УДК 519.22

Р. Є. Майборода, О. В. Сугакова (Київ. нац. ун-т ім. Т. Шевченка)

ОЦІНКА ЕВКЛІДОВИХ ПАРАМЕТРІВ СУМІШІ ДВОХ СИМЕТРИЧНИХ РОЗПОДІЛІВ*

A sample from a mixture of two symmetric distributions is observed. The considered distributions differ by a shift only. Estimates are constructed by the method of estimating equations of parameters of mean locations and concentrations (mixing probabilities) of both components. Conditions of asymptotic normality of these estimates are obtained. Greatest lower bounds for coefficients of dispersion of the estimates are obtained.

Наблюдается выборка со смеси двух симметричных распределений, отличающихся только смещением. Построены оценки метода оценивающих уравнений параметров средних положений и концентраций (вероятностей смешивания) обеих компонент. Получены условия асимптотической нормальности этих оценок. Найдена точная нижняя грань для коэффициентов рассеяния оценок.

1. Вступ. Задача оцінки параметрів моделі суміші двох симетричних розподілів, що відрізняються лише зсувом, вивчалась у роботах [1, 2]. У цих роботах доведено ідентифіковність моделі та побудовано консистентні оцінки параметрів. У роботі [3] запропоновано моментні оцінки для евклідових параметрів і доведено їх асимптотичну нормальність. У даній роботі розглянуто узагальнення моментних оцінок з використанням методу оцінюючих рівнянь. (Загальну теорію оцінюючих рівнянь описано у багатьох підручниках, див. [4].) У п. 2 наведено загальну схему побудови незміщених оцінюючих рівнянь для евклідових параметрів суміші двох симетричних розподілів, у п. 3 — умови асимптотичної нормальності отриманих оцінок та знайдено їх матрицю розсіювання (коваріаційну матрицю граничного нормального розподілу). Оскільки для різних оцінюючих функцій коефіцієнти розсіювання оцінок можуть бути різними, виникає питання про точну нижню межу для цих коефіцієнтів. Цю межу знайдено у п. 4. Питання підрахунку отриманих характеристик та числовий приклад розглянуто у п. 5. Складні доведення вміщено у додатку.

2. Побудова оцінюючих рівнянь. Нехай спостерігається вибірка з незалежних однаково розподілених (н. о. р.) випадкових величин $\xi_1, \ldots, \xi_n$, які мають щільність розподілу

$$\psi(x) = pf(x-a_1) + (1-p)f(x-a_2), \tag{1}$$

де p — ймовірність спостерігати першу компоненту суміші, $0 < p < 1/2$, $a_i \in \mathbb{R}$, $i = 1, 2$ — медіана i-ї компоненти суміші, f — симетрична щільність розподілу відхилення спостереження від медіани (однакова для обох компонент).

Модель (1) можна записати в іншій формі:

$$\xi_j = a_{2-\delta_j} + \eta_j, \tag{2}$$

де δ_j — індикатор того, що спостережуваний об'єкт належить першій компоненті, $\mathrm{P}\{\delta_j = 1\} = p$, $\mathrm{P}\{\delta_j = 0\} = 1 - p$, η_j — н. о. р. випадкові величини з щільністю f.

У цій моделі $\vartheta = (p, a_1, a_2)^T$ є невідомим евклідовим параметром моделі, а симетрична (парна) щільність f — невідомою непараметричною складовою. Таким чином, задача оцінювання ϑ є семіпараметричною.

Для того щоб побудувати оцінюючі рівняння (generalized estimating equations, GEE) для ϑ, розглянемо довільну непарну функцію g таку, що для всіх $\alpha \in \mathbb{R}$ $\mathrm{E}g(\eta_j + \alpha) < \infty$. Тоді

*Виконано за часткової підтримки Swedish Institute grant SI-01424/2007.

$$\mathrm{E}g(\xi_j - a_1) = p\mathrm{E}g(\eta_j) + (1-p)\mathrm{E}g(\eta_j - a_1 + a_2) =$$

$$= (1-p)\int\limits_{-\infty}^{+\infty} g(x - (a_1 - a_2))f(x)dx = -(1-p)\int\limits_{-\infty}^{+\infty} g(x + (a_1 - a_2))f(x)dx$$

та, аналогічно,

$$\mathrm{E}g(\xi_j - a_2) = p\int\limits_{-\infty}^{+\infty} g(x + (a_1 - a_2))f(x)dx.$$

Тому

$$\mathrm{E}[pg(\xi_j - a_1) + (1-p)g(\xi_j - a_2)] = 0. \qquad (3)$$

Щоб використати цю рівність для побудови системи незміщених оцінюючих рівнянь, виберемо три непарні функції g_1, g_2, g_3 і позначимо $\hat{g}_i(\alpha) = \dfrac{1}{n}\sum_{j=1}^{n} g_i(\xi_j - \alpha)$. Тоді система рівнянь відносно невідомих π, α_1, α_2,

$$\pi\hat{g}_i(\alpha_1) + (1-\pi)\hat{g}_i(\alpha_2) = 0, \quad i = 1, 2, 3, \qquad (4)$$

буде незміщеним GEE для параметра ϑ. Статистику $\hat{\vartheta}_n = (\hat{p}_n, \hat{a}_{1,n}, \hat{a}_{2,n})^T$ будемо називати оцінкою евклідового параметра ϑ з оцінюючою трійкою (g_1, g_2, g_3), якщо (3) виконується майже напевно (м. н.) при підстановці $\pi = \hat{p}_n, \alpha_1 = \hat{a}_{1,n}, \alpha_2 = \hat{a}_{2,n}$.

Будемо також використовувати уніфікований векторний запис

$$\hat{\vartheta}_n = \left(\hat{\vartheta}_{1,n}, \hat{\vartheta}_{2,n}, \hat{\vartheta}_{3,n}\right)^T$$

та векторну оцінюючу функцію $h(x,t) = (h_1(x,t), h_2(x,t), h_3(x,t))^T$, $t = (\pi, \alpha_1, \alpha_2)$, $h_i(t) = \pi g_i(x - \alpha_1) + (1-\pi)g_i(x - \alpha_2)$. У цих позначеннях (4) еквівалентно

$$\sum_{j=1}^{n} h(\xi_j, t) = 0. \qquad (5)$$

***Приклад* 1.** Поклавши у (4) $g_i(x) = g_i^{\mu}(x) = x^{2i-1}$, отримаємо GEE-оцінки $\hat{\vartheta}_n^{\mu}$, які дорівнюють моментним оцінкам, описаним у роботі [3]. У цій роботі доведено консистентність та асимптотичну нормальність моментних оцінок.

Дослідження консистентності GEE-оцінок $\hat{\vartheta}_n$ у загальному випадку вимагає перевірки умови контрасту, яка полягає в тому, що справжнє значення $\vartheta = (p, a_1, a_2)^T$ має бути єдиним коренем рівняння

$$\mathrm{E}_{\vartheta} h(\xi_j, t) = 0.$$

Легко бачити, що для багатьох оцінюючих трійок ця умова не виконується. Відповідно у рівняння (5) у таких ситуаціях можуть виникати „зайві” корені, що відповідають неконсистентним оцінкам ϑ. Для того щоб за допомогою таких оцінюючих трійок будувати консистентні оцінки, потрібен додатковий алгоритм вибору „правильного” кореня (5). Наприклад, серед усіх можливих розв'язків оцінюючого рівняння можна вибрати той, який є найближчим (у звичайній евклідовій нормі) до консиситентної моментної оцінки $\hat{\vartheta}_n^{\mu}$ (або до консистентних оцінок, описаних у

[1, 2]). Така схема дозволяє будувати консистентні GEE-оцінки для досить широкого класу оцінюючих трійок (g_1, g_2, g_3).

3. Асимптотична нормальність GEE-оцінок. Для дослідження асимптотичного розподілу нормованих GEE-оцінок введемо наступні позначення:

$$b_1(x) = p^2 f'(x) + p(1-p)f'(x - a_2 + a_1),$$

$$b_2(x) = p(1-p)f'(x - a_1 + a_2) + (1-p)^2 f'(x),$$

$$b_3(x) = (1-2p)f(x) + f(x - a_2 + a_1).$$

(Тут і далі штрих позначає диференціювання.) Далі, для будь-яких непарних функцій g_1, g_2 на $\mathbb{R}$ позначимо

$$\langle g_1, g_2 \rangle_Q = \int\limits_{-\infty}^{+\infty} \big[pg_1(x-a_1) + (1-p)g_1(x-a_2)\big] \times$$

$$\times \big[pg_2(x-a_1) + (1-p)g_2(x-a_2)\big] \psi(x)dx.$$

Для будь-якої оцінюючої трійки $g = (g_1, g_2, g_3)$ задамо матриці $Q(g) := (q_{ik})_{i,k=1}^3 = (\langle g_i, g_k \rangle_Q)_{k,i=1}^3$ та $V(g) := (v_{ik})_{i,k=1}^3$, $v_{ik} = \int_{-\infty}^{+\infty} g_i(x) b_k(x) dx$.

Теорема 1. *Нехай* $\hat{\vartheta}_n$ — GEE-*оцінка параметра* ϑ *з оцінюючою трійкою* (g_1, g_2, g_3) *і виконуються наступні умови*:

1) $\hat{\vartheta}_n$ — *консистентна оцінка*;

2) $\mathrm{E} g_i(\xi_1 - a_k)^2 < \infty$ *для* $i = 1, 2, 3$, $k = 1, 2$;

3) *Для деяких* $\varepsilon > 0$, $\delta > 0$ $\mathrm{E} \sup_{\alpha: |\alpha - a_k| < \varepsilon} |g_i(\xi_1 - \alpha)|^{1+\delta} < \infty$, $\mathrm{E} \sup_{\alpha: |\alpha - a_k| < \varepsilon} |g_i'(\xi_1 - \alpha)|^{1+\delta} < \infty$ *для* $i = 1, 2, 3$, $k = 1, 2$;

4) g_i' *при* $i = 1, 2, 3$ *та* f' *є неперервними функціями на* $\mathbb{R}$;

5) $g_i(x + \Delta) f(x) \to 0$ *при* $x \to \pm\infty$ *для всіх* $\Delta \in \mathbb{R}$;

6) *всі елементи* $V(g)$ *є скінченними і* $\det V(g) \neq 0$.

Тоді $\sqrt{n}(\hat{\vartheta}_n - \vartheta) \to N(0, \Sigma)$, *де* $\Sigma = \Sigma(g_1, g_2, g_3) = V(g)^{-1} Q(g) (V(g)^{-1})^T$.

Для доведення теореми слід застосувати теорему 5.14 з [4] до GEE-оцінок, заданих рівнянням (5). При цьому слід враховувати, що $\mathrm{cov}(h(\vartheta)) = Q(g)$, $\partial \mathrm{E} h(\vartheta) / \partial \vartheta = V(g)$.

***Приклад* 2.** Для моментних оцінок, використовуючи теорему 1, можна отримати, наприклад, формулу для коефіцієнта розсіювання $\hat{p}_n^\mu - \Sigma_{33}$. Зокрема, якщо f — гауссова щільність з нульовим середнім із дисперсією s^2, то отримуємо

$$\Sigma_{33} = \frac{270}{b^{10}} + \frac{75}{2\,b^6} + \left(1 + \frac{1350}{b^8} + \frac{450}{b^6} + \frac{225}{2\,b^4}\right)(1-p)\,p,$$

де $b = |a_1 - a_2|/s$.

4. Точна нижня межа для коефіцієнтів розсіювання оцінок. Зрозуміло, що, вибираючи різні оцінюючі трійки, ми отримуємо різні GEE-оцінки з різними, взагалі кажучи, коефіцієнтами розсіювання. Знайдемо точні нижні межі для коефіцієнтів розсіювання, які можуть бути у таких оцінок.

Далі обмежимось розглядом випадку, коли f — неперервно диференційовна функція і $f(x) > 0$ (а отже, $\psi(x) > 0$) для всіх дійсних x. Це обмеження не є принциповим, але воно дозволяє уникнути технічних ускладнень.

Розглянемо клас $C^2_{bo}(\mathbb{R})$ всіх дійснозначних двічі неперервних непарних функцій $g\colon \mathbb{R}\to\mathbb{R}$ з обмеженим носієм (тобто для всіх $g\in C^2_{bo}(\mathbb{R})$ існують $-\infty<a<$ $<b<+\infty$ такі, що $g(x)=0$ $\forall x\notin[a,b]$). Зрозуміло, що $C^2_{bo}(\mathbb{R})$ з скалярним добутком $\langle\cdot,\cdot\rangle_Q$ є передгільбертовим простором (над полем дійсних чисел). Поповнимо його до гільбертового простору L_Q з нормою $\|g\|_Q=\sqrt{\langle g,g\rangle_Q}$.

Лема 1. 1. *L_Q складається з усіх вимірних непарних функцій $g:\mathbb{R}\to\mathbb{R}$ таких, що $\langle g,g\rangle_Q<\infty$.*

2. *Для довільних $f,g\in L_Q$ рівність $f=g$ еквівалентна $f(x)=g(x)$ для $x\in\mathbb{R}$ майже скрізь* (*м. с.*) *за мірою Лебега.*

3. *L_Q зі скалярним добутком $\langle\cdot,\cdot\rangle_Q$ є сепарабельним гільбертовим простором.*

Доведення пп. 1 і 3 можна провести, використовуючи стандартну схему наближення вимірних функцій ступінчастими (див. [5, с. 79]). Для перевірки п. 2 принциповим моментом є встановлення того, що для непарних функцій з $\langle g,g\rangle_Q=0$ випливає $g=0$ м. с. Доведення цього факту див. у додатку (лема 2).

Нехай $\sigma_i^2(g)=\Sigma_{ii}(g_1,g_2,g_3)$ — дисперсія граничного нормального розподілу (коефіцієнт розсіювання) для GEE-оцінки $\hat\vartheta_{i,n}$ з оцінюючою трійкою $g=(g_1,g_2,g_3)$, $\sigma^2_{i*}=\inf\sigma_i^2(g)$, де інфімум береться по всіх оцінюючих трійках, для яких відповідний коефіцієнт розсіювання є визначеним. Через δ_{ik} позначимо символ Кронекера: $\delta_{ii}=1$, $\delta_{ik}=0$ при $i\neq k$. Для $g\in L_Q$ позначимо

$$B_i(g)=\int\limits_{-\infty}^{+\infty}g(x)b_i(x)dx.$$

Зрозуміло, що B_i є лінійним функціоналом на L_Q, можливо, необмеженим. Позначимо $L_i=\{g\in L_Q\colon B_k(g)=\delta_{ik},\ k=1,2,3\}$. Якщо B_i є обмеженими функціоналами у L_Q, то L_i — афінний підпростір у L_Q.

Теорема 2. *Нехай функціонали B_i, $i=1,2,3$, є обмеженими у L_Q. Тоді $\sigma^2_{i*}=\|g_i^*\|_Q^2$, де g_i^* — ортогональна проекція 0 на афінний підпростір L_i.*

Ключовим питанням застосування теореми 2 є перевірка обмеженості (неперервності) функціоналів B_i. Щоб отримати достатні умови цього, позначимо

$$I_1=\int\limits_{-\infty}^{+\infty}\frac{p^2(f'(x-a_1))^2}{\psi(x)}dx,$$

$$I_2=\int\limits_{-\infty}^{+\infty}\frac{(1-p)^2(f'(x-a_2))^2}{\psi(x)}dx,$$

$$I_3=\int\limits_{-\infty}^{+\infty}\frac{(f(x-a_1)-f(x-a_2))^2}{\psi(x)}dx.$$

Теорема 3. *Якщо $\displaystyle\int_{-\infty}^{+\infty}x^4f(x)dx<\infty$ і для деякого $i=1,2,3$ $I_i<\infty$, то B_i — обмежений функціонал у L_Q і $\|B_i\|_Q\le\sqrt{I_i}$.*

Доведення теореми 3 див. у додатку.

Доведемо теорему 2. Нехай $\mathbb{B}=(\beta_{ik})^3_{i,k=1}$ — довільна невироджена матриця. Зазначимо, що множина розв'язків оцінюючого рівняння (4) не змінюється при переході від оцінюючої трійки $g = (g_1, g_2, g_3)$ до трійки $\tilde{g} = (\tilde{g}_1, \tilde{g}_2, \tilde{g}_3)$, де $\tilde{g}_i = \sum_{k=1}^{3} \beta_{ik} g_k$. При цьому $V(\tilde{g}) = \mathbb{B}V(g)$. Таким чином, вибравши $\mathbb{B} = V(g)^{-1}$, можна отримати оцінюючу трійку $\tilde{g}$, еквівалентну початковій, для якої $V(\tilde{g}) = \mathbb{E}$, де $\mathbb{E} = (\delta_{ik})_{i,k=1}^{3}$ — одинична матриця. Тому, шукаючи точну нижню межу для коефіцієнтів розсіювання GEE-оцінок, можна обмежитись трійками, для яких виконується умова нормування

$$V(g) = \mathbb{E}. \tag{6}$$

Для таких трійок $\Sigma(g) = Q(g)$ і, отже, $\sigma_i^2(g) = \|g_i\|_Q^2$. Умова нормування (6) еквівалентна умові $g_i \in L_i$. Як відомо, $\arg\min_{g \in L_i} \|g\|^2 = g_i^*$. Тому $\sigma_{i*}^2 \geq \|g_i^*\|_Q^2$.

Якщо $g^* = (g_1^*, g_2^*, g_3^*)$ задовольняє умови теореми 1, то, вибравши g^* як оцінюючу трійку, отримаємо $\sigma_i^2(g^*) = \|g_i^*\|^2$.

Якщо умови теореми 1 не виконані для g^*, то, враховуючи те, що L_Q є поповненням класу $C_{bo}^2(\mathbb{R})$ у $\|\cdot\|_Q$, для будь-якого $\varepsilon > 0$ можна вибрати таку трійку $g^\varepsilon = (g_1^\varepsilon, g_2^\varepsilon, g_3^\varepsilon)$, яка задовольняє умови теореми 1, умови нормування (6) і $\|g_i^\varepsilon - g_i^*\|_Q < \varepsilon$. Зрозуміло, що $\sigma_{i*}^2 \leq \sigma_i^2(g^\varepsilon) < \|g_i^*\|_Q + \varepsilon$. Внаслідок довільності $\varepsilon > 0$ отримуємо $\sigma_i^2(g^*) = \|g_i^*\|^2$.

Теорему доведено.

5. Обчислення нижньої межі коефіцієнтів розсіювання. Для того щоб обчислити σ_{i*}^2, виберемо деякий ортонормований базис $u_1, \ldots, u_j, \ldots$ в L_Q. Для будь-яких $g_i \in L_Q$ має місце розклад за базисом

$$g_i = \sum_{j=1}^{\infty} \gamma_j^i u_j,$$

де $\gamma_j^i = \langle g_i, u_j \rangle_Q$. Зазначимо, що якщо B_k — обмежений функціонал у L_Q, то $B_k(g_i) = \sum_{j=1}^{\infty} \gamma_j^i \beta_j^k$, де $\beta_j^k = \int_{-\infty}^{+\infty} u_j(x) b_k(x) dx$. (Ця формула, вочевидь, є правильною для $g_i = u_j$ і переноситься на довільні g_i за лінійністю та неперервністю B_k.) Отже, B_k можна ототожнити з елементом L_Q $\tilde{B}_k = \sum_{j=1}^{\infty} \beta_j^k u_j$:

$$B_k(g_i) = \langle \tilde{B}_k, g_i \rangle_Q.$$

Таким чином, задача мінімізації $\|g_i\|^2$ при обмеженні $g_i \in L_i$ еквівалентна задачі

$$\sum_{j=1}^{\infty} (\gamma_j^i)^2 \to \min_{\gamma_j^i},$$

$$\sum_{j=1}^{\infty} \beta_j^k \gamma_j^i = \delta_{ik}, \quad k = 1, 2, 3.$$

Розв'язуючи цю задачу методом множників Лагранжа, отримуємо $\gamma_j^i = \lambda_1^i \beta_j^1 + \lambda_2^i \beta_j^2 + \lambda_3^i \beta_j^3$, де вектор $\lambda^i = (\lambda_1^i, \lambda_2^i, \lambda_3^i)^T$ є розв'язком рівняння $\mathbb{B}\lambda^i = e_i$, $e_i = (\delta_{i1}, \delta_{i2}, \delta_{i3})^T$, $\mathbb{B} = (\langle \tilde{B}_i, \tilde{B}_k \rangle_Q)_{i,k=1}^{3}$, $\langle \tilde{B}_i, \tilde{B}_k \rangle_Q = \sum_{j=1}^{\infty} \beta_j^i \beta_j^k$. Підставляючи цей вираз у формулу для $\|g_i^*\|^2$, отримуємо $\sigma_{i*}^2 = \|g_i^*\|_Q^2 = e_i^T \mathbb{B}^{-1} e_i$, тобто σ_{i*}^2 — i-й діагональний елемент матриці, оберненої до $\mathbb{B}$.

Коефіцієнти розсіювання оцінок
та відносні асимптотичні ефективності

s	2	1,5	1	0,75	0,5	0,25	0,1
σ^2_{1*}	1301,76	164,00	18,73	5,52	1,33	0,25	0,04
σ^2_{2*}	1890,01	113,69	5,77	1,51	0,40	0,08	0,01
σ^2_{3*}	409,32	27,06	1,57	0,46	0,21	0,19	0,19
$\sigma^2_{1\mu}$	1853,48	277,69	32,29	9,28	2,16	0,30	0,04
$\sigma^2_{2\mu}$	3109,46	302,59	17,91	3,58	0,66	0,23	0,01
$\sigma^2_{3\mu}$	666,27	69,10	4,66	1,06	0,30	0,21	0,19
e_1	1,62	2,55	2,97	2,31	1,42	1,13	1,00
e_2	1,42	1,69	1,72	1,68	1,62	1,18	1,03
e_3	1,64	2,66	3,10	2,37	1,65	2,72	1,02

У числових розрахунках послідовність базисних функцій $u_1, u_2, \ldots, u_K$ можна отримати ортогоналізацією у скалярному добутку $\langle\cdot,\cdot\rangle_Q$ деякої системи $z_1, z_2, \ldots, \ldots, z_K$ неперервно диференційовних непарних функцій. В якості таких функцій ми вибрали $z_k(x) = \sin(k\pi x/T)$, де $k = 1, 2, \ldots$, а T доцільно обирати достатньо великим, щоб

$$\langle u_i, u_k\rangle_Q \approx \int\limits_{-T}^{+T} [pu_i(x-a_1)+(1-p)u_i(x-a_2)]\times[pu_k(x-a_1)+(1-p)u_k(x-a_2)]\psi(x)dx.$$

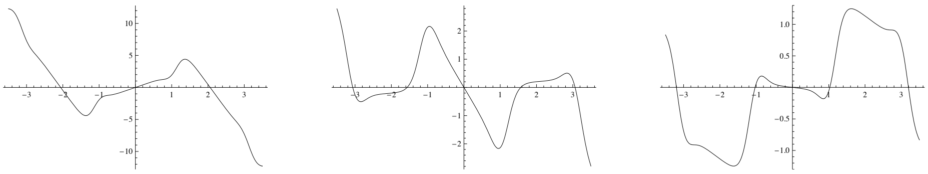
При цьому нескінченні суми $\sum_{j=1}^{\infty} \beta^i_j\beta^k_j$ в означенні $\mathbb{B}$ наближаються частковими сумами $\sum_{j=1}^{K} \beta^i_j\beta^k_j$. Величину K можна вибирати за правилом Рунге.

Цю техніку було використано для обчислення нижніх меж коефіцієнтів розсіювання GEE-оцінок у випадку, коли f є нормальною щільністю з нульовим середнім та дисперсією s^2. Результати для різних s та $a_1 = 0$, $a_2 = 2$, $p = 0.25$ наведено у таблиці. Тут σ^2_{i*} — точні нижні межі для коефіцієнтів розсіювання GEE-оцінок ϑ_i, $\sigma^2_{i\mu}$ — коефіцієнти розсіювання моментних оцінок ϑ_i, $e_i = \sigma^2_{i\mu}/\sigma^2_{i*}$ — відносні ефективності моментних оцінок.

Як видно з цієї таблиці, моментні оцінки є близькими до найкращих можливих GEE-оцінок лише коли s значно менше за $|a_2 - a_1|$. Вибираючи оцінюючі функції більш близькими до оптимальних, можна у деяких випадках втричі покращити дисперсію оцінок.

Графіки оптимальних оцінюючих функцій g^*_i для випадку $s = 0,5$ наведено на рисунку.

6. Висновки. Ми побудували нові GEE-оцінки для евклідових параметрів моделі суміші двох симетричних розподілів, довели їх асимптотичну нормальність і знайшли точні нижні межі для коефіцієнтів розсіювання. Числовий приклад показує, що використання GEE-оцінок з оцінюючими функціями, близькими до



g_1^* g_2^* g_3^*

Оптимальна оцінююча трійка

оптимальних, може суттєво покращити точність оцінювання. На жаль, оптимальні оцінюючі функції є різними для різних значень оцінюваних параметрів та різних щільностей розподілу f. Тому є доцільним застосування адаптивної техніки, аналогічної запропонованій у [6], для оцінювання за спостереженнями з домішкою. При використанні такої техніки на першому кроці за даними будується оцінка для оптимальної оцінюючої трійки, а на другому — ця оцінена трійка використовується для побудови „адаптованих” GEE-оцінок. Питання про ефективність таких адаптованих оцінок має бути предметом подальших досліджень.

Додаток.

Лема 2. *Нехай g — непарна вимірна функція і $\langle g, g\rangle_Q = 0$. Тоді $g = 0$ м. с. за мірою Лебега на $\mathbb{R}$.*

Доведення. Позначимо $\Delta = a_2 - a_1$, $\beta = -(1-p)/p$. З $\langle g, g\rangle_Q = 0$ випливає, що $pg(x - a_1) + (1-p)g(x - a_2) = 0$ м. с. по $x \in \mathbb{R}$, тобто $g(x + \Delta) = \beta g(x)$ м. с. Тому

$$g(x + k\Delta) = \beta^k g(x) \tag{7}$$

м. с. для всіх $k \in \mathbb{Z}$ і, отже, функція $g(x)$ м. с. на $\mathbb{R}$ визначається своїми значеннями на інтервалі $x \in [0, \Delta]$. Покажемо, що функція, отримана за формулою (7), дорівнює 0 м. с., якщо вона непарна.

Припустимо протилежне. Нехай існує непарна функція g, для якої виконується (7) і g не дорівнює 0 на деякій множині додатної міри. Тоді існують $c > 0$ і $A_c \subseteq [0, \Delta]$ такі, що $|g(x)| > c$ для всіх $x \in A_c$ і $\varepsilon := \lambda(A_c) > 0$. Крім того, оскільки функція g є вимірною, то існують $C < \infty$ та $A^C \subseteq [0, \Delta]$ такі, що $|g(x)| < C$ для всіх $x \in A^C$ і $\lambda(A^C) > \Delta - \varepsilon$.

Для довільного $k \in \mathbb{Z}$ розглянемо $x \in k\Delta + A^C$. Згідно з (7), $|g(x)| = |\beta|^k |g(x - k\Delta)| < C|\beta|^k$ для всіх таких x. Отже, для $B_{k-} = \{x \in [k\Delta, (k+1)\Delta] \colon |g(x)| < C|\beta|^k\}$ $\lambda(B_{k-}) > \Delta - \varepsilon$.

Аналогічно, для $x \in -(k+1)\Delta + A_c$ $|g(x)| > c|\beta|^{-(k+1)}$ і, враховуючи непарність g, $|g(-x)| > c|\beta|^{-(k+1)}$. Отже, для $B_{k+} = \{x \in [k\Delta, (k+1)\Delta] \colon |g(x)| > c|\beta|^{-(k+1)}\}$, $\lambda(B_{k+}) \geq \varepsilon$.

Вибравши $k \in \mathbb{Z}$ так, щоб $c|\beta|^{-(k+1)} > C|\beta|^k$, отримаємо $B_{k+} \cap B_{k-} = \emptyset$ і, врахувавши отримані вище нерівності, $\lambda(B_{k+} \cup B_{k-}) = \lambda(B_{k+}) + \lambda(B_{k-}) > \varepsilon + \Delta - \varepsilon = \Delta$. Але за побудовою і B_{k+}, і B_{k-} належать інтервалу $[k\Delta, (k+1)\Delta]$. Отже, $\lambda(B_{k+} \cup B_{k-}) \leq \Delta$. Ця суперечність доводить лему.

***Доведення теореми* 3** проведемо для $i = 1$ (для $i = 2, 3$ доведення аналогічне). Спочатку покажемо, що для будь-якої функції $g \in C^2_{bo}(\mathbb{R})$ виконано

$$\int_{-\infty}^{+\infty} g(x)b_1(x)dx \le \sqrt{I_1\langle g, g\rangle_Q}. \tag{8}$$

Для цього розглянемо задачу оцінки параметра a_1 у випадку, коли p, a_2 і f є відомими. Це задача параметричного оцінювання, для якої інформація Фішера, що припадає на одне спостереження ξ_j, дорівнює I_1. За наслідком з теореми 11.2 розділу 2 в [7], якщо $\check{a}_1$ — регулярна, асимптотично нормальна оцінка a_1, то $\sqrt{n}(\check{a}_1 - a_1) \Rightarrow N(0, \sigma^2_{\check{a}_1})$ і

$$\sigma^2_{\check{a}_1} > 1/I_1. \tag{9}$$

Ми побудуємо регулярну оцінку $\check{a}_1$, для якої $\sigma^2_{\check{a}_1} = \|g\|^2_Q/(B_1(g))^2$. Тоді з (9) випливатиме (8).

Почнемо з побудови незміщеної оцінки для a_1. Позначимо $\bar{\xi} = \dfrac{1}{n}\sum_{j=1}^n \xi_j$, $\tilde{a}_1 = \dfrac{1}{p}(\bar{\xi} - (1-p)a_2)$. Легко бачити, що $\mathrm{E}\tilde{a}_1 = a_1$, тобто $\tilde{a}_1$ — незміщена оцінка a_1.

Покладемо тепер $H(x, \alpha) = pg(x-\alpha) + (1-p)g(x-a_2)$, $\hat{h}(\alpha) = \dfrac{1}{n}\sum_{j=1}^n H(\xi_j, \alpha)$, $b_1^*(x) = p^2 f'(x) + p(1-p)f'(x - a_2 + a_1^*)$, де $a_1^* \in \mathbb{R}$ — деяке фіксоване число, $B_1^*(g) = \displaystyle\int_{-\infty}^{+\infty} g(x)b_1^*(x)dx$,

$$\check{a}_1 = \tilde{a}_1 - \frac{1}{B_1^*(g)}\hat{h}(\tilde{a}_1).$$

Покажемо, що $\check{a}_1$ — регулярна асимптотично нормальна оцінка. Для цього зазначимо, що

$$\sqrt{n}(\check{a}_1 - a_1) = S_n + \varepsilon_n, \tag{10}$$

де

$$S_n = \frac{1}{B_1^*(g)} \frac{1}{\sqrt{n}} \sum_{j=1}^n H(\xi_j, a_1),$$

ε_n — послідовність, що збігається до 0 за ймовірністю рівномірно по a_1 у деякому околі $[a_-, a_+]$ точки a_1^* (тобто для всіх $x > 0$ $\sup_{a_1 \in [a_-, a_+]} \mathrm{P}\{|\varepsilon_n| > x\} \to 0$).

Дійсно, оскільки g — двічі неперервно диференційовна функція, то $\hat{h}(\tilde{a}_1) = \hat{h}(a_1) + \hat{h}'(a_1)(\tilde{a}_1 - a_1) + \dfrac{1}{2}\hat{h}''(\tau)(\tilde{a}_1 - a_1)^2$. Тут штрих позначає диференціювання по α, а τ — проміжна точка між a_1 та $\tilde{a}_1$.

Отже,

$$\sqrt{n}(\check{a}_1 - a_1) = \sqrt{n}\left(\tilde{a}_1 - a_1 - \frac{\hat{h}(\tilde{a}_1)}{B_1^*(g)}\right) = S_n + \varepsilon_n^1 + \varepsilon_n^2,$$

де

$$\varepsilon_n^1 = \sqrt{n}(\tilde{a}_1 - a_1)\left(1 - \frac{\hat{h}'(\tilde{a}_1)}{B_1^*(g)}\right),$$

$$\varepsilon_n^2 = -\frac{1}{2B_1^*(g)} \frac{1}{n} \sum_{j=1}^n H''(\xi_n, \tau)\sqrt{n}(\tilde{a}_1 - a_1)^2.$$

Враховуючи, що $|H''(\xi_j,\tau)| < C$, де $C < \infty$ — деяка константа, отримуємо

$$\mathrm{E}|\varepsilon_n^2| \leq \frac{1}{4(B_1^*(g))^2} C\sqrt{n}\mathrm{E}(\tilde{a}_1 - a_1)^2 \leq \frac{C}{4(B_1^*(g))^2}\frac{1}{\sqrt{n}}\operatorname{Var}\frac{\xi}{p}.$$

Отже, за нерівністю Чебишова ε_n^2 збігається до 0 за ймовірністю рівномірно по a_1.

Для ε_n^1 аналогічну збіжність отримуємо з нерівностей

$$\mathrm{E}\left(\sqrt{n}(\tilde{a}_1 - a_1)\right)^2 \leq \operatorname{Var}\frac{\xi_1}{p}$$

та

$$\mathrm{E}\left(1 - \frac{\hat{h}'(a_1)}{B_1^*(g)}\right)^2 = \frac{1}{nB_1^*(g)}\operatorname{Var} h'(\xi_1, a_1).$$

Рівність (10) доведено. З неї безпосередньо одержуємо

$$\sqrt{n}(\check{a}_1 - a_1) \Rightarrow N(0, \sigma^2_{\check{a}_1}(a)),$$

де

$$\sigma^2_{\check{a}_1}(a) = \frac{\operatorname{Var} h(\xi, a_1)}{(B_1^*(g))^2},$$

до того ж, за теоремами 7 і 10 з [7], ця збіжність є рівномірною на $[a_-, a_+]$. Тому $\check{a}_1$ є регулярною асимптотично нормальною оцінкою і при $a_1^* = a_1$ виконується (9), а отже, і (8).

Таким чином, (8) доведено для $g \in C^2_{bo}(\mathbb{R})$. Оскільки клас таких функцій є щільним у L_Q, то (8) виконується для всіх $g \in L_Q$.

Теорему доведено.

1. *Bordes L., Mottelet S.* Vandekerkhove Semiparametric estimation of a two-component mixture model // Ann. Statist. – 2006. – **34**. – P. 1204 – 1232.
2. *Hunter D. R., Wang S., Hettmansperger T. R.* Inference for mixtures of symmetric distributions // Ibid. – 2007. – **35**. – P. 224 – 251.
3. *Maiboroda R.* Estimation of locations and mixing probabilities by observations from two-component mixture of symmetric distributions // Theory Probab. and Math. Statist. – 2008. – **78** – P. 133 – 141.
4. *Shao J.* Mathematical statistics. – New York: Springer, 1998. – 530 p.
5. *Рид М., Саймон Б.* Методы современной математической физики. Т. 1. Функциональный анализ. – М.: Мир, 1977. – 357 с.
6. *Майборода Р., Сугакова О.* Адаптивні оцінюючі рівняння для середнього положення за спостереженнями з домішкою // Теорія ймовірностей та мат. статистика. – 2009. – Вип. 80. – С. 91 – 99.
7. *Ибрагимов И. А., Хасьминский Р. З.* Асимптотическая теория оценивания. – М.: Наука, 1979. – 527 с.

Одержано 03.11.09