsilon) &= \sigma_n^{-4/3} \sum_{i < j} \mathbf{E}|g_{ij}|^{4/3} I(|g_{ij}| > \epsilon \sigma_n) \rightarrow 0 \end{aligned} \quad (2)$$

для любого фиксированного $\epsilon > 0$ и $n \rightarrow \infty$ и существует константа $a > 0$ такая, что $\Lambda_{2,n} < a$. Тогда $\Delta_n \rightarrow 0$.

При $p = \frac{4}{3}$ утверждение следствия остается верным, если в условии (2) на g_{ij} положить $\epsilon = 0$.

В следующих теоремах получены оценки в центральной предельной теореме.

Теорема 2. Пусть $\sigma_n^2 > 0$ и, кроме того, при некоторых $0 < \delta_1 \leq 1$ и $0 < \delta_2 \leq \frac{1}{3}$ $\mathbf{E}|g_j|^{2+\delta_1} < \infty$, $j = 1, \dots, n$, и $\mathbf{E}|g_{ij}|^{4/3+\delta_2} < \infty$, $1 \leq i < j \leq n$. Тогда при $n \geq 2$

$$\begin{aligned} \Delta_n \leq & 154.5 \sigma_n^{-2-\delta_1} \sum_{j=1}^n \mathbf{E}|g_j|^{2+\delta_1} + 15408 \sigma_n^{-4/3-\delta_2} \sum_{i < j} \mathbf{E}|g_{ij}|^{4/3+\delta_2} \\ & + 78.5 \left(\sigma_n^{-2-\delta_1} \sum_{j=1}^n \mathbf{E}|g_j|^{2+\delta_1} \cdot \sigma_n^{-4/3-\delta_2/2} \sum_{i < j} \mathbf{E}|g_{ij}|^{4/3+\delta_2/2} \right)^{2/3} \end{aligned}$$

Теорема 3. Пусть $0 < \sigma_n^2 < \infty$ и $\mathbf{E}|g_{ij}|^{4/3} < \infty$, $1 \leq i < j \leq n$. Тогда

$$\begin{aligned} \Delta_n \leq & 2.1 \sigma_n^{-2} \sum_{j=1}^n \mathbf{E}g_j^2 I(|g_j| > \sigma_n) + 183.9 \sigma_n^{-3} \sum_{j=1}^n \mathbf{E}|g_j|^3 I(|g_j| \leq \sigma_n) \\ & + 1417.5 \sigma_n^{-4/3} \sum_{i < j} \mathbf{E}|g_{ij}|^{4/3} I(|g_{ij}| > \sigma_n) + 14037.6 \sigma_n^{-5/3} \sum_{i < j} \mathbf{E}|g_{ij}|^{5/3} I(|g_{ij}| \leq \sigma_n). \end{aligned}$$

Следствие 2. Пусть $\sigma_n^2 > 0$. Тогда

$$\Delta_n \leq \min \left\{ 1.186 \sigma_n^{-3} \sum_{j=1}^n \mathbf{E}|g_j|^3 + 15456 \sigma_n^{-5/3} \sum_{i < j} \mathbf{E}|g_{ij}|^{5/3} \right\}. \quad (3)$$

З а м е ч а н и е. Пусть $\tilde{\Phi}: \mathbf{X}^2 \rightarrow \mathbf{R}$ — симметрическая функция, не зависящая от i, j . Если $X_1, \dots, X_n$ — независимые одинаково распределенные случайные величины, то

$$\tilde{\Phi}_{ij}(X_i, X_j) = \frac{2}{n(n-1)} \tilde{\Phi}(X_i, X_j), \quad 1 \leq i < j \leq n,$$

и канонические функции принимают следующий вид:

$$\begin{aligned} g_j(X_j) &= \frac{2}{n} \tilde{g}(X_j), \quad j = 1, \dots, n, \\ g_{ij}(X_i, X_j) &= \frac{2}{n(n-1)} \tilde{g}(X_i, X_j), \quad 1 \leq i < j \leq n. \end{aligned}$$

Если $\mathbf{E}\tilde{g}^2(X_1) = 1$, то

$$\begin{aligned} \Delta_n &\leq 154.5 \mathbf{E}|\tilde{g}(X_1)|^{2+\delta_1} \cdot n^{-\delta_1/2} + 12229 \mathbf{E}|\tilde{g}(X_1, X_2)|^{4/3+\delta_2} \cdot n^{-3\delta_2/2} \\ &\quad + 67.3 \left(\mathbf{E}|\tilde{g}(X_1)|^{2+\delta_1} \mathbf{E}|\tilde{g}(X_1, X_2)|^{4/3+\delta_2/2} \right)^{2/3} n^{-\delta_1/3-\delta_2/2} \end{aligned}$$

при некоторых $0 < \delta_1 \leq 1$ и $0 < \delta_2 \leq \frac{1}{3}$ или

$$\Delta_n \leq \left(185.9 \mathbf{E}|\tilde{g}(X_1)|^3 + 10929 \mathbf{E}|\tilde{g}(X_1, X_2)|^{5/3} \right) \frac{1}{\sqrt{n}}.$$

3. Схема доказательства. Введем следующие обозначения:

$$\begin{aligned} v_{ij}(X_j) &= \mathbf{E}[g_{ij}^n | X_j] = -\mathbf{E}[g_{ij} I(|g_{ij}| > \sigma_n) | X_j], \\ g_{ijn} &= g_{ij}^n - v_{ij}(X_i) - v_{ij}(X_j) + \mathbf{E}g_{ij}^n \end{aligned}$$

для всех $1 \leq i < j \leq n$ и $g_{jn} = g_j^n - \mathbf{E}g_j^n$,

$$\begin{aligned} v_j &= v_j(X_j) = \sum_{i=1}^{j-1} v_{ij}(X_j) + \sum_{i=j+1}^n v_{ij}(X_j), \\ v_j^n &= v_j I(|v_j| \leq \sigma_n), \quad v_{jn} = v_j^n - \mathbf{E}v_j^n \end{aligned}$$

для всех $j = 1, \dots, n$. Тогда $\mathbf{E}g_{jn} = \mathbf{E}v_{jn} = \mathbf{E}g_{ijn} = \mathbf{E}[g_{ijn} | X_i] = \mathbf{E}[g_{ijn} | X_j] = 0$.

Далее используем схему из [1], откуда получаем

$$\Delta_n \leq \sup_x \left| \mathbf{P}\{s_n + w_n < x\} - \Phi(x) \right| + \frac{1}{\sqrt{2\pi}} |a_n| + \frac{1}{\sqrt{2\pi}e} \left| \frac{\bar{\sigma}_n^2}{\sigma^2} - 1 \right| \frac{\sigma_n}{\bar{\sigma}_n} + d_n, \quad (4)$$

где

$$\begin{aligned} \bar{\sigma}_n^2 &= \sum_{j=1}^n \mathbf{E}(g_{jn} + v_{jn})^2, \quad z_j = \bar{\sigma}_n^{-1}(g_{jn} + v_{jn}), \quad j = 1, \dots, n, \\ G_{ij} &= \bar{\sigma}_n^{-1} g_{ijn}, \quad 1 \leq i < j \leq n, \quad w_n = \sum_{i < j} G_{ij}, \quad s_n = \sum_{j=1}^n z_j, \\ a_n &= \sigma_n^{-1} \sum_{j=1}^n \mathbf{E}g_j^n + \sigma_n^{-1} \sum_{j=1}^n \mathbf{E}v_j^n - \sigma_n^{-1} \sum_{i < j} \mathbf{E}g_{ij}^n, \\ d_n &= \sum_{j=1}^n \mathbf{P}\{|g_j| > \sigma_n\} + \sum_{i < j} \mathbf{P}\{|g_{ij}| > \sigma_n\} + \sum_{j=1}^n \mathbf{P}\{|v_j| > \sigma_n\}. \end{aligned}$$

Оценивая a_n и d_n , получаем

$$|a_n| + d_n \leq 2\sigma_n^{-2} \sum_{j=1}^n \mathbf{E}g_j^2 I(|g_j| > \sigma_n) + 6\sigma_n^{-1} \sum_{i < j} \mathbf{E}|g_{ij}| I(|g_{ij}| > \sigma_n).$$

Для оценки первого слагаемого в (4) воспользуемся неравенством Эссеена, обозначив $h = \bar{\sigma}_n^3 (\sum_{j=1}^n \mathbf{E}|g_{jn} + v_{jn}|^3)^{-1} = (\sum_{j=1}^n \mathbf{E}|z_j|^3)^{-1}$ и положив $h > 15$. Тогда

$$\begin{aligned} \sup_x \left| \mathbf{P}\{s_n + w_n < x\} - \Phi(x) \right| &\leq \frac{2}{\pi} \int_0^h |\mathbf{E}e^{its_n} - e^{-t^2/2}| t^{-1} dt \\ &+ \frac{2}{\pi} \int_0^7 |\mathbf{E}e^{it(s_n+w_n)} - \mathbf{E}e^{its_n}| t^{-1} dt \\ &+ \frac{2}{\pi} \int_7^h |\mathbf{E}e^{it(s_n+w_n)} - \mathbf{E}e^{its_n}| t^{-1} dt + \frac{24}{\pi\sqrt{2\pi}} h^{-1} = \varepsilon_1 + \varepsilon_2^4 + \varepsilon_3 + \frac{24}{\pi\sqrt{2\pi}} h^{-1}. \end{aligned}$$

Оценку ε_1 см. в [4], оценка ε_2 элементарна. Рассмотрим ε_3 . Воспользуемся методом рандомизации (см., например, [6]). Зафиксируем $t \in [7, h]$ и рассмотрим последовательность независимых между собой и не зависящих от $X_1, \dots, X_n$ случайных величин $\alpha_1, \dots, \alpha_n$, имеющих распределение Бернулли и таких, что

$$\mathbf{P}\{\alpha_j = 1\} = 1 - \mathbf{P}\{\alpha_i = 0\} = m(t) = 18t^{-2} \ln t, \quad j = 1, \dots, n. \quad (5)$$

Заметим, что $m(t) \in [0, 1]$ для всех t . Тогда случайные величины X_j , $j = 1, \dots, n$, можно представить в виде:

$$X_j \stackrel{d}{=} \alpha_j \bar{X}_j + (1 - \alpha_j) \tilde{X}_j, \quad j = 1, \dots, n,$$

где $\bar{X}_j, \tilde{X}_j$ — независимые копии X_j при всех $j = 1, \dots, n$. Поэтому

$$s_n + w_n \stackrel{d}{=} s_{n1} + s_{n2} + w_{n1} + w_{n2} + w_{n3},$$

где

$$\begin{aligned} s_{n1} &= \sum_{j=1}^n \alpha_j z_j(\bar{X}_j) = \sum_{j=1}^n \alpha_j \bar{z}_j, \quad s_{n2} = \sum_{j=1}^n (1 - \alpha_j) z_j(\tilde{X}_j) = \sum_{j=1}^n (1 - \alpha_j) \tilde{z}_j, \\ w_{n1} &= \sum_{i < j} \alpha_i \alpha_j G_{ij}(\bar{X}_i, \bar{X}_j), \quad w_{n2} = \sum_{i \neq j} \alpha_i (1 - \alpha_j) G_{ij}(\bar{X}_i, \tilde{X}_j), \\ w_{n3} &= \sum_{i < j} (1 - \alpha_i)(1 - \alpha_j) G_{ij}(\tilde{X}_i, \tilde{X}_j), \end{aligned}$$

откуда получаем, что подынтегральная функция в ε_3 совпадает со следующей функцией:

$$\begin{aligned} I &= \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2} + w_{n1} + w_{n2} + w_{n3}) \right\} - \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2}) \right\} \\ &= \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2} + w_{n1}) \right\} - \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2}) \right\} \\ &\quad + \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2} + w_{n1} + w_{n3}) \right\} - \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2} + w_{n1}) \right\} \\ &\quad + \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2} + w_{n1} + w_{n2} + w_{n3}) \right\} \\ &\quad - \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2} + w_{n1} + w_{n3}) \right\} = I_1 + I_2 + I_3. \end{aligned}$$

Пусть α обозначает последовательность случайных величин, определенных в (5). Для оценки I_1 и I_2 воспользуемся тем, что при фиксированных α случайная величина s_{n2} не зависит от s_{n1} и w_{n1} , а s_{n1} не зависит от w_{n3} , приемами, использованными в [1], [8], и следующей леммой, которая является уточнением леммы из [6].

Лемма 1. При $7 \leq t \leq h$ и $h^{-1} \leq 15^{-1}$

$$\mathbf{E}|\mathbf{E}[e^{its_{n1}} | \alpha]|^2 \leq t^{-6}, \quad \mathbf{E}|\mathbf{E}[e^{its_{n2}} | \alpha]|^2 \leq t^{-6} e^{-t^2/3},$$

$$\mathbf{E} \left| \mathbf{E} \left[\exp \left\{ it \sum_{k \neq i, j} \alpha_k z_k \right\} \middle| \alpha \right] \right|^2 \leq t^{-0.08},$$

$$\mathbf{E} \left| \mathbf{E} \left[\exp \left\{ it \sum_{k \neq i, j} (1 - \alpha_k) z_k \right\} \middle| \alpha \right] \right|^2 \leq t^{-0.09} e^{-0.0045t^2}.$$

Чтобы оценить I_3 , пишем представление

$$\begin{aligned} I_3 = & \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2} + w_{n3}) \right\} (e^{itw_{n1}} - 1)(e^{itw_{n2}} - 1) \\ & + \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2} + w_{n3}) \right\} (e^{itw_{n2}} - 1 - itw_{n2}) \\ & + it \mathbf{E} \exp \left\{ it(s_{n1} + s_{n2} + w_{n3}) \right\} w_{n2} = I_{31} + I_{32} + I_{33}, \end{aligned}$$

где I_{31} и I_{32} оцениваются по аналогии с I_1 и I_2 . Для оценки I_{33} снова используем метод рандомизации, представив $\bar{X}_j$, $j = 1, \dots, n$, следующим образом: $\bar{X}_j \stackrel{d}{=} \eta_j Y_j + (1 - \eta_j) \bar{Y}_j$, где $Y_j, \bar{Y}_j$ — независимые копии $\bar{X}_j$ при всех $j = 1, \dots, n$ и $\eta_1, \dots, \eta_n$ — независимые одинаково распределенные случайные величины, не зависящие от α и от $X_1, \dots, X_n$ и имеющие распределение Бернулли: $\mathbf{P}\{\eta_j = 1\} = 1 - \mathbf{P}\{\eta_j = 0\} = \frac{1}{2}$, $j = 1, \dots, n$.

Тогда

$$\begin{aligned} s_{n1} & \stackrel{d}{=} \sum_{j=1}^n \alpha_j \eta_j z_j(Y_j) + \sum_{j=1}^n \alpha_j (1 - \eta_j) z_j(\bar{Y}_j), \\ w_{n2} & = \sum_{i=1}^n \alpha_i \eta_i \sum_{j=1, j \neq i}^n (1 - \alpha_j) G_{ij}(Y_i, \tilde{X}_j) + \sum_{i=1}^n \alpha_i (1 - \eta_i) \sum_{j=1, j \neq i}^n (1 - \alpha_j) G_{ij}(\bar{Y}_i, \tilde{X}_j). \end{aligned}$$

Тем самым, снова выделяем независимую компоненту при фиксированных α и $\eta_1, \dots, \eta_n$. Далее схема доказательства во многом повторяет [1], [6] и [8].

СПИСОК ЛИТЕРАТУРЫ

1. Боровских Ю. В. О нормальной аппроксимации U -статистик. — Теория вероятн. и ее примен., 2000, т. 45, в. 3, с. 469–488.
2. Королюк В. С., Боровских Ю. В. Теория U -статистик. Киев: Наукова думка, 1989, 384 с.
3. Малевич Т. Л., Абдурахманов Г. Р. Центральная предельная теорема для обобщенных (неоднородных) U -статистик от различно распределенных случайных величин. — Изв. АН УзССР, сер. физ.-матем. наук, 1986, № 2, с. 28–33.
4. Феллер В. Введение в теорию вероятностей и ее приложения. Т. 2. М.: Мир, 1984, 738 с.
5. Хашимов Ш. А., Абдурахманов Г. Р. Об оценке скорости сходимости в центральной предельной теореме для обобщенных U -статистик. — Теория вероятн. и ее примен., 1998, т. 43, в. 1, с. 69–81.
6. Alberink I. B. A Berry–Esseen bound for U -statistics in non-i.i.d. case. — J. Theoret. Probab., 2000, v. 13, № 2, p. 519–533.
7. Bentkus V., Götze F., Zitikis R. Lower estimates of the convergence rate for U -statistics. — Ann. Probab., 1994, v. 22, № 4, p. 1707–1714.
8. Friedrich K. O. A Berry–Esseen bound for functions of independent random variables. — Ann. Statist., 1989, v. 17, № 1, p. 170–183.
9. Ghosh M., Dasgupta R. Berry–Esseen theorem for U -statistics in non-i.i.d. case. — Colloq. Math. Soc. János Bolyai, 1982, v. 32, p. 293–313.
10. Hoeffding W. A class of statistics with asymptotically normal distribution. — Ann. Math. Statist., 1948, v. 19, p. 293–325.

Поступила в редакцию
16.VII.2001

Сокращенный вариант
28.IX.2001