УДК 519.21

В. С. Королюк, Ю. В. Боровских

U-статистики от бутстрэп-выборки

Рассматривается асимптотика распределений U-статистик типа бутстрэп. Указаны условия сходимости к нормальному распределению.

Розглядається асимптотика розподілів U-статистик типу бутстреп. Вказані умови збіжності до нормального розподілу.

1. Определения. Алгебраическая лемма. Пусть $X_1, \dots, X_N$ — независимая случайная выборка из совокупности с функцией распределения F. Рассмотрим U-статистику [1]

$$U_N = \binom{N}{2}^{-1} \sum_{1 \leqslant i < j \leqslant N} \Phi(X_i, X_j)$$

с симметрическим ядром Φ. Если $E|\Phi| < \infty$, то $EU_N = E\Phi(X_1, X_2) = \theta$. По выборке $X_1, \dots, X_N$ определим эмпирическую функцию распределения $F_N(x)$. Зафиксируем $F_N(x)$. Затем произведем новую случайную выборку $X_1^*, \dots, X_n^*$ из совокупности с функцией распределения $F_N(x)$. По этой бутстрэп-выборке [2] определим U-статистику

$$U_n^* = \binom{n}{2}^{-1} \sum_{1 \leqslant i < j \leqslant n} \Phi(X_i^*, X_j^*).$$

В последующем P и E будут соответствовать распределению F, а P^* и E^* — распределению F_N. Так что

$$\theta^* = E^* U_N^* = E^* \Phi(X_1^*, X_2^*) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \Phi(X_i, X_j).$$

Лемма. *Справедлива формула*

$$U_n^* = \frac{2}{n(n-1)} \sum_{1 \leqslant i < j \leqslant N} \gamma_i \gamma_j \Phi(X_i, X_j) + \frac{1}{n(n-1)} \sum_{i=1}^{N} \gamma_i (\gamma_i - 1) \Phi(X_i, X_i), \quad (1)$$

где при фиксированном векторе $(X_1, \dots, X_N)$ целочисленный случайный вектор $(\gamma_1, \dots, \gamma_N)$ имеет мультиномиальное распределение с параметрами $(n; N^{-1}, \dots, N^{-1})$.

Доказательство. Пусть $X_1, \dots, X_N$ фиксированы. Распределение бутстрэп-выборки $X_1^*, \dots, X_n^*$ сосредоточено на $\{X_1, \dots, X_N\}$. В соответствии с [2] условная вероятность события, состоящего в том, что $X_1, \dots, X_N$ встретится в бутстрэп-выборке $\gamma_1, \dots \gamma_N$ раз, определяется формулой

$$\frac{n!}{\gamma_1! \dots \gamma_N!} \left(\frac{1}{N}\right)^n, \quad \gamma_1 + \dots + \gamma_N = n,$$

что совпадает с мультиномиальным вектором с параметрами $(n; N^{-1}, \dots$

$\ldots, N^{-1})$. Далее, учитывая определение вектора $(\gamma_1, \ldots, \gamma_N)$, имеем

$$\sum_{1 \leqslant i < j \leqslant n} \Phi(X_i^*, X_j^*) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Phi(X_i^*, X_j^*) -$$

$$- \frac{1}{2} \sum_{i=1}^{n} \Phi(X_i^*, X_i^*) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i \gamma_j \Phi(X_i, X_j) - \frac{1}{2} \sum_{i=1}^{N} \gamma_i \Phi(X_i, X_i).$$

Отсюда следует (1).

2. Сходимость к нормальному распределению. Асимптотическая нормальность U_n^* с невырожденным ядром Φ при условиях $E\Phi^2(X_1, X_2) < \infty$, $E\Phi^2(X_1, X_1) < \infty$ доказана в [3] и затем этот результат уточнялся в [4, 5]. В действительности моментные условия на ядро Φ можно ослабить с сохранением асимптотической нормальности U_n^*. Обозначим

$$g_1(x) = E(\Phi(X_1, X_2) - \theta \mid X_1 = x),$$
$$g_2(x, y) = \Phi(x, y) - \theta - g_1(x) - g_1(y)$$

и, соответственно, введем в рассмотрение бутстрэп-вариант этих g-функций:

$$g_{1*}(x) = E^*(\Phi(X_1^*, X_2^*) - \theta^* \mid X_1^* = x) =$$

$$= \frac{1}{N} \sum_{j=1}^{N} \Phi(x, X_j) - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \Phi(X_i, X_j),$$

$$g_{2*}(x, y) = \Phi(x, y) - \theta^* - g_{1*}(x) - g_{1*}(y).$$

Теорема 1. *Пусть* $\sigma_1^2 = Eg_1^2(X_1)$, $\sigma_{1*}^2 = E^* g_{1*}^2(X_1^*)$ *и выполнены условия*

$$0 < Eg_1^2(X_1) < \infty, \tag{2}$$

$$E|g_2(X_1, X_2)|^{4/3} < \infty, \tag{3}$$

$$E|\Phi(X_1, X_1)|^{1/2} < \infty. \tag{4}$$

Тогда при $\min(n, N) \to \infty$ *справедливо*

$$\frac{\sqrt{n}}{2\sigma_{1*}}(U_n - \theta^*) \xrightarrow{d^*} \tau \ (P\text{-п. н.}), \tag{5}$$

где τ *— стандартная нормальная случайная величина, символ* $\xrightarrow{d^*}$ *обозначает слабую сходимость распределений относительно вероятностной меры* P^*.

Доказательство. Для U_n^* напишем бутстрэп-разложение Гефдинга (ср. с (1.1.19) из [1, с. 21])

$$U_n^* - \theta^* = \frac{2}{n}(g_{1*}(X_1^*) + \ldots + g_{1*}(X_n^*)) + \binom{n}{2}^{-1} \sum_{1 \leqslant i < j \leqslant n} g_{2*}(X_i^*, X_j^*). \tag{6}$$

Так что

$$\frac{\sqrt{n}}{2\sigma_{1*}}(U_n^* - \theta^*) = \frac{1}{\sigma_{1*}\sqrt{n}}(g_{1*}(X_1^*) + \ldots + g_{1*}(X_n^*)) + R_{n*}, \tag{7}$$

где

$$R_{n*} = \frac{1}{\sigma_{1*}(n-1)\sqrt{n}} \sum_{1 \leqslant i < j \leqslant n} g_{2*}(X_i^*, X_j^*),$$

$$\sigma_{1*}^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{N} \sum_{j=1}^{N} \left[\Phi(X_i, X_j) - \frac{1}{N} \sum_{k=1}^{N} \Phi(X_k, X_j) \right] \right)^2.$$

Отсюда согласно усиленному закону больших чисел для U-статистик (см. [1], гл. 3) следует, что при $N \to \infty$

$$\sigma_{1*}^2 \to \sigma_1^2 \quad (P\text{-п. н.}). \tag{8}$$

По условной центральной предельной теореме из [3] для бутстрэп-выборочного среднего значения с учетом (8) имеем слабую сходимость

$$\frac{1}{\sigma_{1*}\sqrt{n}}\,(g_{1*}(X_1^*) + \ldots + g_{1*}(X_n^*)) \xrightarrow{d^*} \tau \ (P\text{-п. н.}), \tag{9}$$

когда $\min(n, N) \to \infty$. Случайная величина R_{n*} в (6) обладает свойством

$$R_{n*} \xrightarrow{P^*} 0 \ (P\text{-п. н.}). \tag{10}$$

Покажем, что (10) можно установить методом из работы [1, с. 138—139]. Пусть $R_{n*}(g_{2*}) = R_{n*}$. Функция $g_{2*}(x, y)$ обладает свойством вырожденности

$$\int g_{2*}(x, y)\,F_n(dx) = 0 \ (P^*\text{-п. н.}). \tag{11}$$

Запишем $g_{2*}(x, y)$ в виде суммы $g_{2*} = \psi_{1*} + \psi_{2*}$ с функциями ψ_{1*} и ψ_{2*}, которые определяются так же, как и в [1, с. 138]. Следовательно,

$$R_{n*} = R_{n*}(\psi_{1*}) + R_{n*}(\psi_{2*}).$$

Отсюда по неравенству Чебышева при любом $\varepsilon > 0$

$$P^*(|R_{n*}| > \varepsilon) \leqslant 4\varepsilon^{-2}E^*R_{n*}^2(\psi_{1*}) + (2\varepsilon^{-1})^{4/3}E^*|R_{n*}(\psi_{2*})|^{4/3}. \tag{12}$$

Функции ψ_{1*} и ψ_{2*} обладают свойством (11). Поэтому математические ожидания справа в (12) можно оценить с помощью неравенств из [1, с. 138—139] и показать, что

$$E^*R_{n*}^2(\psi_{1*}) \to 0 \ (P\text{-п. н.}), \ E^*|R_{n*}(\psi_{2*})|^{4/3} \to 0 \ (P\text{-п. н.}).$$

Это совместно с (12) приводит к (10). Из (7) и (9), (10) вытекает утверждение (5).

3. Сходимость в случае вырожденного ядра. Ради простоты положим $n = N$ и предположим, что в бутстрэп-разложении Гефдинга (6) функция $g_{1*}(x) = 0$ (P^*-п.н.). Поэтому (6) можно записать как

$$u(U_n^* - \theta^*) = \frac{1}{n-1}\sum_{i \neq j} g_{2*}(X_i^*, X_j^*). \tag{13}$$

Пусть $\{\lambda_{nj}\}$ и $\{\varphi_{nj}\}$ обозначают собственные числа и собственные функции интегрального оператора, порожденного бутстрэп-функцией g_{2*}, т. е.

$$\int g_{2*}(x, y)\,\varphi_{nj}(y)\,F_n(dy) = \lambda_{nj}\varphi_{nj}(x).$$

Очевидно, что λ_{nj} и φ_{nj} зависят от $X_1, \ldots, X_n$.

Теорема 2. *Пусть выполнено условие* (11) *и*

$$E|g_2(X_1, X_2)|^2 < \infty, \tag{14}$$

$$\lim_{n\to\infty} \lambda_{nj} = \lambda_j \ (P\text{-п. н.}), \tag{15}$$

где неслучайная последовательность $\{\lambda_j\}$ *удовлетворяет условию*

$$\sum_{j=1}^{\infty} \lambda_j^2 < \infty. \tag{16}$$

Тогда при $n \to \infty$ *имеет место сходимость*

$$n(U_n^* - \theta^*) \xrightarrow{d^*} \eta_\infty \ (P\text{-п. н.}), \tag{17}$$

где $\eta_\infty = \sum_{j=1}^{\infty} \lambda_j(\tau_j^2 - 1)$, *а* $\tau_1, \tau_2, \ldots$ — *независимые стандартные нормальные случайные величины.*

Доказательство. Функцию $g_{2*}(x, y)$ запишем как

$$g_{2*}(x, y) = \sum_{j=1}^{\infty} \lambda_{nj} \varphi_{nj}(x) \varphi_{nj}(y).$$

После подстановки этого ряда в (13) получим

$$n(U_n^* - \theta^*) = \frac{n}{n-1} \sum_{j=1}^{\infty} \lambda_{nj} \left\{ \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \varphi_{nj}(X_k^*) \right)^2 - \frac{1}{n} \sum_{k=1}^{n} \varphi_{nj}^2 (X_k^*) \right\}. \quad (18)$$

Система функций $\{\varphi_{nj}\}$ ортонормирована относительно распределения F_n. Учитывая это, при $n \to \infty$ имеем

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \varphi_{nj}(X_k^*) \xrightarrow{d^*} \tau_j \ (P\text{-п. н.}),$$

$$\frac{1}{n} \sum_{k=1}^{n} [\varphi_{nj}^2(X_k^*) - E^* \varphi_{nj}^2(X_1^*)] \xrightarrow{P^*} 0 \ (P\text{-п. н.}),$$

$$E\varphi_{nj}^2(X_1^* = \frac{1}{n} \sum_{k=1}^{n} \varphi_{nj}^2(X_k) \to 1 \quad (P\text{-п. н.})$$

для любого фиксированного $j = 1, 2, \ldots$. Из этих соотношений и из (18) с помощью аргументов, приведенных в [1, с. 97 — 98], в условиях данной теоремы вытекает (17).

Возникает проблема исследования асимптотических свойств распределений с помощью формулы (1). При таком подходе дело сводится к изучению квадратичных форм от мультиномиального вектора $(\gamma_1, \ldots, \gamma_N)$ с матрицей коэффициентов $\Phi(X_i, X_j)$, $i, j = 1, \ldots, N$.

1. *Королюк В. С., Боровских Ю. В.* Теория U-статистик.— Киев : Наук. думка, 1989.— 384 с.
2. *Эфрон Б.* Нетрадиционные методы многомерного статистического анализа.— М. : Финансы и статистика, 1988.— 263 с.
3. *Bickel P. J., Freedman D. A.* Some asymptotic theory for the bootstrap // Ann. Statist.— 1981.— **9**, N 6.— P. 1196—1217.
4. *Laws* of large numbers for bootstrapped U-statistics / K. B. Athreya, C. Ghosh, L. J. Low, P. K. Sen // J. Statist. Plann. and Inference.— 1984.— **9**, N 2.— P. 185—194.
5. *Shi Xi Quan.* Some asymptotic properties of bootstrapping U-statistics // J. Syst. Sci. and Math. Sci.— 1987.—7, N 1.— P. 14—22.

Ленингр. ин-т текстил. и лег. пром-сти

Получено 29.11.89

УДК 517.987

М. С. Матвейчук

Индефинитная мера в J-пространствах

Предложен аналог теоремы Глисона для логики всех J-проекторов аппроксимативного W^*J-фактора в J-пространстве.

Запропоновано аналог теореми Глісона для логіки всіх J-проекторів апроксимативного W^*J-фактору в J-просторі.

Хорошо известна проблема описания мер на квантовых логиках [1]. Исчерпывающее решение проблемы получено для решетки П всех ортопроекторов W^*-алгебры $\mathfrak{A}$. Существенное продвижение имеется и для логики $\mathscr{L}$ всех проекторов $\mathfrak{A}$. Предложим описание мер на логиках проекторов J-пространств.

Пусть H —пространство с индефинитной метрикой $[\cdot, \cdot]$ и канонической симметрией J (H — J-пространство). Относительно скалярного произведе-