Академія наук України
Інститут математики

На правах рукопису

МАЙБОРОДА Ростислав Євгенович

НЕПАРАМЕТРИЧНА СТАТИСТИКА НЕОДНОРІДНИХ СПОСТЕРЕЖЕНЬ

01.01.05 — теорія ймовірностей
та математична статистика

АВТОРЕФЕРАТ
дисертації на здобуття наукового ступеня
доктора фізико-математичних наук

Київ — 1994

Дисертацією є рукопис.

Робота виконана у Київському університеті імені Тараса Шевченка.

Офіційні опоненти	-	доктор фізико-математичних наук, професор БУЛДИГІН В.В. доктор фізико-математичних наук, старший науковий співробітник КНОПОВ П.С. доктор технічних наук, професор ПОПОВ Ю.Д.
Провідна організація	-	Інститут прикладної математики та механіки АН України

Захист відбудеться "11" жовтня 1994р. о 15 годині на засіданні спеціалізованої ради Д 016.50.01 при Інституті математики АН України за адресою: 252601, Київ, вул. Терещенківська 3, конференц-зал.

З дисертацією можна ознайомитись у бібліотеці інституту.

Автореферат розіслано "____"____________ 1994р.

Вчений секретар
спеціалізованої ради ГУСАК Д.В.

Актуальність теми

Непараметрична статистика і теорія емпіричних функцій, започатковані дослідженнями А.Н. Колмогорова, Р. фон-Мізеса, В.І. Глівенко, Дж.Л. Дуба, розвинуті у роботах М. Донскера, Й.І. Гіхмана, В.С. Королюка, П. Білінгслі, Р.М. Дадлі, В.І. Колчинського та інших, нині являють собою одну з основних галузей математичної статистики. Класичні роботи в цій області присвячені випадку незалежних, однаково розподілених спостережень. Із задач непараметричного оцінювання по неоднаково розподілених спостереженнях найбільш докладно досліджене знаходження моментів розладки. В цій задачі розподіл даних є сталим до деякого моменту, а потім стрибком змінюється, після чого знов залишається сталим. Оцінки моменту зміни розподілу (моменту розладки) досліджувались Дж. Пейджем, К. Кемпом, Ж. Дее та Д. Пікаром, А.Н.Ширяєвим, Д.В. Хінклі, Б.С. Дарховським та Б.Е. Бродським, Л. Телкснісом та Н. Клігене, А.Ф. Ронжиним, Б. Джеймсом, К.Л. Джеймсом і Д. Зигмундом, М. Чорго та Л. Хорватом, Л. Гірайтисом та Р. Лейпусом та ін.

При статистичному аналізі біологічних, геологічних та соціологічних даних часто виникає потреба оцінки характеристик по спостереженнях над сумішами кількох компонент із різними ймовірнісними розподілами, причому концентрації компонент змінюються протягом спостережень. Задачі такого типу можна розглядати як узагальнення оцінки моменту розладки на випадок "поступової" або "повільної" розладки. З другого боку, ця задача є узагальненням на випадок неоднаково розподілених спостережень задачі розщеплення сумішей. У випадку однаково розподілених даних цю задачу досліджували Х. Тейчер, С.А. Якович та Дж. Спрагінс, М.І. Шлезінгер, Ю.М.Рижов та інші. Слід відмітити, що для однаково розподілених даних ця задача практично не допускає непараметричної постановки.

У даній роботі розглядаються задачі непараметричного оцінювання розподілів компонент суміші з концентраціями, що змінюються, і параметричного та непараметричного оцінювання функцій концентрації. Для оцінки розподілів використовуються зважені емпіричні міри, запропоновані С.Дж. Стоуном для задач непараметричної регресії. Такі міри в зв'язку з задачами класифікації розглядались Л. Девроєм та Л. Дерфі. Звичайні емпіричні функції розподілу для неоднорідних спостережень вивчав М.С.А. ван-Зуйлен. З останніх робіт по зважених емпіричних функціях неоднорідних спостережень відмітимо статті Гуан Жонга та М. Харела і М.Л. Пурі.

При оцінці функцій концентрації разом із зваженими емпіричними мірами використовуються зважені функціонали фон-Мізеса. Дослідженням таких функціоналів, розпочатим В. Гьофдингом та Р. фон-Мізесом, присвячено багато літератури, основні результати якої можна знайти у монографії В.С. Королюка та Ю.В. Боровських. Техніка побудови та дослідження оцінок параметрів функцій концентрації близька до методу найменших квадратів у нелінійних регресійних задачах (див. роботи А.Я.Дороговцева, Н.Н. Леоненка та А.В. Іванова, Б.Л.С. Пракаса Рао, Дж. Шао).

При аналізі реальних даних не можна не враховувати, що результатами спостережень часто є значення, виміряні з деякою похибкою. Якщо вважати, що ця похибка є незалежною і адитивною, при оцінюванні розподілу даних виникає задача деконволюції (розгортки). Для незалежних, однаково розподілених спостережень таку задачу розглядали Л. Деврой, Л.А. Стефанські, А.Дж. ван-Ес, Дж. Фан. При цьому замість емпіричних мір природно використовувати емпіричні характеристичні функції та емпіричні генератриси моментів. Дослідження поведінки таких функцій для однорідних спостережень (без похибок) знаходимо у роботах Д.Дж. Кенделла, Дж.Т. Кента, А. Февервегера та Р.А. Мурейки, М. Леду, С. Чорго, Дж.Е. Юкіча. Емпіричні характеристичні функції для неоднорідних спостережень з неоднорідними похибками раніше, наскільки нам відомо, не розглядались.

Інша модель даних виникає у випадку, коли похибка спостережень залежить від значення попередніх вимірювань. Ця модель пов'язана з необхідністю враховувати інерційність вимірюючого приладу. Такий підхід приводить до оцінювання функціональних характеристик випадкового процесу, що є розв'язком деякого стохастичного диференційного рівняння. Відомі дослідження у цій галузі Р.Ш. Ліпцера та А.Н. Ширяева, П.С. Кнопова, Ю.Н. Лінькова. Дискретним аналогом таких процесів є процеси авторегресії, для яких оцінюванням моменту розладки займались Л. Телксніс, И. Клігене. Відмітимо роботу Л. Телксніса, де розглянуто задачу оцінки "поступової" розладки.

Застосований у даній роботі підхід спирається на використання для оцінювання теорем бакстерівського типу (які вивчались П. Леві, Г. Бакстером, Е.Г. Гладишевим, Ю.М. Рижовим, Ю.В. Козаченко, В.В. Булдигіним та ін.) Вперше оцінки такого типу були застосовані (до стаціонарних процесів) у роботі М. Арато, А.Н. Колмогорова та Я.Г. Сіная.

У практичних задачах часто важливо не тільки оцінити концентрації та розподіли компонент у суміші, а й вміти класифікувати об'єкти, тобто відносити їх до тої чи іншої компоненти. Ця задача (інколи її звуть статистичним розпізнаванням образів) тісно пов'язана з оцінюванням щільностей розподілів компонент. Непараметричне оцінювання щільностей та класифікація були предметом великої кількості робіт, починаючи від М. Розенблата та Е. Парзена, огляд яких див. у монографіях В.Н. Вапника та А.Я. Червоненкіса, Л. Девроя та Л. Дерфі. Г.К. Голубев запропонував для оцінки щільностей (в однорідному випадку) використовувати емпіричну характеристичну функцію, "фільтруючи" її подібно до того, як це робиться у спектральному оцінюванні. Ця ідея використана у даній роботі для оцінки щільності розподілу по неоднорідних даних з неоднорідною похибкою.

Для доведення асимптотичних результатів у роботі використано загальну теорію випадкових елементів функціональних просторів, розроблену, зокрема, у роботах А.Н. Колмогорова, Й.І. Гіхмана, А.В. Скорохода, Дж. Келбса, П.Дж. Бікела та М.Дж. Вічури, В.В. Булдигіна, Е.І. Островського та Ю.В. Козаченка.

Мета роботи

Розробити методи непараметричного статистичного оцінювання по спостереженнях над сумішами із концентраціями, що змінюються, включаючи випадок вимірювань з похибками. Побудувати оцінки функцій розподілу, концентрацій компонент, класифікатори спостережень, та дослідити їх асимптотичну поведінку. Побудувати оцінки інтенсивностей для вимірювань з інерційною похибкою.

Наукова новизна

Для оважених неоднорідних емпіричних мір та емпіричних функцій доведено нові асимптотичні та неасимптотичні результати (рівномірна збіжність м.н., оцінка швидкості збіжності м.н., асимптотична нормальність, нерівності для розподілів супремуму), що узагальнюють відомі теореми для однорідного випадку (теорема Гливенко-Кантеллі, теорема Донскера, нерівність Вапника-Червоненкіса). Отримано нормальну асимптотику для оважених функціоналів фон-Мізеса від неоднорідних спостережень.

На основі цих результатів побудовано незміщені непараметричні оцінки для розподілів, характеристичних функцій та генератрис моментів компонент сумішей з змінними концентраціями. Досліджено умови їх рівномірної спроможності та ефективності. Розроблено L_2 та sup методи оцінювання параметрів функцій концентрації, доведено спроможність оцінок, отриманих цими методами. Для оцінок L_2-методом доведено умови асимптотичної нормальності у випадку точних спостережень. Побудовано рівномірні надійні проміжки для розподілів компонент сумішей і асимптотичні та неасимптотичні надійні проміжки для параметрів функцій концентрації. Для двокомпонентних сумішей побудовано непараметричні оцінки функцій концентрації.

У задачі пошуку моментів розладки запропоновано нові квантильні методи, що дозволяють поєднувати оцінювання із зниженням розмірності вибірки. Побудовано непараметричний критерій однорідності вибірки, досліджено його асимптотичні властивості. Доведено умови спроможності квантильних оцінок моментів розладки. Для квантильної оцінки отримано негаусів асимптотичний розподіл.

Побудовано асимптотично баєсівські класифікатори на основі спостережень по сумішах із концентраціями, що змінюються, та оцінки щільностей розподілу, спроможні у L_1, L_2 та sup нормі.

Побудовано спроможні оцінки інтенсивності шуму у задачі оцінювання по спостереженнях з інерційною похибкою, доведено їх асимптотичну нормальність.

Апробація роботи

Результати роботи доповідались на I Всесвітньому конгресі товариства ім. Бернулі (Ташкент, 1986), II та III Всесоюзних семінарах по виявленню змін властивостей випадкових процесів (Звенигород, 1988; Воронеж, 1990), VI Всесоюзному семінарі з непараметричних та робастних статистичних методів у кібернетиці та інформатиці (Іркутськ, 1990), IV Всесоюзній школі-семінарі "Статистичний та дискретний аналіз даних та експертне оцінювання" (Одеса, 1991), Міжнародній конференції пам'яті М.П.Кравчука (Київ–Луцьк, 1992), Міжнародній конференції Chan'De-92 (Київ, 1992), Українсько-угорській конференції "Нові напрями у теорії ймовірностей та математичній статистиці" (Мукачеве, 1992), V та VI Міжнародній Вільнюській конференції з теорії ймовірностей та математичної статистики (Вільнюс, 1989, 1993), III Донецькій міжнародній конференції "Ймовірнісні моделі процесів в управлінні та надійності" (Мелекіно, 1993), на семінарах з теорії ймовірностей та математичної статистики в Київському, Львівському, Московському, Донецькому університетах, Київському політехнічному інституті, інститутах математики та кібернетики АН України, Інституті математики Російської академії наук.

Публікації

Основні результати опубліковані у роботах [1–28].

Об'ем і структура роботи

Робота складається з вступу, шести розділів і списку літератури. Об'ем роботи — 190с машинопису.

Зміст роботи

Перший розділ присвячено дослідженню асимптотики зважених емпіричних мір, побудованих по неоднорідній вибірці. Нехай $\Xi = \{\xi_j^N, j = 1 \div N, N \in \mathbb{N}\}$ — схема серій незалежних у кожній серії випадкових елементів деякого вимірного простору $(\Delta, \mathfrak{A})$ таких, що

$$\mathsf{P}\{\xi_j^N \in A\} = \mu(A, t_j^N), \tag{1}$$

де $\mu : \mathfrak{A} \times [0,1] \to [0,1]$ — деяка функція, яку ми називатимемо розподілом Ξ, t_j^N — "час спостереження" ξ_j^N, $0 \le t_1^N < t_2^N < \ldots < t_N^N \le 1$. Ми будемо казати, що сітка часу для спостережень Ξ рівномірна, якщо $t_j^N = \frac{j}{N}$.

Якщо у (1) функція μ може бути представлена у вигляді

$$\mu(A,t) = \sum_{i=1}^{M} H_i(A) w_i(t), \tag{2}$$

де H_i — ймовірнісні міри на Δ, $w_i : [0,1] \to [0,1]$, $\sum_i w_i(t) = 1$, будемо казати, що Ξ являє собою вибірку з суміші M компонент, H_i назвемо розподілами, w_i — концентраціями компонент у суміші.

Зваженою емпіричною мірою з вагою $a(t)$, побудованою по Ξ, назвемо

$$\hat{\mu}_N(A,a) = \frac{1}{N} \sum_{j=1}^{N} a(t_j^N) \chi\{\xi_j^N \in A\}$$

($\chi(A)$ — індикатор події A).

У п. 1.1 $\mu_N(A,a^i)$ розглядається як оцінка невідомих H_i по $\Xi_N = \{\xi_j^N, j = 1 \div N\}$ у випадку, коли w_i відомі повністю. Доведено, що коли функції w_i лінійно незалежні, то $\hat{\mu}_N(A,a^l)$ з ваговою функцією

$$a^l(t) = \frac{1}{\det \Gamma_N} \sum_{k=1}^{M} (-1)^{l+k} \gamma_{lk}^N w_k(t), \tag{3}$$

де $\Gamma_N = (\langle w_k, w_l \rangle_N)_{k,l=1}^M$ – матриця Грама системи функцій (векторів) $w_k, k = 1 \div M$, γ_{lk}^N – (l,k)-й мінор матриці Γ_N, є незміщеною оцінкою H_l, ефективною в класі всіх незміщених оцінок. Як приклад розглянуто двокомпонентну суміш з $w_1(t) = t$, $w_2(t) = 1 - t$ і критерій для перевірки гіпотези $H_1 = H_2$ для цієї суміші.

У п. 1.2 для $\hat{\mu}(A,a)$ доведено узагальнену теорему Глівенко-Кантеллі. У п. 1.3 для емпіричних функцій розподілу (тобто для випадку, коли $\Delta = \mathbb{R}^d$, $A \in \mathbb{D}$, де $\mathbb{D}$ — клас прямокутників у $\mathbb{R}^d$) отримано оцінку швидкості збіжності м.н. у теоремі

Глівенко-Кантеллі та асимптотичну нормальність рівномірно по $A \in \mathbf{D}$ та $a(\cdot) = a(\cdot, v)$, $v \in V$. У п. 1.4. для сумішей з концентраціями, що змінюються, доведено аналог нерівності Вапника-Червоненкіса.

У другому розділі розглядаються асимптотичні властивості емпіричних функцій. У п. 2.1 розглянуто функції, що є інтегральними перетвореннями емпіричних мір, і отримано оцінки швидкості їх збіжності, що випливають з результатів розділу 1. П. 2.2. має допоміжний характер, тут вміщено відомості про соболевські простори, які використовуються у п. 2.3.

У п. 2.3. досліджується поведінка емпіричних характеристичних функцій та емпіричних генератрис моментів, побудованих по спостереженнях з похибкою. Вважається, що спостерігаються не безпосередньо величини $\xi_j^N \in \mathbb{R}^d$, а результати їх вимірювання з незалежною адитивною похибкою η_j^N, тобто $\zeta_j^N = \xi_j^N + \eta_j^N$. Характеристичні функції η_j^N — $\varphi_j^N(s)$ вважаються відомими, $\mathbf{E}\eta_j^N = 0$. Емпірична характеристична функція з вагою $a(t)$, побудована по спостереженнях ζ_j^N, задається виразом

$$k_N^*(s,a) = \frac{1}{N}\sum_{j=1}^N a(t_j^N)\frac{\exp(i(s,\zeta_j^N))}{\varphi_j^N(s)},$$

$s \in \mathbb{R}^d$, а емпірична генератриса моментів — $f_N^*(s,a) = k_N^*(-is,a)$, $s \in \mathbb{R}^d$. Поведінка $k_N^*(s,a)$ та $f_N^*(s,a)$ досліджується у просторах $L_2(\mathbb{R}^d, \pi)$, де π — деяка ймовірнісна міра, та у $C[c,d]$, $C_p(\mathbb{R}^d)$, де p — вагова функція на $\mathbb{R}^d$. Як приклад наведемо теорему 2.3.4.

Теорема 2.3.4. *Нехай існують такі матриці Q_1, Q_2, S, що для всіх j, N $\mathbf{E}\exp((s,\xi_j^N)) \le C\exp\left(\frac{1}{2}(Q_1 s,s)\right)$, $\mathbf{E}\exp((s,\eta_j^N)) \le C\exp\left(\frac{1}{2}(Q_2 s,s)\right)$, $p^0(s) = \exp\left(-\frac{(Ss,s)}{2}\right)$, $\bar{f}_N(s,a) = \mathbf{E}f_N^*(s,a)$. Якщо*

$$\sup_{v\in V}\operatorname{Var}_t a_N(t,v) \le A_N < \infty, \sup_{v,t}|a_N(t,v)| < A_N,$$

і матриця $S - 2Q_1 - (d+2)Q_2$ є (строго) додатно визначеною, *то для деякого* $\gamma > 0$ *та деякого* $\beta < \infty$

$$\mathbf{P}\{ \sup_{s\in\mathbb{R}^d, v\in V} |p^0(s)(f_N^*(s,a_N(\cdot,v)) - \bar{f}_N(s,a_N(\cdot,v)))| \le A_N\beta\sqrt{\frac{\ln\Lambda}{N}}\}$$

$$= O(N^{-\gamma}),$$

$N \to \infty$.

Якщо матриця $S - 4Q_1 - (d+4)Q_2$ *— строго додатно визначена, то існує така в.в.* $\Lambda < \infty$, *що*

$$\sup_{s\in\mathbb{R}^d, v\in V} |p^0(s)(f_N^*(s,a_N(\cdot,v)) - \bar{f}_N(s,a_N(\cdot,v)))| \le \Lambda A_N\sqrt{\frac{\ln N}{N}}.$$

У п. 2.4. розглядаються зважені функціонали фон-Мізеса

$$U_N(a_0,a_1) = \frac{1}{N^2}\sum_{i,j=1}^{N} a_0(t_i^N)a_1(t_j^N)\Phi(\xi_j^N,\xi_i^N),$$

де $\Phi : \Delta\times\Delta \to \mathbb{R}$ — ядро статистики U_N. Такі функціонали виникають при розгляді скалярних добутків зважених емпіричних функцій розподілу у $L_2(\pi)$. Доведено умови асимптотичної нормальності $U\ (a_0(\cdot,v), a_1(\cdot,v))$ рівномірно по $v \in V$ у випадку, коли ξ_j^N — вибірка із суміші з змінними концентраціями.

Третій розділ присвячений оцінкам функцій концентрації $w_i(t)$ по спостереженнях ξ_j^N або ζ_j^N. Сітка часу тут вважається рівномірною. Розподіли компонент H_k вважаються повністю невідомими, отже ця задача є непараметричною навіть тоді, коли концентрації w_k задано параметрично. У п. 3.1. розглянуто загальну схему побудови оцінок параметрів ϑ для функцій $w_i(t) = w_i(t,\vartheta)$, $\vartheta\in\Theta$. Вважається, що при кожному $\alpha\in\Theta$, $\{w_i(\cdot,\alpha)\}_{i=1}^M$ являє собою систему лінійно незалежних функцій. Тоді можна вибрати ортонормований базис $\{u(t,\alpha), t \in [0,1]\}_{k=0}^{M-1}$ замкненої лінійної оболонки $\{w_i(\cdot,\alpha)\}_{i=1}^M$ у $L_2([0,1],dt)$, такий, що $u_0(t,\alpha)\equiv 1$. Нехай спостерігаються ξ_j^N. Виберемо деякий підклас $\mathfrak{F}\subseteq\mathfrak{A}$. Задамо деяку σ-алгебру $\mathfrak{S}$ підмножин $\mathfrak{F}$ та π — $\mathfrak{S}$-вимірну міру на $\mathfrak{F}$. Позначимо

$$R_N(\alpha) = \sum_{k=1}^{M-1}\int_{\mathfrak{F}} (\hat{\mu}_N(A, u_k(\cdot,\alpha)))^2\pi(dA), \qquad (4)$$

$$R_N^{\max}(\alpha) = \sup_{A\in\mathfrak{F}} \sum_{k=1}^{M-1} (\hat{\mu}_N(A, u_k(\cdot,\alpha)))^2, \qquad (5)$$

$$\vartheta_N \in \operatorname*{argmax}_{\alpha\in\Theta}(R_N(\alpha), \lambda_N),\ \vartheta_N^{\max} \in \operatorname*{argmax}_{\alpha\in\Theta}(R_N^{\max}(\alpha), \lambda_N),$$

де $\operatorname{argmax}_{\alpha\in\Theta}(R(\alpha), \lambda) = \{\alpha \in \Theta :\ R(\alpha) \geq \sup_{\beta\in\Theta} R(\beta) - \lambda\}$, λ_N — деяка числова пос. довність, $\lambda_N \to 0$.

Евристичні міркування, які приводять до оцінки ϑ_N, зрозумілі: $R_N(\alpha)$ є емпіричним аналогом функціоналу

$$R(\alpha) = \int_{\mathfrak{F}} \sum_{k=1}^{M-1} \left(\int_0^1 \mu(A,t) u_k(t,\alpha) dt \right)^2 \pi(dA). \qquad (6)$$

Цей функціонал являє собою проінтегровану по $\pi(dA)$ суму квадратів коефіцієнтів Фур'є у розкладі функцій $\mu(A,\cdot)$ по базису $\{u_k(\cdot,\alpha)\}_{k=1}^{M-1}$. Зрозуміло, що $R(\alpha)$ досягає максимуму при $\alpha = \vartheta$. Статистику ϑ_N будемо називати L_2-оцінкою ϑ, $\vartheta_N^{\max}$ — максимізаційною оцінкою. Якщо оцінювання проводиться по спостереженнях з похибкою ζ_j^N, то у (4) і (5) $\hat{\mu}_N$ замінюється на $k_N^*(s,a)$, $\mathfrak{F}$ на деяку область у $\mathbb{R}^d$. У п. 3.1 доведено умови спроможності ϑ_N та $\vartheta_N^{\max}$ і побудовано (грубі) оцінки швидкості їх збіжності. Дослідження ϑ_N продовжено у п. 3.2, де розглянуто можливість використання випадкової міри π, залежної від вибірки Ξ_N, наведено алгоритми обчислення $R_N(\alpha)$, а також приклади застосування L_2-оцінок у випадку оцінювання моменту мутації, моменту злому та параметрів $\vartheta \in \mathbb{R}$, по яких $w_i(\cdot,\vartheta)$ є гладенькими функціями. Показано, що оцінки швидкості збіжності з п. 3.1, як правило, не дають найкращої по порядку величини швидкості. Тим не менше, для оцінки моменту мутації вони дозволяють будувати неасимптотичні надійні проміжки.

У п. 3.3. доводиться асимптотична нормальність нормованих L_2-оцінок $\sqrt{N}(\vartheta_N - \vartheta)$ у випадку, коли $u(t,\alpha)$ є гладенькими функціями $\alpha \in \Theta \subset \mathbb{R}^m$ ($\lambda_N = 0$). Нормування $\sqrt{N}$ має "правильний" порядок для гладеньких функцій концентрації, як це показано у п. 4.1. Наведемо умови асимптотичної нормальності з п. 3.3.

(I) $\mathfrak{F}$ є класом Вапника-Червоненкіса і виконуються умови емпіричної вимірності.

(II) Параметрична множина Θ є компактом відносно евклідової метрики у $\mathbb{R}^p$ і для будь-якого $\alpha \in \Theta$, $\alpha \neq \vartheta$ функції $\{u_k(\cdot,\alpha),\ k = 1 \div M-1,\ u_j = (\cdot,\vartheta),\ j = 1 \div M-1\}$ — лінійно незалежні.

(III) Функції $u_k(t,\alpha)$ є двічі неперервно диференційовними по α, причому для довільних $i, j = 1 \div p$, $\exists L > 0$ та $0 < \gamma < 1$, такі, що для всіх $t \in [0,1]$, $\alpha, \beta \in \Theta$, $\left|\frac{\partial u_k(t,\alpha)}{\partial \alpha^i}\right| \leq L$ і для $f(t,\alpha) = \frac{\partial^2 u_k(t,\alpha)}{\partial \alpha^i \partial \alpha^j}$, $|f(t,\alpha)| \leq L$, $|f(t,\alpha) - f(t,\beta)| \leq L|\alpha - \beta|^\gamma$.

Наступна умова стосується поведінки критерію $R(\alpha)$, визначеного (6) в околі точки $\alpha = \vartheta$:

(IV) Матриця $R''(\vartheta)$ невироджена (тобто $\det R''(\vartheta) \neq 0$).

(Штрих означає диференціювання по α)

(V) $\sup_{\alpha\in\Theta} \operatorname{Var} u_i(\cdot,\alpha) < \infty$

Теорема 3.3.1. *Якщо виконуються умови (I)–(V) і ϑ є внутрішньою точкою Θ, то $\sqrt{N}(\vartheta_N - \vartheta)$ слабко збігається до деякого гаусового вектора.*

Граничний розподіл можна виразити через розподіли компонент, функції u_k та міру π. При доведенні теореми використано результати з асимптотики емпіричних мір та функціоналів фон-Мізеса, отримані у розділах 1 та 2. Асимптотична нормальність використовується при побудові асимптотичного надійного проміжку у задачі оцінки фази для гармонійної моделі функції концентрації.

L_2 та sup-оцінки, описані у п. 3.1, визначаються по $w_i(t,\vartheta)$ хоча й неоднозначно, але дуже жорстко. Щоб збільшити запас можливих оцінок у п. 3.4 для двокомпонентних сумішей розглядається метод уточнення. Уточнена оцінка $\tilde{\vartheta}_N$ будується на основі заданої оцінки ϑ_N за допомогою асимптотичних розкладів функціоналів від $w_i(\cdot,\vartheta)$. Показано, що коли ϑ_N має степеневий порядок збіжності до ϑ (тобто $\vartheta_N - \vartheta = O(N^{-\gamma})$, $\gamma > 0$), то рядом послідовних уточнень можна побудувати оцінку $\tilde{\vartheta}_N$, яка є асимптотично нормальною при нормуванні $\sqrt{N}(\tilde{\vartheta}_N - \vartheta)$. Цим методом можна отримувати асимптотично нормальні оцінки і в тому випадку, коли функції концентрації не є гладенькими по оцінюваному параметру.

У п. 3.5. розглядаються оцінки, які використовують sup-норму. Тут побудовано рівномірні надійні проміжки для розподілів компонент суміші, асимптотичний надійний проміжок для оцінки моменту мутації, оцінка для моменту "зламу" з правильною швидкістю збіжності.

Задачу непараметричного оцінювання функції концентрації $w_1(t)$ двокомпонентної суміші розглянуто у п. 3.6. Використовується оцінка, основана на розкладі w_1 у ряд Фур'є та оцінюванні коефіцієнтів цього ряду L_2-методом. Показано, що така оцінка є спроможною для ліпшицевих функцій концентрації і отримано оцінку швидкості її збіжності у L_2 та sup нормах.

Розділ четвертий містить дослідження важливого частково випадку аналізу сумішей із змінними концентраціями, а саме, — задачі оцінювання моменту розладки, коли у (2) $M = 2$, $w_1(t) = 1$ при $t < \vartheta$, $w_1(t) = 0$ при $t > \vartheta$.

У п. 4.1. показано, що на відміну від регулярних задач, розглянутих у третьому розділі, для оцінки моменту розладки правильною швидкістю збіжності є не $\frac{1}{\sqrt{N}}$, а $\frac{1}{N}$. У п. 4.2. розглянуто квантильний критерій однорідності, тобто критерій для перевірки гіпотези $H_1 \equiv H_2$ проти альтернативи $H_1 \not\equiv H_2$. Цей критерій можна розглядати як узагальнення на випадок невідомого моменту розладки медіанного критерію однорідності двох вибірок. Принцип побудови критерію та оцінок ϑ пояснимо на простому прикладі медіанної оцінки. Нехай $\Delta = \mathbb{R}$, N — парне, $\mathrm{med}(\Xi_N)$ — медіана вибірки Ξ_N. Розібємо Ξ_N на дві підвибірки: Ξ_N^+ та Ξ_N^-, причому до Ξ_N^+ віднесем ті ξ_j^N, для яких $\xi_j^N > \mathrm{med}(\Xi_N)$, а до Ξ_N^- — ті, для яких $\xi_j^N < \mathrm{med}(\Xi_N)$. Покладемо $\kappa_j^1 = \chi\{\xi \in \Xi_N^+\} - \chi\{\xi \in \Xi_N^-\}$, $S_i^1 = \sum_{j=1}^{i} \kappa_j^1$, $V_1 = \max_j |S_j^1|$, $\hat{k}_N = \mathrm{argmax}_j |S_j^1|$, $\hat{\vartheta}_N = \hat{k}_N / N$. Легко переконатись, що коли $\mathrm{med}(H_1) \neq \mathrm{med}(H_2)$, то при $N \to \infty$, $\hat{\vartheta}_N \to \vartheta$ м.н. і $V_1 \sim CN$. Тому $\hat{\vartheta}_N$ можна в цьому випадку розглядати як оцінку для ϑ, а V_1 використовувати при побудові критерію для перевірки однорідності (тобто гіпотези $H_1 = H_2$) проти альтернативи $\mathrm{med}(H_1) \neq \mathrm{med}(H_2)$. Гіпотеза приймається, коли V_1 менше або дорівнює деякому порогу λ і відхиляється, коли

$V_1 > \lambda$. Оскільки при $H_1 = H_2$ $V_1 \sim C\sqrt{N}$, поріг λ можна підібрати так, щоб критерій був спроможним. Якщо розладка існує ($H_1 \not\equiv H_2$), але $\mathrm{med}(H_1) \neq \mathrm{med}(H_2)$, то медіанний критерій (оцінка) її не помітить. Можна поліпшити цю оцінку, розбиваючи окремо вибірки Ξ_N^+ та Ξ_N^- їх медіанами, будуючи для них відповідні суми S_i^2 і S_i^3, а потім використовуючи

$$\hat{k}_N = \operatorname*{argmax}_j \max_l |S_j^l| \tag{7}$$

$V = \max_{j,l} |S_j^l|$. Критерій, побудований по V, буде помічати зміну квантилей рівня 1/4, 1/2 і 3/4. Подрібнення розбиття можна продовжувати і далі, забезпечуючи, при досить великій вибірці, виявлення будь-яких змін розподілу. Якщо $\Delta \neq \mathbb{R}$, для використання медіанного розбиття можна спочатку застосувати який-небуть метод багатовимірного шкалювання.

Критерії та оцінки такого типу є частковим випадком квантильних оцінок та критеріїв, які розглянуто у п. 4.2–4.4. У п. 4.2 знайдено ймовірність помилки першого роду для цього критерію та її асимптотику при $N \to \infty$. Доведено спроможність критерію. У п. 4.3. доводиться спроможність квантильної оцінки моменту розладки.У п. 4.4. розглянуто асимптотичну поведінку величини $|\hat{k}_N - k_N|$, де k_N — ціла частина числа ϑN. Це приблизно відповідає нормуванню оцінки моменту розладки $N(\hat{\vartheta}_N - \vartheta)$. (Оскільки $\hat{k}_N$ (7) визначено неоднозначно, під $|\hat{k}_N - k_N|$ розуміємо відстань від k_N до найдальшого можливого $\hat{k}_N$). Доведено, що за деяких додаткових умов $|\hat{k}_N - k_N|$ слабко збігається до $|\tau - 0|$, де $\tau = \operatorname*{argmax}_{-\infty<j<+\infty} \sum_{i=0}^{j} \nu_i$, ν_i — незалежні в.в., що приймають значення $-1, 0, 1$. Розподіли ν_i рівні при $i < 0$ та $i > 0$ і залежать від H_k.

У п'ятому розділі розв'язується задача класифікації даних за допомогою навчаючої вибірки, що є вибіркою із суміші з змінними концентраціями. Для цього у п. 5.1. для випадку, коли Δ — сепарабельний метричний простір, застосовується метод k найближчих сусідів, у п. 5.2. для спостережень у $\mathbb{R}^d$ — метод ядерного оцінювання щільностей розподілу. Доводиться спроможність побудованих оцінок щільностей розподілу та відповідних квазибаєсівських методів класифікації.

Розглянемо, наприклад, метод k найближчих сусідів. Нехай (Δ, ρ) — сепарабельний метричний простір. У розпорядженні статистика є вибірка Ξ_N з розподілом (2), $w_l(t)$ — відомі, H_k — невідомі, ξ — результат спостережень об'єкта, який з апріорною ймовірністю $\bar{w}_l$ належить до l-ї компоненти суміші. Потрібно класифікувати цей об'єкт. Позначимо $\rho_k(\xi)$ — відстань від ξ до його k-го найближчого (у метриці ρ) сусіда серед елементів Ξ_N, $B(\xi, \rho_k(\xi)) = B_k$ — куля з центром ξ та радіусом $\rho_k(\xi)$. Тоді класифікатор $g_N(\xi, k_N)$ за методом k_N найближчих сусідів відносить ξ до l-ї компоненти, якщо $\bar{w}_l \hat{\mu}_N(B_k, a^l) \geq \bar{w}_m \hat{\mu}_N(B_k, a^m)$ для всіх $m = 1 \div M$ (a^l задається (3)).

Позначимо $L(g)$ — ймовірність помилки при використанні класифікатора g, L^* — інфінум $L(g)$ по всіх можливих g.

Теорема 5.1.1 *Якщо (Δ, ρ) – сепарабельний метричний простір, сітка часу рівномірна,* $\mathrm{Var}\, w^l(\cdot) < \infty, \int w_l(t)dt > 0$, $l = 1 \div M$, *функції w_l — лінійно незалежні, $k_N/N \to 0$ і $\sqrt{N \ln N}/k_N \to 0$ при $N \to \infty$, то $L(g_N) \to L^*$ м.в.*

У п. 5.3. для спостережень з похибками використовуються оцінки щільності, що є оберненими перетвореннями Фур'є від неоднорідних емпіричних характеристичних функцій. При цьому застосовуються методи фільтрації, підібні до фільтрації спектральних оцінок. Знайдено достатні умови збіжності таких оцінок щільностей у $L_2(\mathbb{R}^d)$.

Шостий розділ присвячено задачі аналізу спостережень, в яких наявна похибка, що залежить від попередніх значень спостережень. Розглядаються дані $x(t_j^N)$, які є спостереженнями в моменти $t_j^N = \frac{j}{N}$ випадкового процесу $x(t)$, що задовольняє стохастичне диференційне рівняння

$$dx(t) = f(x(t), t)dt + g(t)dt, \tag{8}$$

де $f(x, t)$ та $g(t)$ — невипадкові функції, $w(t)$ — стандартний вінерівський процес. Член $f(x(t), t)dt$ у (8) інтерпретується як інерційна характеристика вимірювального приладу, $g(t)$ — як інтенсивність флуктуацій вимірюваної величини. Задача полягає в оцінюванні $g^2(t)$, коли відомо, що $g^2 \in \mathcal{G}$ належить деякому класу функцій $\mathcal{G}$, а f невідома. (Незалежність g від x має тут принципове значення: якщо g є функцією x, добре оцінити Π по $x(t)$, $t \in [0,1]$, взагалі кажучи, неможна).

Для оцінювання використовуються зважені бакстерівські суми

$$B_N(x,a) = \sum_{j=1}^{N} a(t_{j-1}^N)(x(t_j^N) - x(t_{j-1}^N))^2.$$

Розглядається функціонал

$$R_N^*(h) = \int_0^1 h^2(t)dt - 2B_N(x,h)$$

і як оцінку для g^2 обрано

$$\hat{g}_N \in \underset{h\in\mathcal{G}}{\operatorname{argmin}}(R_N^*(h), \lambda_N).$$

У п. 6.1 на основі модифікованої бакстерівської теореми показано, що ця оцінка є спроможною у $L_2([0,1],dt)$, якщо функції класу $\mathcal{G}$ рівномірно обмежені з рівномірно обмеженою варіацією, $f(x,t)$ обмежена на обмежених множинах.

Теорема 6.1.2. *Нехай* $\exists A$, $V \le \infty$, *такі, що* $\forall h \in \mathcal{G}$

$$\sup_t |h(t)| \le A, \quad \operatorname*{Var}_t h(t) \le V,$$

$\hat{g}_N$ *визначено вище,* $\lambda_N = \varepsilon_N = \sqrt{\ln N/N}$. *Тоді для деякої в.в.* $\Lambda < \infty$ *м.н.,*

$$\int_0^1 (\hat{g}_N(t) - g^2(t))^2 dt \le \Lambda\varepsilon_N.$$

Для класів $\mathcal{G}$ з нерівномірно обмеженими варіаціями розглянуто спроможну модифікацію оцінки $\hat{g}_N$, яка базується на використанні "штрафних функцій": замість функціоналу R_N^* максимізується

$$R_N^0(h,\alpha_N) = R_N^*(h) + \alpha_N U(h).$$

Тут $\alpha_N U(h)$ — штрафна функція, що зростає при зростанні варіації h. Оцінено швидкості збіжності для відповідних максимізаційних оцінок.

У п. 6.2. розглядається задача параметричної оцінки $g^2(t,\vartheta)$, де $\vartheta \in \Theta \subseteq \mathbb{R}^d$ — невідомий параметр. Доводиться асимптотична нормальність нормованих оцінок $\sqrt{N}(\vartheta_N - \vartheta)$, де

$$\vartheta_N \in \operatorname*{argmax}_{\alpha\in\Theta}(R_N^*(g^2(\cdot,\alpha),0).$$

Основні положення дисертації опубліковані у наступних роботах:

1. Майборода Р.Е. Оценка производящей функции моментов случайной величины по результатам наблюдений // Теория вероятностей и мат. статистика.— 1985.— Вып. 32.— С.121–131.

2. Майборода Р.Е. Оценки производящей функции моментов для стационарных процессов // Укр. мат. журн.— 1986.— 38, N4.-С.504-509.

3. Майборода Р.Е. Закон повторного логарифма для эмпирических производящих функций моментов (ЭПФМ) // 1 Всемирный конгресс о-ва мат. статистики и теории вероятностей им. Бернулли, Ташкент, 9–15 сент. 1986г.: Тез. докл.–Ташкент, 1986.— Т.2.-С.591.

4. Майборода Р.Е. Об эмпирических производящих функциях моментов случайных величин // Теория вероятностей и ее применения.— 1987.— 32, Вып. 1.-С. 195–196.

5. Майборода Р.Е. Оценивание производящей функции моментов по наблюдениям с погрешностью // Теория вероятностей и мат. статистика.— 1988.— Вып. 39.-С.87–90.

6. Майборода Р.Е. Квантильные методы поиска разладок // Стат. проблемы упр.— 1988.— Вып. 85.— С.96-101.

7. Каплан Е.И., Майборода Р.Е. Алгоритмы обработки неоднородных данных .- Киев: РДЭНТП, 1989.— 20с.

8. Майборода Р.Е. Центральная предельная теорема для эмпирических производящих функций моментов // Теория вероятностей и ее применения.— 1989.— **34**.— Вып. 2.— С.375–380.

9. Майборода Р.Е. Оценка момента возникновения мутации // Статистический анализ данных на ЭВМ. Респ. шк.-семинар, Ужгород, 22–26 мая 1989г : Тез. докл.- Ужгород, 1989.— С.33–34.

10. Майборода Р.Е. Понижение размерности в задаче обнаружения разладки // IV Всесоюзная конф. по многом. стат. анализу, Тарту, 5–7 сент. 1989г : Тез. докл.- Тарту, 1989.— С.100-101.

11. Майборода Р.Е. Об асимптотике производящих функций моментов случайных величин //Укр. мат. журн.-1989.— 41, N 7.— С.992–994.

12. Майборода Р.Е. Непараметрическое оценивание характеристик переменных смесей // Непараметрические и робастные методы в кибернетике и информатике: Иркутск, 14–21 апр. 1990г: Материалы VII Всесоюз. семинара.-Томск.ГУ, 1990.— **2**.— С.341–342.

13. Майборода Р.Е. Медианный критерий однородности выборки // Теория вероятностей и мат. статистика.— 1990.— Вып. 42.— С.82–86.

14. Майборода Р.Е. Эмпирические функции в задачах оценивания изменения свойств временных рядов // Стат. проблемы упр.— 1990.— Вып. 89.— С.60–65.

15. Майборода Р.Е. Медианная оценка разладки в случае слабозависимых наблюдений // Теория вероятностей и мат. статистика.— 1990.— Вып. 43.— С.78–83.

16. Майборода Р.Е. Непараметрический метод поиска разладок для многомерных наблюдений // Теория вероятностей и ее применения.— 1990.— **35**, Вып. 3.— С.582–586.

17. Бесклинская Е.П., Майборода Р.Е. О скорости сходимости некоторых оценок параметров стационарных гауссовских случайных процессов // Теория вероятностей и мат. статистика.— 1990.— Вып. 43.— С.13–19.

18. Майборода Р.Е. Непараметрическое обнаружение разладки по наблюдениям с погрешностью // Укр. мат. журн.— 1991.— 43, N 5,-С.706–709.

19. Майборода Р.Е. Об оценивании параметров переменных смесей // Теория вероятностей и мат. статистика.— 1991. — Вып. 44,-С.87–92.

20. Заложенков О.А. Майборода Р.Е. Задачи статистической обработки неоднородных данных экологического мониторинга // Материалы IV Всесоюзной школы-семинара "Статистический и дискретный анализ данных и экспертное оценивание", Одесса, 2–7 сент. 1991г.— Одесса, 1991.— С.12–13.

21. Майборода Р. Є. Уточнення оцінок концентрацій двокомпонентних сумішей, що змінюються // Тези міжнародної конференції, присвяченої пам'яті академіка М.П.Кравчука, Київ–Луцьк, 22–28 верес. 1992р.— Київ–Луцьк, 1992.— С.122.

22. Майборода Р. Анализ смеси компонентов с временно-зависимыми концентрациями // Международная конференция ChanDe'92, Киев, 29сент. - 2окт. 1992г.: Тез. докл.— Киев, 1992.— С.42–43.

23. Майборода Р.Є. Проекційні оцінки концентрацій сумішей, що змінюються // Теорія імовірностей та мат. статистика. — 1992.— Вип. 46.— С.70–76.

24. Maiboroda R.E. Estimators for parameters of time-dependent mixture consentrations // Доп. АН України.— 1993.— N 4.— P. 17–21.

25. Mayboroda R. Estima... time-dependent concentrations// VI International Vilnius conference on probability theory and math. statistics: Abstracts of communications.— 1993.— 2.— P. 24–25.

26. Майборода Р.Є. Аналіз сумішей по спостереженням з похибкою // Теорія ймовірностей та мат. статистика.— 1993.— Вип. 48.— С.125–134.

27. Майборода Р.Є. Асимптотична нормальність оц... параметрів функцій концентрації сумішей // Теорія ймовірностей та мат. статистика.— 1993.— Вип.49.— С.155-160.

28. Вовк Л.Б., Майборода Р.Е. Об оценив...и момента раскладки для процесса типа Орнштейна-...бека //Укр. мат. журн.–1993.— 45, N 9.— С.1198–1205

НЧЕ

Підп. до друку 24.05.94 Формат 60x84/16. Папір друк. Офс. друк.
Ум. друк. арк. 1,16. Ум. фарбо-відб. 1,16 Обл.-вид. арк. 0,9
Тираж 100 пр. Зам. 133 Безкоштовно.

Віддруковано в Інституті математики АН України
252601 Київ 4, ГСП, вул. Терещенківська, 3