Записки научных
семинаров ПОМИ
Том 260, 1999 г.

Н. В. Грибкова

БУТСТРЕП-АППРОКСИМАЦИЯ РАСПРЕДЕЛЕНИЙ L-СТАТИСТИК

Изучаются условия состоятельности бутстреп-аппроксимации распределений линейных комбинаций порядковых статистик $L_n = \sum_{i=1}^n c_{n,i} X_{n,i}$, построенных по выборке объёма n независимых одинаково распределённых случайных величин, с коэффициентами $c_{n,i}$, определяемыми весовой функцией $J(u)$, $u \in (0,1)$, по формулам: (i) $c_{n,i} = \frac{1}{n} J\left(\frac{i}{n+1}\right)$ или $c_{n,i} = \int_{(i-1)/n}^{i/n} J(u)du$, $i = 1, \ldots, n$. Показано различие в асимптотических свойствах бутстреп-распределений L-статистик этих двух, обычно применяемых в задачах статистического оценивания, типов весов. При одних и тех же естественных моментных предположениях в случае (ii) для состоятельности бутстреп-аппроксимации (сходимости) достаточна непрерывность функции J, в случае (i) требуется, чтобы J удовлетворяла условию Гельдера порядка больше $1/2$, причем в случае (i) из-за смещенности бутстреп-математического ожидания относительно L_n существенную роль для состоятельности играет предельное поведение величины отношения объёмов исходной и искусственной выборок, в то время как в случае (ii) эта величина не оказывает влияния на состоятельность.

Ключевые слова и фразы: бутстрепированные L-статистики, линейные комбинации порядковых статистик, асимптотическая нормальность, состоятельность бутстреп-аппроксимации.

1. Введение

Рассмотрим последовательность $X_1, X_2, \ldots$ независимых одинаково распределенных случайных величин (с.в.) с функцией распределения (ф.р.) F. Обозначим $X_{n,1} \leqslant \cdots \leqslant X_{n,n}$ порядковые статистики, соответствующие первым n элементам последо-

---

Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант 99-01-00724.

вательности, F_n-эмпирическую ф.р. Введем линейную комбинацию порядковых статистик $L_n = \sum_{i=1}^n c_{n,i} X_{n,i}$ с коэффициентами $c_{n,i}$, определяемыми с помощью весовой функции $J(u)$, $0 < u < 1$, по одной из следующих формул

$$\text{(i) } c_{n,i} = \frac{1}{n} J\left(\frac{i}{n+1}\right); \quad \text{(ii) } c_{n,i} = \int\limits_{(i-1)/n}^{i/n} J(u)du, \quad i = 1, \dots, n.$$

Воспользуемся интегральным представлением с.в. L_n, причем в случае (i) будем обозначать ее S_n, а в случае (ii) – T_n. Имеем

$$S_n = \int\limits_0^1 F_n^{-1}(u) J_n(u) du; \quad T_n = \int\limits_0^1 F_n^{-1}(u) J(u) du. \tag{1}$$

Здесь $F_n^{-1}(u) = \inf\{x : F_n(x) > u\}$ – инверсия эмпирической ф.р., $J_n(u) = J\left(\frac{i}{n+1}\right)$, $\frac{i-1}{n} \leqslant u < \frac{i}{n}$, $i = 1, \dots, n$.

Введем теперь бутстреп-эквиваленты статистик S_n и T_n. Пусть $X_1^*, \dots, X_m^*$ – выборка объёма m независимых (условно, при условии $X_1, X_2, \dots, X_n$) с.в. с ф.р. F_n. Символом $*$ будем отмечать величины, относящиеся к условному распределению. Пусть $X_{m,1}^* \leqslant \dots \leqslant X_{m,m}^*$ – вариационный ряд, построенный по бутстреп-выборке, F_m^* – эмпирическая бутстреп-ф.р. Тогда бутстреп- варианты статистик S_n и T_n определяются формулами

$$S_m^* = \int\limits_0^1 (F_m^*)^{-1}(u) J_m(u) du, \quad T_m^* = \int\limits_0^1 (F_m^*)^{-1}(u) J(u) du \tag{2}$$

(для введения в бутстреп см. [6]). Обозначим $\sigma^2(S_n)$, $\sigma^2(T_n)$ дисперсии с.в. S_n и T_n, $\sigma^{*2}(S_m^*)$, $\sigma^{*2}(T_m^*)$ – дисперсии их бутстреп-двойников. Введем величину

$$\tau(J, F) = \int\limits_0^1 F^{-1}(u) J(u) du, \tag{3}$$

оценками которой часто служат статистики S_n и T_n, и величину

$$\sigma^2(J, F) = \int\limits_0^1 \int\limits_0^1 J(u) J(v) [\min(u, v) - uv] dF^{-1}(u) dF^{-1}(v), \tag{4}$$

называемую асимптотической дисперсией. При определенных условиях (см. [9, 4]) величины $n\sigma^2(S_n)$ и $n\sigma^2(T_n)$ сходятся к $\sigma^2(J,F)$ при $n \to \infty$.

В случае среднего ($J \equiv 1$, $L_n = \overline{X}_n$, $E^*(S_m^*) = E^*(T_m^*) = \overline{X}_n$, $m\sigma^{*2}(S_m^*) = m\sigma^{*2}(T_m^*) = S^2$, где $\overline{X}_n$, S^2 – эмпирические среднее и дисперсия) асимптотические свойства бутстреп-распределений изучались в [2, 8] (см. также [1, 3–6]). В [3] исследована возможность бутстреп-аппроксимации для усеченного среднего и для экстремальных порядковых статистик. В [4] рассматривался класс обобщенных L-статистик Серфлинга, исследовались асимптотические свойства обобщенных L-статистик, среди других результатов, в [4] найдены условия состоятельности бутстреп-аппроксимации их распределений. Результаты [4] применимы к обыкновенным L-статистикам вида (ii) (с.в. T_n), но условия на функцию J и моментные ограничения [4] для обычных L-статистик являются избыточными.

В этой работе рассматривается случай обычных неусеченных L-статистик с ограниченной весовой функцией J. В п. 2 изучаются асимптотические свойства естественных нормирующих постоянных бутстрепированных L-статистик, в п. 3 – условия состоятельности бутстреп-аппроксимации распределений L-статистик и точность аппроксимации. В п. 2 будет показано, что при определенных естественных предположениях $E^*(S_m^*) - S_n = o(\max(n^{-1/2}, m^{-1/2}))$, $E^*(T_m^*) - T_n = o(m^{-1/2})$ при $n \to \infty$, $m \to \infty$ с вероятностью 1 (п.н. $[P]$), причем даже в случае гладкой весовой функции, такой, например, как $J(u) = u$, величина смещения $E^*(S_m^*) - S_n$ имеет компоненту порядка $1/n$ п.н. $[P]$ (см. п. 2, доказательство теоремы 1), поэтому при нормировке с.в. $S_m^* - S_n$ средним квадратическим отклонением $\sigma^*(S_m^*)$, имеющим порядок $m^{-1/2}$ (см. п. 2), сказывается предельное поведение величины отношения m/n, и если порядок m преобладает над n при стремлении обеих величин к бесконечности, то аппроксимация распределения с.в. $(S_n - \tau(J,F))/\sigma(S_n)$ распределением ее бутстреп-варианта $(S_m^* - S_n)/\sigma^*(S_m^*)$ оказывается неверной (примеры, показывающие, что бутстреп-аппроксимация распределений статистик в ряде случаев может не быть состоятельной, имеются в [1–3, 6]). В случае статистики T_n проблемы, связанной со смещенностью, не возникает.

2. Асимптотические свойства нормирующих постоянных

Теоремы 1–4 п. 2 касаются свойств бутстреп-математических ожиданий, теорема 5 – бутстреп-дисперсий.

Теорема 1. *Предположим, что* $E|X_1| < \infty$.

(i) *Если функция* J *ограничена на* $(0,1)$ *и непрерывна за исключением может быть множества точек меры Лебега нуль, то при* $n \to \infty$, $m \to \infty$

$$E^*(S_m^*) - S_n \to 0 \quad \text{п.н. } [P]. \tag{5}$$

(ii) *Если* J *ограничена на* $(0,1)$, *то при* $n \to \infty$, $m \to \infty$

$$E^*(T_m^*) - T_n \to 0 \quad \text{п.н. } [P]. \tag{6}$$

Доказательство. Соотношения (5) и (6) будем доказывать прямым вычислением математических ожиданий $E^*(S_m^*)$ и $E^*(T_m^*)$, взятых по условному распределению P^*, для которого $P^*(X_i^* = X_j) = 1/n$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. Для с.в. S_m^* имеем

$$E^*(S_m^*) = \frac{1}{m} \sum_{i=1}^{m} J\left(\frac{i}{m+1}\right) E^* X_{m,i}^*$$

$$= \frac{1}{m} \sum_{i=1}^{n} J\left(\frac{i}{m+1}\right) \sum_{j=1}^{n} X_{n,j} P^*(X_{m,i}^* = X_{n,j})$$

$$= \frac{1}{n} \sum_{j=1}^{n} X_{n,j} \sum_{i=1}^{m} J\left(\frac{i}{m+1}\right) p_{ij}^*, \tag{7}$$

где $p_{ij}^* = \frac{n}{m} P^*(X_{m,i}^* = X_{n,j})$. Заметим, что

$$\sum_{i=1}^{m} p_{ij}^* = \frac{n}{m} \sum_{i=1}^{m} P^*(X_{m,i}^* = X_{n,j}) = \frac{n}{m} \sum_{i=1}^{m} P^*(X_i^* = X_{n,j}) = 1,$$

поэтому величины p_{ij}^*, $i = 1, \ldots, m$ при всех $j = 1, \ldots, n$ образуют вероятностное распределение. Если мы при фиксированном j введем с.в. Y_j, принимающую значения $\frac{i}{m+1}$ с вероятностями p_{ij}^*, $i = 1, \ldots, m$, то внутренняя сумма в правой части (7) есть математическое ожидание с.в. $J(Y_j)$, вычисленное по этому распределению. Обозначим $b(p, n, k) = \binom{n}{k} p^k (1-p)^{n-k}$ вероятности биномиального распределения с параметрами (p, n), тогда

$$P^*(X^*_{m,i} = X_{n,j}) = \sum_{k=i}^{n} b\left(\frac{j}{n}, m, k\right) - \sum_{k=i}^{m} b\left(\frac{j-1}{n}, m, k\right) \tag{8}$$

(при $j = 1$ в правой части (8) есть только первое слагаемое). Отсюда находим

$$p^*_{ij} = \frac{n}{m} \sum_{k=i}^{m} \left[b\left(\frac{j}{n}, m, k\right) - b\left(\frac{j-1}{n}, m, k\right) \right]. \tag{9}$$

При этом, очевидно,

$$\sum_{i=1}^{m} p^*_{ij} = \frac{n}{m} \sum_{i=1}^{m} i \left[b\left(\frac{j}{n}, m, i\right) - b\left(\frac{j-1}{n}, m, i\right) \right]$$

$$= \frac{n}{m} \left(\frac{i}{n}\, m - \frac{j-1}{n}\, m \right) = 1.$$

Таким образом,

$$E^*(S^*_m) - S_n = \frac{1}{n} \sum_{j=1}^{n} X_{n,j} \sum_{i=1}^{m} \left[J\left(\frac{i}{m+1}\right) - J\left(\frac{j}{n+1}\right) \right] p^*_{ij}, \tag{10}$$

где p^*_{ij} определяются по формулам (9). Для с.в. T^*_m аналогично находим

$$E^*(T^*_m) - T_n$$

$$= \frac{1}{n} \sum_{j=1}^{n} X_{n,j} \sum_{i=1}^{m} \left(m \int\limits_{(i-1)/m}^{i/m} J(u)du - n \int\limits_{(j-1)/n}^{j/n} J(u)du \right) p^*_{ij}. \tag{11}$$

Найдем первые два момента вспомогательной с.в. Y_j. Пусть E_j обозначает математическое ожидание по распределению p^*_{ij}, тогда

$$E_j(Y_j) = \sum_{i=1}^{m} \frac{i}{m+1}\, p^*_{ij}$$

$$= \frac{n}{m(m+1)} \sum_{i=1}^{m} \frac{i(i+1)}{2} \left[b\left(\frac{j}{n}, m, i\right) - b\left(\frac{j-1}{n}, m, i\right) \right], \tag{12}$$

$$E_j(Y_j^2) = \sum_{i=1}^{m} \left(\frac{i}{m+1}\right)^2 p_{ij}^*$$

$$= \frac{n}{m(m+1)^2} \sum_{i=1}^{m} \frac{1}{6} i(i+1)(2i+1) \left[b\left(\frac{j}{n}, m, i\right) - b\left(\frac{j-1}{n}, m, i\right) \right]. \tag{13}$$

Используя теперь производящую функцию факториальных моментов биномиального распределения, из (12) и (13) находим

$$E_j Y_j = \frac{m-1}{m+1} \cdot \frac{j-1/2}{n} + \frac{1}{m+1}, \tag{14}$$

$$\sigma_j^2(Y_j) = E_j[Y_j - E_j Y_j]^2$$

$$= \frac{m-1}{(m+1)^2} \left[\frac{j(n-j)}{n^2} - \frac{1}{n}\left(\frac{1}{2} - \frac{j}{n}\right) + \frac{m-5}{12n^2} \right]. \tag{15}$$

Из (14) и (15) следует

$$\left| E_j Y_j - \frac{j}{n+1} \right| < 3 \left(\frac{1}{2n} + \frac{1}{m+1} \right), \tag{16}$$

$$\sigma_j^2(Y_j) < \frac{1}{m+1} \left(\frac{j(n-j)}{n^2} + \frac{1}{2n} \right) + \frac{1}{12n^2}. \tag{17}$$

при всех $j = 1, \dots, n$. Если функция J непрерывна и ограничена на (0,1), то из соотношения (10), оценок (16) и (17) по неравенству Чебышева и усиленному закону больших чисел для среднего $n^{-1}\sum_{j=1}^{n} X_j$, позволяющему вместе с ограниченностью J перейти к пределу в (10), следует (5). Аналогично из (11) следует (6). Теорема доказана.

Теорема 2. *Предположим, что* $A = \sup_{u \in (0,1)} |J(u)| < \infty$.

(i) *Если* J *удовлетворяет условию Гельдера порядка* $1/2 + \alpha$, $0 < \alpha \leqslant 1/2$, *на* $(0,1)$ *за исключением может быть множества точек меры Лебега нуль и*

$$\int_{-\infty}^{\infty} [F(x)(1-F(x))]^{\frac{3+2\alpha}{4}} dx < \infty, \tag{18}$$

то

$$ES_n - \tau(J,F) = O\big(n^{-\frac{1}{2}-\alpha}\big), \quad n \to \infty; \tag{19}$$

$$E^*(S_m^*) - S_n = O\big(\max(n^{-\frac{1}{2}-\alpha}, m^{-\frac{1}{2}-\alpha})\big) \ \text{п.н.}\ [P], \tag{20}$$

$n \to \infty$, $m \to \infty$.

(ii) *Если J удовлетворяет условию Гельдера порядка $0 < \alpha \leqslant 1$ на $(0,1)$ за исключением может быть множества точек меры Лебега нуль и*

$$\int\limits_{-\infty}^{\infty} [F(x)(1-F(x))]^{\frac{1+\alpha}{2}} dx < \infty, \tag{21}$$

то

$$ET_n - \tau(J,F) = O\big(n^{-\frac{1}{2}-\frac{\alpha}{2}}\big), \quad n \to \infty; \tag{22}$$

$$E^*(T_m^*) - T_n = O\big(m^{-\frac{1}{2}-\frac{\alpha}{2}}\big) \ \text{п.н.}\ [P], \ n \to \infty, \ m \to \infty. \tag{23}$$

Замечание 1. Условие (18) при $\alpha = 1/2$ равносильно условию $E|X_1| < \infty$ и оно выполняется, очевидно, при любом $\alpha > 0$, если $E|X_1|^{4/3} < \infty$.

Замечание 2. Условие (21) при $\alpha = 1$ равносильно условию $E|X_1| < \infty$ и оно выполняется, очевидно, при любом $\alpha > 0$, если $EX_1^2 < \infty$.

Замечание 3. Условие гладкости для функции J в утверждении (i) согласуется с предположениями о J в теореме 4 [9].

Доказательство. (i) По условию (18) $E|X_1| < \infty$, поэтому

$$\tau(J,F) = \int\limits_0^{\infty} \int\limits_{F(x)}^{1} J(u)du\,dx - \int\limits_{-\infty}^{0} \int\limits_{0}^{F(x)} J(u)du\,dx \tag{24}$$

(см. формулу (11) [9]). Для S_n аналогично имеем

$$S_n = \int\limits_0^{\infty} \int\limits_{F_n(x)}^{1} J_n(u)du\,dx - \int\limits_{-\infty}^{0} \int\limits_{0}^{F_n(x)} J_n(u)du\,dx. \tag{25}$$

Из (24), (25) и того, что $E\left(\int_{-\infty}^{\infty} \int_{F_n(x)}^{F(x)} J(F(x))du\,dx\right) = 0$, следует

$$E(S_n - \tau(J,F)) = E\left(\int\limits_{-\infty}^{\infty} \int\limits_{F_n(x)}^{F(x)} [J(u) - J(F(x))]du\,dx\right.$$

$$+\int\limits_{0}^{\infty}\int\limits_{F_n(x)}^{1}(J_n(u)-J(u))du\,dx-\int\limits_{-\infty}^{0}\int\limits_{0}^{F_n(x)}(J_n(u))-J(u))du\,dx\Bigg). \quad (26)$$

Так как $\sup_{u\in(0,1)}|J_n(u)-J(u)|\leqslant Bn^{-\frac{1}{2}-\alpha}$, где B – постоянная из условия Гельдера, из (26) следует

$$E|S_n-\tau(J,F)|\leqslant B\int\limits_{-\infty}^{\infty}E|F_n(x)-F(x)|^{3/2+\alpha}dx$$

$$+Bn^{-1/2-\alpha}\left(\int\limits_{0}^{\infty}(1-F(x))dx+\int\limits_{-\infty}^{0}F(x)dx\right). \quad (27)$$

Для оценки первого слагаемого правой части (27) имеем

$$\int\limits_{-\infty}^{\infty}E|F_n(x)-F(x)|^{3/2+\alpha}dx\leqslant\int\limits_{-\infty}^{\infty}[E(F_n(x)-F(x))^2]^{3/4+\alpha/2}dx=$$

$$=n^{-3/4-\alpha/2}\int\limits_{-\infty}^{\infty}[F(x)(1-F(x))]^{(3+2\alpha)/4}dx.$$

По условию (18) последнее выражение есть $O(n^{-3/4-\alpha/2})$, $n\to\infty$. Соотношение (19) доказано. Для доказательства (20) используем (25) и формулу

$$S_m^*=\int\limits_{0}^{\infty}\int\limits_{F_m^*(x)}^{1}J_m(u)du\,dx-\int\limits_{-\infty}^{0}\int\limits_{0}^{F_m^*(x)}J_m(u)du\,dx.$$

Так как $\sup_{u\in(0,1)}|J_m(u)-J_n(u)|\leqslant B\max(n^{-1/2-\alpha},m^{-1/2-\alpha})$, по условию Гельдера аналогично (27) находим

$$E^*|S_m^*-S_n|\leqslant B\int\limits_{-\infty}^{\infty}E^*|F_m^*(x)-F_n(x)|^{3/2+\alpha}dx$$

$$+B\max(n^{-1/2-\alpha},m^{-1/2-\alpha})\left(\int\limits_{0}^{\infty}(1-F_n(x))dx+\int\limits_{-\infty}^{0}F_n(x)dx\right), \quad (28)$$

и поскольку

$$E^*|F^*_m(x) - F_n(x)|^{3/2+\alpha} \leqslant (E^*(F^*_m(x) - F_n(x))^2)^{(3+2\alpha)/4}$$

$$= m^{-(3+2\alpha)/4}[F_n(x)(1 - F_n(x))]^{(3+2\alpha)/4},$$

в условиях (i) правая часть (28) есть $O\big(\max(n^{-1/2-\alpha}, m^{-1/2-\alpha})\big)$, $n \to \infty$, $m \to \infty$, п.н. $[P]$, что доказывает (20).

(ii) Для T_n имеем

$$T_n = \int\limits_0^\infty \int\limits_{F_n(x)}^1 J(u)du\,dx - \int\limits_{-\infty}^0 \int\limits_0^{F_n(x)} J(u)du\,dx, \tag{29}$$

что вместе с (24) приводит к соотношению

$$T_n - \tau(J, F) = \int\limits_{-\infty}^\infty \int\limits_{F_n(x)}^{F(x)} J(u)du\,dx. \tag{30}$$

Следовательно,

$$E(T_n - \tau(J, F)) = \int\limits_{-\infty}^\infty E\left(\int\limits_{F_n(x)}^{F(x)} (J(u) - J(F(x)))du\right) dx. \tag{31}$$

Отсюда по условию Гельдера находим

$$E|T_n - \tau(J, F)| \leqslant B \int\limits_{-\infty}^\infty E|F_n(x) - F(x)|^{1+\alpha}dx$$

$$\leqslant Bn^{-(1+\alpha)/2} \int\limits_{-\infty}^\infty [F(x)(1 - F(x))]^{(1+\alpha)/2}dx = O(n^{-1/2-\alpha/2}), \quad n \to \infty.$$

(22) доказано. Для доказательства (23) аналогично имеем

$$T^*_m - T_n = \int\limits_{-\infty}^\infty \int\limits_{F^*_m(x)}^{F_n(x)} J(u)du\,dx.$$

Отсюда по условию Гельдера следует

$$E^*|T_m^* - T_n| \leqslant Bm^{-(1+\alpha)/2} \int\limits_{-\infty}^{\infty} [F_n(x)(1-F_n(x))]^{(1+\alpha)/2} dx, \tag{32}$$

и правая часть (32) есть $O(m^{-(1+\alpha)/2})$ п.н. $[P]$, $n \to \infty$, $m \to \infty$. Теорема доказана.

Теорема 3. *Предположим, что* $A = \sup_u |J(u)| < \infty$.

(i) *Если* $E|X_1|^{4/3} < \infty$ *и функция* J *удовлетворяет условию Гельдера порядка* $1/2 + \alpha$, $0 < \alpha \leqslant 1/2$, *на* $(0,1)$ *за исключением может быть множества точек меры Лебега нуль, то*

$$ES_n - \tau(J,F) = o(n^{-1/2}), \quad n \to \infty, \tag{33}$$

$$E^*(S_m^*) - S_n = o\big(\max(n^{-1/2}, m^{-1/2})\big) \text{ п.н. } [P], \tag{34}$$

$n \to \infty$, $m \to \infty$.

(ii) *Если* $EX_1^2 < \infty$ *и* J *удовлетворяет условию Гельдера порядка* $\alpha > 0$ *на* $(0,1)$ *за исключением может быть множества точек меры Лебега нуль, то*

$$ET_n - \tau(J,F) = o(n^{-1/2}), \quad n \to \infty, \tag{35}$$

$$E^*(T_m^*) - T_n = o\big(m^{-1/2}\big) \text{ п.н. } [P], \ n \to \infty, \ m \to \infty. \tag{36}$$

Соотношения (33)–(36) с учетом замечаний 1 и 2 непосредственно вытекают из предыдущей теоремы.

Замечание 4. Если в утверждении (ii) теоремы 3 моментное условие $EX_1^2 < \infty$ заменить немного более сильным:

$$\int\limits_{-\infty}^{\infty} [F(x)(1-F(x))]^{1/2} dx < \infty,$$

то (35)–(36) остаются в силе, если функция J только лишь непрерывна на (0,1) за исключением может быть множества точек меры Лебега нуль. Действительно, из (30) в силу непрерывности J следует

$$E(T_n - \tau(J,F)) = \int\limits_{-\infty}^{\infty} E\{(J(u_\theta) - J(F(x)))(F_n(x) - F(x))\} dx, \tag{37}$$

где $u_\theta = F(x) + \theta(F_n(x) - F(x))$, $|\theta| < 1$. Тогда

$$E|T_n - \tau(J, F)|$$
$$\leqslant \frac{1}{\sqrt{n}} \int\limits_{-\infty}^{\infty} [(F(x)(1 - F(x))]^{1/2} \big(E[J(u_\theta) - J(F(x))]^2\big)^{1/2} dx. \qquad (38)$$

Так как J ограничена, интеграл в правой части (38) конечен. В то же время подынтегральное выражение при почти всех x сходится к нулю при $n \to \infty$ в силу непрерывности функции J и известного неравенства $P\{\|F_n - F\|_\infty > \varepsilon\} \leqslant C \exp(-2n\varepsilon^2)$, $\varepsilon > 0$, $n \geqslant 1$, где C – независящая от F и n постоянная. Таким образом, правая часть (38) есть $o(n^{-1/2})$, $n \to \infty$, что доказывает (35). Для T_m^* имеем бутстреп-вариант последнего соотношения

$$E^*|T_m^* - T_n|$$
$$\leqslant \frac{1}{\sqrt{m}} \int\limits_{-\infty}^{\infty} [(F_n(x)(1 - F_n(x))]^{1/2} \big(E^*[J(u_\theta) - J(F_n(x))]^2\big)^{1/2} dx,$$

из которого следует (36).

Теорема 4. *Пусть $E|X_1| < \infty$. Если J удовлетворяет условию Липшица порядка 1 на $(0, 1)$ за исключением может быть множества точек меры Лебега нуль, то*

$$\text{(i)} \quad ES_n - \tau(J, F) = O(n^{-1}), \quad n \to \infty; \qquad (39)$$
$$E^*(S_m^*) - S_n = O\big(\max(n^{-1}, m^{-1})\big), \quad \textit{п.н.}\ [P], \qquad (40)$$

$n \to \infty$, $m \to \infty$;

$$\text{(ii)} \quad ET_n - \tau(J, F) = O(n^{-1}), \quad n \to \infty; \qquad (41)$$
$$E^*(T_m^*) - T_n = O\big(m^{-1}\big), \quad \textit{п.н.}\ [P], \ n \to \infty, \ m \to \infty. \qquad (42)$$

Соотношения (39)–(42) с учетом замечаний 1 и 2 непосредственно следуют из теоремы 2.

Введем для краткости следующие обозначения: $\sigma^2 = \sigma^2(J, F)$, $\sigma_n^2 = \sigma^2(J, F_n)$, $\sigma^{*2} = \sigma^2(J, F_m^*)$, где

$$\sigma^2(J, F) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} J(F(x))J(F(y))[F(\min(x, y)) - F(x)F(y)]\,dx\,dy \quad (43)$$

– асимптотичекая дисперсия.

Теорема 5. *Предположим, что* $EX_1^2 < \infty$, $A = \sup_u |J(u)| < \infty$, *функция* J *непрерывна почти всюду по мере* F^{-1} *в интервале* $(0,1)$. *Тогда при* $n \to \infty$, $m \to \infty$ *с вероятностью* 1

$$m\sigma^{*2}(S_m^*) \to \sigma^2, \tag{44}$$

$$m\sigma^{*2}(T_m^*) \to \sigma^2, \tag{45}$$

$$\sigma_n^2 \to \sigma^2, \tag{46}$$

$$\sigma_m^{*2} \to \sigma^2 \text{ п.н. } [P^*]. \tag{47}$$

Доказательство. Соотношение (44) докажем методом из [9]. Для этого доказательство соотношения (9) теоремы 1 [9] достаточно повторить с небольшими изменениями. Обозначим

$$F_i^*(x) = P^*(X_{m,i}^* < x), \quad F_{ij}^*(x,y) = P^*(X_{m,i}^* < x, X_{m,j}^* < y).$$

Тогда

$$\operatorname{cov}^*(X_{m,i}^*, X_{m,j}^*) = \iint_{R^2} [F_{ij}^*(x,y) - F_i^*(x)F_j^*(y)]dx\,dy$$

и

$$m\sigma^{*2}(S_m^*) = \iint_{R^2} B_m^*(x,y)dx\,dy, \tag{48}$$

где

$$B_m^*(x,y) = \frac{1}{m}\sum_{i=1}^m\sum_{j=1}^m J\left(\frac{i}{m+1}\right) J\left(\frac{j}{m+1}\right)[F_{ij}^*(x,y) - F_i^*(x)F_j^*(y)].$$

Нужно показать, что при $n \to \infty$, $m \to \infty$

$$B_m^*(x,y) \to J(F(x))J(F(y))[F(\min(x,y)) - F(x)F(y)] \tag{49}$$

п.н. $[P]$. Обозначим $M_{ij} = (i/(m+1), j/(m+1))$, $M_{xy}^n = (F_n(x), F_n(y))$ точки единичного квадрата $0 \leqslant u, v \leqslant 1$. Положим $d(M_{ij}, M_{xy}^n) = \max\left(\left|F_n(x) - \frac{i}{m+1}\right|, \left|F_n(y) - \frac{j}{m+1}\right|\right)$. Фиксируем x и y так, чтобы функция $J(u)$ была непрерывна в точках $u = F(x)$, $v = F(y)$. Тогда при достаточно больших n $F_n(x)$ и $F_n(y)$ также являются точками непрерывности J п.н. $[P]$. Введем, как в [9], множество $A_{m,n}^* = \{(i,j) : d(M_{ij}, M_{xy}^n) \leqslant m^{-1/8}\}$. Положим

$$U_m^* = \sum_{i=1}^m I_{\{X_i^* < x\}}, \quad U_m^{*'} = \sum_{i=1}^m I_{\{X_i^* \geqslant x\}},$$

где I_A – индикатор множества A. U_m^*, $U_m^{*'}$ – биномиальные с.в. с параметрами $(m, F_n(x))$ и $(m, 1-F_n(x))$ соответственно. Если $(i,j)\notin A_{m,n}^*$, то $|r/(m+1)-F_n(x)|>m^{-1/8}$ для $r=i$ или $r=j$. Пусть для определенности последнее неравенство выполняется для $r=i$. Тогда, если $i>mF_n(x)$, то

$$F_{ij}^*(x,y)-F_i^*(x)F_j^*(y)\leqslant F_i^*(x)=P^*\{U_m^*\geqslant i\}\leqslant$$

$$\leqslant\frac{E^*(U_m^*-mF_n(x))^4}{(i-mF_n(x))^4}<\frac{3m^2F_n(x)(1-F_n(x))+m}{m^4\left(\frac{i}{m}-F_n(x)\right)^4}<4m^{-3/2}. \quad (50)$$

Если $i\leqslant mF_n(x)$, то $m-i\geqslant m(1-F_n(x))$, и используя так же неравенство Чебышева для четвертых моментов с.в. $U_m^{*'}$ [9], приходим снова к неравенству (50). Таким образом,

$$\left|\frac{1}{m}\sum_{(i,j)\notin A_{m,n}^*}J\left(\frac{i}{m+1}\right)J\left(\frac{j}{m+1}\right)[F_{ij}^*(x,y)-F_i^*(x)F_j^*(y)]\right|$$

$$\leqslant 4A^2m^{-1/2},$$

где $A=\sup_{0,1}|J|$. Следовательно,

$$\left|B_m^*(x,y)-\frac{1}{m}J(F_n(x))J(F_n(y))\sum_{i=1}^m\sum_{j=1}^m[F_{ij}^*(x,y)-F_i^*(x)F_j^*(y)]\right|\leqslant$$

$$\leqslant\max_{(i,j)\in A_{m,n}^*}\left|J\left(\frac{i}{m+1}\right)J\left(\frac{j}{m+1}\right)-J(F_n(x))J(F_n(y))\right|\times$$

$$\times\frac{1}{m}\sum_{i=1}^m\sum_{j=1}^m[F_{ij}^*(x,y)-F_i^*(x)F_j^*(y)]+8A^2m^{-1/2}. \quad (51)$$

В силу непрерывности функции J в точках $F(x)$, $F(y)$

$$J(F_n(x))J(F_n(y))\to J(F(x))J(F(y)) \quad (52)$$

при $n\to\infty$ п.н. $[P]$. Для двойной суммы в правой части (51) имеем

$$\frac{1}{m}\sum_{i=1}^m\sum_{j=1}^m[F_{ij}^*(x,y)-F_i^*(x)F_j^*(y)]=F_n(\min(x,y))-F_n(x)F_n(y) \quad (53)$$

(см. [9]). Так как правая часть (53) при $n \to \infty$ сходится к $F(\min(x,y)) - F(x)F(y)$ п.н. $[P]$, из (51)–(53) следует (49). Возможность предельного перехода под знаком интеграла (48) обеспечивают следующие соотношения:

$$|B^*_m(x,y)| \leqslant A^2 \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{m} [F^*_{ij}(x,y) - F^*_i(x) F^*_j(y)]$$

$$= A^2[F_n(\min(x,y)) - F_n(x)F_n(y)];$$

$$\iint\limits_{R^2} [F_n(\min(x,y)) - F_n(x)F_n(y)]dx\,dy = \operatorname{cov}^*(X^*_1, X^*_1) = \sigma^{*2}(X^*_1),$$

где $\sigma^{*2}(X^*_1) = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \to \sigma^2(X_1) < \infty$ п.н. $[P]$ при $n \to \infty$. Соотношение (44) доказано. Учитывая предположение о непрерывности J, соотношение (45) доказывается так же (см. [9]). Соотношения (46)–(47) получаются предельным переходом под знаком интеграла (43), возможность которого обеспечена условиями теоремы.

3. Бутстреп-аппроксимация распределений

Положим $\tau = \tau(J, F)$,

$$\begin{aligned} F_{L_n}(x) &= P\left\{\frac{L_n - EL_n}{\sigma(L_n)} < x\right\}; \\ F^*_{L_m}(x) &= P^*\left\{\frac{L^*_m - E^*L^*_m}{\sigma^*(L^*_m)} < x\right\}, \end{aligned} \tag{54}$$

где L_n – объединяющее обозначение для S_n и T_n, L^*_m – для S^*_m и T^*_m. Таким образом, $F_{S_n}(x)$ и $F^*_{S^*_m}(x)$ обозначают ф.р. естественным образом нормированных с.в. S_n и S^*_m соответственно, а $F_{T_n}(x)$ и $F^*_{T^*_m}(x)$ – ф.р. таким же образом нормированных с.в. T_n и T^*_m.

Теорема 6. *Предположим, что $EX_1^2 < \infty$, $A = \sup_{u\in(0,1)} J(u) < \infty$ и выполнены следующие условия:*

1) *J удовлетворяет условию Гельдера порядка $\alpha > 1/2$ п.в. по мере F^{-1} на $(0,1)$;*

2) *$\limsup m/n < \infty$, $m \to \infty$, $n \to \infty$;*

3) *$\sigma = \sigma(J, F) > 0$.*

Тогда

$$\sup_x |P\{\sqrt{n}(S_n-\tau)<x\} - P^*\{\sqrt{m}(S_m^*-S_n)<x\}| = o(1), \qquad (55)$$

$$\sup_x \left| P\left\{\frac{\sqrt{n}(S_n-\tau)}{\sigma}<x\right\} - P^*\left\{\frac{\sqrt{m}(S_m^*-S_n)}{\sigma_n}<x\right\}\right| = o(1), \quad (56)$$

$$\sup_x \left| P\left\{\frac{\sqrt{n}(S_n-\tau)}{\sigma_n}<x\right\} - P^*\left\{\frac{\sqrt{m}(S_m^*-S_n)}{\sigma_m^*}<x\right\}\right| = o(1) \quad (57)$$

п.н. $[P]$, $n\to\infty$, $m\to\infty$.

Доказательство. Для доказательства (55) имеем неравенство

$$\sup_x |P\{\sqrt{n}(S_n-\tau)<x\} - P^*\{\sqrt{m}(S_m^*-S_n)<x\}| \leqslant \Delta_1+\Delta_2+\Delta_3,$$

где $\Delta_1=\sup_x |F_{S_n}(x)-\Phi(x)|$, $\Delta_2=\sup_x |F^*_{S_m^*}-\Phi(x)|$,

$$\Delta_3=\sup_x \left|\Phi\left(\frac{\tau-ES_n}{\sigma(S_n)}+\frac{x}{\sqrt{n}\,\sigma(S_n)}\right)-\Phi\left(\frac{S_n-E^*S_m^*}{\sigma^*(S_m^*)}+\frac{x}{\sqrt{m}\sigma^*(S_m^*)}\right)\right|,$$

Φ – стандартная нормальная ф.р. В условиях теоремы $\Delta_1\to 0$ при $n\to\infty$ (см. теорему 2 [9]), $\Delta_2\to 0$ п.н. $[P]$ при $n\to\infty$, $m\to\infty$. Для доказательства последнего достаточно повторить с небольшими изменениями доказательство теоремы 2 [9]. Эти изменения сводятся фактически к замене ф.р. F при доказательстве асимптотической нормальности с.в. S_n эмпирической ф.р. F_n. При этом утверждение об эквивалентности в среднем квадратическом с.в. S_n и ее проекции $\hat{S}_n$ (см. [9]) заменяется утверждением об эквивалентности с вероятностью 1 в среднем квадратическом по условной вероятностной мере P^* с.в. S_m^* и ее проекции $\hat{S}_m^*$. Для Δ_3 справедлива оценка

$$\Delta_3 \leqslant \frac{|\tau-ES_n|}{\sigma(S_n)}+\frac{|S_n-E^*(S_m^*)|}{\sigma^*(S_m^*)}+\max\left(\frac{\sqrt{m}\,\sigma^*(S_m^*)}{\sqrt{n}\,\sigma(S_n)},\ \frac{\sqrt{n}\,\sigma^*(S_n)}{\sqrt{m}\,\sigma^*(S_m^*)}\right)-1.$$

Из теорем 3 и 5 следует, что $\sqrt{n}\,\sigma(S_n)\to\sigma$, $n\to\infty$; $\sqrt{m}\,\sigma^*(S_m^*)\to\sigma$ п.н. $[P]$, $n\to\infty$, $m\to\infty$; $|ES_n-\tau|=o(n^{-1/2})$, $n\to\infty$; $|S_n-E^*(S_m^*)|=o(\max(m^{-1/2},n^{-1/2}))$ п.н. $[P]$, $n\to\infty$, $m\to\infty$. Отсюда, учитывая условие 2, находим, что $\Delta_3\to 0$ п.н. $[P]$, $n\to\infty$,

$m \to \infty$. Соотношение (55) доказано. Справедливость соотношения (56) устанавливается аналогично. Для доказательства (57) имеем

$$\sup_x \left| P\left\{ \frac{\sqrt{n}(S_n - \tau)}{\sigma_n} < x \right\} - P^*\left\{ \frac{\sqrt{m}(S_m^* - S_n)}{\sigma_m^*} < x \right\} \right| \leqslant \Delta_1 + \Delta_2 + \Delta_3,$$

где Δ_1 и Δ_2 обозначают те же величины, что и прежде, а

$$\Delta_3 \leqslant \frac{|\tau - ES_n|}{\sigma(S_n)}$$

$$+ \frac{|S_n - E^*(S_m^*)|}{\sigma^*(S_m^*)} + \max\left(\frac{\sqrt{m}\sigma_n \sigma^*(S_m^*)}{\sqrt{n}\sigma_m^* \sigma(S_n)}, \frac{\sqrt{n}\sigma_m^* \sigma(S_n)}{\sqrt{m}\sigma_n \sigma^*(S_m^*)} \right) - 1.$$

По теореме 5 третье слагаемое здесь стремится к 1 п.н. $[P]$, п.н. $[P^*]$ при $n \to \infty$, $m \to \infty$, поэтому $\Delta_3 \to 0$ п.н. $[P]$, п.н. $[P^*]$ при $n \to \infty$, $m \to \infty$, что доказывает (57).

Теорема 7. *Предположим, что $EX_1^2 < \infty$, $A = \sup_{u \in (0,1)} < \infty$ и выполнены следующие условия:*

1) J удовлетворяет условию Гельдера порядка $\alpha > 0$ п.в. по мере F^{-1} на $(0,1)$;

2) $\sigma = \sigma(J, F) > 0$.

Тогда

$$\sup_x |P\{\sqrt{n}(T_n - \tau) < x\} - P^*\{\sqrt{m}(T_m^* - T_n) < x\}| = o(1),$$

$$\sup_x \left| P\left\{ \frac{\sqrt{n}(T_n - \tau)}{\sigma} < x \right\} - P^*\left\{ \frac{\sqrt{m}(T_m^* - T_n)}{\sigma_n} < x \right\} \right| = o(1), \quad (58)$$

$$\sup_x \left| P\left\{ \frac{\sqrt{n}(T_n - \tau)}{\sigma_n} < x \right\} - P^*\left\{ \frac{\sqrt{m}(T_m^* - T_n)}{\sigma_m^*} < x \right\} \right| = o(1)$$

п.н. $[P]$, $n \to \infty$, $m \to \infty$.

Теорема 8. *Предположим, что $\int_{-\infty}^{\infty} [F(x)(1 - F(x))]^{1/2} dx < \infty$, функция J ограничена на $(0,1)$ и непрерывна п.в. по мере F^{-1} на $(0,1)$. Тогда, если $\sigma = \sigma(J, F) > 0$, то имеют место соотношения (58).*

Теоремы 7–8 доказываются так же, как теорема 6, но с использованием утверждений (ii) теорем 3, 5 и замечания 4 (см. п. 2).

Теорема 9. *Пусть* $E|X_1|^3 < \infty$. *Предположим, что выполняются следующие условия:*

1) J *удовлетворяет условию Липшица порядка* 1 *п.в. по мере* F^{-1} *на* $(0,1)$;

2) F^{-1} *удовлетворяет условию Липшица порядка* 1 *в окрестностях точек разрыва функции* J;

3) $\sigma = \sigma(J,F) > 0$.

Тогда

$$\text{(i)} \qquad \sup_x \left| P\left\{ \frac{T_n - \tau}{\sigma(T_n)} < x \right\} - P^*\left\{ \frac{T_m^* - T_n}{\sigma^*(T_m^*)} < x \right\} \right| = O(\max(n^{-1/2}, m^{-1/2})) \tag{59}$$

п.н. $[P]$, $n \to \infty$, $m \to \infty$.

(ii) *если* $\limsup m/n = O(1)$, $n \to \infty$, $m \to \infty$, *то*

$$\sup_x \left| P\left\{ \frac{S_n - \tau}{\sigma(S_n)} < x \right\} - P^*\left\{ \frac{S_m^* - S_n}{\sigma^*(S_m^*)} < x \right\} \right| = O(m^{-1/2}) \tag{60}$$

п.н. $[P]$, $n \to \infty$, $m \to \infty$.

Доказательство. Соотношения (59)–(60) докажем методом Сингха [8] (см. также [4]) на основе теоремы ван Цвета [10] и теорем 4 и 5 п. 2. Пусть снова L_n – объединяющее обозначение для с.в. S_n, T_n. Обозначим B постоянную из условия Липшица для функции J и пусть $K = \sup |F^{-1}(u) - F^{-1}(v)| / |u-v|$, где супремум берется по всем u, v, из окрестностей точек разрыва J, в которых F^{-1} удовлетворяет условию Липшица. Имеем

$$\sup_x \left| P\left\{ \frac{L_n - \tau}{\sigma(L_n)} < x \right\} - P^*\left\{ \frac{L_m^* - L_n}{\sigma^*(L_m^*)} < x \right\} \right| \leqslant \frac{|E(L_n) - \tau|}{\sigma(L_n)} + \frac{|E^*(L_m^*) - L_n|}{\sigma^*(L_m^*)} + \sup_x |F_{L_n}(x) - \Phi(x)| + \sup_x |F^*_{L_m^*}(x) - \Phi(x)|, \tag{61}$$

где $F_{L_n}(x)$, $F^*_{L_m}(x)$ определены (54). По теоремам 4 и 5 п. 2 первое слагаемое в правой части (61) есть $O(n^{-1/2})$, $n \to \infty$, второе слагаемое есть $O(m^{-1/2})$, п.н. $[P]$, $n \to \infty$, $m \to \infty$, в случае $L_n = T_n$, а в случае $L_n = S_n$ оно есть $O(m^{-1/2})$ при дополнительном условии, что m стремится к ∞ не быстрее, чем n (см. п. 2). Для третьего слагаемого (61) по теореме ван Цвета [10] имеем

$$\sup_x |F_{L_n}(x) - \Phi(x)|$$

$$\leqslant \frac{C}{\sqrt{n}} \left[\left(\frac{B}{\sqrt{n}\,\sigma(L_n)} \right)^3 E|X_1|^3 + \left(\frac{A}{\sqrt{n}\,\sigma(L_n)} \right)^2 EX_1^2 + K^3 + K^2 \right], \tag{62}$$

где C – абсолютная положительная постоянная, $A = \sup_{(0,1)} |J|$. Для четвертого слагаемого (61) по той же теореме [10]

$$\sup_x |F^*_{L^*_m}(x) - \Phi(x)| \leqslant \frac{C}{\sqrt{m}} \left[\left(\frac{B}{\sqrt{m}\,\sigma^*(L^*_m)} \right)^3 E^*|X_1^*|^3 + \right.$$

$$\left. \left(\frac{A}{\sqrt{m}\,\sigma^*(L^*_m)} \right)^2 E^*(X_1^*)^2 + K_n^{*3} + K_n^{*2} \right], \tag{63}$$

где $K_n^* = \sup_{u,v} |F_n^{-1}(u) - F_n^{-1}(v)|/|u - v|$, где супремум берется по тому же множеству, что и в определении K. В силу гладкости F^{-1} на этом множестве с вероятностью P, равной единице, $K_n^* \to K$, $n \to \infty$ [2]. По теореме 5 $\sqrt{m}\,\sigma^*(L^*_m) \to \sigma > 0$ п.н. $[P]$, $n \to \infty$, $m \to \infty$. По усиленному закону больших чисел $E^*|X_1^*|^3 = n^{-1} \sum_{i=1}^n |X_i|^3 \to E|X_1|^3$, аналогично $E^*(X_1^*)^2 \to EX_1^2$. Таким образом, из (61)–(63) следуют (59)–(60). Теорема доказана.

Замечание 5. Для того, чтобы соотношения (59) и (60) являлись оценками точности бутстреп-аппроксимации, необходимы замена в них нормирующих постоянных $\sigma(L_n)$ и $\sigma^*(L^*_m)$ на σ и σ_n (или σ_n и σ^*_m) соответственно и оценки близости всех этих нормирующих постоянных между собой почти наверное (или по вероятности) требуемого порядка. Получение такого рода оценок предполагается в другой работе.

Литература

1. K. B. Athreya, *Bootstrap of the mean in the infinite variance case.* — Ann. Statist. **15** (1987), 724–731.
2. P. J. Bickel, D. Freedman, *Some asymptotic theory for the bootstrap.* — Ann. Statist. **9** (1981), 1196–1217.
3. P. Deheuvels, D. M. Mason, G. R. Shorack, *Some results on the influence of extremes on the bootstrap.* — Ann. Inst. Henri Poincare **29**, No. **1** (1993), 83–103.
4. R. Helmers, P. Janssen, R. Serfling, *Berry-Essen and bootstrap results for generalized L-statistics.* — Scand. J. Statist. **17**, No. **1** (1990), 65–77.
5. R. Helmers, H. Putter, *Bootstrap resampling: a survey of recent research in the Netherlands.* Amsterdam CWI report, BS-R9519 1995, 10.

6. B. Efron, R. J. Tibshirani, *An introduction to the bootstrap*. Chapman and Hall, New York, 1993.
7. N. V. Gribkova, *On bootstrap L-statistics*. In Abstr. of Commun. Intern. Conf. "Asymptotic Methods in Probability and Mathematical Statistics" – St. Petersburg (1998), 103–106.
8. K. Singh, *On asymptotic accuracy of Efron's bootstrap*. — Ann. Statist. **9** (1981), 1187–1195.
9. S. M. Stigler, *Linear functions of order statistics with smooth weight functions*. — Ann. Statist. **2**, No. **4** (1974), 676–693.
10. W. R. van Zwet, *A Berry-Essen bound for symmetric statistic*. — Z. Wahrsh. verv. Gebiete **66**, No. **3** (1984), 425–440.

Gribkova N. V. Bootstrap approximation of the distributions of L-statistics.

Consistency conditions of the bootstrap approximation of the distributions of order statistics defined by various ways are studied.

С.-Петербургский государственный университет путей сообщения

Поступило 25 марта 1999 г.