УДК 519.21

Г. В. Шуренков (Київ. нац. ун-т ім. Т. Шевченка)

АСИМПТОТИЧНО ОПТИМАЛЬНІ ОЦІНКИ МОМЕНТІВ ЗМІНИ

We consider the problem of finding asymptotically optimal estimates of many moments of change in the case of incomplete information about distributions. We prove that if the estimate of maximal probability is asymptotically optimal, then, under certain conditions, it preserves this property after the replacement of values of denseness by their estimates. We solve the problem for the case of one moment of the variation and generalize the results obtained to the case of several moments of change.

Розглянуто задачу знаходження асимптотично оптимальних оцінок багатьох моментів зміни у випадку неповної інформації про розподіли. Доведено, що оцінка максимальної вірогідності, яка є асимптотично оптимальною, за певних умов залишається такою при підстановці оцінок щільності замість справжніх значень. Задачу розв'язано для випадку одного моменту зміни, й отримані результати узагальнено на випадок кількох моментів зміни.

1. Вступ. Задача оцінювання моментів зміни виникає, зокрема, при аналізі геологічних даних та телеметричної інформації. Існує декілька підходів до пошуку моментів зміни [1].

При першому підході розглядають апостеріорні оцінки моментів зміни. В рамках цього підходу задачу пошуку моментів зміни можна розглядати як задачу перевірки гіпотез щодо положення моменту зміни. Тоді оптимальною можна вважати оцінку, яка мінімізує похибку вибору гіпотези. Такою оцінкою буде оцінка методу найбільшої вірогідності (у випадку одного моменту зміни). При другому варіанті цього підходу розглядають задачу оцінки невідомого параметра (а саме, моменту зміни). Оптимальною оцінкою в цьому випадку можна вважати оцінку, що мінімізує середньоквадратичне відхилення. Тоді асимптотично оптимальною буде баєсова оцінка [1].

При другому підході розглядають задачу якнайшвидшого відшукання моменту зміни за даними, що надходять. Така задача зводиться до відшукання оптимального моменту зупинки в деякому класі моментів Маркова [2, 3]. Крім цього, можлива постановка задачі пошуку моментів зміни як пошуку розриву в кривій, що спостерігається, з шумом [4, 5]. Методи оцінювання моментів зміни застосовуються при розпізнаванні образів [6].

У даній роботі розглядається апостеріорна задача оптимальної оцінки багатьох моментів зміни, при цьому використовується так званий швидкий алгоритм пошуку моментів зміни [7], в якому задача зводиться до відшукання оптимальної послідовності номерів розподілів („траєкторії”) за допомогою алгоритму динамічного програмування, запропонованого в [8].

У статті [1] знайдено асимптотично оптимальні оцінки моментів зміни для випадку одного моменту, в [9] ці результати узагальнено на випадок багатьох моментів зміни. У даній роботі розглядається задача відшукання оптимальної оцінки, коли відомо, що дані можуть мати лише два розподіли, але інформація про них є неповною. В даному випадку можливо, використовуючи медіанну оцінку [10], встановити приблизні оцінки моментів зміни, а потім, побудувавши за цими оцінками оцінки щільності, використати їх для більш точного оцінювання моментів зміни.

В роботі показано, що асимптотичний розподіл відхилення такої оцінки збігається з розподілом оптимальної оцінки у випадку одного моменту зміни, й отриманий результат узагальнено на випадок багатьох моментів зміни.

2. Постановка задачі. Будемо розглядати послідовність незалежних випадкових величин $\{\zeta_1, \ldots, \zeta_N\}$, які, взагалі кажучи, не є однаково розподіленими, а можуть мати один із двох розподілів: $F = F_1$, $G = F_2$. Вважаємо, що ці розподіли мають щільності $f = f_1$ та $g = f_2$ відповідно. $\mathrm{P}(\zeta_j \in A) = F_{h_j^0}(A)$, де $h^0 = \{h_j^0,\ j = 1, \ldots, N\}$ — невипадкова послідовність номерів $h_j^0 \in \{1, 2\}$, яка має вигляд $h_j^0 = \text{const}$ при $k_i = [\theta_i N] < j \le [\theta_{i+1} N]$, де $0 = \theta_0 < \theta_1 < \ldots \ldots < \theta_R < \theta_{R+1} = 1$ — фіксовані невипадкові числа, які називають моментами зміни, k_i — точки зміни. Вектор з точок зміни позначимо через $\mathbf{k}$. Послідовності номерів розподілів h будемо називати траєкторіями, h^0 — істинною траєкторією послідовності $\{\zeta_1, \ldots, \zeta_N\}$.

Як оцінку для h^0 розглянемо

$$\hat{h} = \arg\max_h \sum_{i=1}^N \Big(\ln f_{h_i}(\zeta_i) - \pi_N(h_i, h_{i-1}) \Big), \qquad \pi_N(g, l) = \pi_N \mathbb{1}_{g \ne l}, \tag{1}$$

де $\pi_N > 0$ — невипадкова величина.

Оцінки для моментів зміни будуються за траєкторією $\hat{h}$ таким чином:

$$\hat{k}_1 = k_1(\hat{h}) = \min\Big\{ l \mid \hat{h}_l \ne \hat{h}_j,\ 1 \le j < l \Big\},$$

$$\hat{k}_i = k_i(\hat{h}) = \min\Big\{ l \mid \hat{h}_l \ne \hat{h}_j,\ k_{i-1}(\hat{h}) \le j < l \Big\}$$

— i-та точка зміни у траєкторії $\hat{h}$, $\hat{R} = R(\hat{h})$ — кількість точок зміни у $\hat{h}$.

При виконанні відповідних умов на π_N така оцінка є в певному розумінні оптимальною [4].

Нехай $\hat{f}$ та $\hat{g}$ — оцінки f та g, що можуть залежати від спостережень $\zeta_1, \ldots, \zeta_N$ (побудову таких оцінок ми розглянемо в п. 5). Позначимо

$$\phi(x, 1) = \ln \hat{f}(x), \qquad \phi(x, 2) = \ln \hat{g}(x).$$

Введемо функціонал

$$J(h) = \sum_{i=1}^N \big(\phi(\zeta_i, h_i) - \pi_N(h_i, h_{i-1}) \big)$$

і оцінку для h^0

$$\hat{\hat{h}} = \arg\max_h J(h). \tag{2}$$

Оцінками моментів зміни будуть $\hat{\hat{k}}_{j,N} = k_j(\hat{\hat{h}})$, оцінкою кількості змін — $\hat{\hat{R}} = R(\hat{\hat{h}})$. Вектор з оцінок позначимо через $\hat{\hat{k}}$. Основний результат даної роботи полягає в тому, що при певних умовах асимптотичні розподіли $\hat{k}_{j,N}$ та $\hat{\hat{k}}_{j,N}$ є однаковими:

$$\lim_{N \to \infty} \mathrm{P}\Big(\hat{\hat{k}}_{j,N} - k_j = n \Big) = \lim_{N \to \infty} \mathrm{P}\Big(\hat{k}_{j,N} - k_j = n \Big).$$

Позначимо $\operatorname{supp} f = \{x \in \mathbb{R},\ f(x) > 0\}$. Накладемо такі умови на щільності:

1) $\operatorname{supp} f = \operatorname{supp} g$, існує $\delta > 0$ таке, що $f(x) > \delta$, $g(x) > \delta$ для всіх $x \in \operatorname{supp} f$;

2) для будь-якого $\alpha > 0$ міра Лебега множини $\{x \in \mathbb{R} : f(x)/g(x) = \alpha\}$ дорівнює нулю;

3) існує $M < \infty$ таке, що $f(x) < M$, $g(x) < M$.

Введемо позначення: $\|\cdot\|$ — рівномірна напівнорма на $\operatorname{supp} f = \operatorname{supp} g$,

$$\|u\| = \sup_{x \in \operatorname{supp} f} |u(x)|.$$

Позначимо $\Delta_{\hat f}^{\varepsilon} = \mathrm{P}(\|f - \hat f\| > \varepsilon)$, $\Delta_{\hat g}^{\varepsilon} = \mathrm{P}(\|g - \hat g\| > \varepsilon)$.

Розглянемо спочатку цю задачу для випадку одного моменту зміни.

3. Випадок одного моменту зміни. Нехай відомо, що у даних $\zeta_1, \ldots, \zeta_N$ лише один момент зміни. Тоді можна вважати, що

$$\zeta_j = \begin{cases} \eta_{k-j+1}, & j \le k, \\ \xi_{j-k}, & j > k, \end{cases}$$

тобто ми спостерігаємо послідовність $\eta_k, \ldots, \eta_1, \xi_1, \ldots, \xi_{N-k}$; η_j мають розподіл G та щільність g, ξ_j — розподіл F та щільність f.

Як оцінку для $k = k_1$ можна розглядати

$$\hat k_N = \operatorname*{arg\,max}_{1 \le l \le N} \left(\sum_{j=1}^{l} \ln g(\zeta_i) + \sum_{j=l+1}^{N} \ln f(\zeta_i) \right)$$

при відомих щільностях g і f та

$$\hat{\hat k}_N = \operatorname*{arg\,max}_{1 \le l \le N} \left(\sum_{j=1}^{l} \ln \hat g(\zeta_i) + \sum_{j=l+1}^{N} \ln \hat f(\zeta_i) \right)$$

при невідомих. Ці оцінки є аналогами оцінок (1) та (2), якщо відомо, що момент зміни лише один.

Теорема 1. *Нехай виконуються умови* 1 – 3, $\|\hat f_N - f\| \xrightarrow{\mathrm{P}} 0$, $\|\hat g_N - g\| \xrightarrow{\mathrm{P}} 0$ *при* $N \to \infty$, *тоді*

$$\lim_{N\to\infty} \mathrm{P}\left(\hat{\hat k}_N - k = n\right) = \lim_{N\to\infty} \mathrm{P}\left(\hat k_N - k = n\right).$$

Доведення. Позначимо

$$S_l = \sum_{j=1}^{l} \ln \frac{g(\xi_{n+1-j})}{f(\xi_{n+1-j})}, \qquad T_l = \sum_{j=1}^{l} \ln \frac{g(\xi_{n+j})}{f(\xi_{n+j})}, \qquad U_l = \sum_{j=1}^{l} \ln \frac{f(\eta_j)}{g(\eta_j)}.$$

Розглянемо випадок $n > 0$. При $n > 0$ асимптотичний розподіл $\hat k_N - k$ дорівнює

$$\lim_{N\to\infty} \mathrm{P}\left(\hat k_N - k = n\right) = \mathrm{P}\left(S_j > 0,\ 1 \le j \le n;\ \max_j T_j \le 0;\ S_n > \max_j U_j \right).$$

Позначимо

$$\hat S_l = \sum_{j=1}^{l} \ln \frac{\hat g(\xi_{n+1-j})}{\hat f(\xi_{n+1-j})}, \qquad S_l^{\varepsilon} = \sum_{j=1}^{l} \ln \frac{g(\xi_{n+1-j}) + \varepsilon}{f(\xi_{n+1-j}) - \varepsilon}, \qquad \hat T_l = \sum_{j=1}^{l} \ln \frac{\hat g(\xi_{n+j})}{\hat f(\xi_{n+j})},$$

$$T_l^{\varepsilon} = \sum_{j=1}^{l} \ln \frac{g(\xi_{n+j}) + \varepsilon}{f(\xi_{n+j}) - \varepsilon}, \qquad \hat U_l = \sum_{j=1}^{l} \ln \frac{\hat f(\eta_j)}{\hat g(\eta_j)}, \qquad U_l^{\varepsilon} = \sum_{j=1}^{l} \ln \frac{f(\eta_j) + \varepsilon}{g(\eta_j) - \varepsilon}.$$

Розподіл $\hat{\hat k}_N - k$ можна записати у вигляді

$$\mathrm{P}\left(\hat{\hat{k}}_N - k = n\right) = \mathrm{P}\left(\hat{S}_j > 0,\ 1 \le j \le n;\ \hat{T}_j \le 0,\ 1 \le j \le N-k-n;\ \hat{S}_n > \hat{U}_j,\ 1 \le j \le k\right).$$

Оцінимо розподіл через розподіли сум S_j^ε, T_j^ε та U_j^ε:

$$\mathrm{P}\left(S_j^{-\varepsilon} > 0,\ 1 \le j \le n;\ T_j^\varepsilon \le 0,\ 1 \le j \le N-k-n;\ S_n^{-\varepsilon} > U_j^\varepsilon,\ 1 \le j \le k\right) - \Delta_{\hat{f}}^\varepsilon - \Delta_{\hat{g}}^\varepsilon \le$$

$$\le \mathrm{P}\left(\hat{\hat{k}}_N - k = n\right) \le$$

$$\le \mathrm{P}\left(S_j^\varepsilon > 0,\ 1 \le j \le n;\ T_j^{-\varepsilon} \le 0,\ 1 \le j \le N-k-n;\ S_n^\varepsilon > U_j^{-\varepsilon},\ 1 \le j \le k\right) + \Delta_{\hat{f}}^\varepsilon + \Delta_{\hat{g}}^\varepsilon.$$

Асимптотичні розподіли сум, що фігурують в обмеженнях, дорівнюють

$$\lim_{N\to\infty} \mathrm{P}\left(S_j^\varepsilon > 0,\ 1 \le j \le n;\ T_j^{-\varepsilon} \le 0,\ 1 \le j \le N-k-n;\ S_n^\varepsilon > U_j^{-\varepsilon},\ 1 \le j \le k\right) =$$

$$= \mathrm{P}\left(S_j^\varepsilon > 0,\ 1 \le j \le n;\ \max_j T_j^{-\varepsilon} \le 0;\ S_n^\varepsilon > \max_j U_j^{-\varepsilon}\right).$$

Доведемо дві леми щодо неперервності сум S_j^ε; аналогічні леми мають місце для сум T_j^ε та U_j^ε.

Лема 1. *Ймовірності*

$$\mathrm{P}\left(S_j^\varepsilon > 0,\ 1 \le j \le n,\ S_n^\varepsilon > x\right) \quad \text{та} \quad \mathrm{P}\left(S_j^\varepsilon > 0,\ 1 \le j \le n,\ S_n^\varepsilon \ge x\right)$$

неперервні по ε *при* $\varepsilon = 0$.

Доведення. Оскільки $f(x)/g(x) \ne \alpha$ майже напевно, то S_j є неперервними випадковими величинами. Позначимо $f^\varepsilon(x) = f(x) + \varepsilon$, $g^\varepsilon(x) = g(x) + \varepsilon$. Зауважимо, що $S_j^\varepsilon \ge S_j$ при $\varepsilon \ge 0$. Отже,

$$\mathrm{P}\left(S_j^\varepsilon > 0,\ 1 \le j \le n,\ S_n^\varepsilon > x\right) \ge \mathrm{P}\left(S_j > 0,\ 1 \le j \le n,\ S_n > x\right)$$

та

$$\mathrm{P}\left(S_j^\varepsilon > 0,\ 1 \le j \le n,\ S_n^\varepsilon > x\right) - \mathrm{P}\left(S_j > 0,\ 1 \le j \le n,\ S_n > x\right) \le$$

$$\le \sum_{j=1}^{n} \mathrm{P}\left(S_j^\varepsilon > 0,\ S_j \le 0\right) + \mathrm{P}\left(S_n^\varepsilon > x,\ S_n \le x\right) \le$$

$$\le \sum_{j=1}^{n} \mathrm{P}\left(\prod_{i=1}^{j} g^\varepsilon(\xi_i) > \prod_{i=1}^{j} f^{-\varepsilon}(\xi_i),\ \prod_{i=1}^{j} g(\xi_i) \le \prod_{i=1}^{j} f(\xi_i)\right) +$$

$$+ \mathrm{P}\left(\prod_{i=1}^{n} g^\varepsilon(\xi_i) > e^x \prod_{i=1}^{n} f^{-\varepsilon}(\xi_i),\ \prod_{i=1}^{n} g(\xi_i) \le e^x \prod_{i=1}^{n} f(\xi_i)\right).$$

Оцінимо кожен доданок у сумі, взявши до уваги, що

$$\prod_{i=1}^{j}\left(g(\xi_i) + \varepsilon\right) \le \prod_{i=1}^{j} g(\xi_i) + \sum_{i=1}^{j} \varepsilon^i C_i^j \left(\max\left(g(\xi_i)\right)\right)^{j-i}.$$

Звідси при малих ε маємо

$$\prod_{i=1}^{j} g(\xi_i) + \sum_{i=1}^{j} \varepsilon^i C_i^j (\max(g(\xi_i)))^{j-i} \le \prod_{i=1}^{j} g(\xi_i) + \varepsilon 2^j \max(g(\xi_i), 1)^j,$$

аналогічно

$$\prod_{i=1}^{j} (f(\xi) - \varepsilon) \ge \prod_{i=1}^{j} f(\xi) - \varepsilon 2^j \max(f(\xi_i), 1)^j.$$

Отже,

$$\mathrm{P}\left(\prod_{i=1}^{j} g^{\varepsilon}(\xi_i) > e^x \prod_{i=1}^{j} f^{-\varepsilon}(\xi_i),\ \prod_{i=1}^{j} g(\xi_i) \le e^x \prod_{i=1}^{j} f(\xi_i)\right) \le$$

$$\le \mathrm{P}\left(0 \le \prod_{i=1}^{j} f(\xi_i) - e^x \prod_{i=1}^{j} g(\xi_i) < (1+e^x) 2^j \sqrt{\varepsilon}\right) +$$

$$+ \sum_{i=1}^{j} \left(\mathrm{P}\left(f(\xi_i) > 1/\sqrt[2j]{\varepsilon}\right) + \mathrm{P}\left(g(\xi_i) > 1/\sqrt[2j]{\varepsilon}\right)\right) \to \mathrm{P}(S_j = x) = 0,$$

оскільки S_j є неперервною. При $\varepsilon < 0$ доведення проводиться аналогічно.

Лему доведено.

З доведення леми випливає, що $\mathrm{P}(S_n^{\varepsilon} > 0)$ теж неперервні по ε у нулі.

Лема 2. $\mathrm{P}(\max_j S_j^{\varepsilon} > x)$ *та* $\mathrm{P}(\max_j S_j^{\varepsilon} \ge x)$ *неперервні по* ε *при* $\varepsilon = 0$.

Доведення. Доведемо, що при $\varepsilon \to 0$

$$\mu^{\varepsilon} = \mathrm{E} \ln \frac{g(\xi_j) + \varepsilon}{f(\xi_j) - \varepsilon} \to \mu = \mathrm{E} \ln \frac{g(\xi_j)}{f(\xi_j)} < 0.$$

За нерівністю Йєнсена

$$\mu^{\varepsilon} < \ln \mathrm{E} \frac{g(\xi_j) + \varepsilon}{f(\xi_j) - \varepsilon}.$$

Нехай $\varepsilon \le \delta/2$, тоді

$$\mathrm{E} \frac{g(\xi_j) + \varepsilon}{f(\xi_j) - \varepsilon} = \int \frac{g(x) + \varepsilon}{f(x) - \varepsilon} f(x) dx \le \int \frac{g(x) + \delta/2}{f(x) - \delta/2} f(x) dx \le \frac{2}{\delta} \mathrm{E} g(\xi_j) + 1.$$

За теоремою про мажоровану збіжність $\mathrm{E} \ln \dfrac{g(\xi_j) + \varepsilon}{f(\xi_j) - \varepsilon} \to \mu$. Отже, починаючи з деякого $\varepsilon = \gamma$ $\mu^{\varepsilon} < 0$.

За законом великих чисел при $\varepsilon \le \gamma$ $\max_j S_j^{\varepsilon}$ є скінченною випадковою величиною і його розподіл має вигляд [11, с. 458, 463]

$$\mathrm{P}(\max_j S_j^{\varepsilon} \in I) = \left(1 - \sum_{n=1}^{\infty} \tau_n^{\varepsilon}\right)\left(\mathbb{1}_{0 \in I} + \sum_{n=1}^{\infty} \mathrm{P}\left(S_j^{\varepsilon} > 0,\ 1 \le j \le n,\ S_j^{\varepsilon} \in I\right)\right),$$

де

$$\tau_n^{\varepsilon} = \mathrm{P}\left(S_j^{\varepsilon} \le 0,\ 1 \le j \le n-1,\ S_n^{\varepsilon} > 0\right),$$

I в даному випадку дорівнює (x, ∞) або $[x, \infty)$. За теоремою Спарре – Андерсена [11, с. 481]

$$\sum_{n=1}^{\infty} \tau_n^{\varepsilon} = 1 - \exp\left(-\sum_{n=1}^{\infty} \frac{\mathrm{P}(S_n^{\varepsilon} > 0)}{n}\right). \tag{3}$$

Ряд $\sum_{n=1}^{\infty} \mathrm{P}(S_n^{\gamma} > 0) < \infty$, оскільки $\mu^{\gamma} < 0$ [11, с. 483]. Тоді $\mathrm{P}(S_n^{\varepsilon} > 0)/n \le$ $\le \mathrm{P}(S_n^{\gamma} > 0)/n$ при $\varepsilon \le \gamma$ і ряд у правій частині (3) збігається рівномірно по ε в околі 0. $\mathrm{P}(S_j^{\varepsilon} > 0)$ неперервні по ε в нулі, сума ряду в правій частині неперервна по ε в нулі, отже, ряд у лівій частині неперервний по ε в нулі. Другий ряд теж збігається рівномірно по ε

$$\sum_{n=1}^{\infty} \mathrm{P}\left(S_j^{\varepsilon} > 0,\ 1 \le j \le n,\ S_j^{\varepsilon} \in I\right) \le \sum_{n=1}^{\infty} \mathrm{P}\left(S_j^{\gamma} > 0,\ 1 \le j \le n,\ S_n^{\gamma} \in I\right) < \infty,$$

тому, оскільки кожний з елементів неперервний по ε при $\varepsilon = 0$, сума ряду є неперервною по ε при $\varepsilon = 0$.

Лему доведено.

Лема 3. $\mathrm{P}\left(S_j^{\varepsilon} > 0,\ 1 \le j \le n;\ \max_j T_j^{\varepsilon} \le 0;\ S_n^{\varepsilon} > \max U_j^{-\varepsilon}\right)$ *неперервна по* ε *при* $\varepsilon = 0$.

Доведення. Маємо

$$\mathrm{P}\left(S_j^{\varepsilon} > 0,\ 1 \le j \le n;\ \max_j T_j^{\varepsilon} \le 0,\ S_n^{\varepsilon} > \max_j U_j^{-\varepsilon}\right) =$$

$$= \mathrm{P}\left(\max_j T_j^{\varepsilon} \le 0\right) \mathrm{P}\left(S_j^{\varepsilon} > 0,\ 1 \le j \le n;\ S_n^{\varepsilon} > \max_j U_j^{-\varepsilon}\right).$$

За лемою 2 $\mathrm{P}\left(\max_j T_j^{\varepsilon} \le 0\right)$ неперервна по ε у нулі. Доведемо, що

$$\mathrm{P}\left(S_j^{\varepsilon} > 0,\ 1 \le j \le n;\ S_n^{\varepsilon} > \max_j U_j^{-\varepsilon'}\right) \to \mathrm{P}\left(S_j > 0,\ 1 \le j \le n;\ S_n > \max_j U_j\right)$$

при $\varepsilon \to 0$, $\varepsilon' \to 0$.

Маємо

$$\left|\mathrm{P}\left(S_j^{\varepsilon} > 0,\ 1 \le j \le n;\ S_n^{\varepsilon} > \max_j U_j^{-\varepsilon'}\right) - \mathrm{P}\left(S_j > 0,\ 1 \le j \le n;\ S_n > \max_j U_j\right)\right| \le$$

$$\le \int \left|\mathrm{P}\left(S_j^{\varepsilon} > 0,\ 1 \le j \le n;\ S_n^{\varepsilon} > x\right) - \mathrm{P}\left(S_j > 0,\ 1 \le j \le n;\ S_n > x\right)\right| \mathrm{P}\left(U_j^{-\varepsilon'} \in dx\right) +$$

$$+ \left|\mathrm{P}\left(S_n > \max_j U_j^{-\varepsilon'}\right) - \mathrm{P}\left(S_n > \max_j U_j\right)\right|.$$

Перший доданок прямує до нуля за лемою 1 та теоремою про мажоровану збіжність. Запишемо і другий доданок в інтегральному вигляді:

$$\left|\mathrm{P}\left(S_n > \max_j U_j^{-\varepsilon'}\right) - \mathrm{P}\left(S_n > \max_j U_j\right)\right| \le$$

$$\le \int \left|\mathrm{P}\left(\max_j U_j^{-\varepsilon'} < x\right) - \mathrm{P}\left(\max_j U_j < x\right)\right| \mathrm{P}(S_n \in dx).$$

Цей доданок прямує до нуля за лемою 2 та теоремою про мажоровану збіжність.

Лему 3 доведено.

Завершимо доведення теореми. З леми 3 випливає, що при $n > 0$

$$\mathrm{P}\left(\hat{\hat{k}}_N - k = n\right) \to \mathrm{P}\left(S_j > 0,\ 1 \le j \le n;\ \max_j T_j \le 0;\ S_n > \max_j U_j\right).$$

Для випадків $n = 0$ та $n < 0$ доведення проводиться аналогічно.

Теорему доведено.

4. Випадок багатьох моментів зміни. Щоб поширити доведену теорему на випадок декількох змін, слід обмежити клас можливих оцінок щільності. Для цього нам знадобиться поняття функції зростання.

Означення 1. *Нехай S — клас підмножин множини X. Підпослідовністю, породженою множиною $A \in S$ з послідовності $\{x_1, \ldots, x_l\}$, називається множина, яка складається з x_j, що належать A. Позначимо через $\Delta^S(x_1, \ldots, x_l)$ кількість різних підпослідовностей послідовності $x_1, \ldots, x_l$, які породжують множини з S. Функцією зростання класу S називається*

$$m^S(l) = \max_{x_1,\ldots,x_l} \Delta^S(x_1, \ldots, x_l),$$

де $x_j \in X$.

Потрібна умова має вигляд:

4) нехай V_N — послідовність класів оцінок щільностей, $\hat{f}_N(\cdot, \omega)$, $\hat{g}_N(\cdot, \omega) \in V_N$, S_N — клас множин вигляду $\{f^*(x) > a\}$, $f^* \in V_N$, $a \in R$; тоді повинно виконуватись співвідношення

$$\exists\, 0 < \beta < 1 \quad \exists\, C < \infty: \quad \ln m^{S_N}(2N)/N^\beta < C.$$

Також слід посилити умову на штраф:

5) $\pi_N \sim CN^\alpha$, $(1 + \beta)/2 < \alpha < 1$.

Прикладом класу V_N, для якого виконується умова 4, може бути клас кусково-неперервних функцій, кількість інтервалів монотонності яких не перевищує $\nu_N \sim CN^\beta / \ln N$. Можна довести, що для відповідного класу S_N функція зростання обмежена $m^{S_N}(2N) \le 2^{\nu_N}(2N+1)^{2\nu_N}$, що достатньо для виконання умови 4 (доведення аналогічне наведеному в [12, с. 211].

Теорема 2. *Нехай виконано умови* 1 – 5, $\|f - \hat{f}\| \to 0$, $\|g - \hat{g}\| \to 0$. *Тоді* $\hat{\hat{\mathbf{k}}}$ *є консистентною за ймовірністю оцінкою* $\mathbf{k}$.

Доведення. Введемо позначення $\Phi_\zeta(1) = \mathrm{E} \ln f^*(\zeta)\big|_{f^* = \hat{f}}$, $\Phi_\zeta(2) = \mathrm{E} \ln g^*(\zeta)\big|_{g^* = \hat{g}}$, де f^* та g^* — невипадкові. Доведемо таку лему.

Лема 4. *Існують такі $\kappa > 0$ та $\varepsilon > 0$, що коли $\|f - f^*\| < \varepsilon$ та $\|g - g^*\| < \varepsilon$, то виконується*

$$\mathrm{E} \ln f^*(\xi) - \mathrm{E} \ln g^*(\xi) > \kappa, \qquad \mathrm{E} \ln g^*(\eta) - \mathrm{E} \ln f^*(\eta) > \kappa,$$

де ξ має розподіл F, η — розподіл G.

Доведення. Доведемо першу нерівність (друга доводиться аналогічно). Візьмемо $\varepsilon < \delta/2$, тоді, якщо $\|f - f^*\| < \varepsilon$, $\|g - g^*\| < \varepsilon$,

$$\left|\mathrm{E} \ln f^*(\xi) - \mathrm{E} \ln f(\xi)\right| \le \left\|\ln \frac{f^*}{f}\right\| < \max\left(\left\|\frac{f^*}{f} - 1\right\|, \left\|\frac{f}{f^*} - 1\right\|\right) < \frac{2}{\delta}\left\|f^* - f\right\|,$$

$$\left|\mathrm{E} \ln g^*(\xi) - \mathrm{E} \ln g(\xi)\right| < \frac{2}{\delta}\left\|g^* - g\right\|.$$

Отже, для будь-якого $\kappa > 0$ існує $\varepsilon = \min(\kappa\delta/4, \delta/2)$ таке, що коли $\| f - f^* \| < \varepsilon$, $\| g - g^* \| < \varepsilon$, то

$$\left| \mathrm{E} \ln f^*(\xi) - \mathrm{E} \ln f(\xi) \right| < \frac{\kappa}{2}, \qquad \left| \mathrm{E} \ln g^*(\xi) - \mathrm{E} \ln g(\xi) \right| < \frac{\kappa}{2}.$$

А оскільки $\mathrm{E} \ln f(\xi) > \mathrm{E} \ln g(\xi)$, то існує $\kappa > 0$ таке, що $\mathrm{E} \ln f(\xi) - \mathrm{E} \ln g(\xi) > 2\kappa$. Отже, якщо $\| f - f^* \| < \varepsilon$, $\| g - g^* \| < \varepsilon$, шукане співвідношення виконується.

Лему доведено.

З леми випливає, що коли існують ε та κ, які залежать від δ, такі, що коли $\| f - \hat{f} \| < \varepsilon$ та $\| g - \hat{g} \| < \varepsilon$, то $\Phi_\xi(1) - \Phi_\xi(2) > \kappa$ та $\Phi_\eta(2) - \Phi_\eta(1) > \kappa$.

Далі вважатимемо $\varepsilon < \delta/2$. Нехай для послідовностей $a_N > 0$ та π_N виконуються такі співвідношення:

$$a_N < \frac{\pi_N}{2R+3}, \qquad \pi_N < \frac{\kappa\tilde{\theta}N}{2(R+1)}. \tag{4}$$

Зауважимо, що при $\pi_N \sim CN^\alpha$, $\alpha < 1$, останнє співвідношення буде виконуватися починаючи з деякого N. Позначимо через B_N множину

$$\left\{ R = \hat{\hat{R}},\ \hat{\hat{h}}_{\hat{k}_i} = h_{k_i},\ \left| \theta_i - \hat{\hat{\theta}}_i \right| \le \frac{2(R+1)a_N}{\kappa N},\ i = 0, \ldots, R \right\}.$$

Скористаємося доведенням теореми з [7]. Воно стосується функцій ϕ, які не залежать від вибірки, але викладки, наведені там, від цього не змінюються. З нього випливає, що за умов (4) та леми 4 ймовірність

$$\mathrm{P}(B_N) \ge \mathrm{P}\left(B_N, \| f - \hat{f} \| < \varepsilon, \| g - \hat{g} \| < \varepsilon\right) \ge \mathrm{P}(A_N),$$

де

$$A_N = \bigcap\nolimits_{h=1,2} A_N^h,$$

$$A_N^h = \left\{ \max_{1 \le l_1 \le l_2 \le N} \left| \sum_{j=l_1}^{l_2} (\phi(\zeta_j, h) - \Phi_j(h)) \right| < a_N,\ \| f - \hat{f} \| < \varepsilon,\ \| g - \hat{g} \| < \varepsilon \right\}.$$

Оцінимо $\mathrm{P}(\overline{A}_N^h)$:

$$\mathrm{P}(\overline{A}_N^h) \le \sum_{1 \le l_1 \le l_2 \le N} p_h(l_1, l_2, N) + \Delta_{\hat{f}}^\varepsilon + \Delta_{\hat{g}}^\varepsilon,$$

де

$$p_h(l_1, l_2, N) = \mathrm{P}\left(\sup_{\| f^* - f_h \| < \varepsilon,\ f^* \in V_N} \left| \sum_{j=l_1}^{l_2} \left(\ln f^*(\zeta_j) - \mathrm{E} \ln f^*(\zeta_j) \right) \right| \ge a_N \right).$$

Нагадаємо, що $f_1 = f$, $f_2 = g$. Щоб оцінити $p_h(l_1, l_2, N)$, застосуємо нерівність Вапніка – Червоненкіса [12, с. 229]

$$\mathrm{P}\left(\left| \sup_{A \in S} \left(\frac{1}{N} \sum_{j=1}^{N} \xi_{j \in A} - \mathrm{P}(A) \right) \right| > \varepsilon \right) < 6m^S(2N)e^{-\varepsilon^2(N-1)/2},$$

де $\xi_1, \ldots, \xi_N$ — незалежні однаково розподілені випадкові величини, S — клас підмножин множини X, на якій набувають значень ξ_j.

Лема 5. *Починаючи з деякого* a_N *виконується нерівність*

$$p_h(l_1, l_2, N) \le 12 m^{S_N}(2N) e^{-a_N^2/(16N(M+\ln|\delta/2|+\delta)^2)}.$$

Доведення. Доведемо лему для $p_1(l_1, l_2, N)$ (для $p_2(l_1, l_2, N)$ виконується таке ж співвідношення). Використаємо **теорему 13.1** з [12, с. 286, 287]:

Нехай $F(x, \alpha)$, $\alpha \in \Omega$, *— сім'я вимірних на* X *функцій, причому виконано умову* $0 \le F(x, \alpha) \le a$. *Розглянемо систему* S *подій вигляду* $A = \{x : F(x, \alpha) \ge C\}$. *При цьому виконується співвідношення*

$$\sup_{\alpha} \left| \mathrm{E} F(\xi_1, \alpha) - \frac{1}{n} \sum_{j=1}^{n} F(\xi_j, \alpha) \right| \le a \sup_{\alpha \in S} \left| \mathrm{P}(A) - \frac{1}{n} \sum_{j=1}^{n} 1\!\!1_{\xi_j \in A} \right|,$$

де $\xi_1, \ldots, \xi_n$ *— незалежні однаково розподілені величини.*

Використовуючи незалежність $\zeta_1, \ldots, \zeta_N$, можна записати

$$\mathrm{P}\left(\sup_{\|f^* - f\| < \varepsilon,\, f^* \in V_N} \left| \sum_{j=1}^{N} \left(\ln f^*(\zeta_j) - \mathrm{E} \ln f^*(\zeta_j) \right) \right| > a_N \right) \le$$

$$\le \mathrm{P}\left(\sup_{\|f^* - f\| < \varepsilon,\, f^* \in V_N} \left| \sum_{j=1}^{s} \left(\ln f^*(\xi'_j) - \mathrm{E} \ln f^*(\xi'_j) \right) \right| > \frac{a_N \sqrt{s}}{\sqrt{s} + \sqrt{N-s}} \right) +$$

$$+ \mathrm{P}\left(\sup_{\|f^* - f\| < \varepsilon,\, f^* \in V_N} \left| \sum_{j=s+1}^{N} \left(\ln f^*(\eta'_j) - \mathrm{E} \ln f^*(\eta'_j) \right) \right| > \frac{a_N \sqrt{N-s}}{\sqrt{s} + \sqrt{N-s}} \right),$$

де ξ'_j мають розподіл F, η'_j — розподіл G, s — кількість величин з розподілом F. Застосуємо теорему 13.1 до обох доданків, взявши за S множину рівнів функцій $\ln f^*(x) - \ln(\delta/2) > 0$, обмежених зверху $L = M + \varepsilon - \ln(\delta/2)$. Отже, умови теореми виконуються. Тоді

$$\mathrm{P}\left(\sup_{\|f^* - f\| < \varepsilon,\, f^* \in V_N} \left| \sum_{j=1}^{N} \left(\ln f^*(\xi'_j) - \mathrm{E} \ln f^*(\xi'_j) \right) \right| > \frac{a_N \sqrt{s}}{\sqrt{s} + \sqrt{N-s}} \right) \le$$

$$\le \mathrm{P}\left(\sup_{\{f^*(\xi) > a_N\}} \left| \sum_{j=1}^{N} \left(1\!\!1_{f^*(\xi'_j) > a_N} - \mathrm{P}(f^*(\xi'_j) > a_N) \right) \right| > \frac{a_N \sqrt{s}}{(\sqrt{s} + \sqrt{N-s}) L} \right).$$

Для ймовірностей з $f^*(\eta'_j)$ міркування аналогічні. При $a_N / (M + |\ln \delta/2| + \varepsilon) > 2$ (це співвідношення буде виконуватись при досить великому N) ці ймовірності можна оцінити за допомогою нерівності Вапніка – Червоненкіса [12, с. 229 – 231]:

$$\mathrm{P}\left(\sup_{\{f^*(\xi) > a_N\}} \left| \sum_{j=1}^{N} \left(1\!\!1_{f^*(\xi'_j) > a_N} - \mathrm{P}(f^*(\xi'_j) > a_N) \right) \right| > \frac{a_N \sqrt{s}}{(\sqrt{s} + \sqrt{N-s}) L} \right) +$$

$$+ \mathrm{P}\left(\sup_{\{f^*(\eta) > a_N\}} \left| \sum_{j=1}^{N} \left(1\!\!1_{f^*(\eta'_j) > a_N} - \mathrm{P}(f^*(\eta'_j) > a_N) \right) \right| > \frac{a_N \sqrt{N-s}}{(\sqrt{s} + \sqrt{N-s}) L} \right) \le$$

$$\le 12 m^{S_N}(2N) e^{-a_N^2/(16NL^2)}.$$

Лему доведено.

Оцінимо ймовірність A_n^h:

$$\mathrm{P}(\overline{A}_N^h) \le \sum_{l_1 \le l_2} p_h(l_1, l_2, N) + \Delta_{\hat{f}}^{\varepsilon} + \Delta_{\hat{g}}^{\varepsilon} \le$$

$$\le 12 N^2 m^{S_N}(2N) e^{-a_N^2/(16N(M+\ln|\delta/2|+\varepsilon)^2)} + \Delta_{\hat{f}}^{\varepsilon} + \Delta_{\hat{g}}^{\varepsilon}.$$

Оскільки за умов теореми відношення $\ln m^{S_N}(2N)/N^{\beta}$ обмежене, то за умовою на штраф права частина прямує до нуля, отже, $\mathrm{P}(\overline{A}_N) \to 0$ і $\hat{\hat{\mathbf{k}}}$ є консистентною оцінкою $\mathbf{k}$.

Теорему 2 доведено.

Теорема 3. *Нехай виконуються умови* 1 – 5, *тоді при* $1 \le j \le R$

$$\lim_{N\to\infty} \mathrm{P}\Big(\hat{\hat{k}}_{j,N} - k_j = p\Big) = \lim_{N\to\infty} \mathrm{P}\big(\hat{k}_{j,N} - k_j = p\big).$$

Доведення. Скористаємося методом, описаним у [9]. Позначимо

$$\mathcal{H}(d_N) = \big\{ h = h\,(h_1, \ldots, h_N) \,\big|\, R(h) = R,\ k_j(h) \in (k_j - d_N,\ k_j + d_N],\ j \in [1, R] \big\},$$

де $d_N = 2(R+1)a_N/\kappa$ взято з попередньої теореми. За допомогою траєкторії $\check{h} = \arg\min_{h \in \mathcal{H}(d_N)} J(h)$ побудуємо допоміжні оцінки $\check{k}_j = k_j(\check{h})$. Розглянемо події

$$B_N = \left\{ \hat{R} = R,\ \hat{\hat{h}}_{\hat{k}_j} = h_{k_j}^0,\ \hat{\hat{k}}_j \in [k_j - d_N,\ k_j + d_N] \right\}.$$

З доведення попередньої теореми випливає, що $\mathrm{P}(B_N) \to 1$, відповідно $\mathrm{P}(\overline{B}_N) \to$ $\to 0$. Зауважимо, що $\{\check{k}_i \ne \hat{\hat{k}}_i\} \subset \overline{B}_N$, тобто достатньо знайти асимптотичний розподіл $\check{k}_{j,N} - k_j$, який і буде шуканим розподілом. Як доведено в [2], розподіли $\check{k}_{j,N} - k_j$ можна шукати окремо один від одного, і вони дорівнюють відповідним розподілам для випадку одного моменту зміни. Отже, оскільки найкращий розподіл теж має таку властивість, то за теоремою 1

$$\lim_{N\to\infty} \mathrm{P}\Big(\hat{\hat{k}}_{j,N} - k_j = n\Big) = \lim_{N\to\infty} \mathrm{P}\big(\hat{k}_{j,N} - k_j = n\big).$$

Теорему 3 доведено.

5. Оцінювання щільностей. Наведемо приклад побудови оцінок щільності $\hat{f}$, $\hat{g}$ за даними $\zeta_1, \zeta_2, \ldots, \zeta_N$. Розглянемо гістограмні оцінки щільності.

Нехай консистентна оцінка моментів зміни (наприклад, медіанна, якщо медіани розподілів не збігаються [13]) є відомою. Побудуємо за нею оцінки точок зміни. Оскільки ми вважаємо, що дані мають лише два розподіли, можна утворити з них дві вибірки. До цих вибірок дані попадають відповідно до того, на якому проміжку між оцінками моментів зміни вони знаходяться. Потім за цими вибірками побудуємо гістограмні оцінки щільності. Зрозуміло, що дані у вибірках взагалі не є однаково розподіленими, але у випадку консистентності оцінки моментів зміни можна довести, що такі оцінки будуть консистентними.

Також потрібно розглянути умову на клас щільностей, якому повинні належати оцінки. Можна довести, що для класу V_l гістограмних оцінок, побудованих на не більш ніж l проміжках розбиття, функція зростання $m^{S_l}(N) \le$ $\le 2^l(N+1)^l$ (доведення аналогічне наведеному в [12, с. 211]). Отже, для виконання умови 4 необхідно на кожному етапі брати не більш ніж $CN^{\beta}/\ln N$, $0 < \beta < 1$, відрізків розбиття. Як штраф можна взяти $\pi_N \sim CN^{\alpha}$, $(1+\beta)/2 < \alpha < 1$.

6. Висновки. У статті розглянуто задачу знаходження асимптотично оптимальних оцінок багатьох моментів зміни у випадку неповної інформації про розподіли. Доведено, що оцінка максимальної вірогідності, яка є асимптотично оптимальною, залишається такою при підстановці оцінок щільності замість справжніх значень.

1. *Боровков А. А.* Асимптотически оптимальные решения в задаче о разладке // Теория вероятностей и ее применения. – 1998. – **43**, вып. 4. – С. 625 – 654.
2. *Yakir B., Krieger A. M., Pollak M.* Detecting change in regression: first order optimality // Ann. Statist. – 1999. – **27**, № 6. – P. 1896 – 1913.
3. *Vincent Poor H.* Quickest detection with exponential penalty for delay // Ibid. – 1998. – **26**, № 6. – P. 2179 – 2205.
4. *Raimondo M.* Minimax estimation of sharp change points // Ibid. – № 4. – P. 1379 – 1397.
5. *Loader C. R.* Change point estimation using nonparametric regression // Ibid. – 1996. – **24**, № 4. – P. 1667 – 1678.
6. *Hall P., Rau C.* Tracking a smooth fault line in a response surface // Ibid. – 2000. – **28**, № 3. – P. 713 – 733.
7. *Сугакова О. В.* Пошук моментів зміни в потоці незалежних спостережень // Теорія ймовірностей та мат. статистика. – 1996. – Вип. 55. – С. 181 – 186.
8. *Моттль В. В., Мучник И. Б., Яковлев В. Г.* Оптимальная сегментация экспериментальных кривых // Автоматика и телемеханика. – 1983. – **8**. – С. 84 – 95.
9. *Майборода Р. Є., Сугакова О. В.* Граничний розподіл ДП-оцінок багатьох моментів зміни // Теорія ймовірностей і мат. статистика. – 2003. – Вип. 69. – С. 96 – 105.
10. *Майборода Р. Е.* Медианная оценка разладки в случае слабозависимых наблюдений // Теория вероятностей и мат. статистика. – 1990. – Вып. 43. – С. 87 – 91.
11. *Феллер В.* Введение в теорию вероятностей и ее приложения: В 2 т. – М.: Наука, 1967. – Т. 2. – 752 с.
12. *Вапник В. Н., Червоненкис А. Я.* Теория распознавания образов (статистические проблемы изучения). – М.: Наука, 1974. – 416 с.
13. *Шуренков Г. В.* Асимптотика медіанної оцінки моментів зміни // Теорія ймовірностей і мат. статистика. – 2004. – Вип. 70. – С. 149 – 156.

Одержано 07.07.2004,
після доопрацювання — 18.02.2005