Теор. ймовiрност. та матем. статист.
Вип. 54, 1996, стор. 99–108

Teor. Imovirnost. ta Matem. Statyst.
No. 54, 1996, pp. 99–108

КОРРЕЛЯЦІЙНИЙ АНАЛІЗ СУМІШЕЙ. I

УДК 519.21

Р. Є. МАЙБОРОДА

РЕЗЮМЕ. Для спостережень, що є вибіркою із суміші зі змінними концентраціями, побудовано аналоги корреляційних коефіцієнтів Кендала та Спірмена. Доведено їх слушність та асимптотичну нормальність. При доведенні використано асимптотичну теорію неоднорідних емпіричних мір та гладких функціоналів від них.

1. Вступ

Міри залежності (корреляції) випадкових величин широко застосовуються у прикладній статистиці. З непараметричних коефіцієнтів корреляції найстарішим і, мабуть, найпоширенішими є спірменівське ρ ([3, 10]). Якщо $\eta_1, \dots, \eta_N$ — вибірка з незалежних, однаково розподілених в.в. у $\boldsymbol{R}^2$ $(\eta_j = (\eta_j(1), \eta_j(2)))$,

$$\widehat{H}_N(x) = \frac{1}{N}\sum_{j=1}^{N} I\{\eta_j < x\}$$

— однорідна емпірична функція розподілу вибірки η_j, то спірменівський коефіцієнт корреляції між першою та другою координатами η_j можна визначити, як

$$\rho_N(\widehat{H}_N) = 1 - \frac{6N^2}{N^2-1}\widetilde{\rho}(\widehat{H}_N),$$

де для будь-якої ф.р. G на $\boldsymbol{R}^2$

$$\hat{\rho}(G) = \int \big(G(x,\infty) - G(\infty,y)\big)^2 \, G(dx,dy). \tag{1}$$

Інший загальний коефіцієнт — кендалівське τ задається як $\tau_N(\widehat{H}_N) = \frac{N}{N-1}\widetilde{\tau}(\widehat{H}_N)$,

$$\begin{aligned}\widetilde{\tau}(G) = \iint & \operatorname{sign}\big(G(x,\infty) - G(u,\infty)\big) \\ & \times \operatorname{sign}\big(G(\infty,y) - G(\infty,v)\big)\, G(dx,dy)\, G(du,dv).\end{aligned} \tag{2}$$

У даній роботі розглядається випадок, коли спостереження становлять не однорідну вибірку, а суміш кількох компонент, концентрації яких змінюються від спостереження до спостереження [5]. Такі дані часто зустрічаються у медицині, біології, соціології. Як правило, міру залежності в цьому випадку треба будувати для кожної

1991 *AMS Mathematics Subject Classification.* Primary 62G30, 60F17; Secondary 62G20, 60F15.

компоненти суміші окремо. Якщо існує деяка оцінка $\widehat{H}_k^N$ для справжньої ф.р. H_k k-тої компоненти суміші, то як аналог $\rho_N(\widehat{H}_N)$ і $\tau_N(\widehat{H}_N)$ можна використати $\rho_N(\widehat{H}_k^N)$ і $\tau_N(\widehat{H}_k^N)$. При цьому виникає задача дослідження асимптотичної поведінки таких функціоналів при $N \to \infty$.

Ми вивчимо цю поведінку у випадку, коли $\widehat{H}_k^N$ — зважена емпірична функція розподілу, побудована за вибіркою із суміші зі змінними концентраціями. Перше, що треба перевірити, це слушність, тобто збіжність $\rho_N(\widehat{H}_k^N) \to \rho^*(H_k) = 1 - 6\widetilde{\rho}(H_k)$, $\tau_N(\widehat{H}_k^N) \to \widetilde{\tau}(H_k)$ при $N \to \infty$. Потім ми доведемо (за відповідних умов) асимптотичну нормальність ρ: $\sqrt{N}(\rho_N(\widehat{H}_k^N) - \rho^*(H_k)) \Rightarrow \mathcal{N}(0, v)$. Оскільки гранична дисперсія v залежіть від невідомих розподілів компонент H_l, виникає задача оцінки v за вибіркою. Ми покажемо, що слушну оцінку v можна побудувати методами складаного ножа та бутстрепу.

Взагалі кажучи, з некоррельованності за Спірменом ($\rho^*(G) = 0$) або Кендалом ($\widetilde{\tau}(G) = 0$) ще не випливає незалежність координат ($G(x, y) = G(x, \infty)G(\infty, y)$). Можна запропонувати міру залежності, яка не має цієї вади, неприклад,

$$\varkappa(G) = \int \big(G(x,y) - G(x,\infty)G(\infty,y)\big)^2 G(dx,\infty)\, G(\infty,dy). \tag{3}$$

Легко переконатись, що для довільних ф.р. G, $\varkappa(G) = 0$ рівносильно $G(x, y) = G(x, \infty)G(\infty, y)$, $\forall(x, y) \in \boldsymbol{R}^2$. Тому коефіцієнт $\varkappa(\widehat{H}_k^N)$ буде точнішою мірою залежності ніж $\rho_N(\widehat{H}_k^N)$ та $\tau_N(\widehat{H}_k^N)$. Для однорідної вибірки τ та ρ мають ту перевагу над $\varkappa$, що у випадку незалежності $\eta_j(1)$ та $\eta_j(2)$ їх розподіли не залежать від розподілу η_j. Це дозволяє будувати відповідні критерії для перевірки гіпотези незалежності проти альтернативи — некоррельованості за Спірменом або Кендаллом. На жаль, для суміші зі змінними концентраціями $\rho_N(\widehat{H}_k^N)$ та $\tau_N(\widehat{H}_k^N)$ такої властивості не має: їхні розподіли залежать від розподілів усіх компонент суміші. Тому застосування $\varkappa$ більш доцільне, хоча обчислення τ і ρ може бути корисним для порівняння з аналогічними результатами, отриманими за однорідними вибірками.

Для $\varkappa(\widehat{H}_k^N)$ ми виконаємо ту ж програму, що для ρ — доведемо слушність, асимптотичну нормальність, застосованість методів складаного ножа та бутстрепу. Ці результати базуються на загальній теорії неоднорідних емпіричних функцій та методах асимптотичного аналізу функціоналів від емпіричних функцій, які для однорідного випадку розвинуто в роботах [9, 11] та ін.

У даній роботі вміщено основні відомості з теорії неоднорідних емпіричних мір, що будуть потрібні для дослідження асимптотики мір залежності та умови їх слушності і асимптотичної нормальності. У продовженні цієї роботи [6] розглянуто застосованість бутстрепу та складаного ножа.

2. Неоднорідні емпіричні міри

Надалі використовуватимемо трикутні нескінченні масиви вигляду $\underline{a} = (a_j^N, j = 1, \dots, N, N \in \boldsymbol{N})$, де a_j^N — числа, випадкові елементи, d-вимірні вектори (в останньому випадку $a_j^N = (a_j^N(1), \dots, a_j^N(d))$ і $\underline{a}(l) = (a_j^N(l), j = 1, \dots, N, N \in \boldsymbol{N})$), a^N означає $(a_j^N, j = 1, \dots, N)$. Всі операції типу додавання, множення та ін. виконуються над масивами покомпонентно: $\underline{ab} = (a_j^N b_j^N, j = 1, \dots, N, N \in \boldsymbol{N})$. Кутовими дужками з індексом N позначимо усереднення N-того рядка: $\langle \underline{a} \rangle_N = n^{-1} \sum_{j=1}^N a_j^N$ і $\langle \underline{a} \rangle = \lim_{N\to\infty} \langle a \rangle_N$ (якщо така границя існує), $\sup \underline{a} = \sup_{j,N} a_j^N$. Для функцій f, визначених на деякій множині T, позначатимемо $|f|_\infty = \sup_{t \in T} |f(t)|$.

Нехай $\underline{\xi}$ — трикутний масив двовимірних випадкових векторів $\xi_j^N = (\xi_j^N(1), \xi_j^N(2))$, таких, що ξ_j^N незалежні при фіксованому N і для будь-якого $x = (x(1), x(2)) \in \boldsymbol{R}^2$,

$$\mathsf{P}\{\xi_j^N < x\} = \sum_{k=1}^{M} w_j^N(k) H_k^N(x), \tag{4}$$

де $M \in \mathbb{N}$ — фіксоване число, $\underline{w}$ — заданий (відомий) масив дійсних векторів, такий що $w_j^N(k) \geq 0$, $\sum_{k=1}^{M} w_j^N(k) = 1$, H_k^N — (невідома) функція розподілу на $\boldsymbol{R}^2$. Число M можна трактувати як кількість компонент у суміші, $w_j^N(k)$ — як концентрацію k-тої компоненти у суміші під час j-того спостереження, H_k^N — ф.р. k-тої компоненти. Нас цікавитимуть оцінки H_k^N за спостереженнями $\xi^N = (\xi_j^N, j = 1, \ldots, N)$ при фіксованому N. Асимптотика оцінок досліджуватиметься при $N \to \infty$.

Як показано у [4], добрими оцінками неоднорідної е.ф.р.

$$\widehat{H}^N(x, \underline{a}(k)) - \widehat{H}_k^N(x) = \frac{1}{N} \sum_{j=1}^{N} a_j^N(k) I\{\xi_j^N < x\}, \tag{5}$$

де $I\{A\}$ — індикатор події A, $\underline{a}$ — деякий набір чисел.

Приклад 1. Якщо у (5) покласти

$$\underline{a}(k) = \sum_{l=1}^{M} (-1)^{l+k} \gamma_{lk}^N \underline{w}(k) / \det \Gamma_N, \tag{6}$$

де $\Gamma_N = (\langle \underline{w}(l) \underline{w}(k) \rangle_N)_{k,l=1}^M$, γ_{kl}^N — (k,l)-тий мінор Γ_N і $\det \Gamma_N \neq 0$, то $\widehat{H}^N(w, \underline{a}(k))$ буде незсуненою оцінкою H_k^N, ефективною в класі всіх незсунених оцінок.

Приклад 2. Нехай $t_j^N = j/N$, $w_j^N(k) = \widetilde{w}_k(t_j^N)$, де $\widetilde{w}_k \colon [0,1] \to [0,1]$ — фіксовані відомі функції. Вважатимемо, що w_k є інтегровним за Ріманом. Тоді $\langle \underline{w}(l) \underline{w}(k) \rangle = \int_0^1 \widetilde{w}_k(t) \widetilde{w}_l(t)\, dt$. Позначивши $\Gamma_\infty = (\langle \underline{w}(l) \underline{w}(k) \rangle)_{k,l=1}^M$, $\gamma_{k,l=1}^M$ — (k,l)-тий мінор Γ_∞, при $\det \Gamma_\infty \neq 0$ можна покласти

$$\underline{a}(k) = \sum_{l=1}^{M} (-1)^{l+k} \gamma_{lk}^\infty \underline{w}(k) / \det \Gamma_\infty. \tag{7}$$

Позначимо

$$\overline{H}_k^N(x) = \mathsf{E}\, \widehat{H}_k^N(x) = \frac{1}{N} \sum_{j=1}^{N} a_j^N(k) \mathsf{P}\{\xi_j^N < x\} = \sum_{l=1}^{M} \langle \underline{a}(k) \underline{w}(l) \rangle_N\, H_l^N(x),$$

$$Y_N(x) = (Y_N(x,1), \ldots, Y_N(x,p)),$$

де $Y_N(x,j) = \sqrt{N}\big(\widehat{H}_j^N(x) - \overline{H}_j^N(x)\big)$, $x \in \boldsymbol{R}^2$. Наступна теорема характеризує асимптотичну поведінку нормованих емпіричних ф.р. при $N \to \infty$.

Теорема 2.1. *Якщо*

(i) $\sup \underline{a}(l) < A < \infty$, $l = 1, \ldots, p$;

(ii) $\exists\ \langle \underline{w}(k) \underline{a}(l) \underline{a}(m) \rangle$, $\langle \underline{w}(k) \underline{w}(i) \underline{a}(l) \underline{a}(m) \rangle$, $k, i = 1, \ldots, M$, $l, m = 1, \ldots, p$;

(iii) H_k^N — *неперервні ф.р. на* $\boldsymbol{R}^2$, *і існують неперервні на* $\boldsymbol{R}^2$ *ф.р.* $H_1, \ldots, H_M$ *такі, що* $H_k^N(x) \to H_k(x)$ $\forall x \in \boldsymbol{R}^2$, $k = 1, \ldots, M$, $N \to \infty$, *то* Y_N *слабко збігається у просторі* $(D(R^2), |\cdot|_\infty)$ *до неперервного гауссового випадкового поля* $Y(x,l))$ *з нульовим середнім та корреляційною функцією*

$$\mathsf{E}Y(x,l)Y(y,m) = \sum_{k=1}^{M} \langle \underline{w}(k), \underline{a}(l)\underline{a}(m)\rangle H_k(\min(x,y)) - \sum_{k,i=1}^{M} \langle \underline{w}(k)\underline{w}(i)\underline{a}(l)\underline{a}(m)\rangle H_k(x)H_i(y).$$

Зауваження. В умовах прикладу 1 (i) виконано, якщо $\det \Gamma_N > c > 0$ для всіх N, а для виконання (ii) крім цього потрібно існування $\underline{\langle w(k)w(i)a(l)a(m)\rangle}$. В умовах прикладу 2 (i) та (iii) виконано при $\det \Gamma_\infty \neq 0$.

Доведення. Спочатку доведемо компактність розподілів $Y_N(x,l)$ у $(D(R^2), |\cdot|_\infty)$, а потім збіжність скінченномірних розподілів. Компактність досить довести окремо при фіксованому l для кожного $Y_N(x,l)$. Позначимо $\sigma_N: \{1,\dots,N\} \to \{1,\dots,N\}$ таку перестановку індексів, для якої $a_1^N(l) \le a_2^N(l) \le \dots \le a_N^N(l)$,

$$\widetilde{Y}_N(x,t) = N^{-1/2} \sum_{j=1}^{[tN]} I\{\xi^N_{\sigma_N(j)} < x\}.$$

Компактність розподілів $\widetilde{Y}_N$ на $(D(R^2), |\cdot|_\infty)$ доводиться за допомогою критерію Бікела–Вічури [8] так само, як у [5]. Але

$$Y_N(x,l) = \frac{1}{\sqrt{N}} = -\mathsf{E}I\{\xi^N_{\sigma_N(j)} < x\} = \sum_{j=1}^{N} a_{\sigma_N(j)} I\{\xi^N_{\sigma_N(j)} < x\}$$
$$= \sum_{j=1}^{N} a_{\sigma_N(j)} \left(\widetilde{Y}_N\left(x, \frac{j}{N}\right) - \widetilde{Y}_N\left(x, \frac{j-1}{N}\right)\right)$$
$$= a^N_{\sigma_N(N)} \widetilde{Y}_N(x,1) - \sum_{j=2}^{N} \widetilde{Y}_N\left(x, \frac{j}{N}\right)\left(a_{\sigma_N(j)} - a^N_{\sigma_N(j-1)}\right).$$

Отже

$$|Y_N(x,l) - Y_N(y,l)| \le \sup_t \left|\widetilde{Y}_N(x,t) - \widetilde{Y}_N(y,t)\right| \left(a_{\sigma_N(N)} + a^N_{\sigma_N(N)} - a^N_{\sigma_N(1)}\right)$$

(внаслідок монотонності $a^N_{\sigma_N(j)}$). З умови (i) отримуємо

$$|Y_N(x,l) - Y_N(y,l)| \le 3A \sup_t \left|\widetilde{Y}_N(x,t) - \widetilde{Y}_N(y,t)\right|,$$

отже з компактності розподілів $\widetilde{Y}_N(x,t)$ випливає компактність $Y_N(x,l)$.

Збіжність скінченномірних розподілів випливає з того факту, що $\mathsf{E}Y_N(x,l) = \mathsf{E}Y(x,l) = 0$,

$$\mathsf{E}Y_N(x,l)Y_N(y,m) = \sum_{k=1}^{M} \langle \underline{w}(k), \underline{a}(l)\underline{a}(m)\rangle H_k^N(\min(x,y)) - \sum_{k,i=1}^{M} \langle \underline{w}(k)\underline{w}(i)\underline{a}(l)\underline{a}(m)\rangle_N H_k^N(x)H_i^N(y)$$

і, за умови (ii), (iii), $\mathsf{E}\, Y_N(x,l)Y_N(y,m) \to \mathsf{E}\, Y(x,l)Y(y,m)$, $N \to \infty$, причому $Y_N(x,l)$ є нормованими сумами незалежних рівномірно обмежених випадкових величин.

Теорему доведено. $\square$

Позначення $X \stackrel{\mathrm{d}}{=} Y$ надалі означатиме, що X та Y мають однакові розподіли.

Наслідок 2.1. *Якщо виконуються умови теореми 2.1, то, розширивши при необхідності основний імовірнісний простір, можна задати послідовність випадкових процесів $Y^*(x,l)$, $Y_N^*(x,l)$ таких, що $Y^*(\cdot,\cdot) \stackrel{\mathrm{d}}{=} Y(\cdot,\cdot)$, $Y_N^*(\cdot,\cdot) \stackrel{\mathrm{d}}{=} Y_N(\cdot,\cdot)$ і $|Y_N^*(\cdot,\cdot) - Y^*(\cdot,\cdot)|_\infty \to 0$ м.н.*

Це твердження безпосередньо випливає з теореми 2.1 та теореми Скорохода (див. [9]).

Лема 2.1. *Нехай $\underline{x} = (x_j^N)$ — масив випадкових величин, незалежних при фіксованому N, таких, що для деякого $\delta > 0$ $\sup_{j,N} \mathsf{E}\,|x_j^N|^{2+\delta} < \infty$, $\mathsf{E}\, x_j^N = 0$. Тоді*

$$A_N = \frac{1}{N}\sum_{j=1}^N x_j^N \to 0, \qquad N \to \infty \quad \text{м.н.}$$

Доведення. За теоремою 19 розділу 3 [7], для деякої константи $C > 0$,

$$\begin{aligned}\mathsf{E}\,|S_N|^{2+\delta} &\le \frac{C}{N^{2+\delta}}\left(\sum_{j=1}^N \mathsf{E}\,|x_j^N|^{2+\delta} + \left(\sum_{j=1}^N \mathsf{E}(x_j^N)^2\right)^{(2+\delta)/2}\right) \\ &\le \frac{C}{N^{2+\delta}}(N + N^{1+\delta/2}) \le \frac{C}{N^{1+\delta/2}}.\end{aligned}$$

Отже, за нерівністю Чебишова, для будь-якого $\lambda > 0$,

$$\mathsf{P}\{|S_N| > \lambda\} \le C\lambda^{-2-\delta}N^{-1-\delta/2},$$

і $\sum_{N=1}^\infty \mathsf{P}\{|S_N| > \lambda\} < \infty$ звідси й випливає твердження леми. $\square$

Позначимо ${}^+a_j^N(l) = a_j^N(l)I\{a_j^N(l) > 0\}$, ${}^-a_j^N(l) = a_j^N(l)I\{a_j^N(l) < 0\}$.

Теорема 2.2. *Якщо*

(i) $\sup|\underline{a}| < A < \infty$,
(ii) $H_i^N = H_i$, *не залежіть від* N,
(iii) *існують* $\langle {}^+\underline{aw}(k)\rangle < \infty$, $\langle {}^-\underline{aw}(k)\rangle < \infty$,

то $|\widehat{H}^N(\cdot,a) - \overline{H}(\cdot,a)|_\infty \to 0$ *м.н.,* $N \to \infty$, *де* $\overline{H}(\cdot,a) = \sum_{k=1}^M \langle \underline{aw}(k)\rangle\, H_k(x)$.

Зауваження. Для прикладу 2 умови (i) та (iii) виконуються при $\det\Gamma_\infty \neq 0$

Доведення. Ми доведемо, що $|\widehat{H}^N(\cdot,{}^+\underline{a}) - \overline{H}(\cdot,{}^+\underline{a})|_\infty \to 0$ м.н. (збіжність $|\widehat{H}^N(\cdot,{}^-\underline{a}) - \overline{H}(\cdot,{}^-\underline{a})|_\infty \to 0$ доводиться аналогічно). Позначимо

$$\overline{H}_N(\cdot,{}^+\underline{a}) = \sum_{k=1}^M \left\langle {}^+\underline{a}w(k)\right\rangle H_k^N(x).$$

Тоді $|\overline{H}_N(\cdot,{}^+\underline{a}) - \overline{H}(\cdot,{}^+\underline{a})|_\infty \to 0$ за умовою (iii), $\mathsf{E}\,\widehat{H}^N(\cdot,{}^+\underline{a}) = \overline{H}_N(\cdot,{}^+\underline{a})$. Позначимо

$$S_N(x) = \widehat{H}_k^N(x) - \overline{H}_N(x,{}^+\underline{a}) = \frac{1}{N}\sum_{j=1}^N x_j^N,$$

де $x_j^N = {}^+a_j^N(I\{\xi_j^N < x)\} - \mathsf{P}\{\xi_j^N < x\})$. Збіжність $S_N(x) \to 0$ м.н. при фіксованих x випливає з леми 2.1, та умови (i), перехід до рівномірної збіжності $|S_N|_\infty \to 0$ м.н.

проводиться так само, як у звичайній теоремі Глівенко–Кантеллі (див. теорему 1 додатку 1 [1]), оскільки ${}^+\underline{a} > 0$ і, отже, $\widehat{H}_k^N$ та $\overline{H}_N(\cdot, {}^+\underline{a})$ є ф.р. деяких скінченних мір на $\boldsymbol{R}^2$.

Теорему доведено. □

3. Слушність та асимптотична нормальність

Нехай $\mathcal{M}$ — клас усіх імовірнісних мір на $\boldsymbol{R}^2$, які мають неперервну ф.р., $\mathcal{M}^*$ — клас ф.р. мір з $\mathcal{M}$. Для $A \in \boldsymbol{R}$, $Z_A(\boldsymbol{R}^d)$ — клас усіх зарядів на $\boldsymbol{R}^d$, варіація яких не перевищує A, $Z_A^*(\boldsymbol{R}^d)$ — клас ф.р. зарядів з $Z_A^*(\boldsymbol{R}^d)$, $Z_A^* = Z_A^*(\boldsymbol{R}^2)$. Всюди в цьому параграфі вважаємо, що розподіл ξ_j^N задано (4) з розподілами компонент $H_k^N = H_k$, що не залежать від N. Ми будемо розглядати функціонали $\varphi\colon Z_A^* \to \boldsymbol{R}$ та вивчати умови, коли $\varphi(\widehat{H}_k^N) \to \varphi(H_k)$ та $\sqrt{N}(\varphi(\widehat{H}_k^N) - \varphi(H_k)) \Rightarrow \mathcal{N}(0, v)$.

Теорема 3.1. *Якщо виконано умови теореми* 2.2,

$$|\overline{H}(\cdot, \underline{a}) - H_k(\cdot)|_\infty \to 0, \tag{8}$$

і φ — неперервне у точці H_k відображення $(Z_A^*, |\cdot|_\infty) \to \boldsymbol{R}$, то $\varphi(\widehat{H}_k^N) \to \varphi(H_k)$ м.н.

Зауваження. Умова (8) означає, що $\widehat{H}_k^N$ є асимптотично незсуненою оцінкою H_k. Вона виконується у прикладі 1, а у прикладі 2 випливає з інтегровності за Ріманом $\widetilde{w}_k$ та умови $\det \Gamma_\infty \neq 0$.

Доведення. З теореми 2.2 випливає, що $|\widehat{H}_k^N - H_k|_\infty \to 0$ м.н., отже, з неперервності φ одержуємо твердження теореми. □

Лема 3.1. *Функціонали $\widetilde{\rho}$, $\varkappa$, $\widetilde{\tau}$ є неперервними у точках з $\mathcal{M}^*$.*

Доведення. Почнемо з $\widetilde{\rho}$. Розкриваючи дужки та двічі інтегруючи частинами, отримуємо (для неперервних H),

$$\widetilde{\rho}(H) = \frac{2}{3} + 2\int H(x,y) H(dx, \infty)\, H(\infty, dy), \tag{9}$$

отже достатньо довести, що для довільних $H_n \in Z_a^*$, таких, що $|H_n - H|_\infty \to 0$,

$$J_n = \int H_n(x,y)\, H_n(dx,\infty)\, H_n(\infty, dy) - \int H(x,y)\, H(dx,\infty)\, H(\infty,dy) \to 0.$$

Зауважимо, що $J_n = J_n^1 + J_n^2$, де

$$J_n^1 = \int H(x,y)\big(H_n(dx,\infty)H_n(\infty,dy) - H(dx,\infty)H(\infty,dy)\big) \to 0$$

за першою теоремою Хеллі (теорема 5 у [2], с. 215), а

$$J_n^2 = \int \big(H_n(x,y) - H(x,y)\big) H_n(dx,\infty) H_n(\infty,dy) \le A^2 |H_n - H|_\infty \to 0.$$

Отже $\widetilde{\rho}$ — неперервний функціонал.

Аналогічно доводиться неперервність $\varkappa$.

Доведемо неперервність $\widetilde{\tau}$. Зауважимо, що коли $H \in \mathcal{M}^*$, то для довільного $\delta > 0$ знайдуться $\varepsilon_\delta > 0$ і множина $K_\delta = \bigcup_{j=1}^m B_j \subset \boldsymbol{R}^4$ такі, що B_j — (можливо необмежені) прямокутники у $\boldsymbol{R}^4$, $m < \infty$, $H \oplus H(K_\delta) < \delta$ (тут і далі $H_1 \oplus H_2$ — декартів добуток мір з ф.р. H_1 та H_2) і для всіх $(x,y,u,v) \in \boldsymbol{R}^4 \setminus K_\delta$, $|H(x,\infty) - H(u,\infty)| > \varepsilon_\delta$, $|H(\infty,y) - H(\infty,v)| > \varepsilon_\delta$. Дійсно, легко бачити, що $H(\{(x,u)\colon |H(x,\infty) - H(u,\infty)| <$

$\varepsilon\}) = \mathsf{P}\{|\zeta_1 - \zeta_2| < \varepsilon\} \to 0$, $\varepsilon \to 0$ (ζ_j — незалежні в.в., рівномірно розподілені на $[0,1]$). Отже, обираючи ε_δ^1 достатньо малим, отримуємо

$$H \oplus H(\{(x,y,u,v): |H(x,\infty) - H(u,\infty)| < \varepsilon_\delta^1\}) < \delta/4.$$

Аналогічно можна обрати ε_δ^2 так, щоб

$$H \oplus H(\{(x,y,u,v): |H(\infty,y) - H(\infty,v)| < \varepsilon_\delta^2\}) < \delta/4$$

і апроксимувати об'єднання цих двох множин скінченним набором прямокутників K_δ так, щоб $H \oplus H(K_\delta) < \delta$. При цьому можна покласти $\varepsilon_\delta = \min(\varepsilon_\delta^1, \varepsilon_\delta^2)$. Позначимо

$$J(H,S) = \int \operatorname{sign}(H(x,\infty) - H(u,\infty)) \operatorname{sign}(H(\infty,y) - H(\infty,v)) H(dx,dy) H(du,dv).$$

Фіксуємо довільне $\delta > 0$. Нехай $|H_n - H|_\infty$. Маємо $|\widetilde{\tau}(H_n) - \widetilde{\tau}(H)| < J_1 + J_2$, де $J_1 = |J(H_n,K_\delta) - J(H,K_\delta)| \le H_n \oplus H_n(K_\delta) + H \oplus H(K_\delta) \le 2\delta + o(1)$ при $n \to \infty$, оскільки $H_n \oplus H_n(B_j) \to H \oplus H(B_j)$ на кожному B_j. Далі $J_2 = |J(H_n, \boldsymbol{R}^4 \setminus K_\delta) - J(H, \boldsymbol{R}^4 \setminus K_\delta)$. При достатньо великих n, $|H_n(x,\infty) - H(x,\infty)|_\infty < \varepsilon_\delta/3$, отже $\operatorname{sign}(H_n(x,\infty) - H_n(u,\infty)) = \operatorname{sign}(H(x,\infty) - H(u,\infty))$ на $\boldsymbol{R}^4 \setminus K_\delta$. Тому при великих n, $J_2 = |H_n \oplus H_n(\boldsymbol{R}^4 \setminus K_\delta) - H \oplus H(\boldsymbol{R}^4 \setminus K_\delta)| \to 0$. Внаслідок довільності δ, маємо $\widetilde{\tau}(H_n) \to \widetilde{\tau}(H)$, $n \to \infty$.

Лему доведено. □

Наслідок 3.1. *Якщо виконані умови теореми* 2.2, (8), *і* $H_l \in \mathcal{M}$, $l = 1, \ldots, M$, *то* $\tau_N(\widehat{H}_k^N) \to \widetilde{\tau}(H_k)$, $\rho_N(\widehat{H}_k^N) \to 1 - 6\widetilde{\rho}(H_k)$, $\varkappa(\widehat{H}_k^N) \to \varkappa(H_k)$, $N \to \infty$ *м.н.*

Перейдемо до вивчення асимптотичної нормальності. Для цього нам знадобиться поняття неперервної диференційовності функціоналу за Адамаром.

Означення. Функціонал $\varphi: Z_A^* \to \boldsymbol{R}$ назвемо неперервно диференційовним за Адамаром у точці $H \in Z_A^*$ на парі $(\mathcal{M}_1, \mathcal{M}_2)$, $\mathcal{M}_j \subseteq Z_A^*$, якщо існує такий лінійний функціонал $D_H: Z_A^* \to \boldsymbol{R}$, що для довільних $t_n \to 0$, $h, h_n \in \mathcal{M}_2$, $H_n \in \mathcal{M}_1$ таких, що $|H_n - H|_\infty \to 0$, $|h_n - h|_\infty \to 0$ справедливо

$$\left| \frac{1}{t_n}\big(\varphi(H_n + t_n h_n) - \varphi(H_n)\big) - D_H h \right|_\infty \to 0, \qquad n \to \infty.$$

Функціонал D_H зветься похідною (за Адамаром) від φ в точці H.

Нехай задано деякий функціонал $\varphi: Z_A^* \to \boldsymbol{R}$ з похідною D_H. Позначимо η^k, η_j^k — незалежні випадкові величини з ф.р. H_k, $\delta_j^k = I\{\eta_j^k < x\}$, $\delta^k = I\{\eta^k < x\}$, $D = D_{H_k}$, $\overline{\delta}^l = D(\delta^l - H_l(\cdot))$,

$$v^\varphi = v_k = \sum_{l=1}^{M} \langle \underline{a}(k)\underline{a}(k)\underline{w}(l)\rangle\, \mathsf{E}(\overline{\delta}^{\,l})^2 - \sum_{m,l=1}^{M} \langle \underline{a}(k)\underline{a}(k)\underline{w}(m)\underline{w}(l)\rangle\, DH_m DH_l. \qquad (10)$$

Теорема 3.2. *Нехай виконуються умови* (i)–(iii) *теореми* 2.1, H_k *неперервні*,

$$|\overline{H}_N^k - H_k|_\infty = o(N^{-1/2}), \qquad (11)$$

φ — *неперервно диференційовний за Адамаром у точці* H_k *на парі* $(Z_A^*, C(\boldsymbol{R}^2))$, *його похідна неперервна на* $(D(\boldsymbol{R}^2), |\cdot|_\infty)$ *і існує таке* $\alpha > 0$, *що для всіх* $l = 1, \ldots, M$,

$$\mathsf{E}(\overline{\delta}^{\,l})^{2+\alpha} < \infty. \qquad (12)$$

Тоді $J_N = \sqrt{N}(\varphi(\widehat{H}_k^N) - \varphi(H_k))$ *слабко збігається до гауссового розподілу з нульовим середнім і дисперсією* v_k, *визначеною* (10).

Зауваження. Умова (11) виконується в прикладі 2, якщо $\det \Gamma \neq 0$ і w_k є функціями з обмеженою варіацією.

Доведення. Покладемо $t_N = N^{-1/2}$. Скориставшись наслідком 2.1, маємо

$$J_N = t_N^{-1}\left(\varphi(H_k + t_N Y_N(\cdot,k)) - \varphi(H_k)\right) \overset{\mathrm{d}}{=} t_N^{-1}\left(\varphi(H_k + t_N Y_N^*(\cdot,k)) - \varphi(H_k)\right) = J_N^*$$

і $|Y_N^* - Y|_\infty \to 0$, $N \to \infty$. Використовуючи (11), потраєкторну неперервність Y, та диференційовність φ за Адамаром, маємо $|Y_N^* - DY| \to 0$ м.н., звідки й випливає $J_N \Rightarrow DY$.

Підрахуємо $\widetilde{v} = \mathsf{E}(DY)^2$. За теоремою 2.1,

$$\widetilde{v} = \mathsf{E}\Big(\Big(D \lim_{N\to\infty} Y_N\Big)\Big)^2 = \mathsf{E}\Big(\Big(\lim_{N\to\infty} DY_N\Big)\Big)^2 = \lim_{N\to\infty} \mathsf{E}(DY_N)^2,$$

де перша рівність випливає з неперервності оператора D, а друга з умови (12) та центральної граничної теореми для величин $\lambda_j^N D(I\{\xi_j^N < \cdot\} - \mathsf{P}\{\xi_j^N < \cdot\})$. Отже

$$\widetilde{v} = \lim_{N\to\infty} \frac{1}{N^2} \sum_{i,j=1}^{N} a_j^N(k) a_i^N(k)\, \mathsf{E}\, \lambda_j^N \lambda_i^N$$

$$= \lim_{N\to\infty} \sum_{l=1}^{M} \langle \underline{a}(k)\underline{a}(k)\underline{w}(l)\rangle_N\, \mathsf{E}(\overline{\delta}^{\,l})^2 - \sum_{i,j=1}^{M} \langle \underline{a}(k)\underline{a}(k)\underline{w}(l)\underline{w}(m)\rangle_N\, \mathsf{E}\,\overline{\delta}^{\,l}\, \mathsf{E}\,\overline{\delta}^{\,m} = v.$$

Теорему доведено. □

Щоб використати цю теорему для перевірки асимптотичної нормальності ρ_N і τ_N нам знадобляться дві леми.

Для $E, E_n' \in D(\boldsymbol{R}^2)$, $B, B_n', C, C_n' \in D(\boldsymbol{R})$ позначимо $U = (E, B, C)$ і, відповідно, $U_n' = (E_n', B_n', C_n')$ і т.д., $|U|_\infty = |E|_\infty + |B|_\infty + |C|_\infty$,

$$\Psi(U) = \int E(x,y) B(dx) C(dy), \tag{13}$$

$\mathcal{U}$ — клас U, таких, що $E \in Z_A^*$, $B, C \in Z_A^*(\boldsymbol{R})$, $\mathcal{U}'$ — клас U, таких, що E, B, C — неперервні функції.

Лема 3.2. *Якщо* $U \in \mathcal{U} \cap \mathcal{U}'$, *то* Ψ *є неперервно диференційовним за Адамаром у точці* U *на парі* $(\mathcal{U}, \mathcal{U}')$.

Доведення. Ми доведемо, що для будь-яких трійок $U_n, U_n' \in \mathcal{U}$, $u = (e, b, c) \in \mathcal{U}'$, $u_n = (e_n, b_n, c_n)$ таких, що $U_n \to U$, $U_n' = U_n + t_n u_n$, $t_n \to 0$, $u_n \to u$

$$J^n = \frac{1}{t_n}\left|\Psi(U_n') - \Psi(U_n) - t_n D_U u\right| \to 0,$$

де

$$D_U u = \int e(x,y) B(dx) C(dy) + \int E(x,y) b(dx) C(dy) + \int E(x,y) B(dx) c(dy) \tag{14}$$

(це і означає неперервну диференційовність за Адамаром з похідною D_U).

Зауважимо, що

$$\frac{1}{t_n}(\Psi(U_n') - \Psi(U_n)) = \int e_n(x,y)B_n'(dx)C_n'(dy) + \int E_n(x,y)b_n(dx)C_n'(dy) + \int E_n(x,y)B_n(dx)c_n(dy),$$

отже $J^n \le J_1^n + J_2^n + J_3^n$, де

$$J_1^n = \left|\int E_n(x,y)B_n'(dx)C_n'(dy) - \int e(x,y)B(dx)C(dy)\right|,$$

$$J_2^n = \left|\int E_n(x,y)b_n(dx)C_n'(dy) - \int E(x,y)b(dx)C(dy)\right|,$$

$$J_3^n = \left|\int E_n(x,y)B_n(dx)c_n(dy) - \int E(x,y)B(dx)c(dy)\right|.$$

Оцінимо $J_1^n \le J_{11}^n + J_{12}^n$, де

$$J_{11}^n = \left|\int (e_n(x,y) - e(x,y))B_n'(dx)C_n'(dy)\right| \le |e_n - e|_\infty \operatorname{Var} B_n' \oplus C_n' \le A^2|e_n - e|_\infty \to 0, \qquad n \to \infty,$$

$$J_{12}^n = \left|\int e(x,y)B_n'(dx)C_n'(dy) - \int e(x,y)B(dx)C(dy)\right|.$$

Оскільки $|B_n' - B|_\infty \to 0$, $|C_n' - C|_\infty \to 0$, то ф.р. $B_n' \oplus C_n'$ збігається до ф.р. $B \oplus C$ в усякому випадку поточково. Тому з неперервності $e(x,y)$ за першою теоремою Хеллі маємо $J_{12}^n \to 0$. Отже, $J_1^n \to 0$ при $n \to 0$.

Оцінимо $J_2^n \le J_{21}^n + J_{22}^n + J_{23}^n$, де

$$J_{21}^n = \left|\int E_n(x,y)b_n(dx)C_n'(dy) - \int E_n(x,y)b^*(dx)C_n'(dy)\right|,$$

$$J_{22}^n = \left|\int E_n(x,y)b^*(dx)C_n'(dy) - \int E(x,y)b^*(dx)C(dy)\right|,$$

$$J_{23}^n = \left|\int E(x,y)b^*(dx)C(dy) - \int E(x,y)b(dx)C(dy)\right|,$$

де b^* — деяка функція обмеженої варіації, що буде обрана пізніше.

Позначимо $\psi_n(x) = \int E_n(x,y)C_n'(dy)$, $\psi(x) = \int E(x,y)C(dy)$. Легко бачити, що $\operatorname{Var}\psi_n \le KA^2$, $\operatorname{Var}\psi \le KA^2$. (Тут і далі K — деякі, можливо різні, константи.)

Використовуючи стандартні оцінки, маємо: $J_{21} \le |b_n - b^*|_\infty \operatorname{Var}\psi_n \le K|b - b^*|_\infty + |b - b_n|_\infty$, $J_{22} \le |\psi_n - \psi|_\infty \operatorname{Var} b^*$, $J_{23} \le |b - b^*|_\infty \operatorname{Var}\psi \le K|b - b^*|_\infty$. Оскільки $|\psi_n - \psi|_\infty \le |E_n' - E|_\infty \operatorname{Var} C_n' + |C - C_n'|_\infty \sup_x \operatorname{Var} E(x,\cdot) \to 0$, $n \to \infty$, то для довільного $\delta > 0$, вибираючи спочатку b^* так, щоб $|b - b^*|_\infty < \delta$, а потім n_0 так, щоб $|b - b_n|_\infty < \delta$, $|\psi_n - \psi|_\infty < \delta$ для всіх $n > n_0$, отримуємо $J_2^n \le K(3 + \operatorname{Var} b^*)\delta$. Отже $J_2^n \to 0$, $n \to \infty$.

Аналогічно доводиться, що $J_3^n \to 0$, $n \to 0$.

Лему доведено. □

Лема 3.3. *Якщо H — неперервні ф.р., то $\widetilde{\rho}$ та ж неперервно диференційовні за Адамаром у точці H на парі $(Z_A^*, C(\boldsymbol{R}^2))$.*

Доведення. Згідно з (9) для диференційовності τ і ρ достатньо, щоб був диференційовним функціонал $\overline{\rho}(H) = \Psi(H, H(\cdot,\infty), H(\infty,\cdot))$. Оскільки для диференціювання за Адамаром справедливе ланцюжкове правило, то диференційовність ρ випливає з леми 3.2.

Для $\varkappa$ маємо

$$\varkappa(H)=\Psi(H^2,H(\cdot,\infty),H(\infty,\cdot))+\frac{1}{2}\Psi(H,H^2(\cdot,\infty),H^2(\infty,\cdot)) + \int H^2(x,\infty)H(dx,\infty)\int H^2(\infty,y)H(dy,\infty).$$

Для того, щоб застосувати ланцюжкове правило, досить зауважити, що функція $x\to x^2$ є диференційовною, а функції $H^2(x,y)$, $H^2(x,\infty)$, $H^2(\infty,y)$ є неперервними ф.р. $\square$

Наслідок 3.2. *Якщо виконуються умови* (i)–(ii) *теореми* 2.1 *та* (11), H_l, $l=1,\dots,M$ — *неперервні ф.р., то*

$$\sqrt{N}\big(\widetilde{\rho}(\widehat{H}_a^N)-\widetilde{\rho}(H_k)\big)\Rightarrow\mathcal{N}(0,v_k^\rho),$$
$$\sqrt{N}\big(\varkappa(\widehat{H}_a^N)-\varkappa(H_k)\big)\Rightarrow\mathcal{N}(0,v_k^\varkappa),$$

де v_k^ρ *і* $v_k^\varkappa$ *обчислюються за* (10).

Доведення. Для перевірки умов теореми 3.2 можна скористатись лемою 3.3. Неперервність похідних $\widetilde{\rho}$ і $\varkappa$ та (12) перевіряється за допомогою (14). $\square$

Література

1. А. А. Боровков, *Математическая статистика*, "Наука", Москва, 1984.
2. И. И. Гихман, А. В. Скороход, М. И. Ядренко, *Теория вероятностей и математическая статистика*, "Выща школа", Киев, 1979.
3. М. Кендэл, *Ранговые корреляции*, "Статистика", Москва, 1975.
4. Р. Є. Майборода, *Непараметрична статистика неоднорідних спостережень*, Дисертація на здобуття вченого ступеня доктора фіз.-мат. наук (1994), Київ.
5. Р. Е. Майборода, *Некоторые задачи статистики смесей с переменными концентрациями*, Теор. вероятност. и применен. **39** (1995), № 4, 840–845.
6. Р. Є. Майборода, *Кореляційний аналіз сумішей*. II, Теор. ймовірност. та матем. статист. (1995) (подано до друку).
7. В. В. Петров, *Предельные теоремы для сумм независимых случайных величин*, "Наука", Москва, 1987.
8. P. J. Bickel and M. J. Wichura, *Convergence criteria for multiparameter stochastic processes and some applications*, Ann. Math. Statist. **42** (1971), № 5, 1656–1670.
9. R. D. Gill, *Lectures on survival analysis* (*Part I*), Preprint, University of Utrecht, Dept. of Math. (1993), № 764, 48.
10. C. Spearman, *The proof and measurement of assjciacion between two things*, Am. J. Psych **15** (1904), 72–101.
11. J. Shao, *Differentiability of statistical functions and consistency of the jackknife*, Ann. Statist. **21** (1993), № 1, 61–75.

252127, КИЇВ-127, ВУЛ. АКАДЕМІКА ГЛУШКОВА, 6, НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМ. Т. Г. ШЕВЧЕНКА, МЕХАНІКО-МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ, КАФЕДРА ТЕОРІЇ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ

Надійшла 12.02.95