

# Зміст

<b>Передмова</b>	<b>5</b>
<b>1 Моделі сумішей зі змінними концентраціями</b>	<b>6</b>
1.1 Вступ . . . . .	6
1.2 Прикладні задачі аналізу сумішей . . . . .	8
1.3 Приклади модельованих даних . . . . .	17
<b>2 Оцінювання функцій розподілу</b>	<b>25</b>
2.1 Незміщені мінімаксні оцінки . . . . .	26
2.2 Асимптотика емпіричних мір . . . . .	33
2.3 Виправлені зважені емпіричні функції розподілу . . . . .	46
2.4 Асимптотично ефективна оцінка розподілу . . . . .	56
<b>3 Оцінки числових характеристик розподілів компонент</b>	<b>65</b>
3.1 Лінійні оцінки функціональних моментів . . . . .	66
3.2 Адаптивні оцінки моментів . . . . .	72
3.3 Виправлені оцінки для моментів . . . . .	76
3.4 Оцінювання квантилів . . . . .	83
3.5 Оцінка екстремальних точок розподілів компонент . . . . .	89
<b>4 Оцінювання щільностей розподілів компонент</b>	<b>95</b>
4.1 Ядерні оцінки щільності . . . . .	96
4.2 Асимптотична нормальність ядерних оцінок . . . . .	100
4.3 Вибір параметра згладжування . . . . .	105
4.4 Неядерні оцінки щільностей розподілів . . . . .	113
<b>5 Аналіз спостережень з домішкою</b>	<b>123</b>
5.1 Оцінки щільності по спостереженнях з домішкою . . . . .	123

5.2	Адаптивні оцінки для параметрів . . . . .	133
<b>6</b>	<b>Задачі класифікації</b>	<b>147</b>
6.1	Баєсова класифікація . . . . .	148
6.2	Метод найближчого сусіда . . . . .	154
6.3	Асимптотика порогових класифікаторів . . . . .	157
6.4	Класифікація на основі єдиного індекса . . . . .	169
6.5	Швидкість збіжності класифікаторів єдиного індекса . . . . .	179
<b>7</b>	<b>Допоміжні відомості</b>	<b>197</b>
7.1	Формули інтегрування і пов'язані з ними нерівності . . . . .	198
7.2	Нерівність для визначників . . . . .	200
7.3	Ймовірнісні нерівності і граничні теореми . . . . .	205
7.4	Слабка збіжність випадкових функцій . . . . .	209
7.5	Ефективність. Мінімаксність. Інформація. . . . .	212
7.6	Оцінювання щільності за кратними вибірками . . . . .	214
	<b>Література</b>	<b>216</b>
	<b>Список позначень</b>	<b>223</b>

# Передмова

Моделі сумішей природно виникають у задачах статистичного аналізу даних медико-біологічних, соціологічних, економічних досліджень. Математична теорія статистичного оцінювання у рамках таких моделей успішно розвивається починаючи з кінця XIX століття. У класичній моделі скінченної суміші вважається, що концентрації компонент є сталими, а дані є незалежними, однаково розподіленими випадковими величинами.

Дана книга присвячена аналізу сумішей, у яких концентрації компонент змінюються від спостереження до спостереження. Ми розглядаємо задачі оцінювання розподілів компонент суміші та їх характеристик: моментів, квантилів, щільностей розподілу. Отримані оцінки дають змогу побудувати алгоритми класифікації на основі спостережень з суміші зі змінними концентраціями.

Основна увага у книзі приділена дослідженню асимптотичних властивостей побудованих алгоритмів при необмеженому зростанні обсягу вибірки. Ми застосовуємо адаптивний підхід для побудови асимптотично оптимальних оцінок і порівнюємо асимптотичні властивості класифікаторів, отриманих на основі різних технік класифікації.

Книга містить в основному результати, опубліковані за останні роки, після виходу монографії [26]. Для зручності читачів основні твердження з цієї роботи, що використовуються у даній книзі, вміщено у п. 2.1-2.2. Крім власних результатів ми розглядаємо також результати О. Кубайчук, Ю. Іванька та А. Лодатко.

Автори вдячні проф. В.В. Булдігіну, Ю.В. Козаченку, Ю.С. Мішурі та М.П. Моклячуку за підтримку у роботі.

Книга підготовлена та опублікована за підтримки програми Tempus у рамках проекту TEMPUS PROJECT IB-JEP-25054-2004.

# Розділ 1

## Моделі сумішей зі змінними концентраціями

### 1.1 Вступ

У цій книжці розглядаються методи статистичного аналізу даних, розподіл яких задається моделлю суміші зі змінними концентраціями. Ця модель придатна для опису даних різних медико-біологічних, соціологічних, економічних досліджень. Деякі приклади застосування аналізу сумішей до прикладних задач наведено у книжці [26]. У п. 1.2 описано два таких приклади: біологічний та соціально-психологічний. Вони досить умовні, оскільки мають на меті дати загальне враження про можливості аналізу сумішей, без заглиблення у детальний опис реального статистичного дослідження.

Для того, щоб читач отримав уявлення про те, як можуть виглядати дані, що задаються моделлю суміші зі змінними концентраціями, ми розглянемо декілька прикладів модельованих даних у п. 1.3.

У наступних розділах книги описано методи оцінки функцій розподілу компонент суміші (розділ 2) та таких характеристик розподілів, як функціональні моменти, квантілі (розділ 3) та щільності розподілу (розділ 4). Всі ці оцінки отримані непараметричними методами, тобто жодних припущень про параметричну модель розподілів компонент ми не використовуємо. Концентрації компонент, тобто ймовірності з якими певні компоненти будуть спостерігатись у даному спостереженні, вважаються відомими. (Про оцінювання концентрацій див. у книжці [26]).

Задачі аналізу сумішей двох компонент у випадку, коли розподіл однієї

компоненти повністю невідомий, а для іншої задана параметрична модель, розглянуто у розділі 5. Тут побудовані семіпараметричні оцінки невідомих параметрів та непараметричні оцінки для щільності компоненти, яка не має параметричної моделі.

Нарешті у розділі 6 розглядається задача побудови класифікаторів, які дозволяють розділити суміш на компоненти.

В останньому розділі вміщено допоміжні відомості з теорії ймовірностей, теорії випадкових процесів та математичної статистики та інші твердження, які використовуються у основних розділах книги.

Моделі сумішей кількох розподілів для опису статистичних даних з'явилися ще у XIX столітті у роботах С. Ньюкомба [60] та К. Пірсона [61]. Протягом XX століття розвитку набули дослідження у галузі класичної моделі скінченних сумішей, в якій розподіл спостережуваних даних  $\xi_j$  один і той же для всіх спостережень  $j = 1, \dots, N$  і описується у вигляді

$$P\{\xi_j \in A\} = w_1 H_1(A) + w_2 H_2(A) + \dots + w_M H_M(A),$$

де  $H_i$  — розподіли компонент суміші,  $w_i$  — ймовірності змішування (mixing probabilities) які можна трактувати як концентрації компонент у суміші,  $M$  — кількість компонент суміші.

Огляд результатів у цій області можна знайти у книгах Дж. Маклахлана та К. Басфорда [57], Дж. Маклахлана та Д. Піла [58]. Як правило, у класичній моделі використовують параметричні моделі розподілів компонент. Це пов'язано з тим, що у загальній непараметричній формі задача оцінювання розподілів компонент не є ідентифікованою. Тому для отримання консистентних оцінок потрібно накладати додаткові обмеження на розподіли, що роблять задачу ідентифікованою. Зокрема умови ідентифікованості багатьох параметричних задач оцінювання у класичній моделі сумішей встановлено у роботах Г. Тейчера [63], С. Яковіца та Ж. Спрагінса [73], Г. Гольцмана, А. Мунка, Т. Гнетінга [51].

Останнім часом з'явився ряд робіт, у яких на модель накладаються умови ідентифікованості непараметричного типу. У роботі П. Холла та К. Зоу [47] ідентифіковність досягається внаслідок незалежності координат багатовимірних спостережень у кожній компоненті. Роботи Л. Борде, С. Моттле, П. Вандекерхове [40] та Д. Хантера, С. Ванга, Т. Хетмансергерера [52] присвячені випадку, коли компоненти мають розподіл, симетричний відносно медіани, причому відрізняються одна від одної лише зміщенням.

Інший підхід до отримання консистентних оцінок полягає в тому, щоб відмовитись від припущення про однакову розподіленість даних. У робо-

тах П. Халла, Д. Тіттерінгтона [48] та Д. Тіттерінгтона, А. Сміта, У. Макова [69] розглядається багатовибіркова задача, причому у кожній вибірці дані описуються класичною моделлю суміші, компоненти всіх сумішей однакові, але концентрації компонент у різних вибірках різні. Очевидним узагальненням цієї моделі є суміші зі змінними концентраціями, у яких кожному спостереженню відповідають свої значення концентрацій. Саме дослідженню цієї моделі присвячена дана книжка. Вона спирається на ряд робіт авторів а також О. Кубайчук, Ю. Іванька та А. Лодатко. Нажаль, через обмеження обсягу книги, ми не змогли включити до неї результати Д. Похилько з теорії вейвлет-оцінок щільностей розподілу та А. Рижова з аналізу сумішей за цензурованими спостереженнями. Ми також не розглядаємо задачі оцінки параметрів функцій концентрації, з якими можна ознайомитись в [26]. У цій книжці концентрації завжди вважаються повністю відомими.

## 1.2 Прикладні задачі аналізу сумішей

**Аналіз даних генетичного дослідження.** Нехай спостерігаються деякі живі організми — це можуть бути люди, тварини або рослини. Для визначеності, будемо вважати, що мова йде про мишей. Нас цікавить зв'язок між певною характеристикою фенотипу миші, наприклад — довжиною тіла і наявністю або відсутністю певного варіанта (алеля) деякого гена у її генотипі. Зрозуміло, що такий зв'язок не може бути жорстко детермінованим — той чи інший алель не визначає однозначно довжину миші. Мова може йти лише про те, що різним генотипам відповідають різні розподіли довжини тіла мишей з цими генотипами.

Позначимо досліджуваний об'єкт — мишу через  $O$ , а номер алеля досліджуваного гена, котрий присутній у генотипі  $O$  — через  $\text{ind}(O)$ .<sup>1</sup> Довжину миші  $O$  позначимо через  $\xi(O)$ . З математичної точки зору,  $\xi(O)$  — випадкова величина, розподіл якої залежить від  $\text{ind}(O)$ . Позначимо

$$H_k(x) = \mathbb{P}\{\xi(O) < x \mid \text{ind}(O) = k\}$$

— функція розподілу довжини миші, яка має  $k$ -тий варіант генотипу.

---

<sup>1</sup>Строго кажучи, у кожної миші повинно бути два, можливо різних, алеля одного гена — по одному у кожній з двох гомологічних хромосом. Тому, якщо нас цікавить взаємодія цих алелів, потрібно перенумерувати всі можливі різні пари і під  $\text{ind}(O)$  слід розуміти номер тієї пари, яка присутня у  $O$ .

Нехай в результаті дослідження були виміряні довжини  $N$  мишей —  $O_1, O_2, \dots, O_N$ , які дорівнюють відповідно  $\xi_1 = \xi(O_1), \dots, \xi_N = \xi(O_N)$ . Сукупність всіх даних вимірювання позначимо  $\Xi_N = (\xi_1, \dots, \xi_N)$ . Якщо біолог має можливість однозначно розкласифікувати мишей за їх генотипом, то оцінка  $H_k$  за спостереженнями  $\xi(O_j)$ ,  $j = 1, \dots, N$  не викликає жодних ускладнень. Дійсно, у цьому випадку можна просто сформулювати вибірку, що складається з мишей із даним генотипом і на роль оцінки  $H_k$  обрати емпіричну функцію розподілу, побудовану за цією вибіркою. В результаті отримуємо оцінку

$$\tilde{H}_N^k(x) = \frac{1}{N_k} \sum_{j=1}^N \mathbb{I}\{\xi(O_j) < x, \text{ind}(O_j) = k\}.$$

(Тут  $\mathbb{I}\{A\}$  — індикатор події  $A$ ,  $N_k = \sum_{j=1}^N \mathbb{I}\{\text{ind}(O_j) = k\}$  — кількість досліджених мишей, які мали генотип  $k$ -того типу.)

Ця оцінка є найкращою непараметричною оцінкою для функції розподілу з усіх загальноприйнятих точок зору — вона консистентна, незміщена, має найменшу дисперсію в класі всіх незміщених оцінок, є оцінкою емпіричної найбільшої вірогідності і т.д.

Однак генетичні тести, що використовуються для визначення генотипу, як правило, працюють не безпомилково. Крім того, часто дослідник взагалі не має можливості виявити безпосередньо той ген, який його цікавить і змушений судити про його наявність або відсутність опосередковано, за наявністю так званих маркерів — генів, що знаходяться поруч з досліджуваним на тій же хромосомі. Якщо у такому випадку не враховувати можливі помилки класифікації, то оцінка на основі емпіричної функції розподілу стане зміщеною.

Для того, щоб врахувати ефекти помилок, доцільно замість однозначної класифікації задати ймовірності того, що генотип даної миші належить певному класу:

$$w_j^k = P\{\text{ind}(O_j) = k\}.$$

Ці ймовірності для  $O_j$  визначаються за результатами генетичних тестів (як правило, для таких тестів відомі ймовірності помилкової класифікації), наявністю або відсутністю певних генетичних маркерів та за апіорними ймовірностями того, що навмання обрана з популяції миша має генотип відповідного типу. Зрозуміло, що за таких обставин  $w_j^k$  будуть різними для різних мишей. Фактично, для кожної досліджуваної тварини можна

визначити свої апіорні ймовірності наявності заданого варіанту генотипу<sup>2</sup> Якщо  $w_j^k$  та  $H_k$  задані, то розподіл спостережуваної характеристики  $\xi_j = \xi(O_j)$  визначається за формулою<sup>3</sup>

$$P\{\xi_j \in A\} = \sum_{m=1}^M w_j^m H_m(A). \quad (1.1)$$

При цьому дані, що відповідають різним мишам, природно вважати незалежними між собою.

Дані, які складаються з незалежних випадкових величин (векторів), розподіл яких описується (1.1), будемо називати вибіркою з суміші зі змінними концентраціями. Популяції об'єктів  $O$  (мишей) які відповідають різним значенням  $\text{ind}(O)$  будемо називати компонентами суміші. Функція  $H_k$  — це функція розподілу спостережуваної характеристики (у нашому прикладі — довжини тіла) об'єктів, що належать  $k$ -тій компоненті. Для простоти ми будемо скорочувати цю назву до “функція розподілу  $k$ -тої компоненти”.  $w_j^k$  — це ймовірність, з якою у  $j$ -тому спостереженні спостерігається об'єкт з  $k$ -тої компоненти. Ми будемо називати  $w_j^k$  концентрацією  $k$ -тої компоненти у суміші під час  $j$ -того спостереження. Інша назва для цих величин — ймовірності змішування (mixing probabilities).

У класичній моделі скінченної суміші концентрації компонент є одними і тими самими для всіх спостережень. Цю модель можна назвати сумішшю зі сталими концентраціями. У цій книжці розглядається випадок, коли концентрації змінюються від спостереження до спостереження. Очевидно, що для розглянутого нами прикладу генетичних досліджень модель зі змінними концентраціями є більш адекватною.

Які статистичні задачі виникають перед дослідником при аналізі таких сумішей? Звичайно, біолог цікавиться розподілами різних компонент, тобто оцінками для  $H_k$ . Ці розподіли доцільно оцінювати параметричними методами, якщо існує апіорна параметрична модель для  $H_k$ , наприклад, якщо можна вважати, що  $\xi$  для кожної окремої компоненти має гауссів розподіл. Однак таке припущення часто є занадто обмежувачим. Тому корисними є непараметричні методи оцінювання  $H_k$ . Таким методам присвячений розділ 2. Вони спираються на використання зважених емпіричних

<sup>2</sup>Апіорними ці ймовірності є в тому розумінні, що вони визначаються до вимірювання спостережуваної характеристики  $\xi_j$  лише за результатами генетичного аналізу.

<sup>3</sup>Тут і далі розподіл (міра) та відповідна функція розподілу, як правило, позначаються однією і тією ж літерою: для випадкової величини  $\eta_k$  з функцією розподілу  $H_k$ ,  $H_k(A) = P\{\eta_k \in A\}$ .



функцій розподілу вигляду

$$\hat{H}_k(x) = \frac{1}{N} \sum_{j=1}^N a_j^k \mathbb{I}\{\xi_j < x\}, \quad (1.2)$$

де  $a_j^k$  — деякі вагові коефіцієнти, покликані компенсувати наявність у вибірці спостережень, що належать “непотрібним” (у даний момент) компонентам суміші.

Вагові коефіцієнти можна обирати по різному, виходячи з різних вимог до оцінки. У розділі 2 розглянуті три можливих варіанти. Найпростіший варіант — мінімаксні коефіцієнти, які дозволяють отримати незміщену оцінку, мінімаксну по відношенню до квадратичного ризику у класі всіх незміщених оцінок. Ці коефіцієнти підраховуються за простими формулами (2.10) і визначаються лише за концентраціями  $w_j^m$  (не залежать від значень спостережуваної характеристики  $\xi_j$ ). Недоліком таких мінімаксних коефіцієнтів є те, що вони породжують оцінки  $\hat{H}_k$ , які не є монотонно зростаючими, тобто не можуть вважатися функціями розподілу.

Для усунення цього недоліку, можна запропонувати різні алгоритми виправлення  $\hat{H}_k$ , які перетворюють її на функцію розподілу. Ряд таких алгоритмів розглянуто у п. 2.3. Виявляється, що виправлені оцінки також можна записати у формі, аналогічній (1.2), однак їх вагові коефіцієнти будуть залежати від  $\xi_j$ . Незважаючи на це, асимптотичні властивості виправлених оцінок виявляються цілком аналогічними властивостям мінімаксних — вони так само є консистентними та асимптотично нормальними, а їх коефіцієнт розсіювання (гранична дисперсія) такий самий, як і у мінімаксних оцінок.

Мінімаксні оцінки є найкращими у найгіршому випадку, але це не означає, що не існує оцінок, які могли б переважати мінімаксні у певних ситуаціях. Ми розглядаємо питання про побудову асимптотично ефективних оцінок функцій розподілу у п. 2.4 для випадку, коли розподіли компонент є дискретними. Отримані оцінки також можна записати у вигляді зваженої суми індикаторів, однак вираз є дещо більш складним, ніж (1.2). Асимптотично ефективні оцінки при великих обсягах вибірки у багатьох випадках є більш точними ніж мінімаксні, однак для малих обсягів вибірки більш надійними є мінімаксні оцінки.

Хоча функції розподілу несуть повну інформацію про розподіл даних, у прикладній статистиці оцінки для них використовуються не часто. Біолога скоріше можуть зацікавити середні значення довжин тіла тва-

рин з різними генотипами, їх дисперсії, медіани, квартилі, тощо. Аналогічні характеристики різних компонент сумішей цікавлять і спеціалістів у інших предметних областях. Задачі оцінки таких числових характеристик розподілів розглядаються у розділі 3. Зрозуміло, що такі характеристики можна оцінювати, виходячи з оцінки для відповідної функції розподілу. Наприклад, якщо потрібно оцінити функціональний момент  $\bar{g}_k = \int g(x)H_k(dx) = \mathbb{E}g(\eta_k)$ , то оцінкою, що відповідає  $\hat{H}_k$ , буде зважений емпіричний момент

$$\hat{g}_k = \int g(x)\hat{H}_k(dx) = \frac{1}{N} \sum_{j=1}^N a_j^k g(\xi_j). \quad (1.3)$$

Однак виявляється, що досліджувати такі оцінки інколи зручніше не на основі властивостей зважених емпіричних функцій розподілу, а виходячи з загальних теорем про асимптотику сум випадкових величин. Таке дослідження проведено у п. 3.1–3.2 для оцінок функціональних моментів і в результаті отримані оцінки з мінімальним коефіцієнтом розкиданості.

Якщо спеціалісту у предметній області потрібно “подивитись” на розподіл даних, то для однорідної вибірки він, скоріше за все, побудує гістограму або графік іншої оцінки щільності розподілу. Аналогам таких оцінок для сумішей зі змінними концентраціями присвячено розділ 4. На відміну від розглянутих раніше характеристик розподілу, задача оцінки щільності є “нерегулярною”: якщо емпіричні функції розподілу, емпіричні моменти та квантілі збігаються до оцінюваних характеристик з швидкістю порядку  $1/\sqrt{N}$ , то для оцінок щільності за однорідною вибіркою характерна швидкість збіжності порядку  $1/N^{\beta/(2\beta+1)}$ , де  $\beta$  — порядок гладкості оцінюваної щільності (див. п. 7.6). Виявляється, що такий порядок гладкості є характерним і для зважених ядерних оцінок щільності, побудованих за вибіркою з суміші зі змінними концентраціями. У п. 4.4 коротко описані також проєкційні та гістограмні оцінки і оцінки щільності за методом найближчого сусіда.

Досі ми вважали, що для розподілів всіх компонент суміші використовується непараметрична модель. Інколи додаткова інформація для деяких компонент дозволяє побудувати параметричні моделі розподілу. Для таких випадків можуть стати у пригоді методи, описані у розділі 5. Тут у випадку двокомпонентної суміші розглядаються оцінки методу відсіяної найбільшої вірогідності та методу моментів для невідомих параметрів параметрично заданої компоненти та для щільності розподілу “непараметричної компоненти”.

Нарешті, окрім задач вивчення та опису розподілів компонент, дослідник може мати прикладну мету побудови класифікаційного алгоритму, який дозволить би за спостережуваною характеристикою визначити, до якої компоненти належить об'єкт. У нашому прикладі задача виглядає так: у нової миші  $O$ , для якої не проводився генетичний аналіз, вимірюється довжина тіла —  $\xi(O)$ . Потрібно з'ясувати, який різновид генотипу  $\text{ind}(O)$  вона має. Інформація для прийняття рішення міститься у вибірці  $\Xi_N$ . Це задача статистичного навчання (розпізнавання образів). Для випадку, коли у розпорядженні дослідника є повністю розкласифікована навчаюча вибірка, вона добре досліджена. Ми у розділі 6 розглядаємо випадок, коли навчаюча вибірка вибрана з суміші зі змінними концентраціями. На основі баєсового підходу будуються різні класифікатори і вивчаються їх асимптотичні властивості.

Розглянемо тепер менш очевидний приклад застосування аналізу сумішей зі змінними концентраціями.

**Соціологія дражливих питань.** При проведенні соціологічних та психологічних досліджень часто виникають ускладнення, пов'язані з тим, що питання, котрі цікавлять дослідника, є “болючими”, дражливими для опитуваного внаслідок певної культурної або соціальної специфіки. Дражливими можуть бути питання про особливості сексуального життя, вживання наркотиків та ін. Годі сподіватись відвертої відповіді незнайомих людей на такі питання, особливо, якщо опитуваний є, скажімо, підлітком, а дослідник є тим, хто уособлює для нього авторитет “дорослого світу” — вчителем, лікарем, психологом.

Зрозуміло, що при проведенні соціологічних опитувань з дражливих питань наявність невизначеної кількості хибних відповідей ускладнює статистичний аналіз результатів і приводить до зміщення оцінок. Тому по дражливих питаннях намагаються або проводити анонімні опитування у яких шанси отримати правдиві відповіді вищі, або оцінювати ситуацію за опосередкованими даними. При цьому виникає проблема узгодження таких анонімних/опосередкованих даних з даними, отриманими за допомогою індивідуалізованих методик, скажімо — за даними психологічного тестування.

Як приклад розглянемо дослідження учнів-старшокласників середніх шкіл з метою виявлення зв'язків між епізодичним вживанням наркотичних речовин та психологічними характеристиками особистості, такими як рівень інтелекту, тривожність, інтровертованість-екстравертованість, тощо. Питання “чи вживаєте ви наркотики?” відноситься до дражливих.

Задаючи його психолог ризикує викликати напруження і втратити контакт з опитуваним, причому достовірність отриманої відповіді буде досить сумнівною. Об'єктивні медичні методики обстеження дозволяють виявити лише осіб з цілком сформованою наркотичною залежністю. Крім того примусові обстеження такого роду з дослідницькою метою є етично неприпустимими, а використання добровольців або даних, отриманих в зв'язку з кримінальними порушеннями, очевидно, веде до зміщення вибірки.

В той же час можна оцінити частку учнів, які у даній школі мали досвід вживання наркотиків на основі результатів анонімних опитувань, за експертними оцінками вчителів та шкільних психологів, за даними про кількість зафіксованих випадків виявлення незаконного обороту наркотиків у даній школі.

Нехай дослідження проводиться у різних школах міста, причому кожен опитуваний  $O$  проходить набір стандартизованих тестів, за результатами яких визначаються певні значення його особистісних характеристик, наприклад,  $\xi^1(O)$  — рівень інтелектуального розвитку (IQ),  $\xi^2(O)$  — рівень тривожності (нейротизм),  $\xi^3(O)$  — рівень інтровертованості-екстравертованості і т.д., всього  $d$  різних характеристик. Таким чином, з кожним опитуваним пов'язаний вектор характеристик  $\xi(O) = (\xi^1(O), \dots, \xi^d(O)) \in \mathbb{R}^d$ . В результаті обстеження  $N$  осіб  $O_1, \dots, O_N$  отримано значення  $\xi_j = (\xi_j^1, \dots, \xi_j^d) = \xi(O_j)$ .

Дослідника цікавить, чи відрізняється розподіл характеристик  $\xi(O)$  у популяції осіб, які мають досвід вживання наркотиків, від розподілу  $\xi(O)$  у тих, хто наркотиків не вживав.

Позначимо  $\text{ind}(O)$  — статус особи  $O$  по відношенню до наркотиків:  $\text{ind}(O) = 1$ , якщо  $O$  не вживав наркотиків,  $\text{ind}(O) = 2$  — якщо  $O$  мав досвід вживання наркотиків. (Можна розглянути і більш детальну класифікацію, наприклад, розділити тих, хто обмежився однією спробою і після того не вживав наркотиків, тих, хто вживає їх не регулярно, і тих, хто знаходиться у стані сформованої наркотичної залежності).

Оскільки питання про відношення до наркотиків є дразливим, статус  $O_j$  невідомий. Однак за опосередкованими даними відомо, що у школі, в якій навчається  $O_j$ , частка тих, хто не має досвіду вживання наркотиків, становить  $w_j^1$ , а частка тих, хто має такий досвід —  $w_j^2 = 1 - w_j^1$ . Якщо  $O_j$  був обраний серед учнів школи навмання, то  $P\{\text{ind}(O_j) = m\} = w_j^m$ . Нехай  $H_m(\cdot)$  — розподіл психологічних характеристик особи, що має  $m$ -тий

статус. Тоді

$$P\{\xi_j \in A\} = \sum_{m=1}^2 w_j^m H_m(A),$$

тобто розподіл спостережень описується моделлю суміші зі змінними концентраціями. Для дослідження даних можна тепер використати всю ту техніку, яка була описана вище у контексті аналізу генетичних даних: оцінювання розподілів, моментів, квантилів, щільностей розподілу для психологічних характеристик осіб, що вживають або не вживають наркотики. Можна також будувати класифікатори, які за психологічними характеристиками особистості намагатимуться визначити її статус по відношенню до наркотиків.

При такому підході виникають певні сумніви методологічного характеру.

По-перше, якщо у обстеженні приймає участь велика кількість учнів однієї школи, то вибірку всередині цієї школи слід вважати вибіркою без повернення. У такій ситуації статуси обстежуваних не можна вважати незалежними між собою, отже і спостереження  $\xi_j$  будуть залежними. Це зауваження, безумовно є важливим, і для таких ситуацій потрібно використовувати модифікації відповідних алгоритмів, які враховували б наявність залежності.

Друге зауваження стосується напрямку причинних зв'язків у розглядуваній моделі. У нашому генетичному прикладі цей напрямок очевидний: той чи інший різновид генотипу є причиною, що визначає довжину тіла та інші фенотипічні ознаки тварини. Ситуації, коли зміни довжини тіла впливають на генетичні особливості даної миші, у сучасній генетиці вважаються неможливими.

Для психологічного прикладу напрямок причинності не можна визначити так однозначно, але скоріше він є оберненим: не вживання наркотиків приводить до змін рівня екстравертованості, а екстравертний підліток має інші шанси стати споживачем наркотиків, ніж інтравертний<sup>4</sup>. Для рівня інтелекту можливий і “прямий” зв'язок: споживання наркотиків знижує інтелект (точніше, зменшує здатність правильно виконувати тести, які покликані виміряти рівень інтелекту).

При “оберненому” зв'язку, коли спостережувана характеристика є причиною, що визначає, до якої компоненти популяції потрапить об'єкт, мо-

---

<sup>4</sup>Більші чи менші? Це може залежати від культурних та соціальних особливостей даного міста, країни, народу.

дель суміші не виглядає адекватною. Природніше було б описувати такі дані у термінах регресії з дискретним відгуком (бінарним, якщо компонент лише дві). У простішому випадку це може бути логістична регресія. Насправді принципової протилежності між моделями, що спираються на регресійний підхід та моделями, які використовують логіку класифікації по компонентах суміші, немає. Часто за статистичними даними взагалі не можна визначити напрямок причинного зв'язку, а можна лише стверджувати, що деякий зв'язок існує.

У таких випадках використання моделі суміші зі змінними концентраціями для опису даних можна вважати не менш виправданим, ніж застосування інших технік (наприклад, регресійних). Однак при цьому не можна вкладати в інтерпретацію отриманих результатів зміст, якого вони насправді не мають. Наприклад, не варто сподіватись, що класифікатор, побудований за результатами дослідження, дозволить виявляти прихованих наркоманів на основі їх психологічних характеристик, так, як біологічні діагностичні процедури дозволяють визначати особливості генотипу за його відображенням у фенотипі. У іншому місті, за інших соціальних обставин зв'язок між психологічними змінними та ставленням до наркотиків може бути зовсім іншим, ніж у тих обставинах, у яких проводилось дослідження.

Нарешті третє зауваження, яке варто зробити щодо аналізу наших “наркологічних” даних, полягає в тому, що сам розподіл по компонентах тут виглядає досить штучно. Ми вже відмітили, що, крім розбиття на дві групи можна запропонувати і інші класифікації, побудовані на більш детальному аналізі відношення опитуваного до наркотиків. В принципі, “статус” міг би взагалі бути неперервною змінною, що характеризувала б більший або менший рівень використання наркотиків даною особою. Чи не є у такій ситуації застосування моделі скінченної суміші надмірним спрощенням ситуації?

Зрозуміло, що відповідь на це питання можна дати лише у рамках реального дослідження, причому статистик і спеціаліст у предметній області мають тісно співпрацювати для того, щоб ця відповідь була коректною.

Зроблені нами перестороги мають на меті показати можливі обмеження у застосуванні моделі суміші зі змінними концентраціями. Втім, вони так або інакше стосуються і більшості інших статистичних методів.

## 1.3 Приклади модельованих даних

Як, розглядаючи реальні дані, помітити, що їх природно описувати моделлю суміші зі змінними концентраціями? Для того, щоб продемонструвати це, ми використаємо три приклади модельних даних, згенерованих датчиками псевдовипадкових чисел. Дані перших двох прикладів описуються моделлю суміші зі змінними концентраціями, що містить дві компоненти з розподілами  $H_1$  та  $H_2$ . Концентрація першої компоненти у суміші змінюється лінійно від (майже) 0 до 1:  $w_j^1 = j/N$ , де  $N$  — кількість спостережень. Таким чином, функція розподілу  $j$ -того спостереження —  $\xi_j$  має вигляд

$$F_j(x) = P\{\xi_j < x\} = \frac{j}{N}H_1(x) + (1 - \frac{j}{N})H_2(x).$$

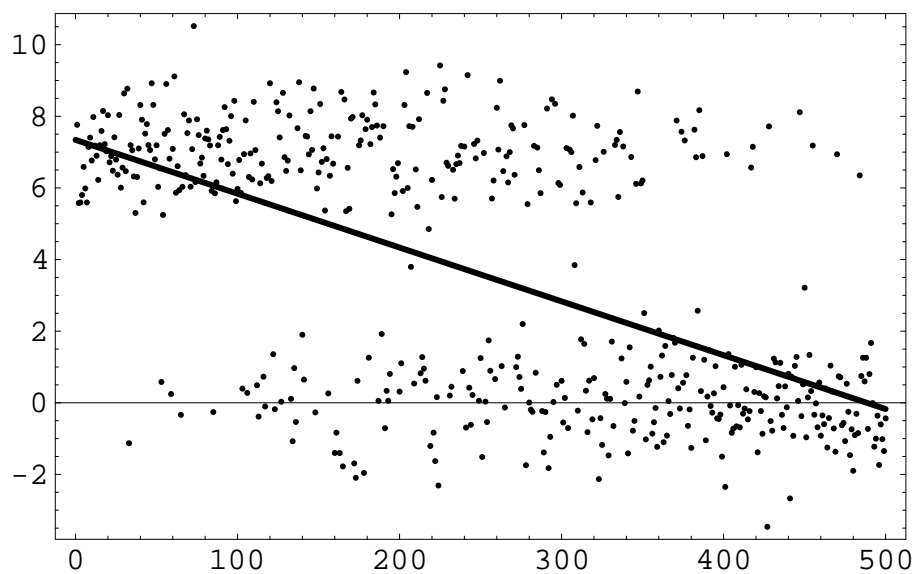
У першому прикладі розподіли обох компонент гауссові:  $H_1 \sim N(0, 1)$ ,  $H_2 \sim N(7, 1)$ . На діаграмі розсіювання (рис. 1.1а) по горизонталі відкладено номер спостереження по порядку —  $j$ , а по вертикалі — значення відповідного  $\xi_j$ . Чудово помітно, як поступово друга компонента у суміші змінюється першою. Зрозуміло, що спроба опису змін розподілу  $\xi_j$  в залежності від  $j$  за допомогою, скажімо, такої стандартної моделі як лінійна регресія є цілком не адекватними. Лінія регресії (зображена на діаграмі) вірно відображає зміну математичних сподівань  $F_j$ , але сам феномен двох компонент ігнорує зовсім.

Розбиття на дві компоненти чудово помітне і на гістограмі даних (рис. 1.1б). У певному розумінні, цей приклад є “нецікавим”: дослідник може практично безпомилково розділити суміш на компоненти і далі досліджувати кожен компонент окремо.

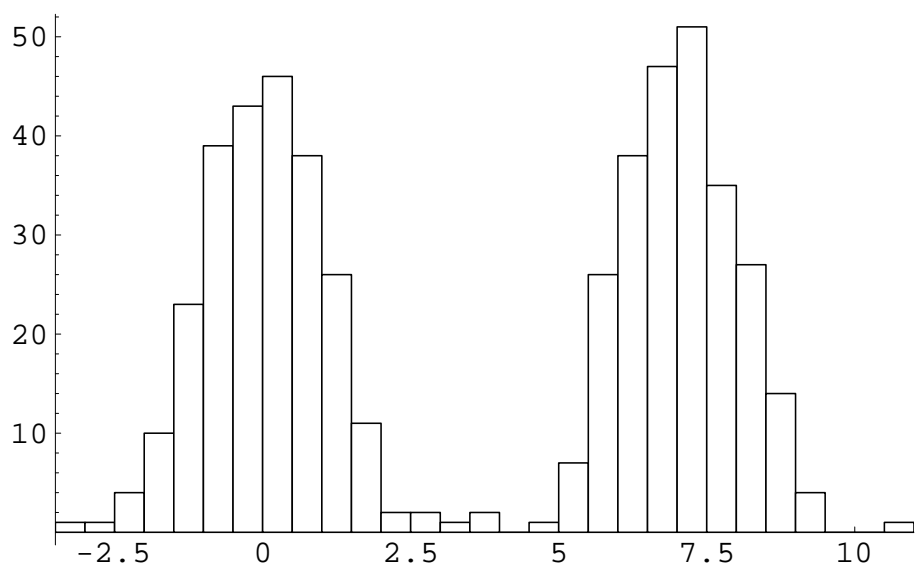
Складніша ситуація зображена на рис.1.2: тут розподіли компонент  $H_1 \sim N(0, 1)$ ,  $H_2 \sim N(2, 1)$ .

Вони значно ближчі один до одного, тому діаграма розсіювання не сприймається як складена з двох кластерів. Скоріше це виглядає як лінійна регресія з сильно розкиданими похибками. Гістограма також не дає можливості помітити суміш.

Але і в цій, складній для аналізу ситуації, зважені емпіричні функції розподілу адекватно оцінюють розподіли компонент (див. рис.1.3). Тут пунктиром зображено графіки справжніх функцій розподілу, а суцільною лінією — їх оцінки. (Оцінка для другої компоненти виглядає помітно гіршою ніж для першої, але це чисто випадковий ефект, на інших даних могло б бути навпаки).



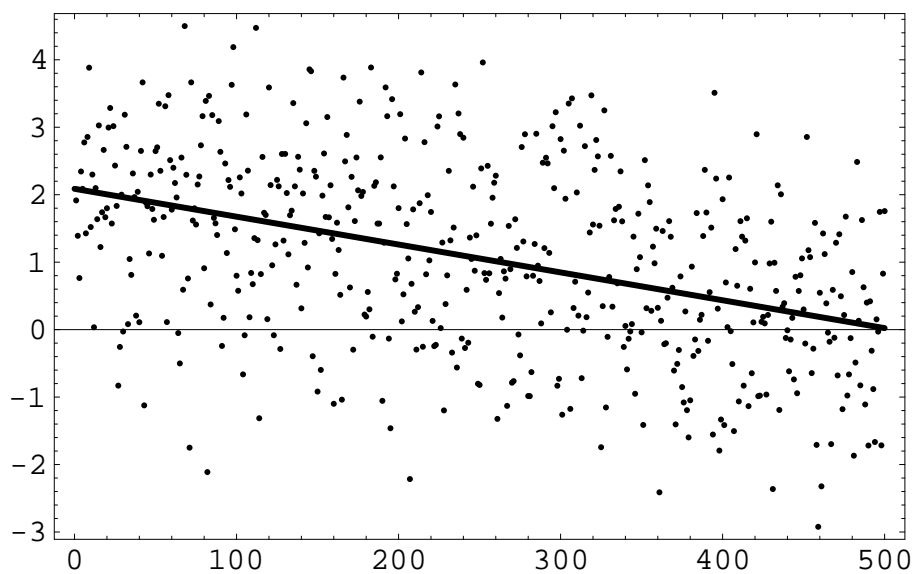
(a)



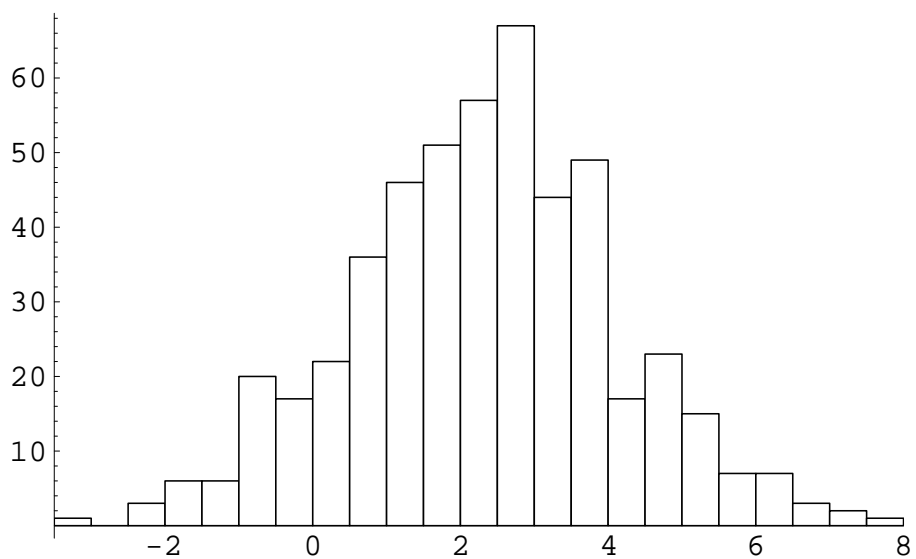
(б)

Рисунок 1.1: “Проста” двокомпонентна суміш.



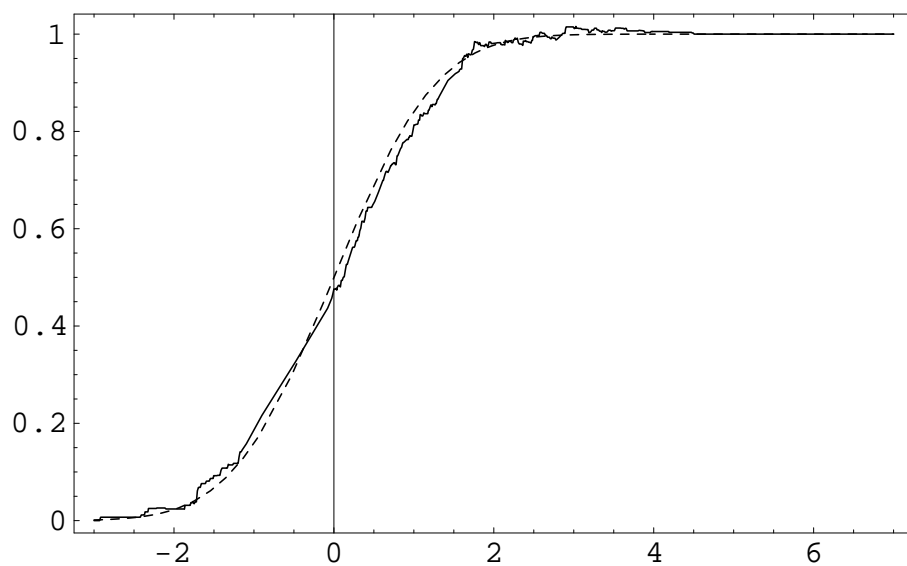


(a)

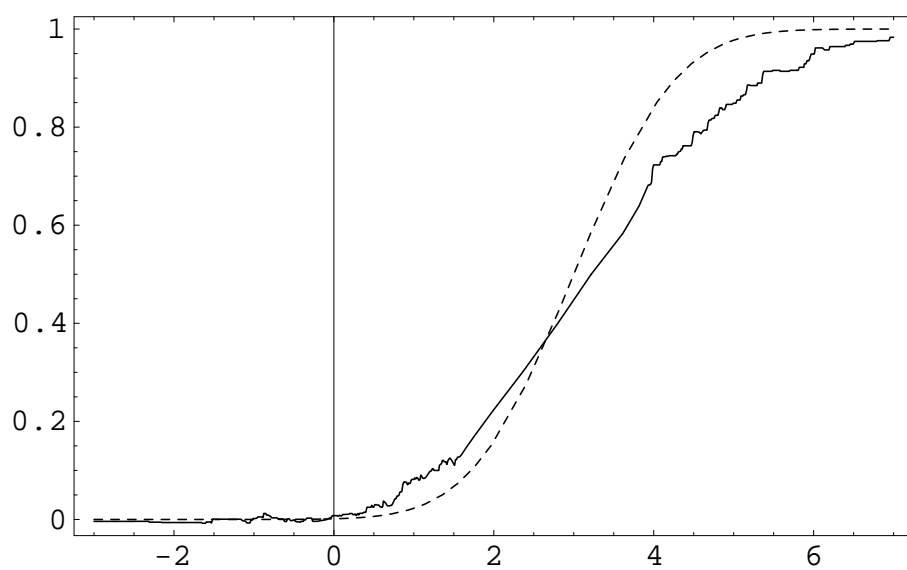


(б)

Рисунок 1.2: “Непомітна” двокомпонентна суміш.



(а)



(б)

Рисунок 1.3: Функції розподілу та їх оцінки для “непомітної” суміші.  
(а) перша компонента; (б) друга компонента

Для порівняння наведемо результат аналізу даних, згенерованих за регресійною моделлю:

$$\xi_j = 3 - 0.00317j + \varepsilon_j,$$

де  $\varepsilon_j$  — незалежні, гауссові, однаково розподілені похибки з нульовим середнім та дисперсією 2.89. Ці параметри обрані, щоб діаграма розсіювання вийшла подібною до діаграми для сумішей. І дійсно, на діаграмі рис. 1.4а не можна помітити принципових відмінностей від діаграми рис. 1.3а.

Більше того, на рис. 1.4б наведено графіки емпіричних функцій розподілу, які мали б бути оцінками справжніх функцій розподілу компонент суміші. Зрозуміло, що для даних, отриманих у регресійній моделі, компонент не існує в принципі. Однак оцінки виглядають цілком природно, не гірше ніж ті, які зображені на рис. 1.3. Вони не є монотонними і їх значення трохи виходять за межі інтервалу  $[0, 1]$ , але те ж саме можна помітити і для оцінок з попереднього прикладу.

Чи є який-небудь простий спосіб відрізнити такі “регресійні” дані від даних, що описуються моделлю сумішей? Можна, наприклад, звернути увагу на дисперсії спостережень  $\xi_j$ . У розглядуваній регресійній моделі  $\text{Var } \xi_j = 2.89$  є константою. У моделі двокомпонентної суміші

$$\begin{aligned} \text{Var } \xi_j &= E \xi_j^2 - (E \xi_j)^2 \\ &= w_j^1(\sigma_1^2 + m_1^2) + w_j^2(\sigma_2^2 + m_2^2) - (w_j^1 m_1 + w_j^2 m_2)^2, \end{aligned}$$

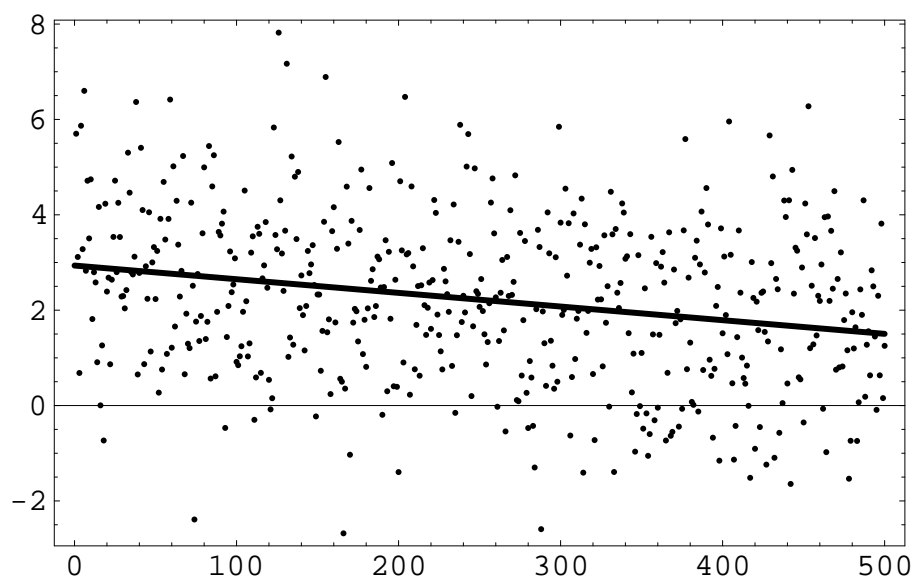
де  $m_i$  та  $\sigma_i^2$  позначають математичне сподівання та дисперсію  $i$ -тої компоненти. Легко бачити, що для  $w_j^1 = j/N$   $\text{Var } \xi_j$  є квадратичною функцією від  $j$ , вигнутою вгору.

Щоб побачити це на рисунку, підрахуємо залишки лінійної регресії для  $X_j$  по  $j$ :  $e_j = \hat{b}_0 + \hat{b}_1 j$ , де  $\hat{b}_i$  — оцінки методу найменших квадратів для коефіцієнтів лінійної регресії. Розглянемо діаграму квадратів залишків —  $(e_j)^2$  (рис. 1.5).

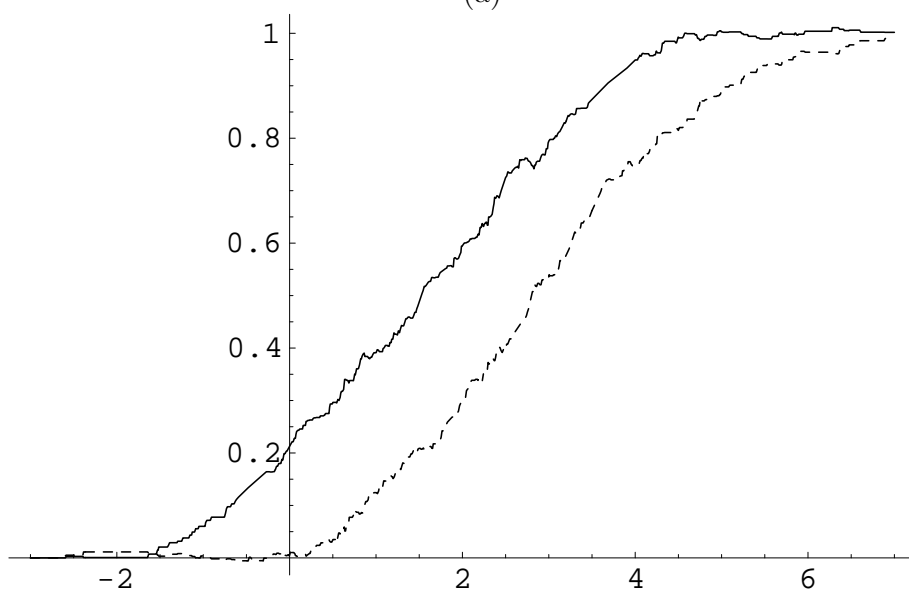
На рис. 1.5а зображені квадрати залишків спостережень з суміші. Помітно, що лінія регресії<sup>5</sup> яка описує залежність середнього  $(e_j)^2$  від  $j$  вигинається вгору. На рис. 1.5б зображено квадрати залишків справжньої регресії — для них залежності від номеру спостереження немає (як і слід було сподіватись).

Якщо відкинути модель суміші і стояти на позиціях регресійної моделі, рис. 1.5а можна інтерпретувати як свідчення гетероскедастичності похибок. Звичайно, коли дослідник має справу з реальними даними, механізм

<sup>5</sup>Це поліноміальна регресія третього порядку.

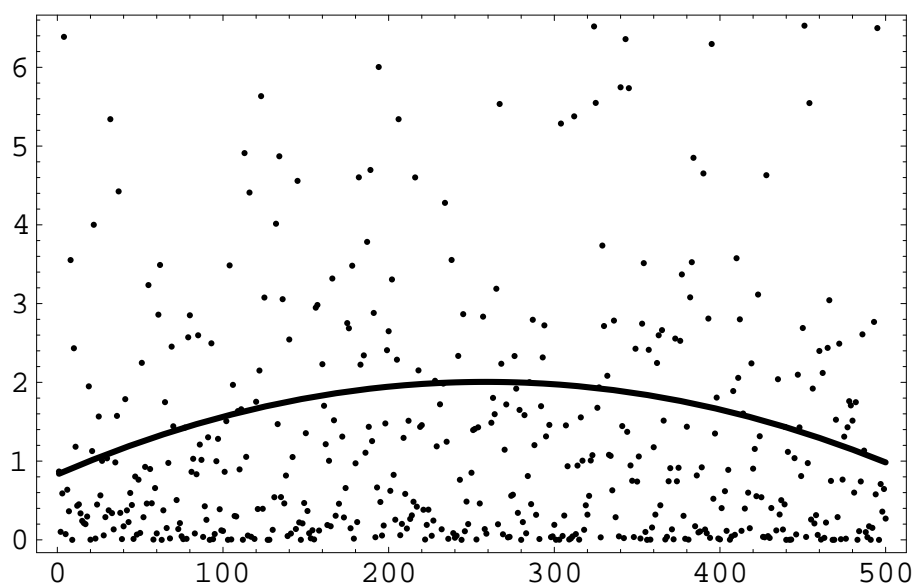


(а)

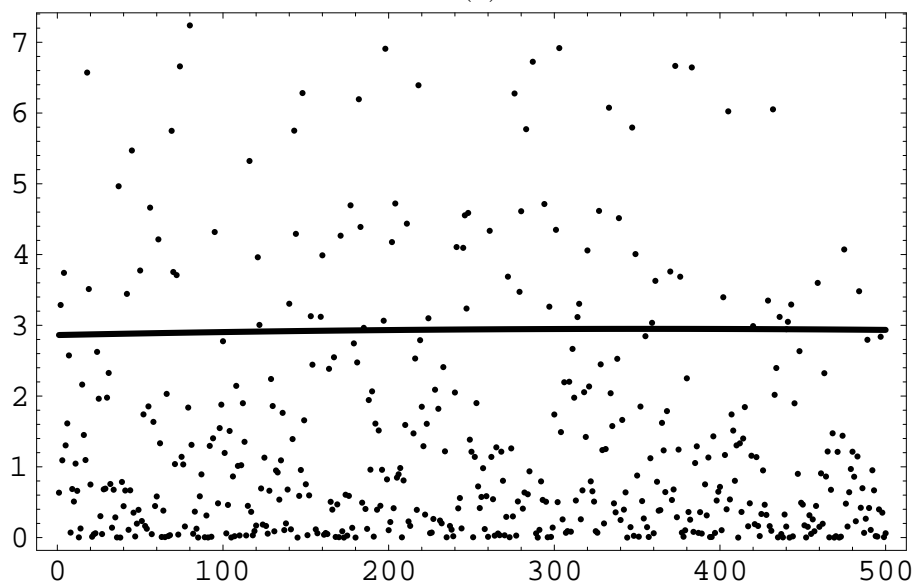


(б)

Рисунок 1.4: Регресія. (а) діаграма розсіювання (б) оцінки розподілів неіснуючих компонент



(а)



(б)

Рисунок 1.5: Діаграма квадратів залишків зі згладжуванням поліноміальною регресією. (а) суміш зі змінними коецентраціями (б) регресія

утворення яких невідомий, припущення про гетероскедастичну регресійну модель є цілком допустимим. Але ця модель на змістовному рівні вже не є такою очевидною, як проста лінійна регресія. В усякому випадку, залежність дисперсії від номера спостережень вимагатиме якогось пояснення. Для суміші зі змінними концентраціями цей ефект впливає безпосередньо з самої моделі.

Зміна дисперсії від спостереження до спостереження буде ознакою моделі суміші лише в тому випадку, коли математичні сподівання і/або дисперсії компонент є різними. Якщо вони однакові, звичайні методи аналізу можуть взагалі не помічати неоднорідність даних. Але спеціальні техніки аналізу сумішей, описані у цій книзі, дозволяють виділити і дослідити особливості розподілів компонент таких даних.

Підсумовуючи, можна сказати, що питання про можливість застосування моделі суміші до певних даних не простіше і не складніше, ніж для більшості класичних статистичних моделей. У простіших випадках наявність суміші є очевидною. У складних — тільки виходячи з певних апріорних міркувань про природу даних доцільно визначати, якою саме з можливих альтернативних моделей слід їх описувати.

## Розділ 2

# Оцінювання функцій розподілу

Функції розподілу є основними характеристиками, що використовуються у теорії ймовірностей та математичній статистиці для опису розподілу випадкових величин та векторів. Їх оцінки можуть бути корисними як самі по собі, так і для побудови алгоритмів оцінювання інших ймовірнісних характеристик та перевірки гіпотез. У цьому розділі ми розглядаємо декілька підходів до оцінювання функцій розподілу, що спираються на використання зважених емпіричних мір. Простіша версія таких оцінок описана у параграфі 2.1. Дані про асимптотику зважених емпіричних мір та функцій розподілу вміщені у параграфі 2.2. Подальші параграфи цього розділу присвячені виправленням та уточненням цих оцінок.

## 2.1 Незміщені мінімаксні оцінки

У цьому розділі розглядаються задачі оцінювання розподілів компонент за спостереженнями з суміші зі змінними концентраціями. Вважаємо, що спостерігаються деякі об'єкти  $O_1, \dots, O_N$ , кожен з яких може належати одній з  $M$  популяцій (компонент). Номер популяції, якій належить об'єкт  $O_j$  позначимо  $\text{ind}(O_j)$ . Справжнє значення  $\text{ind}(O_j)$  вважається невідомим, але відомі ймовірності  $w_{j:N}^k = \mathbf{P}\{\text{ind}(O_j) = k\}$ . Ці ймовірності називають концентраціями або ймовірностями перемішування (mixing probabilities). Концентрації компонент повинні задовольняти наступні умови:  $0 \leq w_{j:N}^m \leq 1$ ,  $\sum_{m=1}^M w_{j:N}^m = 1$ .

У всіх об'єктів спостерігається один і той самий набір характеристик  $\xi$ . Для  $O_j$  цей набір (вектор) позначимо  $\xi_{j:N} = \xi(O_{j:N})$ . Множину всіх можливих значень характеристик  $\xi$  позначимо  $\mathcal{X}$ . У цій книжці, як правило,  $\mathcal{X}$  це або дійсна пряма або дійснозначний вектор. Взагалі кажучи,  $\mathcal{X}$  може бути будь-яким вимірним простором, тобто простором, на якому задана  $\sigma$ -алгебра вимірних підмножин  $\mathfrak{A}$ . Спостережувані характеристики вважаємо випадковими елементами  $\mathcal{X}$ , незалежними для різних  $O_j$ . Розподіл цих характеристик залежить від того, якій компоненті належить об'єкт. Розподіл характеристик  $k$ -тої компоненти будемо позначати  $H_m$ , тобто

$$H_m(A) = \mathbf{P}\{\xi(O) \in A \mid \text{ind}(O) = m\}$$

для всіх вимірних множин з  $\mathcal{X}$ . Надалі, у випадку, коли  $\mathcal{X} = \mathbb{R}^d$  — скінченновимірний векторний простір, будемо також  $H_m$  позначати відповідну функцію розподілу:

$$H_m(x) = \mathbf{P}\{\xi(O) < x \mid \text{ind}(O) = m\}$$

для всіх  $x \in \mathbb{R}^d$ . (Нерівності для векторів слід розуміти покоординатно).

Таким чином, розподіл спостережуваних характеристик має вигляд

$$\mathbf{P}\{\xi_{j:N} \in A\} = \mu_{j:N}(A) = \sum_{m=1}^M w_{j:N}^m H_m(A). \quad (2.1)$$

Надалі ми будемо використовувати схему серій для опису асимптотичної поведінки наших оцінок при необмеженому зростанні обсягу вибірки, тобто коли  $N \rightarrow \infty$ . Тому спостережувані дані  $\Xi_N = (\xi_1, \dots, \xi_N)$  розглядаються як один рядок трикутного масиву  $\Xi = \{\Xi_N : N = 1, 2, \dots\}$ . (Зрозуміло, що реально статистик має справу лише з вибіркою фіксованого обся-



гу — з одним рядочком  $\Xi$ ). Відповідно і концентрації кожної компоненти можна трактувати як трикутні масиви:  $w^m = \{w_{j:N}^m, j = 1, \dots, N, N \in \mathbb{N}\}$ .

Крім концентрацій будуть використовуватись і інші масиви аналогічної структури. Часто до таких масивів буде застосовуватись оператор усереднення по рядочках. Ми будемо позначати його

$$\langle w^m \rangle_N = \langle w^m \rangle_N = \frac{1}{N} \sum_{j=1}^N w_{j:N}^m.$$

Аналогічно, якщо  $a = \{a_{j:N}, j = 1, \dots, N, N \in \mathbb{N}\}$ ,  $b = \{b_{j:N}, j = 1, \dots, N, N \in \mathbb{N}\}$ , то

$$\begin{aligned} \langle a. + b. \rangle_N &= \frac{1}{N} \sum_{j=1}^N (a_{j:N} + b_{j:N}), \quad \langle (a.)^2 \rangle_N = \frac{1}{N} \sum_{j=1}^N (a_{j:N})^2, \\ \langle a.b. \rangle_N &= \frac{1}{N} \sum_{j=1}^N (a_{j:N} b_{j:N}) \end{aligned} \quad (2.2)$$

і т.д. Функціонал  $\langle a.b. \rangle_N$ , визначений (2.2), можна розглядати як скалярний добуток  $N$ -тих рядочків наших масивів. Якщо границя  $\lim_{N \rightarrow \infty} \langle a. \rangle_N$  існує, то ми будемо позначати її  $\langle a. \rangle = \langle a \rangle$ .

Ми будемо використовувати зважені емпіричні міри вигляду

$$\hat{\mu}_N(A, a) = \frac{1}{N} \sum_{j=1}^N a_{j:N} \mathbb{1}\{\xi_{j:N} \in A\}, \quad A \in \mathfrak{A}, \quad (2.3)$$

як оцінки для  $H_m$  за  $\Xi_N$ . Тут  $a$  є деяким не випадковим трикутним масивом вагових коефіцієнтів. (Під не випадковістю ми маємо на увазі незалежність від  $\Xi_N$ , але не від  $w^m$ ). Ці вагові коефіцієнти часто будуть залежати від деякого параметра (параметрів), скажімо,  $\vartheta \in \Theta$ . У таких випадках ми інколи будемо писати просто  $\hat{\mu}_N(A, a(\vartheta)) = \hat{\mu}_N(A, \vartheta)$ .

Якщо  $\mathcal{X} = \mathbb{R}^d$  і  $A(x) = \{y \in \mathbb{R}^d : y < x\}$  то

$$\hat{F}_N(x, a) := \hat{\mu}_N(A(x), a) = \frac{1}{N} \sum_{j=1}^N a_{j:N} \mathbb{1}\{\xi_{j:N} < x\} \quad (2.4)$$

є оцінкою для функції розподілу  $H_k$  і зветься зваженою емпіричною функцією розподілу (з.е.ф.р.).

Якщо вимагати незміщеності  $\hat{\mu}_N(A, a)$  як оцінки  $H_k(A)$ , то з (2.3) отримуємо

$$H_k(A) = \mathbb{E} \hat{\mu}_N(A, a) = \frac{1}{N} \sum_{j=1}^N a_{j:N} \mathbb{P}\{\xi_{j:N} \in A\} = \sum_{m=1}^M \langle aw^m \rangle_N H_m(A)$$

для всіх можливих наборів  $H_m$ ,  $m = 1, \dots, M$ . При  $N > M$  ця умова виконується тоді і тільки тоді, коли

$$\langle aw^m \rangle_N = \mathbb{1}\{m = k\} \text{ для всіх } m = 1, \dots, M. \quad (2.5)$$

Умову (2.5) ми будемо називати умовою незміщеності.

Помітимо, що коли (2.5) виконується, то

$$\langle a \rangle_N = \langle a1 \rangle_N = \langle a \sum_{m=1}^M w^m \rangle_N = \sum_{m=1}^M \langle aw^m \rangle_N = 1,$$

отже, якщо  $\hat{\mu}_N$  є незміщеною оцінкою для  $H_k$ , то  $\hat{\mu}_N(\mathcal{X}) = \frac{1}{N} \sum_{j=1}^M a_{j:N} = 1$ .

Однак, оскільки  $w^m \geq 0$  для всіх  $m$ , то з умови незміщеності випливає, що  $a_{j:N}$  повинні приймати негативні значення для деяких  $j$ . З (2.3) легко бачити, що у цьому випадку  $\hat{\mu}_N$  не може бути ймовірнісною (тобто невід'ємною) мірою на  $\mathcal{X}$  якщо всі  $\xi_{j:N}$  є різними. З іншого боку, для всіх  $A$ ,  $|\hat{\mu}_N(A)| \leq \langle |a_{\cdot:N}| \rangle_N$ . Отже, якщо (2.5) виконано, то  $\hat{\mu}_N$  є знакозмінною мірою (зарядом) зі скінченною варіацією на  $\sigma$ -алгебрі  $\mathfrak{A}$ .

З класу всіх можливих незміщених оцінок вигляду (2.3) доцільно обрати одну, в деякому розумінні найкращу. Як міру якості у цьому параграфі будемо використовувати гарантований ризик при квадратичній функції витрат<sup>1</sup>. Нагадаємо, що це таке.

Нехай є деяка оцінка  $\tilde{H}_k(A)$  для  $H_k(A)$  за спостереженнями  $\Xi_N$ . Ми будемо вважати, що витрати від використання неточної оцінки  $\tilde{H}_k$  замість справжнього значення  $H_k$  задаються квадратичною функцією ризику  $(\tilde{H}_k(A) - H_k(A))^2$ . Відповідно, середні витрати при використанні оцінки  $\tilde{H}_k$  становлять  $\mathbb{E}(\tilde{H}_k(A) - H_k(A))^2$ . Тоді

$$R(\tilde{H}_k) = \sup_{H_m, m=1, \dots, M, A \in \mathfrak{A}} \mathbb{E}(\tilde{H}_k(A) - H_k(A))^2$$

<sup>1</sup>Про інший підхід, що спирається на поняття асимптотичної ефективності, див. п. 2.4

являє собою гарантований ризик оцінки  $\tilde{H}_k$ , тобто максимальні витрати, які в середньому можна понести при використанні оцінки при найгірших значеннях характеристик моделі. Ми будемо брати  $\sup$  по всіх можливих ймовірнісних розподілах на  $(\mathcal{X}, \mathfrak{A})$  оскільки розглядається непараметрична задача оцінювання. Розглянемо зважену емпіричну міру  $\hat{\mu}(A, a)$  як оцінку для  $H_k(A)$ . Тоді гарантований ризик буде функцією вектора коефіцієнтів  $a$ ,  $J(a) = R(\hat{\mu}(\cdot, a))$ . Знайдемо  $J(a)$  якщо виконані умови незміщеності (2.5). Використовуючи ці умови, отримуємо

$$\begin{aligned} J(a) &= \sup_{H_m, A} \mathbb{E} \left( \frac{1}{N} \sum_{j=1}^N a_{j:N} (\mathbb{I}\{\xi_{j:N} \in A\} - \mathbb{P}\{\xi_{j:N} \in A\}) \right)^2 \\ &= \sup_{H_m, A} \frac{1}{N^2} \sum_{j=1}^N (a_{j:N})^2 \mathbb{E} (\mathbb{I}\{\xi_{j:N} \in A\} - \mathbb{P}\{\xi_{j:N} \in A\})^2 \\ &= \frac{1}{4N} \langle (a)^2 \rangle_N, \end{aligned}$$

оскільки  $\sup_{H_m, A} (\mathbb{P}\{\xi_{j:N} \in A\} - (\mathbb{P}\{\xi_{j:N} \in A\})^2) \leq \frac{1}{4}$  причому значення  $\frac{1}{4}$  досягається коли  $H_m(A) = \frac{1}{2}$  для всіх  $m = 1, \dots, M$ .

Отже, ми повинні знайти вектор  $a = (a_{1:N}, \dots, a_{N:N})$ , який мінімізує

$$J(a) = \frac{1}{4N} \langle a^2 \rangle_N, \quad (2.6)$$

при виконанні умов незміщеності

$$\langle w^m a \rangle_N = \mathbb{I}\{m = k\} \quad \forall m = 1, \dots, M. \quad (2.7)$$

Використаємо метод множників Лагранжа для розв'язання цієї задачі мінімізації. Як відомо, необхідною умовою того, що  $a$  є точкою умовного екстремуму  $J$  при обмеженнях (2.7), є

$$\frac{\partial}{\partial a_{j:N}} (J(a) + \sum_{m=1}^M \lambda_m \langle w^m a \rangle_N) = 0, \quad (2.8)$$

де  $j = 1, \dots, N$ ,  $\lambda_m$  — невизначені множники Лагранжа. Умова (2.8) рівнозначна

$$a = \sum_{l=1}^M c_l w_{:,N}^l,$$

де  $c_l$  — довільні константи. Тобто вектор оптимальних вагових коефіцієнтів є лінійною комбінацією векторів навантажень. Підставляючи цей розклад у (2.7), отримуємо систему лінійних рівнянь для  $c_l$ :

$$\sum_{l=1}^M c_l \langle w^l w^m \rangle_N = \mathbb{I}\{m = k\}. \quad (2.9)$$

Припустимо, що матриця  $\Gamma_N = (\langle w^l w^m \rangle_N)_{l,m=1}^M$  є невиродженою. Тоді (2.9) має єдиний розв'язок

$$c_l = \frac{(-1)^{l+k} \gamma_{lk:N}}{\det \Gamma_N},$$

де  $\gamma_{lk:N}$  — це  $lk$ -мінор  $\Gamma_N$ . Відповідний оптимальний (мінімаксний) вектор вагових коефіцієнтів визначається як

$$a_{j:N}^k = \frac{1}{\det \Gamma_N} \sum_{m=1}^M (-1)^{m+k} \gamma_{km:N} w_{j:N}^m. \quad (2.10)$$

Умова  $\det \Gamma_N \neq 0$  еквівалентна лінійній незалежності системи векторів  $w_{\cdot:N}^m$ ,  $m = 1, \dots, M$ , оскільки  $\Gamma_N$  є матрицею Грама цієї системи у скалярному добутку  $\langle w^l w^m \rangle_N$ . Щоб підкреслити, що  $\langle w^l w^m \rangle_N$  є скалярним добутком векторів з  $\mathbb{R}^N$ , інколи будемо записувати його у вигляді  $\langle w^l, w^m \rangle_N = \langle w^l w^m \rangle_N$ .

Підставивши  $a^k$ , визначені (2.10), у (2.6), отримуємо найменше можливе значення гарантованого ризику

$$J(a^k) = \frac{\gamma_{kk:N}}{4N \det \Gamma_N}.$$

Дійсно,  $J(a^k) = \frac{1}{4N} \langle a^k, a^k \rangle_N$ . Помітимо, що  $a_{\cdot:N}^k = e_k^T \Gamma_N^{-1} \vec{w}_{\cdot:N}$ , де  $\vec{w}_{\cdot:N} = (w_{\cdot:N}^1, \dots, w_{\cdot:N}^M)^T$ . Отже маємо

$$\langle a^k, a^k \rangle_N = \langle e_k^T \Gamma_N^{-1} \vec{w}_{\cdot:N}, e_k^T \Gamma_N^{-1} \vec{w}_{\cdot:N} \rangle_N = e_k^T \Gamma_N^{-1} \Gamma_N \Gamma_N^{-1} e_k = e_k^T \Gamma_N^{-1} e_k.$$

Ми показали, що вагові коефіцієнти  $a^k$ , визначені (2.10), забезпечують найкращий (з точки зору гарантованого ризику) результат при оцінюванні  $H_k$  зваженою емпіричною мірою  $\hat{\mu}_N(A, a)$ . Чи можна оцінити  $H_k$  якою-небудь іншою незміщеною оцінкою, що мала б менший гарантований ризик? Відповідь негативна.

**Теорема 2.1.1** Нехай  $\Xi_N$  має розподіл (2.1),  $w^m$   $m = 1, \dots, M$  відомі,  $H_m$  — невідомі. Якщо  $\det \Gamma_N \neq 0$ , то для будь-якої вимірної функції  $\tilde{H}_k : \mathfrak{A} \times \mathcal{X}^N \rightarrow [0, 1]$ , такої, що  $\mathbb{E} \tilde{H}_k(A, \Xi_N) = H_k(A)$  для всіх можливих розподілів  $H_m$ ,

$$R(\tilde{H}_k(\cdot, \Xi_N)) \geq J(a^k) = \frac{\gamma_{kk:N}}{4N \det \Gamma_N}.$$

**Доведення.** Виберемо будь-які  $x_1, x_2 \in \mathcal{X}$ ,  $x_1 \neq x_2$ . Нехай розподіли  $H_m$  мають вигляд

$$H_m(A) = p_m \mathbb{I}\{x_1 \in A\} + (1 - p_m) \mathbb{I}\{x_2 \in A\}. \quad (2.11)$$

Якщо обмежитись розглядом лише таких розподілів, то задача оцінки розподілів  $H_k$  зведеться до оцінки вектора  $p = (p_1, \dots, p_m)$ . Це задача параметричного оцінювання і нижня межа ризику для неї визначається нерівністю Крамера (див. 7.9).

Щоб використати цю нерівність, обчислимо інформаційну матрицю Фішера для стохастичного експерименту по оцінюванню  $H_m(\{x_1\}) = p_m$ ,  $m = 1, \dots, M$  за одним спостереженням  $\xi_{j:N}$  у випадку, коли справжні значення параметрів є  $p_m = p_m^0 = \frac{1}{2}$ . Для цього задамо міру  $\nu(A) = \mathbb{I}\{x_1 \in A\} + \mathbb{I}\{x_2 \in A\}$ . Міри  $H_m$  є абсолютно неперервними відносно  $\nu$  для всіх  $p$  та  $m$  і

$$h_m(x, p) = \frac{dH_m}{d\nu} = p_m \mathbb{I}\{x = x_1\} + (1 - p_m) \mathbb{I}\{x = x_2\}.$$

Елементи інформаційної матриці  $I^j = (I_{kl}^j)_{k,l=1}^M$  для спостереження  $\xi_{j:N}$  обчислюються за формулою

$$I_{kl}^j = \int_{\mathcal{X}} \frac{\partial \bar{h}_j(x, p)}{\partial p_k} \frac{\partial \bar{h}_j(x, p)}{\partial p_l} \frac{\nu(dx)}{h(x, p)} \Big|_{p=p^0},$$

де

$$\bar{h}_j(x, p) = \frac{d\mu_{j:N}}{d\nu} = \sum_{m=1}^M w_{j:N} h_m(x, p).$$

Легко бачити, що

$$I_{kl}^j = \frac{w_{j:N}^k w_{j:N}^l}{\sum_{m=1}^M w_{j:N} p_m^0} + \frac{w_{j:N}^k w_{j:N}^l}{1 - \sum_{m=1}^M w_{j:N} p_m^0} = 4w_{j:N}^k w_{j:N}^l$$

(оскільки  $p_m^0 = 1/2$ ,  $\sum_{m=1}^M w_{j:N}^m = 1$ ). З незалежності спостережень  $\xi_{j:N}$  отримуємо, що інформаційна матриця  $\Xi_N \in I = \sum_{j=1}^N I^j = 4N\Gamma_N$ . За нерівністю Крамера, враховуючи незміщеність оцінки  $\tilde{H}_k$ , отримуємо

$$\mathbb{E}(\tilde{H}_k(\{x_1\}, \Xi_N) - H_k(\{x_1\}))^2 \geq e_k^T I^{-1} e_k = \frac{e_k^T \Gamma_N^{-1} e_k}{4N} = \frac{\gamma_{kk}}{4N \det \Gamma_N},$$

де  $e_k$  є одиничним вектором у  $\mathbb{R}^M$ ,  $k$ -та координата якого дорівнює 1 (а інші є нулями). Символ  $A^T$  позначає транспонування матриці  $A$ .

Супремум в означенні гарантованого ризику береться по класу, який включає всі можливі розподіли  $H_k$  вигляду (2.11) і всі можливі множини  $A = \{x_1\}$ . Тому

$$R(\tilde{H}_k(\cdot, \Xi_N)) \geq \frac{\gamma_{kk}}{4N \det \Gamma_N}. \quad (2.12)$$

Теорема доведена.

**Приклад 1.** Нехай у суміші з двох компонент ( $M = 2$ ) концентрації є константами,  $w_{j:N}^1 = w$  не залежить від  $j$  та  $N$ ,  $w_{j:N}^2 = 1 - w$ . Тоді з умови незміщеності випливає

$$\langle wa^1 \rangle_N = w \langle a^1 \rangle_N = 1$$

$$\langle (1 - w)a^1 \rangle_N = (1 - w) \langle a^1 \rangle_N = 0.$$

Це можливо тільки тоді, коли  $w = 1$ , тобто коли у нашій “суміші” присутня лише одна компонента. Отже, зважена емпірична міра не може бути незміщеною оцінкою розподілу компонент у випадку, коли концентрації є константами.

Більше того, у цьому випадку однозначне оцінювання неможливе. Дійсно, у нашій моделі

$$\mathbb{P}\{\xi_{j:N} \in A\} = wH_1(A) + (1 - w)H_2(A), \quad (2.13)$$

де  $w$  — відоме число, а  $H_1$  і  $H_2$  повністю невідомі. Припустимо, що справжній розподіл  $\xi_{j:N}$  має вигляд

$$\xi_{j:N} = \begin{cases} x_1 & \text{з ймовірністю } \alpha \\ x_2 & \text{з ймовірністю } 1 - \alpha \end{cases}.$$

Нехай  $H_m(A) = p_m \mathbb{I}\{x_1 \in A\} + (1 - p_m) \mathbb{I}\{x_2 \in A\}$ , де  $p_m \in [0, 1]$  — деякі числа. Якщо

$$wp_1 + (1 - w)p_2 = \alpha, \quad (2.14)$$

то легко перевірити, що (2.13) виконується. Тому (2.13) не визначає  $H_1$  та  $H_2$  однозначно і їх неможливо оцінити навіть знаючи розподіли  $\xi_{j:N}$  абсолютно точно. Тим більше, їх не можна оцінити маючи лише вибірку  $\Xi_N$ .

**Приклад 2.** Нехай суміш складається з двох компонент, тобто  $M = 2$ . Тоді  $w_{j:N}^2 = 1 - w_{j:N}^1$ , отже, щоб задати концентрації досить визначити  $w_{j:N}^1$ . Надалі у випадку суміші з двома компонентами ми будемо позначати  $w_{j:N} = w_{j:N}^1$ , пропускаючи індекс 1 для спрощення позначень. Позначимо  $s^k = \langle (w)^k \rangle_N$ . За означенням,

$$\Gamma_N = \begin{pmatrix} \langle w, w \rangle_N & \langle w, 1 - w \rangle_N \\ \langle 1 - w, w \rangle_N & \langle 1 - w, 1 - w \rangle_N \end{pmatrix} = \begin{pmatrix} s^2 & s^1 - s^2 \\ s^1 - s^2 & 1 - 2s^1 + s^2 \end{pmatrix}$$

і  $\det \Gamma_N = \Delta = s^2 - (s^1)^2 = \langle w, w \rangle_N - (\langle w \rangle_N)^2$ . Відмітимо, що  $\Delta$  являє собою вибірккову дисперсію набору концентрацій  $w_{j:N}^1$ ,  $j = 1, \dots, N$  розглядуваних як “вибірка” (вибірккова дисперсія  $w_{j:N}^2$  така ж сама).

Зрозуміло, що  $\det \Gamma_N = \Delta = 0$  тоді і тільки тоді, коли  $w_{j:N}^1 = \text{const}$  не залежить від  $j$ . У цьому випадку непараметричне оцінювання  $H_k$  неможливе, як це показано у прикладі 1. Якщо концентрації не є константами, то  $\gamma_{11:N} = 1 - 2s^1 + s^2$ ,  $\gamma_{22:N} = s^2$ ,

$$a_{j:N}^1 = ((1 - s^1)w_{j:N} + (s^2 - s^1))/\Delta, \quad a_{j:N}^2 = (s^2 - s^1 w_{j:N})/\Delta. \quad (2.15)$$

## 2.2 Асимптотика емпіричних мір

У цьому параграфі ми розглянемо основні результати про поведінку емпіричних мір і емпіричних функцій розподілу при зростанні обсягу вибірки до нескінченності. Вони будуть використовуватись далі у цій книжці для аналізу асимптотичної поведінки різних оцінок та класифікаторів, призначених для аналізу сумішей зі змінними концентраціями. Ці результати можна розділити на три великих групи: (i) твердження про збіжність майже напевне або за ймовірністю до граничного невинядкового значення, (ii) оцінки швидкості цієї збіжності та (iii) твердження про граничний розподіл певним способом нормованих відхилень емпіричних мір від їх граничних значень (ці нормовані відхилення називають емпіричними процесами). Якщо розглядати емпіричні міри як оцінки для невідомих розподілів компонент, то твердження групи (i) описують умови консистентності оцінок, а (iii) — умови асимптотичної нормальності (випадок

негауссових граничних розподілів для емпіричних мір ми не розглядаємо). Однак, оскільки емпіричні міри можна використовувати не тільки як оцінки розподілів, ми наведемо тут децю більш загальні теореми про їх асимптотику, ніж це потрібно для оцінювання. Консистентність та асимптотична нормальність будуть наслідками цих загальних теорем.

Асимптотику емпіричних мір  $\hat{\mu}_N(a, A)$ , визначених (2.3) коли  $N \rightarrow \infty$ , можна вивчати або при фіксованих вагових коефіцієнтах  $a$  та множині  $A$ , або рівномірно по деякому класу множин і/або деякому класу вагових функцій, або, розглядаючи  $\hat{\mu}_N(a, A)$  як функції  $A$  та функціонали  $a$  — у відповідних функціональних просторах. Тут ми, в основному, зосередимося на другому варіанті. Одразу відмітимо, що годі сподіватись, наприклад, рівномірної збіжності  $\hat{\mu}_N(a, A)$  на класі всіх можливих вимірних множин у  $\mathbb{R}$ . Цей клас є занадто великим. Ми будемо обирати такі класи множин, на яких умови рівномірної збіжності є не набагато важчими, ніж умови збіжності при фіксованій множині. Для збіжності майже напевне це умова скінченної апроксимованості, описана нижче. Скінченна апроксимованість класу множин відносно певної міри на  $\mathcal{X} = \mathbb{R}^d$  — досить слабка умова, як показують лема 2.2.1, 2.2.2, подані нижче.

При дослідженні збіжності розподілів емпіричних процесів ми обмежуємось лише функціями розподілу, тобто на роль множин  $A$  вибираємо лише напівнескінченні інтервали (прямокутники) у просторі  $\mathcal{X} = \mathbb{R}^d$ . Це дуже сильне обмеження, якого, в принципі, можна позбутись. Але для потреб асимптотичного аналізу алгоритмів, які розглядаються у цій книжці, нам буде цілком достатньо тверджень про рівномірну асимптотичну нормальність емпіричних функцій розподілу. Під асимптотичною нормальністю звичайно розуміють слабку збіжність розподілів. Відповідно рівномірною асимптотичною нормальністю природно вважати слабку збіжність розподілів досліджуваної послідовності випадкових процесів (полів) розглядуваних як елементи деякого простору функцій з рівномірною метрикою до певного гауссового розподілу на цьому просторі. Чудовим прикладом застосування такого підходу є книжка П. Білінгслі [1].

Однак використання функціональних просторів для опису рівномірної слабкої збіжності зустрічається з певними труднощами, коли виникає потреба досліджувати розривні функції на некомпактних множинах — це ускладнює опис відповідних функціональних просторів, робить результати менш зрозумілими інтуїтивно. Тому ми, поруч з функціональним підходом, будемо використовувати техніку одного ймовірнісного простору, введену А.В. Скороходом [37]. Вона спирається на той фундамен-



тальний факт, що у сепарабельних метричних просторах слабка збіжність розподілів випадкових елементів  $\zeta_n$  до розподілу елемента  $\zeta$  еквівалентна існуванню послідовності  $\zeta'_n$  та елемента  $\zeta'$ , таких, що розподіл  $\zeta'_n$  той самий що у  $\zeta_n$ , а у  $\zeta'$  — той самий, що у  $\zeta$ , причому  $\zeta'_n \rightarrow \zeta'$  майже напевно (теорема Скорохода). Таким чином, слабка збіжність виявляється, парадоксальним чином еквівалентною “сильній” збіжності майже напевно, але не самої досліджуваної послідовності, а послідовності “копій” з тим же розподілом. Часто застосування теореми Скорохода дозволяє зробити техніку асимптотичного аналізу прозорішою і зрозумілішою.

Доведення тверджень (i) спирається на підсилений закон великих чисел, (iii) — на центральну граничну теорему (ЦГТ). Для отримання тверджень групи (ii) ми використаємо варіант класичної нерівності Вапника-Червоненкіса [7], поширений на випадок сумішей зі змінними концентраціями. Нерівність Вапника-Червоненкіса дозволяє отримувати як оцінки швидкості збіжності майже напевно, так і оцінки ймовірностей відхилення емпіричних мір від граничних значень. Ці оцінки уточнюють підсилений закон великих чисел, але не досягають точності ЦГТ — в той час як ЦГТ забезпечує швидкість збіжності порядку  $\sqrt{N}$ , нерівність Вапника-Червоненкіса — тільки  $\sqrt{N/\log(N)}$ . Важливою перевагою цих оцінок є те, що вони рівномірні не тільки по певному класу множин  $A$ , але і по обсягу вибірки  $N$  — вони виконуються для всіх  $N$ . Вони також не залежать від розподілів компонент і концентрацій компонент у суміші. Така універсальність робить нерівності Вапника-Червоненкіса досить грубими — для вибірок помірного обсягу вони часто виявляються не набагато кращими ніж тривіальна нерівність  $P(A) < 1$ . Однак рівномірність оцінок буває дуже корисною для того, щоб доводити прямування до 0 залишкових доданків у асимптотичних формулах.

Для набору вагових коефіцієнтів  $a$  ми будемо розглядати два варіанти умов: коли  $a$  є фіксованою функцією і коли  $a = a(v)$  є функцією деякого параметру. Варіант з фіксованими ваговими коефіцієнтами дозволяє помітно спростити умови граничних теорем, а твердження про асимптотику  $\hat{\mu}_N(a(v), A)$  рівномірно по  $v$  використовуються при дослідженні адаптивних оцінок, коли вагові коефіцієнти обираються залежними від вибірки.

Перейдемо до переліку основних результатів (їх доведення можна знайти в [26], п.2.3–2.5).

**(i) Збіжність майже напевно.** Нехай  $\hat{\mu}_N(A, a)$  — емпірична міра,

визначена (2.3). Позначимо

$$\bar{\mu}_N(A, a) := \mathbb{E} \hat{\mu}_N(A, a) = \frac{1}{N} \sum_{j=1}^N a_{j:N} \mathbb{P}\{\xi_{j:N} \in A\} = \sum_{m=1}^M \langle aw^m \rangle_N H_m(A).$$

Ми розглянемо твердження про збіжність  $\hat{\mu}_N(A, a)$  до  $\bar{\mu}_N(A, a)$  рівномірно по  $A$  на деякому класі множин  $\mathfrak{S} \subseteq \mathfrak{A}$ , тобто

$$\sup_{A \in \mathfrak{S}} |\hat{\mu}_N(A, a) - \bar{\mu}_N(A, a)| \rightarrow 0,$$

$N \rightarrow \infty$ . Для однорідних вибірок така збіжність доводиться у теоремі Глівенка-Кантеллі. Сформулюємо аналогічний результат для сумішей зі змінними концентраціями. Для цього нам буде потрібне означення класу множин скінченно-апроксимованого відносно деякої міри.

Нехай  $(\mathcal{Y}, \mathfrak{Y})$  є вимірним простором з мірою  $\nu$ .

**Означення 2.2.1** Клас множин  $\mathfrak{S} \subset \mathfrak{Y}$  зветься скінченно-апроксимованим відносно міри  $\nu$ , якщо для будь-якого  $\varepsilon > 0$  існує такий скінченний клас множин  $\mathfrak{S}(\varepsilon)$ , що для всіх  $A \in \mathfrak{S}$  знайдуться множини  $A^-$  та  $A^+$  з  $\mathfrak{S}(\varepsilon)$  такі, що  $A^- \subseteq A \subseteq A^+$ ,  $\nu(A \setminus A^-) \leq \varepsilon, \nu(A^+ \setminus A) \leq \varepsilon$ .

Клас  $\mathfrak{S}(\varepsilon)$  назовемо  $\varepsilon$ -мережею для  $\mathfrak{S}$ ,  $A^-$  і  $A^+$  — відповідно, нижньою та верхньою апроксимаціями  $A$ .

Наступні леми показують, що скінченно-апроксимовані класи є достатньо великими для багатьох застосувань.

**Лема 2.2.1** ([2], додаток 1.) Нехай  $\mathcal{Y} = \mathbb{R}^d$ ,  $\mathfrak{S}$  є класом всіх прямокутників вигляду  $S = \{x \in \mathbb{R}^d : y_1 \leq x \leq y_2\}$ , де  $y_1, y_2$  довільні вектори у  $\mathbb{R}^d$ . Тоді клас  $\mathfrak{S}$  є скінченно-апроксимованим відносно будь-якої ймовірнісної міри на  $\mathbb{R}^d$ .

**Лема 2.2.2** ([2], додаток 1.) Нехай  $\mathcal{Y} = \mathbb{R}^d$ ,  $\mathfrak{S}$  є класом всіх опуклих множин на  $\mathcal{Y}$ ,  $H$  є мірою на  $\mathcal{Y}$ , абсолютно неперервною відносно міри Лебега. Тоді  $\mathfrak{S}$  є скінченно-апроксимованим класом відносно  $H$ .

Наступна теорема дає достатні умови збіжності емпіричних мір м.н. у випадку фіксованого набору вагових коефіцієнтів  $a$ .

**Теорема 2.2.1** *Нехай*

(i)  $\mathfrak{S}$  є скінченно-апроксимованим класом відносно всіх розподілів  $H_m$ ,  $m = 1, \dots, M$ ;

(ii) вагові коефіцієнти є рівномірно обмеженими:  $|a_{j:N}| \leq C$  для деякого  $C < \infty$  і всіх можливих  $j$  та  $N$ ;

(iii) для всіх  $m = 1, \dots, M$  існують  $\langle a, \mathbb{I}\{a. > 0\}w^m \rangle$  і  $\langle a, \mathbb{I}\{a. < 0\}w^m \rangle$ .  
Тоді

$$\sup_{A \in \mathfrak{S}} |\hat{\mu}_N(a, A) - \bar{\mu}_N(A, a)| \rightarrow 0$$

м.н. при  $N \rightarrow \infty$ .

**Зауваження.** Умова (i) успадкована від класичної теореми Глівенка-Кантеллі для однорідних вибірок. Умова (ii), в принципі, не є непокритуваною. Використовуючи нерівність Вапника-Червоненкіса, можна отримувати умови збіжності емпіричних мір з коефіцієнтами, що прямують до нескінченності при  $N \rightarrow \infty$ . Але ця умова є інтуїтивно зрозумілою, легко перевіряється і виконується у багатьох застосуваннях. Тому ми її використовуємо.

Умова (iii) виглядає дещо дивно. Насправді, для того, щоб можна було визначити  $\bar{\mu}_N(A, a)$ , потрібне існування границь

$$\begin{aligned} \lim_{N \rightarrow \infty} \langle aw^m \rangle_N &= \lim_{N \rightarrow \infty} \left( \frac{1}{N} \sum_{j=1}^N a_{j:N} \mathbb{I}\{a_{j:N} > 0\} + \frac{1}{N} \sum_{j=1}^N a_{j:N} \mathbb{I}\{a_{j:N} < 0\} \right) \\ &= \langle a, \mathbb{I}\{a. > 0\}w^m \rangle + \langle a, \mathbb{I}\{a. < 0\}w^m \rangle \end{aligned}$$

Таким чином, замість природної умови збіжності  $\langle aw^m \rangle_N$  ми вимагаємо, щоб збігався кожен з двох доданків на які ця величина розбивається. Неважко побудувати приклад, в якому б послідовність  $\langle aw^m \rangle_N$  збігалась, а послідовності  $\langle a, \mathbb{I}\{a. > 0\}w^m \rangle_N$  та  $\langle a, \mathbb{I}\{a. < 0\}w^m \rangle_N$  — ні. Однак у практичних застосуваннях такі випадки нам не зустрічались.

Нагадаємо, що у випадку  $\mathcal{X} = \mathbb{R}^d$  зважена емпірична функція розподілу визначається як  $\hat{F}_N(x, a) = \hat{\mu}((-\infty, x], a)$ . Аналогічно, функція розподілу  $m$ -тої компоненти суміші — це  $H_m(x) = H_m((-\infty, x])$ .

**Наслідок 2.2.1** *Нехай  $|a_{j:N}| < C$ . Якщо для всіх  $m = 1, \dots, M$  існують  $\langle a, \mathbb{I}\{a. > 0\}w^m \rangle$  і  $\langle a, \mathbb{I}\{a. < 0\}w^m \rangle$ , то*

$$\sup_{x \in \mathbb{R}^d} |\hat{F}_N(x, a) - \sum_{m=1}^M \langle aw^m \rangle H_m(x)| \rightarrow 0 \text{ (м.н.)}$$

Доведення зводиться до застосування теореми 2.2.1 та леми 2.2.1.

Нехай  $a^k$  є мінімаксним вектором вагових коефіцієнтів, визначеним (2.10). Позначимо  $\hat{H}_k^N(A) = \hat{\mu}(a^k, A)$ ,  $A \in \mathfrak{A}$  — емпірична міра, що є мінімаксною оцінкою  $H_k$ ,  $\hat{H}_k^N(x)$ ,  $x \in \mathbb{R}^d$  — відповідна зважена емпірична функція розподілу.

**Наслідок 2.2.2** *Нехай для всіх  $m, l = 1, \dots, M$  і для будь-якого  $C \in \mathbb{R}$  існують  $\langle w^m \mathbb{I}\{w^m > C\} \rangle$ ,  $\langle w^m w^l \rangle$ , і матриця  $\Gamma = (\langle w^m w^l \rangle)_{m,l=1}^M$  є невідродженою. Тоді*

$$\sup_{x \in \mathbb{R}^d} |\hat{H}_m^N(x) - H_m(x)| \rightarrow 0 (\text{м.н.})$$

для всіх  $m$ .

Якщо, крім того,  $\mathfrak{S}$  є скінченно-апроксимованим класом відносно всіх  $H_m$ ,  $m = 1, \dots, M$ , то

$$\sup_{A \in \mathfrak{S}} |\hat{H}_m^N(A) - H_m(A)| \rightarrow 0 (\text{м.н.}).$$

Тепер розглянемо випадок, коли вагові коефіцієнти  $a_{j:N}$  є функціями деякого параметра  $v \in V$ :  $a_{j:N} = a_{j:N}(v)$  і потрібна рівномірна по  $v \in V$  збіжність емпіричних мір. Ми накладемо на вагові вектори спеціальні умови, які використовують поняття варіації. Варіацію  $N$ -того рядочка масиву  $a$  визначимо як

$$|a|_{\text{VAR}_{:N}} = \sum_{j=2}^N |a_{j:N} - a_{j-1:N}|,$$

а варіацію всього масиву — як  $|a|_{\text{VAR}} = \sup_N |a|_{\text{VAR}_{:N}}$ .

**Теорема 2.2.2** *Нехай*

- (i) виконуються умови (i) та (ii) теореми 2.2.1;
- (ii) для всіх  $t \in T = [0, 1]$  існують границі

$$\lim_{N \rightarrow \infty} \sum_{j=1}^{tN} a_{j:N} w_{j:N}^m \mathbb{I}\{a_{j:N} > 0\},$$

$$\lim_{N \rightarrow \infty} \sum_{j=1}^{tN} a_{j:N} w_{j:N}^m \mathbb{I}\{a_{j:N} < 0\};$$

- (iii)  $\sup_{v \in V} |a(v)|_{\text{VAR}} < \infty$ .

Тоді

$$\sup_{v \in V, A \in \mathfrak{S}} |\hat{\mu}_N(a(v), A) - \sum_{m=1}^M \langle a(v)w^m \rangle_N H_m(A)| \rightarrow 0 \quad (2.16)$$

м.н. при  $N \rightarrow \infty$ .

**Приклад 1. Стохастичні концентрації.** Нехай вектори концентрацій  $(w_{j:N}^1, \dots, w_{j:N}^M)$  генеруються деяким випадковим механізмом, незалежно при різних  $j$ . Після того, як  $w_{j:N}^m$  були згенеровані, ймовірності  $\mathbf{P}\{\xi_{j:N} \in A\}$  визначаються (2.1). Тобто ці ймовірності розглядаються як умовні при фіксованих  $w_{j:N}^m$ ,  $j = 1, \dots, N$ ,  $m = 1, \dots, M$ . Якщо вектори  $(w_{j:N}^1, \dots, w_{j:N}^M)$  незалежні, однаково розподілені при всіх  $j$  та  $N$ , то за законом великих чисел отримуємо

$$\langle w^m w^k \rangle_N \rightarrow \mathbf{E} w_{1:1}^m w_{1:1}^k = \langle w^m w^k \rangle \quad (2.17)$$

м.н. при  $N \rightarrow \infty$  (адже  $|w_{j:N}^m| < 1$ ).

Така модель зветься моделлю зі стохастичними концентраціями. Як правило, розглядаючи такі моделі, фіксують одну реалізацію  $w^m$ , яка насправді спостерігається і розглядають її як дану, не випадкову. Ігноруючи події, що мають нульові ймовірності, вважаємо, що для моделей зі стохастичними концентраціями (2.17) виконується завжди.

Аналогічно переконуємось у існуванні  $\langle w^m \mathbf{1}\{\sum_m w^m > C\} \rangle$ . Таким чином, для забезпечення збіжності  $\hat{H}_k^N$  до  $H_k$  за наслідком 2.2.2, досить досить вимагати, щоб матриця  $— (\mathbf{E} w_{1:1}^m w_{1:1}^k)_{m,k=1}^M$  була не виродженою.

**(ii) Оцінки швидкості збіжності.** Використаємо тепер підхід Вапника-Червоненкіса для отримання деяких оцінок швидкості збіжності неорідних емпіричних мір. Щоб зробити це, нам будуть потрібні додаткові обмеження на класи множин, на яких будуть отримані рівномірні оцінки швидкості збіжності за ймовірністю. Ці обмеження будуть сформульовані у термінах класів Вапника-Червоненкіса.

Щоб з'ясувати це поняття, сформулюємо спочатку деякі допоміжні означення та твердження. Нехай  $(\mathcal{X}, \mathfrak{A})$  — вимірний простір спостережень,  $\mathfrak{S} \subseteq \mathfrak{A}$  — деякий клас вимірних множин,  $X^l = (x^1, \dots, x^l)$  — послідовність елементів  $\mathcal{X}$ . Кожна множина  $A \in \mathfrak{S}$  породжує підпослідовність  $X^A$  послідовності  $X^l$ , яка складається з усіх  $X^l$ , що належать  $A$ . Послідовність  $X^A$  назвемо породженою (множиною  $A$ ) підпослідовністю послідовності  $X^l$ . Позначимо  $\delta^{\mathfrak{S}}(X^l)$  кількість всіх різних послідовностей  $X^A$ , породжених множинами  $A \in \mathfrak{S}$  на  $X^l$  і

$$g^{\mathfrak{S}}(l) = \max_{X^l} \delta^{\mathfrak{S}}(X^l),$$

де максимум береться по всіх можливих послідовностях  $X^l$  з  $\mathcal{X}$ . Функція  $g^{\mathfrak{S}}$  зветься функцією зростання класу  $\mathfrak{S}$ . Наступне твердження доведено у (Вапник, Червоненкіс, 1974) теорема 10.1.

**Твердження 2.2.1** *Для будь-якого класу множин  $\mathfrak{S}$  функція зростання  $g^{\mathfrak{S}}(l)$  або є тотожно рівною  $2^l$  для всіх  $l \in \mathbb{N}$ , або задовольняє нерівність*

$$g^{\mathfrak{S}}(l) \leq \frac{3l^{n-1}}{2(n-1)!}, \quad (2.18)$$

де  $n$  — це перше значення  $l$ , для якого  $g^{\mathfrak{S}} \neq 2^l$ .

Класи множин, для яких функція зростання не є тотожно рівною  $2^l$  для всіх  $l$ , називають класами Вапника-Червоненкіса (VC-класами). З нерівності (2.18) випливає, що функція зростання VC-класу має не більше, ніж степеневий порядок зростання з показником  $n$ .

Найменше  $n$ , для якого існує таке  $C < \infty$ , що для всіх  $l$ :  $g^{\mathfrak{S}}(l) \leq Cl^n$ , зветься ємністю VC-класу  $\mathfrak{S}$ .

Оцінки швидкості збіжності, отримані у цьому параграфі, мають нетривіальний зміст лише для VC-класів. При цьому повинна виконуватись також наступна умова емпіричної вимірності.

Ми будемо казати, що клас  $\mathfrak{S}$  задовольняє умові емпіричної вимірності, якщо для всіх  $l \in \mathbb{N}$  функція

$$\rho(x_1, \dots, x_l, y_1, \dots, y_l) = \sup_{A \in \mathfrak{S}} \left| \frac{1}{N} \sum_{i=1}^l (\mathbb{I}\{x_i \in A\} - \mathbb{I}\{y_i \in A\}) \right|$$

є вимірною функцією відносно  $\mathfrak{A}^{2l}$ .

Відомо багато VC-класів на  $\mathcal{X} = \mathbb{R}^d$ , для яких умова емпіричної вимірності виконується відносно борелевської  $\sigma$ -алгебри. Такими є, наприклад, клас усіх прямокутників у  $\mathbb{R}^d$ , клас усіх куль в  $\mathbb{R}^d$  (функція зростання цих класів  $g^{\mathfrak{S}}(l) \leq (l+1)^d$ ), клас всіх багатогранників у  $\mathbb{R}^d$ , кількість граней яких не перевищує  $C$  (для цього класу  $g^{\mathfrak{S}}(l) \leq (3(l-1)^d/d!)^C$ ) (див. [7]). Сформулюємо тепер аналог теореми Вапника-Червоненкіса для випадку сумішей зі змінними концентраціями.

**Наслідок 2.2.3** *Нехай  $\mathcal{A}$  — клас всіх можливих трикутних масивів  $a_{j:N}$ . Тоді, для всіх  $\lambda > 2M/N$ ,*

$$\mathbb{P} \left\{ \sup_{A \in \mathfrak{S}} \sup_{a \in \mathcal{A}} \frac{|\hat{\mu}_N(A, a) - \bar{\mu}_N(A, a)|}{2 \sup_{1 \leq j \leq N} |a_{j:N}| + |a| \text{VAR}_{:N}} \geq \lambda \right\}$$

$$\leq M \left( 6Ng^{\mathfrak{S}}(2N) \exp \left( -\frac{\lambda^2 N}{32M^2} \right) + 2 \exp \left( -\frac{\lambda^2 N}{8M^2} \right) \right).$$

У випадку, коли потрібна оцінка ймовірності відхилення емпіричної міри від граничного значення для фіксованого вагового масиву, можна скористатись наступною теоремою.

**Теорема 2.2.3** *Нехай  $a_{j:N}$  — довільний масив вагових коефіцієнтів,  $\mathfrak{S}$  — VC-клас. Тоді, для всіх  $\lambda > 2M/N$ ,*

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{A \in \mathfrak{S}} \frac{|\hat{\mu}_N(A, a) - \bar{\mu}_N(A, a)|}{2 \sup_{1 \leq j \leq N} |a_{j:N}| + \max_{1 \leq j \leq N} a_{j:N} - \min_{1 \leq j \leq N} a_{j:N}} \geq \lambda \right\} \\ & \leq M \left( 6Ng^{\mathfrak{S}}(2N) \exp \left( -\frac{\lambda^2 N}{32M^2} \right) + 2 \exp \left( -\frac{\lambda^2 N}{8M^2} \right) \right). \end{aligned}$$

Ці дві теореми дозволяють оцінити швидкість збіжності за ймовірністю. Використовуючи лему Бореля-Кантеллі, з них можна отримати оцінку швидкості збіжності майже напевне. Ця оцінка подібна до рівномірної версії закону повторного логарифму (ЗПЛ) для емпіричних функцій розподілу однорідних вибірок. Однак ми отримуємо швидкість збіжності  $\sqrt{\frac{\ln N}{N}}$ , а не  $\sqrt{\frac{\ln \ln N}{N}}$ , як у ЗПЛ. Це погіршення швидкості збіжності є наслідком того, що ми використовуємо для опису асимптотичної поведінки нашої вибірки схему серій, а не послідовність незалежних, однаково розподілених випадкових елементів.

**Теорема 2.2.4** *Нехай  $\mathfrak{S}$  є VC-класом. Тоді*

*1. Для будь-якого вагового масиву  $a$  існує така випадкова величина  $\Lambda < \infty$  м.н., що*

$$\sup_{A \in \mathfrak{S}} |\hat{\mu}_N(A, a) - \bar{\mu}_N(A, a)| \leq \Lambda \sqrt{\frac{\ln N}{N}} \sup_{1 \leq j \leq N} |a_{j:N}|, \quad (2.19)$$

*для всіх  $N \in \mathbb{N}$ .*

*2. Існує така випадкова величина  $\Lambda < \infty$  м.н., що для всіх  $N$  та всіх  $A \in \mathfrak{S}$ ,*

$$|\hat{\mu}_N(A, a) - \bar{\mu}_N(A, a)| \leq \Lambda \sqrt{\frac{\ln N}{N}} \sup_{1 \leq j \leq N} (|a_{j:N}| + |a| \text{VAR}_{:N}). \quad (2.20)$$

**Зауваження.** З цієї теореми, вочевидь, випливає збіжність майже напевно емпіричних мір у випадку обмежених вагових коефіцієнтів. Однак клас множин  $\mathfrak{S}$ , на якому збіжність є рівномірною, тут вужчий, ніж у теоремі 2.2.1, оскільки не кожен скінченно-апроксимований клас є класом Валника-Червоненкіса.

Наприклад, за лемою 2.2.2, клас  $\mathfrak{S}$  всіх опуклих множин в  $\mathbb{R}^d$  є скінченно-апроксимованим відносно будь-якої абсолютно неперервної на  $\mathbb{R}^d$  ймовірнісної міри. Отже, за теоремою 2.2.1,  $\sup_{A \in \mathfrak{S}} |\hat{\mu}_N(a, A) - \bar{\mu}_N(a, A)| \rightarrow 0$  якщо розподіли компонент є абсолютно неперервними, а вагові коефіцієнти задовольняють умовам (ii)-(iii) цієї теореми. Але, як показано у [7],  $\mathfrak{S}$  не є VC-класом, тому теорема 2.2.4 нічого не дає для оцінки швидкості цієї збіжності.

У випадку  $\mathcal{X} = \mathbb{R}^d$  для зважених емпіричних функцій розподілу отримуємо наступне твердження про оцінку швидкості рівномірної збіжності до оцінюваної компоненти щільності  $H_k$ .

**Наслідок 2.2.4** *Нехай для вагових коефіцієнтів  $a$  виконана умова незміщеності (2.5). Тоді*

1. *Існує така абсолютна константа  $C$ , що для всіх  $\lambda > 2M/N$ ,*

$$\begin{aligned} & \mathbb{P} \left\{ \frac{\sup_{x \in \mathbb{R}^d} |\hat{F}_N(x, a) - H_k(x)|}{2 \sup_{1 \leq j \leq N} |a_{j:N}| + \max_{1 \leq j \leq N} a_{j:N} - \min_{1 \leq j \leq N} a_{j:N}} \geq \lambda \right\} \\ & \leq M \left( C(2N + 2)^{d+1} \exp \left( -\frac{\lambda^2 N}{32M^2} \right) + 2 \exp \left( -\frac{\lambda^2 N}{8M^2} \right) \right). \end{aligned}$$

2. *Для будь-якого вагового масиву  $a$  існує така випадкова величина  $\Lambda < \infty$  м.н., що*

$$\sup_{x \in \mathbb{R}^d} |\hat{F}_N(x, a) - H_k(x)| \leq \Lambda \sqrt{\frac{\ln N}{N}} \sup_{1 \leq j \leq N} |a_{j:N}|, \quad (2.21)$$

для всіх  $N \in \mathbb{N}$ .

Зрозуміло, що з цього наслідку випливає рівномірна консистентність  $\hat{F}_N(x, a)$  за умови, що  $\sqrt{\frac{\ln N}{N}} \sup_{1 \leq j \leq N} |a_{j:N}|$  прямує до 0.

**(iii) Збіжність розподілів.** Ми будемо розглядати слабку збіжність розподілів випадкових полів у просторах функцій з рівномірною метрикою а також еквівалентну їй збіжність за Скороходом, тобто збіжність



м.н. копій розглядуваних випадкових полів до копії граничного процесу. Нагадаємо основні означення.

Нехай  $S$  — будь-який метричний простір,  $\{\zeta_n\}$  — деяка послідовність його випадкових елементів (в.е.). Кажуть, що  $\{\zeta_n\}$  слабо збігається до випадкового елемента  $\zeta$ , якщо для будь-якої неперервної, обмеженої функції  $g : S \rightarrow \mathbb{R}$  виконується

$$\lim_{n \rightarrow \infty} \mathbb{E} g(\zeta_n) = \mathbb{E} g(\zeta).$$

Слабку збіжність будемо позначати  $\zeta_n \Rightarrow \zeta$ .

Будемо казати, що послідовність  $\{\zeta_n\}$  випадкових елементів деякого метричного простору  $S$  збігається за Скороходом (позначення  $\zeta_n \xrightarrow{\text{Sk}} \zeta$ ), якщо існує такий ймовірнісний простір, на якому можна побудувати послідовність в. е.  $\zeta'_n$  та в.е.  $\zeta'$  такі, що

1.  $\zeta'_n, \zeta'$  мають ті ж самі розподіли, що і  $\zeta_n, \zeta$  відповідно.
2.  $\zeta'_n \rightarrow \zeta'$  у  $S$  м.н. при  $n \rightarrow \infty$ .

**Твердження 2.2.2** *теорема Скорохода (див. [37, 62]). Нехай  $S$  є сепарабельним метричним простором,  $\zeta_n, \zeta$  є в.е. з  $S$  такими, що  $\zeta_n \Rightarrow \zeta$ ,  $n \rightarrow \infty$ . Тоді  $\zeta_n \xrightarrow{\text{Sk}} \zeta$  у  $S$ .*

(Оскільки зі збіжності м.н. випливає слабка збіжність, то з  $\zeta_n \xrightarrow{\text{Sk}} \zeta$  завжди випливає  $\zeta_n \Rightarrow \zeta$ .)

У наступних розділах нам будуть потрібні умови асимптотичної нормальності емпіричних функцій розподілу як оцінок для функцій розподілу компонент суміші. Розглянемо емпіричний процес

$$B_N(x) = \sqrt{N}(\hat{F}_N(x, a) - H_m(x)), \quad (2.22)$$

де  $F_N$  зважена емпірична функція розподілу, визначена (2.4), з ваговими коефіцієнтами  $a$ , що задовольняють умову незміщеності (2.5). Ми будемо вивчати збіжність емпіричних процесів у просторі  $D(\mathbb{R}^d)$  функцій на  $\mathbb{R}^d$  без розривів другого роду з рівномірною нормою

$$|z|_\infty = \sup_{x \in \mathbb{R}^d} |z(x)|.$$

Відмітимо, що цей простір не є сепарабельним, тому твердження про збіжність за Скороходом у ньому є більш сильним, ніж твердження про слабку збіжність.

**Теорема 2.2.5** *Нехай*

1. Для деякого  $A < \infty$ :  $\sup_{j,N} |a_{j:N}| < A$ ;
2. Для всіх  $l, m = 1, \dots, M$  існують границі  $\langle w^l w^m(a)^2 \rangle$ ;
3.  $H_m$  є неперервними функціями на  $\mathbb{R}^d$  при всіх  $m = 1, \dots, M$ ;
4. Виконана умова незміщеності (2.5).

Тоді на деякому випадковому просторі можна побудувати процеси  $B'_N(x)$  та  $B(x)$ , такі, що:

1. Процеси  $B'_N(x)$  мають такий самий розподіл, як і  $B_N(x)$ ;
2.  $B(x)$  є гаусовим випадковим процесом з неперервними траєкторіями, нульовим середнім і коваріаційною функцією

$$\begin{aligned} \mathbb{E} B(x)B(y) &= \sum_{m=1}^M \langle w^m(a)^2 \rangle H_m(\min(x, y)) \\ &\quad - \sum_{i,m=1}^M \langle w^m w^i(a)^2 \rangle H_m(x) H_i(y) \end{aligned}$$

3.  $\sup_{x \in \mathbb{R}^d} |B'_N(x) - B(x)| \rightarrow 0$  м.н. при  $N \rightarrow \infty$ .

(З  $\sum_{m=1}^M w_{j:N}^m = 1$  випливає  $\langle w^m(a)^2 \rangle_N = \sum_{i=1}^M \langle w^m w^i(a)^2 \rangle_N$ , тому існування  $\langle w^m w^i(a)^2 \rangle$  є наслідком умови 2 теореми).

Теорема, по суті, стверджує, що  $B_N \xrightarrow{\text{Sk}} B$  у рівномірній нормі. Вона є тривіальним узагальненням наслідку 2.5.1 з [26].

Тепер наведемо теорему про асимптотичну нормальність емпіричних функцій розподілу, у яких вагові коефіцієнти залежать від деякого параметра  $v$ . Для цього нам будуть потрібні наступні позначення.

Будемо вважати, що  $\mathcal{X} = \mathbb{R}^d$ . Позначимо  $b_{j:N}(t) = \mathbb{I}\{j < tN\}$ . Припустимо, що границі  $\langle b(u)b(v)w^m w^l \rangle$  існують для всіх  $u, v \in [0, 1]$ ,  $m, l = 1, \dots, M$ ,  $t \in [0, 1]$  і введемо у розгляд випадковий процес  $U(x, t)$ ,  $x \in \mathbb{R}^d$ , розподіл якого визначений наступними умовами:

1.  $U(x, t)$  — гауссів процес на  $\mathbb{R}^d \times [0, 1]$ ;
2.  $\mathbb{E} U(x, t) = 0$
3. Коваріаційна функція процесу  $U$  має вигляд

$$\begin{aligned} \mathbb{E} U(x, u)U(y, v) &= \sum_{m=1}^M \langle b(u)b(v)w^m \rangle H_m(\min(x, y)) \\ &\quad - \sum_{m_1, m_2=1}^M \langle b(u)b(v)w^{m_1} w^{m_2} \rangle H_{m_1}(x) H_{m_2}(y). \end{aligned} \quad (2.23)$$

(Можна показати, що для будь-яких функцій розподілу  $H_m$  та концентрацій  $w^m$  ця функція є коваріаційною функцією деякого гауссового процесу. Якщо  $H_m$  — неперервні функції, то  $U(\cdot, t)$  має м.н. неперервну реалізацію на  $\mathbb{R}^d$ ).

Будемо розглядати вагові масиви спеціального вигляду, а саме, вважаємо, що ваговий масив  $a_{j:N}$  можна зобразити у формі

$$a_{j:N}(v) = \tilde{a}\left(\frac{j}{N}, v\right), \quad (2.24)$$

де  $\tilde{a}_N(t, v)$  є деякими неперервними функціями  $\tilde{a}_N : [0, 1] \times V \rightarrow \mathbb{R}$ ,  $v$  — деякий параметр, що належить множині  $V$ .

**Теорема 2.2.6** *Нехай масив  $a_{j:N}(v)$  визначається (2.24) і виконуються наступні умови*

(i) *існує функція  $a : [0, 1] \times V \rightarrow \mathbb{R}$  така, що  $\sup_v \text{VAR}_t a(t, v) < \infty$  і*

$$\sup_v \text{VAR}_t |\tilde{a}_N(t, v) - a(t, v)| \rightarrow 0$$

*при  $N \rightarrow \infty$ ;*

(ii) *Для всіх  $u, v \in [0, 1]$ ,  $m, l = 1, \dots, M$ , існують  $\langle b(u)b(v)w^m w^l \rangle$ ;*

(iii) *функції розподілу  $H_m$   $m = 1, \dots, M$  є неперервними на  $\mathbb{R}^d$ .*

*Тоді випадкове поле*

$$Y_N(x, v) = \sqrt{N}(\hat{F}_N(x, a(v)) - \mathbb{E} \hat{F}_N(x, a(v)))$$

*збігається за Скороходом у рівномірній метриці до*

$$Y(x, v) = \int_0^1 a(t, v) U(x, dt).$$

Зрозуміло, що граничне випадкове поле  $Y(x, v)$  також буде гауссовим з нульовим середнім, а його коваріаційна функція буде мати вигляд

$$\begin{aligned} \mathbb{E} Y(x_1, v_1) Y(x_2, v_2) &= \sum_{m=1}^M \langle a(u) a(v) w^m \rangle H_m(\min(x, y)) \\ &- \sum_{m_1, m_2=1}^M \langle a(u) a(v) w^{m_1} w^{m_2} \rangle H_{m_1}(x) H_{m_2}(y). \end{aligned}$$

**Приклад 2. Функціональні концентрації.** При описі даних прикладних досліджень поруч зі стохастичною моделлю концентрацій, описаною у прикладі 1, зустрічаються моделі іншого типу, які ми назвемо моделями функціональних концентрацій. У цих моделях вважається, що  $w_{j:N}^m = \tilde{w}^m(t_{j:N})$ , де  $\tilde{w}^m : [0, 1] \rightarrow [0, 1]$  деякі функції,  $t_{.:N}$  - послідовні розбиття  $t_{1:N} < t_{2:N} < \dots < t_{N:N}$  інтервалу  $[0, 1]$ . В принципі, ці розбиття можуть бути досить різноманітними, але ми, для простоти викладу, обмежимося рівномірним розбиттям  $t_{j:N} = j/N$ .

Якщо функції  $\tilde{w}^m$  є інтегровними за Ріманом, то

$$\langle b(u)b(v)w^mw^l \rangle = \int_0^{\min(u,v)} \tilde{w}^m(t)\tilde{w}^l(t)dt$$

і аналогічно визначаються всі інші характеристики, пов'язані з застосуванням оператора  $\langle \cdot \rangle$  до концентрацій та вагових коефіцієнтів.

## 2.3 виправлені зважені емпіричні функції розподілу

Нехай спостереження являють собою випадкові величини  $\xi_{j:N}$  (одновимірні, тобто  $\mathcal{X} = \mathbb{R}$ ) що описуються моделлю суміші зі змінними концентраціями (2.1).

У п. 2.1 на роль оцінки для розподілу  $H_k$   $k$ -тої компоненти суміші зі змінними концентраціями запропоновано використовувати зважені емпіричні функції розподілу

$$\hat{F}_N(x, a) = \frac{1}{N} \sum_{j=1}^N a_{j:N} \mathbb{I}\{\xi_{j:N} < x\} \quad (2.25)$$

де  $a = (a_{1:N}, \dots, a_{N:N})$  — деякий не випадковий вектор вагових коефіцієнтів. Якщо покласти  $a = a^k$ , де  $a^k$  визначено (2.10), то ця оцінка буде незміщеною і мінімаксною. У п. 2.2 показано, що за досить широких умов вона буде рівномірно консистентною та асимптотично нормальною. Однак, якщо серед  $a_{j:N}$  є від'ємні, функція  $\hat{F}_N(x, a)$  не є монотонно неспадною і, отже, не може бути функцією розподілу ймовірнісної міри. У деяких застосуваннях на це можна не звертати уваги, але, при використанні, наприклад, бутстреп-техніки, оцінки розподілу даних обов'язково повинні бути

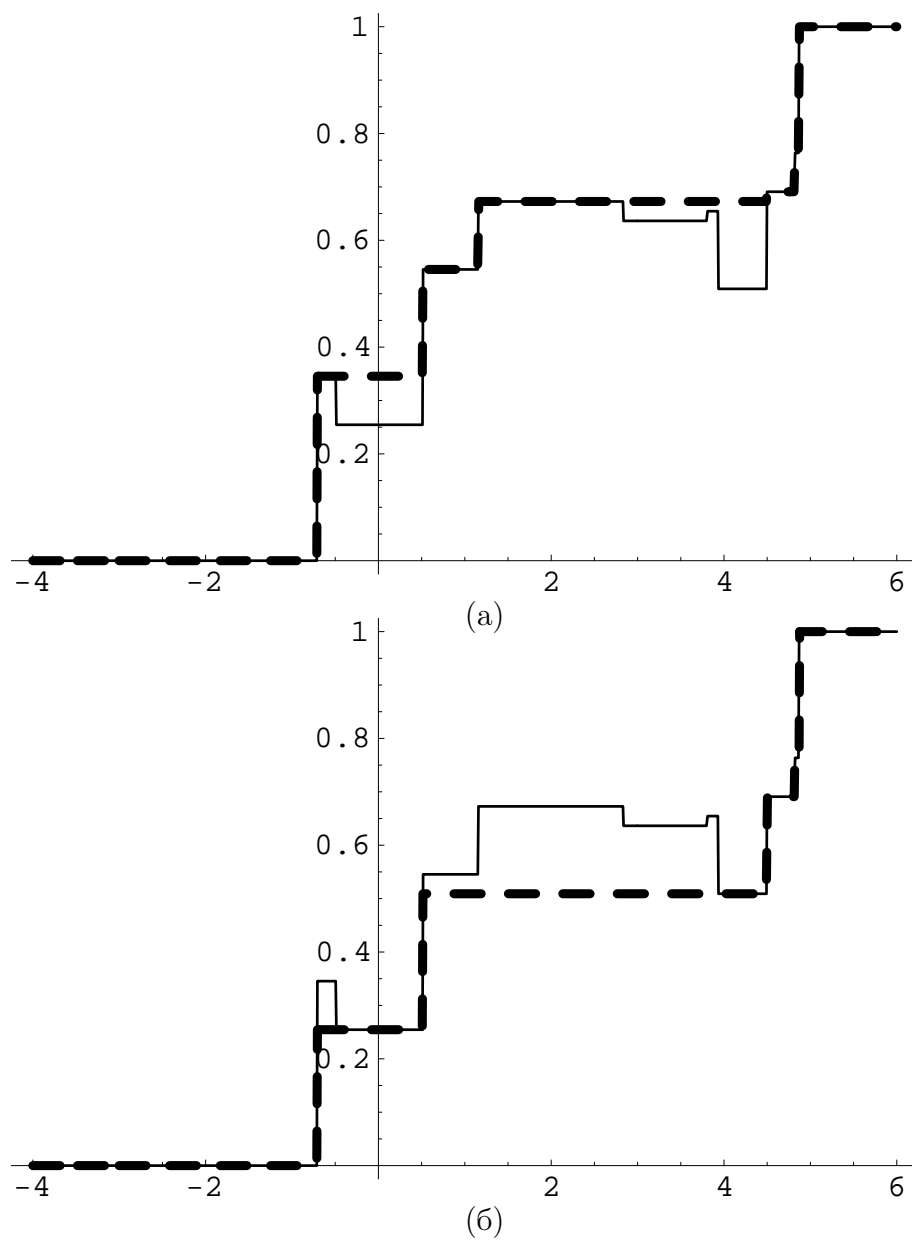


Рисунок 2.1: З.е.ф.р.  $\hat{F}_N(x, a)$  (суцільна лінія) та її виправлення (пунктир):  
 (а)  $\hat{F}_N^+(x, a)$ ; (б)  $\hat{F}_N^-(x, a)$ .

ймовірнісними розподілами, інакше згенерувати бутстреп-вибірку неможливо.

У таких випадках з.е.ф.р.  $\hat{F}_N(x, a)$  можна виправити, поклавши

$$\hat{F}_N^+(x, a) = \sup_{y < x} \hat{F}_N(y, a) \quad (2.26)$$

Функція  $\hat{F}_N^+(x, a)$  приймає лише додатні значення і є монотонно неспадною, але вона може приймати значення більші 1 (див. рис. 2.1). Тому остаточно виправлена зважена емпірична функція розподілу має вигляд

$$\tilde{F}_N^+(x, a) = \min(1, \hat{F}_N^+(x, a)). \quad (2.27)$$

Можна запропонувати і інші алгоритми виправлення, що діють за подібною логікою.

У цьому параграфі ми опишемо різні варіанти виправлення з.е.ф.р., наведемо ефективний алгоритм їх підрахунку та дослідимо їх асимптотичну поведінку. Буде показано, що за певних умов вони є асимптотично нормальними з таким самим граничним розподілом, як і у з.е.ф.р., визначених (2.25). Тобто, асимптотична поведінка емпіричного процесу

$$B_N^+(x) = \sqrt{N}(\tilde{F}_N^+(x, a) - H_k(x)) \quad (2.28)$$

у рівномірній нормі не відрізняється від поведінки емпіричного процесу

$$B_N(x) = \sqrt{N}(\hat{F}_N(x, a) - H_k(x)). \quad (2.29)$$

**Алгоритм обчислення виправленої з.е.ф.р.** Припустимо спочатку, що всі значення у вибірці  $\Xi_N = (\xi_{1:N}, \dots, \xi_{N:N})$  є різними. Позначимо  $\sigma$  перестановку чисел  $1, 2, \dots, N$ , яка забезпечує впорядкування вибірки у порядку зростання:  $\xi_{\sigma(1):N} < \xi_{\sigma(2):N} < \dots < \xi_{\sigma(N):N}$ . (Числа  $\sigma(j)$ ,  $j = 1, \dots, N$  прийнято називати “антирангами”, оскільки  $\sigma^{-1}(j)$  це ранг  $j$ -того спостереження у вибірці). Оскільки функція  $\hat{F}_N(x, a)$  є сталою на інтервалах  $(\xi_{\sigma(j):N}, \xi_{\sigma(j+1):N})$ , то такою ж є і  $\hat{F}_N^+(x, a)$ , визначена (2.26). Отже,

$$\hat{F}_N^+(x, a) = \frac{1}{N} \sum_{j=1}^N b_j^+ \mathbb{I}\{\xi_{j:N} < x\} = \frac{1}{N} \sum_{j=1}^N b_{\sigma(j):N}^+ \mathbb{I}\{\xi_{\sigma(j):N} < x\}$$

де  $b_j^+$  — це деякі коефіцієнти, що (на відміну від  $a_{j:N}$ ) залежать від вибірки  $\Xi_N$ .

Ідея алгоритму полягає в тому, щоб, рухаючись по варіаційному ряду зліва направо, послідовно виправляти коефіцієнти  $a_{\sigma(j):N}$ , які відповідають за те, що сума

$$\hat{S}_j^N = N\hat{F}_N(\xi_{\sigma(j):N}, a) = \sum_{i: \xi_{i:N} \leq \xi_{\sigma(j):N}} a_{i:N}$$

“спускається нижче” своїх попередніх значень.

Алгоритм має наступний вигляд:

1. Обчислити антиранги  $\sigma(j)$ ,  $j = 1, \dots, N$  вибірки  $\Xi_N$ .
2. Покласти  $b_{\sigma(1)}^+ = \max(a_{\sigma(1):N}, 0)$ ,  $\hat{S}_1 = a_{\sigma(1):N}$ ,  $S_1^+ = b_{\sigma(1)}^+$ .
3. Для  $j$  від 2 до  $N$  виконати:
 
$$\begin{aligned} \hat{S}_j &= \hat{S}_{j-1} + a_{\sigma(j):N}; \\ b_{\sigma(j)}^+ &= \max(\hat{S}_j - S_{j-1}^+, 0); \\ S_j^+ &= S_{j-1}^+ + b_{\sigma(j)}^+. \end{aligned}$$

Якщо потрібно обчислити коефіцієнти  $b_j^*$  для функції  $\tilde{F}_N^+$ , визначеної (2.27), то у п.3 алгоритму потрібно ввести додаткову перевірку: поки  $\hat{S}_j < N$ ,  $b_{\sigma(j)}^* = b_{\sigma(j)}^+$ , а як тільки при деякому  $j_0$  виконано  $\hat{S}_{j_0} \leq N$ , то  $b_{\sigma(j_0)}^* = N - \hat{S}_{j_0}$  і  $b_{\sigma(j)}^* = 0$  для всіх  $j > j_0$ .

Помітимо, що знаходження антирангів — це процедура, аналогічна сортуванню вибірки. Швидкі алгоритми сортування вимагають порядку  $CN \ln N$  операцій. Виконання п. 2-3 вимагає порядку  $CN$  операцій. Отже, загальна кількість операцій, потрібних для розрахунку коефіцієнтів  $b^+$  та  $b^*$  має порядок  $CN \ln N$ . Такі алгоритми прийнято вважати швидкими.

Якщо у вибірці наявні декілька рівних між собою значень, скажімо,  $\xi_{j_1:N} = \xi_{j_2:N} = \dots = \xi_{j_l:N}$ , то доцільно замінити їх одним значенням  $\xi_{j_1:N}$ , якому відповідає ваговий коефіцієнт  $a_{j_1:N}^* = a_{j_1:N} + \dots + a_{j_l:N}$ . Коефіцієнти виправленої з.е.ф.р. можна після цього розраховувати за наведеним вище алгоритмом.

Коефіцієнти  $b^+$  та  $b^*$  залежать від спостережень, але не залежать від значення  $x$ , при якому підраховується з.е.ф.р.  $\hat{F}_N^+$  або  $\tilde{F}_N^+$ . Зрозуміло, що коли виправлену з.е.ф.р.  $\tilde{F}_N^+$  потрібно обчислити при багатьох різних значеннях  $x$ , її коефіцієнти доцільно підрахувати один раз і запам'ятати, звертаючись до них кожного разу, коли виникне потреба обчислення виправленої з.е.ф.р.

#### Інші варіанти виправлення з.е.ф.р.

Крім  $\tilde{F}_N^+$  визначеної (2.27), ми розглянемо кілька інших варіантів виправлення з.е.ф.р. Усі функції, отримані в результаті виправлення, ми бу-

демо вважати неперервними зліва, що відповідає означенню ф.р.  $F_\xi(x) = \mathbb{P}\{\xi < x\}$ . У деяких випадках формули виправлення зручно будувати так, що отримана функція виходить неперервною зправа. Для того, щоб “довиправити” такі функції введемо оператор  $L[f](x) = \lim_{y \uparrow x} f(y)$ , який заміняє значення функції  $f$  у точці стрибка на границю  $f$  у цій точці зліва. Зауважимо, що з практичної точки зору, це “довиправлення” ніякої ролі не грає, оскільки всі розглядувані нами виправлені з.е.ф.р. матимуть вигляд

$$\frac{1}{N} \sum_{j=1}^N b_j \mathbb{I}\{\xi_{j:N} < x\},$$

і для їх використання досить тільки обчислити коефіцієнти  $b_j$ .

Отже, введемо

$$\hat{F}_N^+(x, a) := \sup_{y < x} \hat{F}_N(y, a), \quad (2.30)$$

$$\hat{F}_N^-(x, a) := L[\inf_{y > x} \hat{F}_N(y, a)], \quad (2.31)$$

$$\tilde{F}_N^-(x, a) := \max(\hat{F}_N^-(x, a), 0), \quad (2.32)$$

$$\tilde{F}_N^\pm(x, a) := \frac{1}{2}(\tilde{F}_N^+(x, a) + \tilde{F}_N^-(x, a)), \quad (2.33)$$

$$\hat{F}_N^\pm(x, a) = \begin{cases} \hat{F}_N^+(x, a), & \text{якщо } \hat{F}_N^+(x, a) \leq 1/2; \\ \hat{F}_N^-(x, a), & \text{якщо } \hat{F}_N^-(x, a) \geq 1/2; \\ 1/2, & \text{в інших випадках.} \end{cases} \quad (2.34)$$

Функцію  $\hat{F}_N^+(x, a)$  можна назвати “верхньою огибаючою” для функції  $\hat{F}_N(x, a)$ , оскільки це найменша монотонно неспадна функція, графік якої лежить вище  $\hat{F}_N(x, a)$ . Аналогічно,  $\hat{F}_N^-(x, a)$  — нижня огибаюча  $\hat{F}_N(x, a)$  (див. рис. 2.1 на с. 47). Виправляючи зважену емпіричну функцію розподілу, природно вимагати, щоб отримана оцінка лежала між нижньою та верхньою огибаючими. Зрозуміло, що таких оцінок існує нескінченно багато.

Функції  $\tilde{F}_N^+(x, a)$  та  $\tilde{F}_N^-(x, a)$  зрізають оцінки  $\hat{F}_N^+(x, a)$  та  $\hat{F}_N^-(x, a)$  відповідно на правому та на лівому кінцях області можливих значень спостережень, там, де ці оцінки виходять за межі інтервалу  $[0, 1]$ .

Для підрахунку  $\tilde{F}_N^-(x, a)$  та  $\hat{F}_N^-(x, a)$  можна скористатись тим же алгоритмом, який був запропонований для верхньої огибаючої, але з рухом по варіаційному ряду у протилежному напрямку — від найбільших значень до найменших.



Оцінка  $\hat{F}_N^\pm(x, a)$  отримана склеюванням верхньої та нижньої огинаючої  $\hat{F}_N(x, a)$ : ліва нижня частина графіка  $\hat{F}_N^+(x, a)$  утворюється верхньою огинаючою, тобто  $\hat{F}_N^+(x, a)$ , а верхня права частина — нижньою огинаючою, тобто  $\hat{F}_N^-(x, a)$ . Склеювання проведено “по медіані”. Такий спосіб утворення оцінки дозволяє відмовитись від зрізання на кінцях, оскільки  $0 \leq \hat{F}_N^\pm(x, a) \leq 1$  для всіх  $x \in \mathbb{R}$  і для будь-яких вагових коефіцієнтів  $a$ .

**Асимптотика виправлених з.е.ф.р.** Надалі індексом  $*$  позначатимемо будь-яку комбінацію  $\hat{\cdot}$  або  $\tilde{\cdot}$  з  $+$ ,  $-$  або  $\pm$ , тобто  $F^*(x, a)$  може бути будь-якою з функцій, визначених (2.30-2.34). У цьому параграфі ми будемо розглядати лише з.е.ф.р., які використовуються для оцінки розподілу  $k$ -тої компоненти суміші, тобто  $H_k$ .

Враховуючи монотонність  $H_k$ , легко бачити, що

$$\sup_{x \in \mathbb{R}} |\hat{F}_N^+(x, a) - H_k(x)| \leq \sup_{x \in \mathbb{R}} |\hat{F}_N(x, a) - H_k(x)| \quad (2.35)$$

і

$$\sup_{x \in \mathbb{R}} |\hat{F}_N^-(x, a) - H_k(x)| \leq \sup_{x \in \mathbb{R}} |\hat{F}_N(x, a) - H_k(x)|, \quad (2.36)$$

тому нерівність Валника-Червоненкіса без змін переноситься на виправлені з.е.ф.р. і з рівномірної консистентності  $\hat{F}_N(\cdot, a)$  як оцінки  $H_k$  (наприклад, при виконанні умов наслідку 2.2.4) впливає рівномірна консистентність  $F_N^*(x, a)$ .

Перейдемо до дослідження асимптотичної нормальності.

Емпіричним процесом для  $\hat{F}_N(x, a)$  як і раніше будемо називати

$$B_N(x) = B_N(x, a) := \sqrt{N}(\hat{F}_N(x, a) - H_k(x)),$$

і, відповідно, емпіричним процесом для виправленої з.е.ф.р.  $F_N^*(x, a)$  назовемо

$$B_N^*(x) = B_N(x, a) := \sqrt{N}(F_N^*(x, a) - H_k(x)).$$

Спочатку ми дослідимо асимптотичну поведінку  $\hat{B}_N^+$ .

Помітимо, що процес  $\hat{B}_N^+$ , визначений (2.28), можна задати за допомогою  $B_N(x)$ :

$$\hat{B}_N^+(x) = \sqrt{N}(\sup_{y < x} (H_k(y) + B_N(y)/\sqrt{N}) - H_k(x)). \quad (2.37)$$

Для доведення асимптотичної нормальності  $B_N^+$  скористаємося технікою одного ймовірнісного простору з п. 2.2.

Надалі ми будемо ототожнювати процес  $B_N(x)$  з  $B'_N(x)$ , побудованим у теоремі 2.2.5. (оскільки їх розподіли однакові, а нас по суті цікавить слабка збіжність). Під  $\hat{B}_N^+(x)$  будемо розуміти процес, визначений (2.37).

Будемо називати  $x$  точкою росту функції розподілу  $H_k$ , якщо  $\forall \delta > 0$ ,  $H_k(x) - H_k(x - \delta) > 0$ . Множину всіх точок росту функції  $H_k$  позначимо  $\text{supp } H_k$  і назовемо носієм розподілу  $H_k$ .

**Теорема 2.3.1** *Нехай виконані умови теореми 2.2.5 і для всіх  $m = 1, \dots, M$ ,  $\text{supp } H_m \subseteq \text{supp } H_k$ . Тоді*

$$\sup_{x \in \mathbb{R}} |\hat{B}_N^+(x) - B_N(x)| \rightarrow 0 \quad (2.38)$$

при  $N \rightarrow \infty$  за ймовірністю.

**Доведення теореми** проведемо у два кроки. Спочатку доведемо, що має місце поточкова збіжність за ймовірністю, тобто, для всіх  $x \in \mathbb{R}$ ,

$$\mathbf{P}\{|\hat{B}_N^+(x) - B_N(x)| > \varepsilon\} \rightarrow 0 \text{ при } N \rightarrow \infty. \quad (2.39)$$

На другому кроці, використовуючи (2.39), доведемо (2.38).

Покажемо, що при доведенні досить розглянути випадок, коли носії розподілів усіх компонент є обмеженими. Зробимо перетворення  $\xi_{j:N} \rightarrow \check{\xi}_{j:N} = \frac{2}{\pi} \arctan \xi_{j:N}$ . Нехай  $\check{H}_m$  — ф.р. випадкової величини  $\frac{2}{\pi} \arctan \eta_m$ , де  $\eta_m$  — в.в. з розподілом  $H_m$ . Тоді  $(\check{\xi}_{j:N}, j = 1, \dots, N)$  є вибіркою із суміші зі змінними концентраціями з розподілами компонент  $\check{H}_m$  і концентраціями  $w_{j:N}^m$ . При цьому, якщо  $\check{B}_N$  і  $\hat{\check{B}}_N^+$  — відповідні емпіричні процеси, побудовані по  $(\check{\xi}_{j:N}, j = 1, \dots, N)$ , то  $\check{B}_N(\frac{2}{\pi} \arctan(x)) = \check{B}_N(x)$  і  $\sup_x |\check{B}_N(x) - \hat{\check{B}}_N^+(x)| = \sup_x |B_N(x) - \hat{B}_N^+(x)|$ . Оскільки  $\text{supp } \check{H}_m \subseteq [-1, 1]$ , то це означає, що ми можемо надалі обмежитись лише розглядом таких вибірок, у яких  $\text{supp } H_m \subseteq [-1, 1]$ .

Отже, доведемо (2.39). Оскільки, за побудовою, завжди  $\hat{B}_N^+(x) \geq B_N(x)$ , то нам досить переконатись, що  $\forall \varepsilon > 0$ ,  $\mathbf{P}\{\hat{B}_N^+ \geq B_N(x) + \varepsilon\} \rightarrow 0$  при  $N \rightarrow \infty$ .

Помітимо, що коли для деяких  $\delta > 0$  та  $x \in \mathbb{R}$ ,  $(x - \delta, x) \cap \text{supp } H_k = \emptyset$  (і, за умовою теореми,  $(x - \delta, x) \cap \text{supp } H_m = \emptyset$  для всіх  $m = 1, \dots, M$ ), то  $\mathbf{P}\{\xi_{j:N} \in (x - \delta, x)\} = 0$  і, отже,  $\hat{F}_N(x, a) = \hat{F}_N(x - \delta, a)$ ,  $H_k(x) = H_k(x - \delta)$ ,  $F_N^+(x, a) = F_N^+(x - \delta, a)$ ,  $B_N(x) = B_N(x - \delta)$ ,  $\hat{B}_N^+(x) = \hat{B}_N^+(x - \delta)$ . Позначимо  $s(x) = \sup\{x' \in \text{supp } H_k : x' < x\}$ . Тоді  $s(x) \in \text{supp } H_k$  і  $B_N(x) = B_N(s(x))$ ,

$\hat{B}_N^+(x) = \hat{B}_N^+(s(x))$ . Тому (2.39) досить довести для  $x \in \text{supp } H_k$ , що ми і зробимо.

Нехай  $\delta$  — довільне число, таке, що  $0 < \delta < \varepsilon$ ,  $t_0 \in \mathbb{R}$ ,  $r \in \mathbb{N}$ . Позначимо  $t_j = t_0 + \delta j$ ,  $A_N = \{\hat{B}_N^+(x) \geq B_N(x) + \varepsilon\}$ ,  $A_N^- = \{B_N(x) < t_0\}$ ,  $A_N^+ = \{\hat{B}_N^+ \geq t_r + \varepsilon\}$ ,  $A_N^j = \{B_N(x) \in [t_j, t_j + \delta], \hat{B}_N^+ > t_j + \varepsilon\}$ . Тоді  $A_N \subseteq A_N^+ \cup A_N^- \cup \bigcup_{j=0}^{r-1} A_N^j$ .

Зафіксуємо довільні  $z > 0$ ,  $\varepsilon > 0$ . Покладемо  $\delta = \varepsilon/2$ . Оцінимо ймовірності подій  $A_N^+$ ,  $A_N^-$ ,  $A_N^j$ . Оскільки, для будь-якого фіксованого  $\lambda > 0$ ,  $\mathbf{P}\{|B_N(x) - B(x)| > \lambda\} \rightarrow 0$  при  $N \rightarrow \infty$  і  $\mathbf{P}\{|B(x)| > \lambda\} \rightarrow 0$  при  $\lambda \rightarrow \infty$ , то можна обрати  $t_0$  так, щоб  $p_N^- = \mathbf{P}(A_N^-) < \varepsilon/3$  для всіх досить великих  $N$ . Враховуючи, що, крім того,  $\hat{B}_N^+(x) > B_N(x)$ , можна обрати досить велике  $r$  (і, отже,  $t_r$ ) так, щоб  $p_N^+ = \mathbf{P}(A_N^+) < \varepsilon/3$ . Фіксуємо  $t_0$  та  $r$ .

Оцінимо тепер  $p_N^j = \mathbf{P}(A_N^j) \leq \mathbf{P}\{B_N(x) < t_{j+1}, \hat{B}_N^+(x) > t_{j+1} + \delta\}$  (оскільки  $t_{j+1} = t_j + \delta$ ,  $\varepsilon = 2\delta$ ). Але

$$\begin{aligned} \{\hat{B}_N^+(x) > t_{j+1} + \delta\} &= \{\sup_{y < x} (H_k(y) + B_N(y)/\sqrt{N}) > H_k(x) + (t_{j+1} + \delta)/\sqrt{N}\} \\ &= \{\exists y \leq x : B_N(y) > t_{j+1} + \delta + \sqrt{N}(H_k(x) - H_k(y))\} \end{aligned}$$

і

$$A_N^j \subseteq \{B_N(x) < t_{j+1} \text{ і } \exists y \leq x : B_N(y) > t_{j+1} + \delta + \sqrt{N}(H_k(x) - H_k(y))\}.$$

Фіксуємо деяке  $l > 0$ . Помітимо, що остання подія виконується або тоді, коли процес  $B_N(y)$  виходить за рівень  $\sqrt{N}(H_k(x) - H_k(y)) + t_{j+1} + \delta$  при деякому  $y < x - l$ , або за рівень  $t_{j+1} + \delta$  на інтервалі  $[x - l, x]$ . Тому  $p_N^j \leq \mathbf{P}(C_N) + \mathbf{P}(D_N)$ , де  $C_N = \{\sup_y B_N(y) > t_{j+1} + \delta + \sqrt{N}(H_k(x) - H_k(x - l))\}$ ,  $D_N = \{B_N(x) < t_{j+1}, \exists y \in [x - l, x], B_N(y) > t_{j+1} + \delta\}$ .

Оцінимо  $\mathbf{P}\{D_N\} \leq \mathbf{P}\{\sup_{|y-x|<l} |B_N(x) - B_N(y)| > \delta\}$ . Оскільки  $B(x)$  — процес з неперервними траєкторіями, то можна обрати достатньо мале  $l$  так, щоб

$$\mathbf{P}\left\{\sup_{|x-y|<l} |B(x) - B(y)| > \frac{\delta}{3}\right\} \leq \frac{\varepsilon}{18r}, \quad (2.40)$$

а за теоремою 2.2.5, при достатньо великих  $N$ ,

$$\mathbf{P}\left\{\sup_x |B_N(x) - B(x)| > \frac{\delta}{3}\right\} \leq \frac{\varepsilon}{18r} \quad (2.41)$$

Оскільки  $|B_N(x) - B_N(y)| \leq |B_N(x) - B(x)| + |B(x) - B(y)| + |B(y) - B_N(y)|$ , то з виконання (2.40) - (2.41) маємо

$$\mathbf{P}\{D_N\} \leq \frac{\varepsilon}{6r}. \quad (2.42)$$

Фіксуємо  $l$  і оцінимо  $P(C_N)$ . Оскільки при  $l > 0$ ,  $H_k(x) > H_k(x - l)$ , а  $\sup_y B_N(y) \rightarrow \sup_y B(y) < \infty$  ( $N \rightarrow \infty$ ) за теоремою 2.2.5, то  $P(C_N) \rightarrow 0$  при  $N \rightarrow \infty$ . Тому при великих  $N$ ,

$$P(C_N) \leq \frac{\varepsilon}{6r}. \quad (2.43)$$

Об'єднуючи нерівності (2.42) і (2.43) отримуємо  $p_N^j \leq \frac{\varepsilon}{3r}$ .

Остаточно маємо

$$P(A_N) \leq p_N^+ + p_N^- + \sum_{j=1}^r p_N^j \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \sum_{j=1}^r \frac{\varepsilon}{3r} \leq \varepsilon$$

при достатньо великих  $N$ . Отже (2.39) доведено.

Доведемо (2.38). Нехай  $[a, b]$  — довільний інтервал. Оскільки  $\hat{B}_N^+(x) \geq B_N(x)$ , то  $\inf_{x \in [a, b]} \hat{B}_N^+(x) \geq \inf_{x \in [a, b]} B_N(x)$ . Оцінимо  $\sup_{x \in [a, b]} \hat{B}_N^+(x)$  зверху. Оскільки  $\hat{B}_N^+$  — монотонно спадна функція на інтервалах між стрибками  $\hat{F}_N^+(x, a)$ , то цей супремум може досягатись або у момент стрибка функції  $\hat{F}_N^+(x, a)$ , або на лівому кінці інтервала  $[a, b]$ . У першому випадку він співпадає з  $\sup_{x \in [a, b]} B_N(x)$ . Отже

$$\sup_{x \in [a, b]} B_N^+(x) \leq \max(B_N^+(a), \sup_{x \in [a, b]} B_N(x)). \quad (2.44)$$

Внаслідок того, що траєкторії  $B(x)$  неперервні, для будь-яких  $\lambda, \varepsilon > 0$ , можна обрати таке  $\delta$ , що для  $t_j = -1 + \delta j$  буде виконуватись нерівність  $P\{\sup_j |B(t_j) - B(t_{j-1})| > \varepsilon\} < \lambda$ , а за (2.39), для достатньо великих  $N$ ,

$$P\{\sup_j |B_N^+(t_j) - B_N(t_j)| > \varepsilon\} < \lambda. \quad (2.45)$$

За теоремою 2.2.5,  $\delta$  можна обрати так, що

$$P\{\sup_{|x-y|<\delta} |B_N(x) - B_N(y)| > \varepsilon\} < \lambda$$

і, отже,

$$P\{\max_j (\sup_{y \in [t_j, t_{j+1}]} B_N(x) - \inf_{y \in [t_j, t_{j+1}]} B_N(x)) > \varepsilon\} < \lambda. \quad (2.46)$$

Для будь-якого  $x \in [-1, 1]$  знайдеться  $j$  таке, що  $x \in [t_j, t_{j+1}]$  і

$$\hat{B}_N^+(x) \geq \inf_{y \in [t_j, t_{j+1}]} \hat{B}_N^+(y) \geq \inf_{y \in [t_j, t_{j+1}]} B_N(y),$$

а за (2.44),

$$\begin{aligned} B_N^+(x) &\leq \sup_{y \in [t_j, t_{j+1}]} \hat{B}_N^+(y) \leq \max(\hat{B}_N^+(t_j), \sup_{y \in [t_j, t_{j+1}]} B_N(y)) \\ &\leq \max(B_N(t_j) + \varepsilon, \sup_{y \in [t_j, t_{j+1}]} B_N(y)) \leq \sup_{y \in [t_j, t_{j+1}]} B_N(y) + \varepsilon, \end{aligned}$$

якщо виконана подія, що стоїть під знаком ймовірності у (2.45). Тому, враховуючи (2.45), маємо

$$\mathbb{P}\{\forall j, \forall x \in [t_j, t_{j+1}] \inf_{y \in [t_j, t_{j+1}]} B_N(y) \leq \hat{B}_N^+(x) \leq \sup_{y \in [t_j, t_{j+1}]} B_N(y) + \varepsilon\} < \lambda.$$

Звідси, враховуючи (2.46), отримуємо

$$\mathbb{P}\{\sup_x |\hat{B}_N^+(x) - B_N(x)| > 2\varepsilon\} < 2\lambda.$$

Внаслідок довільності  $\varepsilon$  і  $\lambda$  отримуємо твердження теореми.

Тепер переконаємось, що твердження теореми 2.3.1 вірне і для інших методів виправлення зважених емпіричних функцій розподілу.

**Теорема 2.3.2** *Нехай*

1. Для деякого  $A < \infty \sup_{j,N} |a_{j:N}| < A$ ;
2. Для всіх  $l, m = 1, \dots, M$  існують границі  $\langle w^l w^m(a)^2 \rangle$ ;
3.  $H_m$  є неперервними функціями на  $\mathbb{R}$  при всіх  $m = 1, \dots, M$ ;
4. Виконана умова незміщеності (2.5):

$$\langle a w^m \rangle_N = \mathbb{I}\{m = k\} \text{ для всіх } m = 1, \dots, M;$$

5. Для всіх  $m = 1, \dots, M$ ,  $\text{supp } H_m \subseteq \text{supp } H_k$ .

Тоді

$$\sup_{x \in \mathbb{R}} |\hat{B}_N^*(x) - B_N(x)| \rightarrow 0 \quad (2.47)$$

при  $N \rightarrow \infty$  за ймовірністю.

**Зауваження.** Якщо рівномірної збіжності вимагати лише на деякому інтервалі  $I = [x_0, x_1]$ , тобто (2.47) замінити на

$$\sup_{x \in I} |\hat{B}_N^*(x) - B_N(x)| \rightarrow 0, \quad (2.48)$$

то умову 5 теореми можна замінити на

5'. Для всіх  $m = 1, \dots, M$ ,  $(\text{supp } H_m \cap I) \subseteq (\text{supp } H_k \cap I)$ .

Нагадаємо, що за теоремою 2.2.5, емпіричний процес  $B_N$  слабо збігається до гауссового процесу  $B$  з нульовим середнім і коваріаційною функцією

$$\begin{aligned} \mathbb{E} B(x)B(y) &= \sum_{m=1}^M \langle w^m(a)^2 \rangle H_m(\min(x, y)) \\ &\quad - \sum_{i,m=1}^M \langle w^m w^i(a)^2 \rangle H_m(x) H_i(y) \end{aligned}$$

**Наслідок 2.3.1** В умовах теореми 2.3.2 емпіричні процеси  $B_N^*$  слабо збігаються до  $B$  у просторі  $D(\mathbb{R})$  з рівномірною нормою.

**Доведення теореми 2.3.2.** Для  $\hat{B}_N^+$  твердження теореми доведено у теоремі 2.3.1. Оскільки

$$\hat{F}_N^-(x, a) = L[1 - \hat{F}_{-\Xi_N}^+(-x)],$$

(тут через  $\hat{F}_{-\Xi_N}^+(x)$  позначено виправлену з.е.ф.р. (2.26), обчислену за вибіркою  $-\Xi_N = (-\xi_{1:N}, \dots, -\xi_{N:N})$ ) то справедливості твердження теореми для  $\hat{B}_N^-$  випливає з його справедливості для  $\hat{B}_N^+$ . Далі маємо

$$\hat{F}_N^-(x, a) \leq \tilde{F}_N^-(x, a) \leq \tilde{F}_N^\pm(x, a) \leq \tilde{F}_N^+(x, a) \leq \hat{F}_N^+(x, a)$$

і аналогічні нерівності є вірними для відповідних  $B_N^*$ . Звідси робимо висновок, що твердження теореми є вірним для  $\tilde{B}_N^-$ ,  $\tilde{B}_N^+$ ,  $\tilde{B}_N^\pm$ . Вірність твердження теореми для  $\hat{B}_N^\pm$  також випливає з того, що

$$\hat{F}_N^-(x, a) \leq \hat{F}_N^\pm(x, a) \leq \hat{F}_N^+(x, a).$$

## 2.4 Асимптотично ефективна оцінка розподілу

Емпіричні міри  $\hat{\mu}(a^k, \cdot)$  з мінімаксними ваговими коефіцієнтами (2.10) є оцінками розподілів компонент  $H_k(\cdot)$ , найкращими у найгіршому випадку: як показано у теоремі 2.1.1, існує такий (найгірший) набір розподілів компонент, для якого жодна незміщена оцінка не може мати середньоквадратичний ризик, менший ніж  $\hat{\mu}(a^k, \cdot)$ , а для всіх інших розподілів ризик  $\hat{\mu}(a^k, \cdot)$  кращий, ніж для найгіршого.

В той же час ця теорема не заперечує існування оцінок, які могли б бути кращими ніж  $\hat{\mu}(a^k, \cdot)$  на розподілах, відмінних від найгіршого. Хотілося б знайти таку оцінку, яка оцінювала б будь-який розподіл краще, ніж будь-яка інша оцінка. Зрозуміло, що в такій формі це бажання нездійсненне — для будь-якого розподілу  $H_k^0$  найкращою можливою буде оцінка  $\tilde{H}_k$ , яка незалежно від даних тотожно дорівнює  $H_k^0$ . Однак для всіх розподілів  $H_k \neq H_k^0$  ця оцінка буде чи не найгіршою.

Тому у класичній теорії оцінювання по вибірках великого обсягу (див. короткий огляд у п. 7.5, а більш докладно — наприклад, у книзі [10]), найкращими вважаються оцінки, які можна назвати асимптотично локально мінімаксними. У [10] саме такі оцінки називають асимптотично ефективними (АЕ). Для будь-якого значення невідомого параметра<sup>2</sup>, якщо розглядати задачу оцінювання у як завгодно малому відкритому околі цього значення, мінімаксний по цьому околу ризик АЕ оцінки стає меншим ніж мінімаксний ризик будь-якої іншої оцінки.

Нажаль, емпіричні міри  $\hat{\mu}(a^k, \cdot)$  не є асимптотично ефективними оцінками. У цьому параграфі ми розглянемо побудову АЕ оцінок для випадку, коли простір спостережень  $\mathcal{X}$  є скінченним. Без обмеження загальності можна вважати, що  $\mathcal{X} = \{1, \dots, L\}$ , де  $L$  — фіксоване число. Розподіли  $H_k$  у цьому випадку визначаються набором ймовірностей  $H_m(\{l\}) = H_{(l,m)}$ ,  $l = 1, \dots, L$ ,  $m = 1, \dots, M$ . Вибірка з суміші зі змінними концентраціями складається з незалежних спостережень  $\xi_j$  з розподілом

$$P\{\xi_j = l\} = \sum_{m=1}^M w_j^m H_{(l,m)}. \quad (2.49)$$

Тут  $w_j^m$  відомі (як звичайно у цій книзі), а  $H_{(l,m)}$  потрібно оцінити за вибіркою  $\Xi_N = (\xi_1, \dots, \xi_N)$ . (Щоб зменшити і без того велику кількість індексів, ми у цьому параграфі дещо відступимо від загальної схеми позначень, прийнятої у книзі, зокрема, не будемо писати у нижньому індексі спостережень, концентрацій і вагових коефіцієнтів  $_N$ , вважаючи, що всюди йдеться про вибірку обсягу  $N$ .)

У такій постановці задача стає параметричною, а при мінімальних обмеженнях — регулярною. Це дозволяє використовувати асимптотичну теорію параметричного оцінювання. Зокрема, виявляється, що оцінки найбільшої вірогідності (ОНВ) для  $H_{(l,m)}$  будуть асимптотично ефективними. (Відмітимо, що у загальній непараметричній постановці задачі оцінювання

<sup>2</sup>у нас параметром є набір розподілів компонент

розподілів компонент оцінки емпіричного методу найбільшої вірогідності не є навіть консистентними.) Однак у цій задачі ОНВ не записуються у явному вигляді, а знаходження їх чисельними методами може бути пов'язане з труднощами навіть при не дуже великих значеннях  $L$ .

Ми розглянемо інший підхід до оцінювання, який також забезпечує побудову асимптотично ефективних оцінок. Цей підхід спирається на загальну техніку адаптивного оцінювання: спочатку невідомий параметр оцінюється грубо, за допомогою пілотної оцінки, а потім пілотна оцінка використовується для більш акуратного налаштування характеристик оцінюючого алгоритму на локальні особливості задачі.

Спочатку введемо у розгляд клас оцінок, які ми будемо використовувати для побудови адаптивного алгоритму.

Ми побудуємо асимптотично ефективну оцінку набору невідомих параметрів  $(H_{(l,m)})_{l=1,\dots,L;m=1,\dots,M}$ . Оскільки  $\sum_{l=1}^L H_{(l,m)} = 1$ , досить оцінити  $H_{(l,m)}$  при  $l = 1, \dots, L-1$ . Надалі ми будемо розглядати пару  $(l, m)$  як один блочний індекс, еквівалентний індексу  $j(l, m) = (L-1)(m-1) + l$ . Набір  $\vec{H} = (H_{(l,m)})_{l=1,\dots,L-1;m=1,\dots,M}$  у цьому випадку являє собою вектор вимірності  $(L-1)M$ . Розглянемо спочатку статистики вигляду

$$\hat{H}(a) = \frac{1}{N} \sum_{j=1}^N \sum_{l_1=1}^{L-1} a_j^{l_1} \chi_j^{l_1}, \quad (2.50)$$

де  $\chi_j^l = \mathbb{I}\{\xi_j = l\}$  — індикатор події  $\xi_j = l$ ,  $a_j^{l_1}$  — не випадкові вагові коефіцієнти<sup>3</sup>. Ці статистики назвемо лінійними оцінками.

Зважена емпірична міра з мінімаксними коефіцієнтами є частковим випадком лінійної оцінки. Щоб задати її введемо

$$g_{m_1 m_2} = \frac{1}{N} \sum_{j=1}^N w_j^{m_1} w_j^{m_2} = \langle w^{m_1} w^{m_2} \rangle_N. \quad (2.51)$$

Як і раніше позначатимем

$$\Gamma_N = (g_{m_1 m_2})_{m_1, m_2=1}^M.$$

Нехай  $\det \Gamma_N \neq 0$ . Тоді існує обернена матриця  $\Gamma_N^{-1}$ , елементи якої будемо позначати  $\bar{g}_{m_1 m_2}$ , тобто  $(\bar{g}_{m_1 m_2})_{m_1, m_2=1}^M := \Gamma_N^{-1}$ .

<sup>3</sup>Зверніть увагу на те, що у цьому позначенні верхній індекс вагових коефіцієнтів пов'язаний не з номером оцінюваної компоненти, а зі значенням, яке можуть приймати дані.



Тепер мінімаксну оцінку для  $H_{(l_1, m_1)}$  можна зобразити у вигляді (2.50) з ваговими коефіцієнтами

$$a_j^l(m_1) = \begin{cases} \sum_{m=1}^M \bar{g}_{m_1 m} w_j^m & \text{якщо } l = l_1, \\ 0 & \text{якщо } l \neq l_1. \end{cases}$$

Цю оцінку позначимо  $H_{(l_1, m_1)}^{LS}$  (Індекс LS є скороченням для “least squares” — найменші квадрати. Він вказує на те, що оцінка мінімізує середньоквадратичний ризик.)

Як вже було відмічено,  $H_{(l_1, m_1)}^{LS}$  не є ефективною оцінкою — можливі незміщені оцінки, які для певного значення параметру  $\vec{H}$  мають дисперсію, меншу ніж  $H_{(l_1, m_1)}^{LS}$ . Фіксуємо  $\vec{H}$  і знайдемо вагові коефіцієнти, з якими лінійна оцінка (2.51) буде мати найменшу дисперсію в класі всіх незміщених лінійних оцінок для  $\vec{H}$ .

Позначимо  $P_j(l) = P_j(l, \vec{H}) = \mathbb{P}\{\xi_j = l\} = \sum_{m=1}^M w_j^m H_{(l, m)}$ ,

$$p_j^{l_1 l_2}(\vec{H}) = P_j(l_1, \vec{H}) \mathbb{I}\{l_1 = l_2\} - P_j(l_1, \vec{H}) P_j(l_2, \vec{H}),$$

$$\Pi_j = \Pi_j(\vec{H}) = (p_j^{l_1 l_2}(\vec{H}))_{l_1, l_2=1}^{L-1},$$

$$\bar{p}_j^{l_1 l_2} = \bar{p}_j^{l_1 l_2}(\vec{H}) = \begin{cases} \frac{P_j(L, \vec{H}) + P_j(l_1, \vec{H})}{P_j(L, \vec{H}) P_j(l_1, \vec{H})} & \text{якщо } l_1 = l_2, \\ \frac{1}{P_j(L, \vec{H})} & \text{якщо } l_1 \neq l_2. \end{cases}$$

Відмітимо, що матриця  $\bar{\Pi}_j = (\bar{p}_j^{l_1 l_2})_{l_1, l_2=1}^{L-1}$  є оберненою до  $\Pi_j$  (це перевіряється безпосереднім множенням цих матриць).

Тепер позначимо

$$\gamma_{(l_1, m_1)(l_2, m_2)}^N(\vec{H}) = \frac{1}{N} \sum_{j=1}^N \bar{p}_j^{l_1 l_2} w_j^{m_1} w_j^{m_2},$$

$$\Gamma_N(\vec{H}) = (\gamma_{(l_1, m_1)(l_2, m_2)}^N(\vec{H}))_{l_i=1, \dots, L-1; m_i=1, \dots, M}.$$

Набір  $\Gamma_N(\vec{H})$  можна розглядати як квадратну матрицю вимірності  $((L-1) \cdot M) \times ((L-1) \cdot M)$ . Позначимо  $\Lambda_N(\vec{H}) = (\lambda_{(l_1, m_1)(l_2, m_2)}^N(\vec{H}))$  — матрицю, обернену до  $\Gamma_N(\vec{H})$ , якщо вона існує. (Тобто  $\sum_{l, m} \lambda_{(l_1, m_1)(l, m)}^N \gamma_{(l, m)(l_2, m_2)} = \mathbb{I}\{l_1 = l_2, m_1 = m_2\}$ ). Покладемо

$$a_j^{l_2}(\vec{H}, l_1, m_1) = a_j^{l_2}(l_1, m_1) = \sum_{l=1}^{L-1} \sum_{m=1}^M \bar{p}_j^{l_2 l_1} \lambda_{(l_1, m_1)(l, m)}^N(\vec{H}) w_j^m \quad (2.52)$$

**Теорема 2.4.1** *Нехай*

- (i)  $N > L \cdot M$ ,
- (ii) матриця  $\Gamma_N$ , визначена (2.51), є невиродженою,
- (iii) для всіх  $j = 1, \dots, N$ ,  $l = 1, \dots, L$ ,  $0 < P_j(l) < 1$ .

Тоді матриця  $\Gamma_N(\vec{H})$  невироджена і статистика  $\hat{H}(a;(\vec{H}, l_1, m_1))$  є незміщеною оцінкою  $H_{(l_1, m_1)}$  з дисперсією, мінімальною в класі всіх незміщених оцінок вигляду (2.50). Ця дисперсія дорівнює

$$\sigma_{(l_1, m_1)}^N(\vec{H}) = \mathbb{E}(\hat{H}(a;(\vec{H}, l_1, m_1)) - H_{(l_1, m_1)})^2 = \frac{1}{N} \lambda_{(l_1, m_1)(l_1, m_1)}^N(\vec{H}).$$

**Зауваження.** Оскільки  $\Gamma_N$  є матрицею Грама системи векторів  $w^m = (w_1^m, \dots, w_N^m)$ ,  $m = 1, \dots, M$ , то її невиродженість еквівалентна лінійній незалежності цих векторів.

**Доведення.** Легко бачити, що оцінки вигляду (2.50) є незміщеними для  $H_{(l_1, m_1)}$  тоді і тільки тоді, коли

$$\frac{1}{N} \sum_{j=1}^N a_j^l w_j^{m_2} = \mathbb{I}\{m_1 = m_2, l = l_1\}. \quad (2.53)$$

Позначимо  $J(a; \vec{H}) = J(a; \cdot) = \mathbb{E}(\hat{H}(a; \cdot) - H_{(l_1, m_1)})^2$ . Якщо  $\hat{H}(a; \cdot)$  — незміщена оцінка, то

$$J(a; \vec{H}) = \frac{1}{N^2} \sum_{j=1}^N \vec{a}_j^T \Pi_j \vec{a}_j, \quad (2.54)$$

де  $\vec{a}_j = (a_j^1, \dots, a_j^{L-1})^T$ . Розв'язуючи задачу мінімізації функціоналу  $J$ , заданого (2.54) при обмеженні (2.53), отримуємо (2.52), якщо  $\Gamma_N(\vec{H})$  і всі  $\Pi_j$  — невироджені матриці. Безпосереднім обчисленням отримуємо  $J(a; (l_1, m_1), \vec{H}) = \frac{1}{N} \lambda_{(l_1, m_1)(l_1, m_1)}^N(\vec{H})$ . Невиродженість  $\Pi_j$  впливає з того, що при  $0 < P_j(l) < 1$ , існує  $\bar{\Pi}_j = \Pi_j^{-1}$ . Невиродженість  $\Gamma_N(\vec{H})$  впливає з наступної леми.

**Лема 2.4.1** *Нехай  $Z_j = (z_{(l_1, l_2)}^j)_{l_1, l_2=1}^{L-1}$ ,  $j = 1, \dots, N$  — довільні симетричні, додатньовизначені матриці,  $w_j^m$ ,  $j = 1, \dots, N$ ,  $m = 1, \dots, M$  — довільні числа,  $\Gamma_N$  визначено (2.51),  $e_{\min}(Z)$  — найменше власне число матриці  $Z$ ,*

$$\tilde{g}_{(l_1, m_1)(l_2, m_2)} = \frac{1}{N} \sum_{j=1}^N z_{(l_1, l_2)}^j w_j^{m_1} w_j^{m_2}, \quad \tilde{G} = (\tilde{g}_{(l_1, m_1)(l_2, m_2)})_{(l_1, m_1)(l_2, m_2)}.$$

Тоді

$$\det \tilde{G} \geq \left( \min_{1 \leq j \leq N} e_{\min}(Z_j) \right)^{M(L-1)} \det \Gamma.$$

Твердження леми випливає з теореми 7.2.1, якщо у цій теоремі покласти  $T = \{1, \dots, N\}$ ,  $w_m(j) = w_j^m$ ,  $\mu_{l_1, l_2}(A) = N^{-1} \sum_{j \in A} z_{l_1, l_2}^j$ .

Застосовуючи лему 2.4.1 до матриці  $\Gamma_N(\vec{H})$  з урахуванням додатньо-визначеності  $\bar{P}_j$ , отримуємо  $\det \Gamma_N(\vec{H}) \neq 0$ .

Теорема доведена.

Скористатись теоремою 2.4.1 безпосередньо для побудови оцінки не можна, оскільки оптимальні вагові коефіцієнти у (2.52) залежать від невідомого параметра  $\vec{H}$ . Підставляючи в (2.52) замість  $\vec{H}$  його оцінку  $\vec{H}^{LS} = (H_{(l, m)}^{LS})$ , отримуємо адаптивну оцінку

$$\check{H}_{(l_1, m_1)}^N = \hat{H}(a(\vec{H}^{LS}, l_1, m_1)), \quad \check{H}^N = (\check{H}_{(l, m)}^N).$$

Оскільки у цьому випадку вагові коефіцієнти є випадковими і залежать від спостережень, теорему 2.4.1 до адаптивної оцінки застосувати неможна. Однак, як свідчить наступна теорема, граничний розподіл адаптивної оцінки такий самий, як і у найкращої лінійної оцінки.

**Теорема 2.4.2** *Нехай виконані наступні умови*

(i) *Існують  $\delta_1, \delta_2 > 0$ , такі, що для всіх  $j \in N$  і всіх  $\vec{H}_1$ , таких, що  $|\vec{H}_1 - \vec{H}| < \delta_1$ ,*

$$\delta_2 < P_j(\vec{H}_1) < 1 - \delta_2. \quad (2.55)$$

(ii) *Існує  $\lim_{N \rightarrow \infty} \Gamma_N = \Gamma$ ,  $\det \Gamma \neq 0$ .*

(iii) *Існує  $\delta_3$ , таке, що для всіх  $\vec{H}_1, \vec{H}_2$   $|\vec{H}_1 - \vec{H}_2| < \delta_3$  і всіх  $m_1, m_2 = 1, \dots, M$ ,  $l_1, l_2 = 1, \dots, L-1$ , існують границі*

$$\lim_{N \rightarrow \infty} \tilde{\gamma}_{(l_1, m_1)(l_2, m_2)}^N(\vec{H}_1, \vec{H}_2) = \tilde{\gamma}_{(l_1, m_1)(l_2, m_2)}(\vec{H}_1, \vec{H}_2), \quad (2.56)$$

де

$$\tilde{\gamma}_{(l_1, m_1)(l_2, m_2)}^N(\vec{H}_1, \vec{H}_2) = \frac{1}{N} \sum_{j=1}^N \sum_{i_1, i_2=1}^{L-1} \bar{p}_j^{i_1 l_1}(\vec{H}_1) \bar{p}_j^{i_2 l_2}(\vec{H}_2) w_j^{m_1} w_j^{m_2} p_j^{i_1 i_2}(\vec{H}).$$

Тоді при  $N \rightarrow \infty$  розподіл  $\sqrt{N}(\check{H}^N - \vec{H})$  слабо збігається до нормального з нульовим середнім та коваріаційною матрицею  $\Lambda(\vec{H})$ .

**Зауваження.** Умова (i) виконана, якщо  $\delta < H_{(l,m)} < 1 - \delta$  для всіх  $l, m$ , або якщо  $\delta < w_j^k < 1 - \delta$  для всіх  $j, k$ .

Дещо громіздка умова (iii) виконана для звичайних моделей  $w_j^m$  — для стохастичних і для функціональних концентрацій (див. приклади 1 і 2 у п. 2.2).

**Доведення теореми.** Нехай  $D$  — деякий окіл  $\vec{H}$ , такий, що для будь-яких  $\vec{H}_1, \vec{H}_2 \in D$  виконано (2.55) та (2.56). Позначимо

$$F_{(l_1, m_1)}(\vec{H}_1) = \frac{1}{N} \sum_{j=1}^N \sum_{l=1}^{L-1} \bar{p}_j^{ll_1}(\vec{H}_1) w_j^{m_1} \chi_j^l,$$

$$Y_{(l_1, m_1)}^N(\vec{H}_1) = \sqrt{N}(F_{(l_1, m_1)}(\vec{H}_1) - \mathbb{E} F_{(l_1, m_1)}(\vec{H}_1)).$$

Покажемо, що  $\vec{Y}^N(\cdot)$  слабо збігається у  $C(D)$  до гауссового випадкового поля  $\vec{Y}$  з середнім 0 та  $\mathbb{E} \vec{Y}(\vec{H}_1)(\vec{Y}(\vec{H}_2))^T = \tilde{\Gamma}(\vec{H}_1, \vec{H}_2)$ , де  $\tilde{\Gamma}(\vec{H}_1, \vec{H}_2) = (\tilde{\gamma}_{(l_1, m_1)(l_2, m_2)}(\vec{H}_1, \vec{H}_2))$ . Легко бачити, що  $\mathbb{E} \vec{Y}^N(\vec{H}_1) = 0$  і

$$\mathbb{E} Y_{(l_1, m_1)}^N(\vec{H}_1) Y_{(l_2, m_2)}^N(\vec{H}_2) = \tilde{\gamma}_{(l_1, m_1)(l_2, m_2)}^N(\vec{H}_1, \vec{H}_2).$$

Оскільки в.в.  $\chi_j^l$  незалежні і  $|\chi_j^l| \leq 1$ , а  $\bar{p}_j^{l_1 l_2}(H_1)$  рівномірно обмежені на  $D$ , з цього випливає збіжність скінченновимірних розподілів  $\vec{Y}^N$  до  $\vec{Y}$ .

Доведемо компактність розподілів  $\vec{Y}^N$  у  $C(D)$ . Для цього помітимо, що на  $D$ ,

$$Y_{(l_1, m_1)}(\vec{H}_1) - Y_{(l_2, m_2)}(\vec{H}_2) = \frac{1}{\sqrt{N}} \sum_{j=1}^N \sum_{l=1}^{L-1} z_{jl}(\vec{H}_1, \vec{H}_2) (\chi_j^l - P_j(l, \vec{H})),$$

де  $|z_{jl}(\vec{H}_1, \vec{H}_2)| \leq C|\vec{H}_1 - \vec{H}_2|$ ,  $C$  — деяка константа. Тому

$$\mathbb{E} |Y_{(l_1, m_1)}(\vec{H}_1) - Y_{(l_2, m_2)}(\vec{H}_2)|^{2k} \leq C|\vec{H}_1 - \vec{H}_2|^{2k}$$

для будь-якого цілого  $k \geq 1$ . Тому, внаслідок критерію компактності в  $C(D)$  (теорема 7.4.4), отримуємо компактність розподілів  $\vec{Y}^N$  і, отже, слабоку збіжність у  $C(D)$ .

За законом великих чисел з умови (ii) випливає, що  $H_{(l,m)}^{LS}$  збігається м.н. при  $N \rightarrow \infty$  до  $H_{(l,m)}$ . Отже,  $\vec{Z}^N = \vec{Y}^N(\vec{H}^{LS})$  слабо збігається до  $\vec{Y}(\vec{H})$ . Відмітимо, що згідно з (2.52),  $\sqrt{N}(\vec{H}^N - \vec{H}) = \Lambda_N(\vec{H}^{LS})\vec{Z}_N$ . Оскільки  $\mathbb{E} \vec{Y}(\vec{H})(\vec{Y}(\vec{H}))^T = \Gamma(\vec{H}, \vec{H}) = \Lambda^{-1}(\vec{H})$ , то залишилось довести, що

$\Lambda_N(\vec{H}^{LS}) \rightarrow \Lambda(\vec{H})$  м.н. Помітимо, що для  $\vec{H}_1, \vec{H}_2 \in D$ ,  $|\Gamma_N(\vec{H}_1) - \Gamma_N(\vec{H}_2)| \leq C|\vec{H}_1 - \vec{H}_2|$ . Тому

$$\begin{aligned} |\Gamma_N(\vec{H}^{LS}) - \Gamma(\vec{H})| &\leq |\Gamma_N(\vec{H}^{LS}) - \Gamma_N(\vec{H})| + |\Gamma_N(\vec{H}) - \Gamma(\vec{H})| \\ &\leq C|\vec{H}^{LS} - \vec{H}| + |\Gamma_N(\vec{H}) - \Gamma(\vec{H})| \rightarrow 0 \end{aligned}$$

м.н. за умовою (iii) теореми і внаслідок збіжності  $\vec{H}^{LS}$  до  $\vec{H}$ . Враховуючи умову (ii) та лему 2.4.1, отримуємо  $\det \Gamma_N(\vec{H}) \geq C > 0$ . Тому  $\det \Gamma(\vec{H}) > 0$ . Отже, існує  $\Lambda(\vec{H}) = \Gamma^{-1}(\vec{H})$  і  $\Lambda_N(\vec{H}^{LS}) \rightarrow \Lambda(\vec{H})$  м.н.

Теорема доведена.

Покажемо тепер, що гранична поведінка локально-мінімаксного середньоквадратичного ризику  $\check{H}^N$  є найкращою можливою для всіх можливих оцінок.

**Теорема 2.4.3** *Нехай  $\vec{T}^N$  — довільна оцінка  $\vec{H}$  за  $\Xi_N$ . Якщо*

*(i) для  $\vec{H}^0 = (H_{(l,m)}^0)$  виконано  $0 < H_{(l,m)}^0 < 1$  для всіх  $1 \leq l \leq L$ ,  $1 \leq m \leq M$ ,*

*(ii) виконана умова (ii) теореми 2.4.2,*

*(iii) для будь-якого  $\vec{H}_1$  з деякого околу  $\vec{H}^0$  існує*

$$\lim_{N \rightarrow \infty} \Gamma_N(\vec{H}_1) = \Gamma(\vec{H}_1)$$

*(iv) виконано (2.55),*

*то для будь-якого  $\delta > 0$  и будь-якого вектора  $\vec{h}$ ,*

$$\lim_{N \rightarrow \infty} N \sup_{|\vec{H} - \vec{H}^0| < \delta} \mathbb{E}((\vec{T}^N - \vec{H})^T \vec{h})^2 \geq \vec{h}^T \Lambda(\vec{H}^0) \vec{h},$$

*зокрема,*

$$\lim_{N \rightarrow \infty} N \sup_{|\vec{H} - \vec{H}^0| < \delta} \mathbb{E}(T_{(l,m)}^N - H_{(l,m)})^2 \geq \lambda_{(l,m)(l,m)}(\vec{H}^0).$$

**Доведення.** Скористаємось теоремою Гаєка (теорема 7.5.2). Для цього необхідно перевірити, що стохастичний експеримент по оцінюванню  $\vec{H}$  за  $\Xi_N$  є локально асимптотично нормальним (ЛАН) і його інформаційна матриця Фішера  $I_N$  асимптотично еквівалентна  $N(\Lambda(\vec{H}))^{-1} = N\Gamma(\vec{H})$ .

Обчислимо  $I_N = \sum_{j=1}^N I^j$ , де  $I^j$  — інформаційна матриця одного спостереження  $\xi_j$ . За означенням (див. п. 7.5),

$$I_{(l_1, m_1)(l_2, m_2)}^j = \sum_{x=1}^L \frac{\partial P_j(x, \vec{H})}{\partial H_{(l_1, m_1)}} \frac{\partial P_j(x, \vec{H})}{\partial H_{(l_2, m_2)}} \frac{1}{P_j(x, \vec{H})}$$

$$= w_j^{m_1} w_j^{m_2} \left( \frac{\mathbb{I}\{l_1 = l_2\}}{P_j(\vec{H}, l_1)} + \frac{1}{P_j(\vec{H}, L)} \right) = w_j^{m_1} w_j^{m_2} \bar{p}_j^{l_1 l_2}.$$

Отже,  $I_N = N\Gamma_N(\vec{H})$  і, внаслідок умови (ii), отримуємо  $I_N \sim N\Gamma(\vec{H}) = N(\Lambda(\vec{H}))^{-1}$ . Матриця, обернена до  $\Lambda(\vec{H})$  існує, оскільки  $\det G \neq 0$  і, отже, за лемою 2.4.1,  $\det \Gamma(\vec{H}) \neq 0$ .

Доведення ЛАН стохастичного експерименту зводиться до перевірки умов 1 та 2 у теоремі 7.5.1. Вони виконані, оскільки  $\frac{\partial}{\partial \vec{H}} \ln P_j(\chi_j^l, \vec{H})$  та  $\frac{\partial^2}{\partial \vec{H}^2} P_j(\chi_j^l, \vec{H})$  є рівномірно обмеженими функціями в околі  $\vec{H}^0$ .

Теорема доведена.

## Розділ 3

# Оцінки числових характеристик розподілів компонент

У цьому розділі ми розглянемо методи оцінювання таких ймовірнісних характеристик, як функціональні моменти та квантілі розподілів компонент сумішей зі змінними концентраціями. Ці характеристики можна трактувати як функціонали від розподілів (функцій розподілу) і застосувати для побудови та дослідження оцінок загальні методи, які використовують підстановку емпіричних оцінок функцій розподілу у відповідні функціонали (метод підстановки). Техніка підстановки загальновідома (див, наприклад, [2], [64]). Умови консистентності та асимптотичної нормальності оцінок методу підстановки у моделях сумішей зі змінними концентраціями наведені у п. 4.1. [26]. Однак тут ми віддамо перевагу прямим методам дослідження, орієнтованим на конкретний вигляд оцінюваних функціоналів. Завдяки цьому вдається побудувати більш точну теорію та інколи отримати оцінки, які мають кращі властивості, ніж оцінки методу підстановки.

### 3.1 Лінійні оцінки функціональних моментів

У цьому параграфі ми почнемо розгляд задачі оцінювання функціональних моментів розподілів компонент суміші зі змінними концентраціями. Як і раніше, ми вважаємо, що спостережувані дані  $(\xi_{j:N}, j = 1, \dots, N)$  являють собою вибірку з незалежних спостережень з суміші зі змінними концентраціями, тобто

$$P\{\xi_j \leq x\} = \sum_{k=1}^M w_{j:N}^k H_k(x) \quad (3.1)$$

де  $M$  — кількість компонент у суміші,  $H_k$  — невідомі розподіли компонент,  $w_{j:N}^k$  — відомі концентрації компонент. Задача полягає в тому, щоб оцінити заданий функціональний момент  $k$ -тої компоненти у суміші, тобто

$$\bar{g}^k = \int g(x) H_k(dx) \quad (3.2)$$

де  $g$  — деяка фіксована відома вимірна функція,  $g : \mathcal{X} \rightarrow \mathbb{R}$ . (У цьому параграфі  $\mathcal{X}$  може бути довільним вимірним простором).

Функціональні моменти з різними функціями  $g$  грають велику роль як у теорії ймовірностей та математичній статистиці, так і у статистичному аналізі реальних даних. Такі основні характеристики розподілу випадкової величини, як математичне сподівання, дисперсія, асиметрія та ексцес виражаються через перші чотири поліноміальних моменти  $\int x^k H_m(dx)$ ,  $k = 1, 2, 3, 4$ . На використанні різних функціональних моментів побудовані метод оцінюючих рівнянь та метод моментів для оцінювання невідомих параметрів.

У випадку незалежних, однаково розподілених випадкових елементів  $\zeta_1, \dots, \zeta_N$  з розподілом  $H_k$ , “теоретичному” моменту  $\bar{g}^k$  відповідає емпіричний момент

$$\frac{1}{N} \sum_{j=1}^N g(\zeta_j),$$

котрий можна розглядати як результат підстановки звичайної (не зваженої) емпіричної міри, побудованої по  $\zeta_j$ , у формулу (3.2) замість  $H_k$ . Таким чином, емпіричний момент однорідної вибірки можна розглядати як оцінку методу підстановки для теоретичного моменту.

По аналогії з цією класичною оцінкою, у випадку спостережень з суміші зі змінними концентраціями можна для оцінки  $\bar{g}^k$  підставити у (3.2)



замість  $H_k$  зважену емпіричну міру

$$\hat{\mu}_N(A, a) = \frac{1}{N} \sum_{j=1}^N a_{j:N} \mathbb{I}\{\xi_{j:N} \in A\},$$

вибравши вагові коефіцієнти  $a$  так, щоб ця міра була хорошою оцінкою  $H_k$ . В результаті такої підстановки отримуємо зважений емпіричний момент

$$\hat{g}_N(a) = \int g(x) \hat{\mu}_N(a, dx) = \frac{1}{N} \sum_{j=1}^N a_{j:N} g(\xi_j), \quad (3.3)$$

котрий ми будемо розглядати як оцінку  $\bar{g}^k$ . Оцінки вигляду (3.3) з не випадковими (незалежними від  $\xi_{j:N}$ ) ваговими коефіцієнтами будемо називати лінійними оцінками.

Якщо  $\det \Gamma_N \neq 0$ , на роль  $a$  можна обрати вектор мінімських вагових коефіцієнтів  $a^k$ , визначений (2.10). Однак чи буде цей вибір найкращим? Відповідь на це питання, взагалі кажучи, негативна: важко знайти таку точку зору, з якої мінімські вагові коефіцієнти слід було б вважати найкращими можливими.

Але і занадто поганим такий вибір теж не є. Оцінка

$$\hat{g}_N^{k,0} = \hat{g}_N(a^k) \quad (3.4)$$

за досить широких умов буде консистентною та асимптотично нормальною. В той же час, можна підібрати вагові коефіцієнти  $a$ , при яких коефіцієнт розсіювання (асимптотична дисперсія)  $\hat{g}_N(a)$  буде кращим (меншим) ніж у  $\hat{g}_N^{k,0}$ . Нажаль коефіцієнт розсіювання лінійної оцінки залежить від невідомих характеристик розподілу компонент суміші, зокрема, і від самих оцінюваних функціональних моментів. При різних розподілах компонент різні вагові набори забезпечують мінімізацію коефіцієнта розсіювання.

Тому ми розглянемо техніку адаптивного оцінювання, подібну до використаної у п. 2.4: спочатку невідомі характеристики розподілу оцінюються за допомогою не найкращих лінійних оцінок (наприклад, з мінімськими ваговими коефіцієнтами), потім за допомогою таких “пілотних” оцінок розраховуються оцінки для оптимальних коефіцієнтів і остаточно оцінка використовує ці оцінені коефіцієнти. Отримана адаптивна оцінка буде мати коефіцієнт розсіювання, рівний найменшому можливому коефіцієнту розсіювання лінійної оцінки. (Зауважимо, що сама адаптивна оцінка

у нашому розумінні не є лінійною — її вагові коефіцієнти залежать від випадкових даних).

Слід відмітити, що ця схема має і принципову відмінність від схеми адаптації емпіричних розподілів з п. 2.4. Задача, розглянута у п. 2.4 була параметричною: задавши розподіли всіх компонент, ми повністю визначали розподіл спостережуваних даних. Задання функціональних моментів  $\bar{g}^k$  не визначає однозначно розподіл  $\xi_{j:N}$ . З цієї відмінності випливають два наслідки: (i) для побудови адаптивної оцінки ми змушені на першому кроці оцінювати більше параметрів (функціональних моментів) ніж збираємось оцінити остаточною оцінкою; (ii) отримана адаптивна оцінка буде асимптотично кращою (не гіршою), ніж будь-яка лінійна оцінка, але не виключено, що існують оцінки зовсім іншого типу, кращі ніж адаптивні. Тобто у даній ситуації ми не можемо гарантувати асимптотичну ефективність адаптивної оцінки в класі всіх можливих оцінок, як це було у п. 2.4.

Цей параграф присвячено виконанню першої половини описаної програми: тут ми отримуємо умови консистентності та асимптотичної нормальності лінійних оцінок, знайдемо їх коефіцієнти розсіювання, отримуємо точну нижню межу для цих коефіцієнтів і визначимо вагові вектори, на яких ця межа досягається. Побудова адаптивних оцінок та їх асимптотичні властивості розглянуті у наступному параграфі.

Точніше, ми дещо узагальнимо свою задачу при вивченні асимптотичної нормальності. Замість того, щоб використовувати для нормування відхилення оцінок від справжнього значення оцінюваної величини  $\sqrt{N}$ , ми будемо ділити їх на їх дисперсії і отримувати умови збіжності розподілів таких нормованих відхилень до стандартного нормального. Такий підхід дозволяє вивчати асимптотику оцінок у випадку, коли границі  $\langle w^k w^m \rangle_N$  не існують або гранична матриця  $\det \Gamma$  є виродженою. (Матриці  $\det \Gamma_N$ , звичайно, мають бути невиродженими, інакше консистентне оцінювання у нашій непараметричній моделі буде неможливе). Результати для нормування  $\sqrt{N}$  ми, таким чином, отримуємо як наслідки більш загальних теорем. Для формулювання та доведення результатів ми будемо використовувати матрично-векторні позначення, які дещо видозмінюють позначення, прийняті у інших частинах книги. Набір концентрацій всіх компонент для фіксованого  $N$   $W_N = (w_{j:N}^k, j = 1, \dots, N, k = 1, \dots, N)$  будемо розглядати як матрицю з  $k$  стовпчиків та  $N$  рядків. Вектори-стовпчики цієї матриці будемо позначати  $\vec{w}_N^k$ . Відповідно вагові вектори у (3.3) також є векторами-стовпчиками довжини  $N$ :  $\vec{a}_N = (a_{j:N}, j = 1, \dots, N)^T$  Індекс  $N$

для спрощення позначень інколи писати не будемо.

Статистика  $\hat{g}_N(\vec{a})$ , визначена (3.3), є незміщеною оцінкою  $\bar{g}^k$  (тобто  $E \hat{g}_N(\vec{a}) = \bar{g}^k$  для всіх  $H_m$ ,  $m = 1, \dots, M$ ) тоді і тільки тоді, коли

$$\frac{1}{N} \vec{a}^T W_N = e_k^T, \quad (3.5)$$

де  $e_k = (0, \dots, 0, 1, 0, \dots, 0)^T$   $k$ -тий базисний вектор стандартного базису у  $\mathbb{R}^M$ . Це, по суті, матричний запис звичної для нас умови незміщеності (2.5)

Нехай вектори  $\vec{w}_N^1, \dots, \vec{w}_N^M$  лінійно незалежні. Тоді мінімаксні вагові коефіцієнти у матричній формі записуються у вигляді

$$\vec{a} = \vec{a}^{k,0} = W_N \Gamma_N^{-1} e_k, \quad (3.6)$$

де  $\Gamma_N = \frac{1}{N} W_N^T W_N$  — звична для нас матриця Грама системи векторів  $\{\vec{w}_N^m\}_{m=1}^M$  у скалярному добутку  $\langle \vec{w}_N^k, \vec{w}_N^m \rangle_N = \frac{1}{N} (\vec{w}_N^k)^T \vec{w}_N^m$ . (Ми використовуємо тут у позначенні  $\vec{a}^{k,0}$  додатковий верхній індекс 0, щоб підкреслити, що мова йде саме про мінімаксні вагові коефіцієнти на відміну від інших можливих коефіцієнтів лінійних оцінок для  $\bar{g}^k$ ). Як ми вже домовились, будемо позначати  $\hat{g}_N^{k,0} = \hat{g}_N(\vec{a}^{k,0})$ .

Тепер розглянемо лінійні оцінки  $\hat{g}_N(\vec{a})$  визначені (3.3) з ваговим вектором  $\vec{a}$  для якого виконується умова незміщеності (3.5). Клас всіх таких оцінок позначимо  $\mathcal{L}_0$ . Підрахуємо дисперсію  $\hat{g}_N(\vec{a})$ .

Позначимо  $\gamma = (\gamma_1, \dots, \gamma_{2M}) = (\bar{g}^1, \dots, \bar{g}^M, \bar{g}^2, \dots, \bar{g}^M)$ , де  $\bar{g}^{2m} = \int g^2(x) H_m(dx)$ ,

$$d_{j:N}(\gamma) = \text{Var } g(\xi_{j:N}) = E(g(\xi_{j:N}))^2 - (E g(\xi_{j:N}))^2 = \sum_{m=1}^M \bar{g}^{2m} w_{j:N}^m - \left( \sum_{m=1}^M \bar{g}^m w_{j:N}^m \right)^2$$

Далі ми як правило будемо вважати, що  $(\bar{g}^m)^2 < \bar{g}^{2m} < \infty$  для всіх  $m = 1, \dots, M$ . Тоді

$$\text{Var } \hat{g}_N(\vec{a}) = \frac{1}{N^2} \sum_{j=1}^N (a_{j:N})^2 d_{j:N}(\gamma).$$

Мінімум цього виразу по  $\vec{a}$  при виконанні умови (3.5) дорівнює

$$\sigma^2(k, N) = \frac{1}{N} d(k, N),$$

де

$$d(k, N) = e_k^T \Gamma_N^{-1}(\gamma) e_k, \quad (3.7)$$

і досягається на ваговому векторі

$$\vec{a} = \vec{a}^k(\gamma) = D^{-1}(\gamma) W_N \Gamma_N^{-1}(\gamma) e_k, \quad (3.8)$$

де  $D(\gamma) = \text{diag}(d_1(\gamma), \dots, d_N(\gamma))$ ,  $\Gamma_N(\gamma) = \frac{1}{N} W_N^T D^{-1}(\gamma) W_N$  є матрицею Грама системи векторів  $\{\vec{w}_N^m\}_{m=1}^M$  у скалярному добутку

$$\langle \vec{w}_N^m, \vec{w}_N^k \rangle^D = \frac{1}{N} \sum_{j=1}^N d_{j:N}^{-1}(\gamma) w_{j:N}^m w_{j:N}^k.$$

(Це можна довести, використовуючи метод множників Лагранжа так само, як у п. 2.1).

Отже, ефективний ваговий вектор для оцінки  $\vec{g}^k$  визначається (3.8), якщо справжнє значення вектора  $(\bar{g}^1, \dots, \bar{g}^M, \bar{g}^{2^1}, \dots, \bar{g}^{2^M})$  дорівнює  $\gamma$ .

**Асимптотична поведінка оцінок.** Позначимо через  $\eta_k$  деякі випадкові елементи з розподілами  $H_k$ . Надалі літерою  $C$  позначаються деякі константи, можливо різні.

**Теорема 3.1.1** *Нехай*

1. Для всіх  $m = 1, \dots, M$ ,  $\mathbf{E} |g(\eta_m)| < \infty$ .

2. Для деякого  $C > 0$ ,  $\det \Gamma_N > C$  для всіх  $N$ .

Тоді оцінка  $\hat{g}_N^k$  є консистентною, тобто  $\hat{g}_N^k \rightarrow \bar{g}^k$  за ймовірністю.

**Доведення.** З умови 1 випливає, що  $|\mathbf{E} g(\xi_{j:N})| < \infty$ . Позначимо  $g_{j:N} = g(\xi_{j:N}) - \mathbf{E} g(\xi_{j:N})$ ,  $\zeta_{j:N} = \frac{a_{j:N}^{k,0}}{N} g_{j:N}$ . Застосуємо до  $\hat{g}_N^k - \bar{g}^k = S_N = \sum_{j=1}^N \zeta_{j:N}$  закон великих чисел у схемі серій (теорема 7.3.7). За цією теоремою, якщо

(а)  $M_1 := \sum_{j=1}^N \mathbf{E} |\zeta_{j:N}| \leq C < \infty \forall N$ ,

(б)  $M_2(\tau) := \sum_{j=1}^N \mathbf{E} |\zeta_{j:N}| \mathbb{I}\{|\zeta_{j:N}| > \tau\} \rightarrow 0$  при  $N \rightarrow \infty$  для будь-якого  $\tau > 0$ ,

то  $S_N \rightarrow 0$  за ймовірністю.

Відмітимо, що з умови 2 теореми випливає  $\sup_{N,j} |a_{j:N}^{k,0}| < \infty$ . Отже виконання (а) є наслідком умови 1 теореми.

Щоб перевірити (б), оцінимо

$$\mathbf{E} |\zeta_{j:N}| \mathbb{I}\{|\zeta_{j:N}| > \tau\} \leq \frac{C}{N} \mathbf{E} \tilde{\zeta} \mathbb{I}\{\tilde{\zeta} > \frac{N\tau}{C}\},$$

де  $\tilde{\zeta} = \sum_{m=1}^M |g(\eta_m)|$ .

Отже,  $M_2(\tau) \leq C \mathbb{E} \tilde{\zeta} \mathbb{I}\{\tilde{\zeta} > \frac{N\tau}{C}\} \rightarrow 0$  при  $N \rightarrow \infty$ , оскільки  $\mathbb{E} \tilde{\zeta} < \infty$ , за умовою 1.

**Теорема 3.1.2** *Нехай*

1. Для всіх  $m = 1, \dots, M$ ,  $0 < \text{Var } g(\eta_m) < \infty$ .

2. Для деякого  $C > 0$ ,  $\det \Gamma_N > C$  для всіх  $N$ .

Тоді розподіли випадкових величин

$$Y_N := \frac{\sqrt{N}}{\sqrt{d(k, N)}} (\hat{g}_N(\vec{a}^k(\gamma)) - \bar{g}^k)$$

слабко збігаються при  $N \rightarrow \infty$  до стандартного нормального розподілу.

**Доведення.** Відмітимо, що при виконанні умови 1 існують такі  $0 < C_1, C_2 < \infty$  що  $C_1 \leq d_{j:N} \leq C_2$  для всіх  $j$  та  $N$ . Тому

$$C_1 \langle \vec{w}^m, \vec{w}^k \rangle \leq \langle \vec{w}^m, \vec{w}^k \rangle^D \leq C_2 \langle \vec{w}^m, \vec{w}^k \rangle.$$

Оскільки  $\Gamma_N$  та  $\Gamma_N(\gamma)$  є матрицями Грама однієї і тієї ж системи векторів у скалярних добутках  $\langle \cdot, \cdot \rangle$  та  $\langle \cdot, \cdot \rangle^D$  відповідно, то з теореми 7.2.1 отримуємо, що для деяких  $0 < c'_1, c'_2 < \infty$  (незалежних від  $N$  та  $\gamma$ )

$$c'_1 \det \Gamma_N \leq \det \Gamma_N(\gamma) \leq c'_2 \det \Gamma_N.$$

Отже, з виконання умови 2 випливає  $\det \Gamma_N(\gamma) > c > 0$ . Тому матриця  $\Gamma_N^{-1}(\gamma)$  існує і всі її елементи є рівномірно обмеженими по  $N$ . Отже,  $\sup_{j,N} |a_{j:N}^k(\gamma)| < \infty$ .

Позначимо

$$\zeta_{j:N} = \frac{a_{j:N}(\gamma)}{\sqrt{Nd(k, N)}} (g(\xi_{j:N}) - \bar{g}^k).$$

Тоді  $Y_N = \sum_{j=1}^N \zeta_{j:N}$ . Твердження теореми випливає тепер з центральної граничної теореми для  $\zeta_{j:N}$  з умовою Ліндеберга (теорема 5 у п. 4 розділу 8, [3]). Умова Ліндеберга перевіряється аналогічно умові (b) теореми 3.1.1.

Теорема доведена.

### 3.2 Адаптивні оцінки моментів

У цьому параграфі ми розглянемо адаптивні оцінки моментів, побудовані за схемою, описаною у п. 3.1. Тут зберігаються позначення, введені у п. 3.1.

З точки зору мінімізації асимптотичної дисперсії, найкращою лінійною оцінкою для функціонального моменту  $\bar{g}$ , визначеного (3.2), є оцінка з ваговими коефіцієнтами  $a^k(\gamma)$ , визначеними (3.8), де  $\gamma = (\gamma_1, \dots, \gamma_{2M}) = (\bar{g}^1, \dots, \bar{g}^M, \overline{g^2}^1, \dots, \overline{g^2}^M)$  — вектор, який складається з функціональних моментів з функціями  $g$  та  $g^2$  для всіх компонент суміші. Елементи цього вектора нам невідомі, тому реалізувати “найкращу лінійну оцінку” практично неможливо.

Оцінимо  $\gamma$  за допомогою очевидної (але не оптимальної) оцінки,  $\hat{\gamma}_N = (\hat{g}_N^{1,0}, \dots, \hat{g}_N^{M,0}, \hat{g}_N^{2,0}, \dots, \hat{g}_N^{M,0})$ . Підставивши цю оцінку замість справжнього  $\gamma$  у формулу для оптимальної лінійної оцінки отримуємо адаптивну оцінку

$$\check{g}_N^k = \hat{g}_N(\vec{a}^k(\hat{\gamma}_N)). \quad (3.9)$$

Покажемо, що при виконанні умов теореми 3.1.2 асимптотична поведінка адаптивної оцінки така ж сама, як асимптотична поведінка найкращої лінійної оцінки.

**Теорема 3.2.1** *Нехай*

1. Для всіх  $m = 1, \dots, M$ ,  $0 < \text{Var } g(\eta_m) < \infty$ .
2. Для деякого  $C > 0$ ,  $\det \Gamma_N > C$  для всіх  $N$ .

*Тоді розподіли випадкових величин*

$$Z_N := \frac{\sqrt{N}}{\sqrt{d(k, N)}}(\check{g}_N^k - \bar{g}^k)$$

*слабко збігаються при  $N \rightarrow \infty$  до стандартного нормального розподілу.*

Для доведення теореми нам буде потрібна лема, яка описує граничну поведінку випадкової величини

$$\tau_\varepsilon(N) := \sup_{\alpha: |\alpha - \gamma| < \varepsilon} \sqrt{N} |\hat{g}_N(\vec{a}^k(\alpha)) - \hat{g}_N(\vec{a}^k(\gamma))|.$$

**Лема 3.2.1** *В умовах теореми 3.2.1*

$$\sup_N \mathbf{P}\{\tau_\varepsilon(N) > \delta\} \rightarrow 0$$

*при  $\varepsilon \rightarrow 0$  для всіх  $\delta > 0$ .*

Спочатку ми покажемо, як з цієї леми випливає твердження теореми, а потім доведемо саму лему.

**Доведення теореми 3.2.1.** Розглянемо випадкові величини  $Y_N$ , визначені у теоремі 3.1.2. Для будь-якого  $\delta > 0$

$$J_N := \mathbb{P}\{|Z_N - Y_N| > \delta\} \leq \mathbb{P}\{\tau_\varepsilon(N) > \delta\} + \mathbb{P}\{|\hat{\gamma}_N - \gamma| > \varepsilon\}$$

Зафіксуємо довільне  $\lambda > 0$ . За лемою 3.2.1 можна обрати  $\varepsilon > 0$  так, що  $\mathbb{P}\{\tau_\varepsilon(N) > \delta\} < \lambda$  для всіх  $N$ . За теоремою 3.1.1, в умовах теореми 3.2.1  $\hat{\gamma}_N \rightarrow \gamma$  за ймовірністю при  $N \rightarrow \infty$ . Отже, при достатньо великому  $N$ ,  $\mathbb{P}\{|\hat{\gamma}_N - \gamma| > \varepsilon\} < \lambda$ . Тому, для таких  $N$ ,  $J_N < 2\lambda$ . Внаслідок довільності  $\lambda > 0$  звідси випливає, що  $J_N \rightarrow 0$ , тобто  $|Z_N - Y_N| \rightarrow 0$  за ймовірністю.

За теоремою 3.1.2, розподіл  $Y_N$  слабо збігається до стандартного нормального розподілу. Отже, і розподіл  $Z_N$  також збігається до стандартного нормального.

Теорема доведена.

Для доведення леми 3.2.1 нам буде потрібна іще одна допоміжна лема. Для будь-якого  $\alpha \in \{0, 1\}^d$  і будь-якої гладенької функції  $f : K \rightarrow \mathbb{R}$  позначимо  $D^\alpha f(u)$  частинну похідну функції  $f$  по всіх координатах  $u^j$  для яких  $\alpha_j = 1$ , тобто

$$D^\alpha = \prod_{\alpha_j=1} \frac{\partial}{\partial u^j}.$$

**Лема 3.2.2** Нехай  $f_N(u) = \frac{1}{\sqrt{N}} \sum_{j=1}^N b_j(u) \zeta_j$   $u \in \mathbb{R}^d$ , де

1.  $\zeta_j$  — незалежні випадкові величини з нульовим математичним сподіванням і  $\text{Var } \zeta_j \leq S$  для деякого  $S < \infty$ ,

2.  $b_j(u)$  є  $d + 1$ -кратно диференційовними функціями  $u$  і для деякого  $L < \infty$

$$\sup_{u \in K, j, l, \alpha} |D^\alpha \frac{\partial}{\partial u^l} b_j(u)| < L,$$

де  $K$  деякий паралелепіпед у  $\mathbb{R}^d$ ,  $\sup$  береться по  $j = 1, \dots, N$ ,  $l = 1, \dots, d$ ,  $\alpha \in \{0, 1\}^d$ .

Тоді

$$\mathbb{P}\left\{ \sup_{|u-v| < \varepsilon, u, v \in K} |f_N(u) - f_N(v)| > \delta \right\} \leq \frac{C_K S L^2 \varepsilon^2}{\delta^2}$$

де  $C_K$  — константа, що залежить лише від  $K$ .

**Доведення лема 3.2.2.** Помітимо, що

$$\tau := \sup_{|u-v|<\varepsilon, u,v \in K} |f_N(u) - f_N(v)| \leq \sup_{u \in K} \sum_{l=1}^d \left| \frac{\partial}{\partial u^l} f_N(u) \right| \varepsilon =: \mu \varepsilon$$

Отже, за нерівністю Чебишова

$$\mathbf{P}\{\tau \geq \delta\} \leq \mathbf{P}\{\mu \varepsilon \geq \delta\} \leq \frac{\varepsilon^2 \mathbf{E} \mu^2}{\delta^2}. \quad (3.10)$$

Зафіксуємо довільне  $l$  та оцінимо  $f'_N(u) := \frac{\partial}{\partial u^l} f_N(u)$ . Застосовуючи лему 7.1.1 отримуємо, що

$$\begin{aligned} \mathbf{E} \left( \sup_{u \in K} |f'_N(u)| \right)^2 &\leq C \mathbf{E} \left( \sum_{\alpha \in \{0,1\}^d} \int_{K_\alpha} (D^\alpha f'_N(u_\alpha))^2 (du_1)^\alpha \right) \\ &\leq C \sum_{\alpha \in \{0,1\}^d} \int_{K_\alpha} \frac{1}{N} \sum_{i,j=1}^N D^\alpha b'_j(u) D^\alpha b'_i(u) \mathbf{E} \zeta_i \zeta_j (du_1)^\alpha \\ &\leq C S \sum_{\alpha \in \{0,1\}^d} \int_{K_\alpha} \frac{1}{N} \sum_{j=1}^N (D^\alpha b'_j(u))^2 (du_1)^\alpha < C_K S L^2. \end{aligned}$$

Отже  $\mathbf{E} \mu^2 \leq C_K S L^2$ , і за (3.10) отримуємо твердження лема.

Лема доведена.

**Доведення лема 3.2.1** Позначимо  $\zeta_j = g(\xi_{j:N}) - \mathbf{E} g(\xi_{j:N})$ ,  $b_j(u) = a_{j:N}^k(u) - a_{j:N}^k(\gamma)$  ( $u \in \mathbb{R}^{2M}$ ). Тоді

$$\sqrt{N}(\hat{g}_N(\vec{a}^k(u)) - \hat{g}_N(\vec{a}^k(\gamma))) = \frac{1}{\sqrt{N}} \sum_{j=1}^N b_j(u) \zeta_j. \quad (3.11)$$

Так само, як у теоремі 3.1.2 отримуємо, що  $\inf_{u,N} \det \Gamma_N(u) > c' > 0$ . Отже функції  $b_j(u)$  є диференційовними по  $u$  у деякому околі справжнього значення  $\gamma$ . Тому  $D^\alpha b'_j(u)$  є рівномірно обмеженими по  $u$  у цьому околі  $\gamma$ . Застосування лема 3.2.2 до суми у правій частині (3.11) закінчує доведення.

Лема доведена.



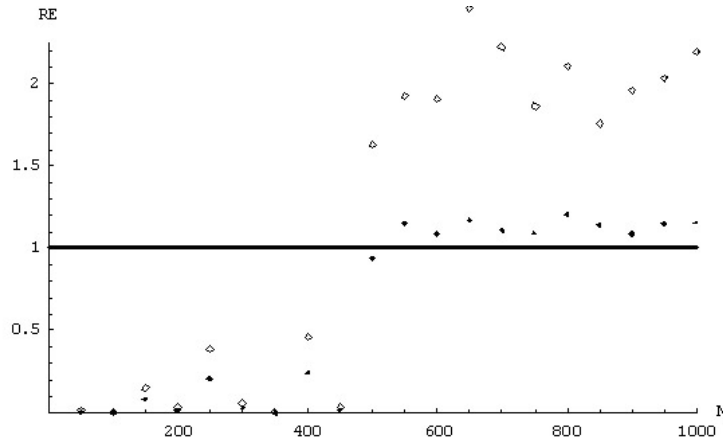


Рисунок 3.1: Відносна ефективність оцінок  $RE_k = MSE(\hat{g}_N^{k,0})/MSE(\check{g}_N^k)$  для різних об'ємів вибірки  $N$ .  $\bullet$  - для першої компоненти ( $k = 1$ );  $\circ$  - для другої компоненти ( $k = 2$ ).

**Приклад.** Поведінка лінійних та адаптивних оцінок досліджувалась на модельованих вибірках. Вибірki були побудовані з суміші двох компонент ( $M = 2$ ), причому їх концентрації змінювались за лінійним законом:  $w_{j:N}^1 = \frac{j}{N}$ ,  $w_{j:N}^2 = 1 - \frac{j}{N}$ . Розподіли обох компонент були нормальними:  $H_1$  була  $N(0, 1)$ ,  $H_2$  була  $N(1, 2)$ . Для вибірок обсягу від 50 до 1000 дві оцінки для середніх (тобто мінімаксна  $\hat{g}_N^{k,0}$  та адаптивна  $\check{g}_N^k$  з  $g(x) = x$ ) підраховано для  $K = 500$  модельованих вибірок. Підраховані середньоквадратичні похибки оцінок (MSE) та їх відносна ефективність  $RE_k = MSE(\hat{g}_N^{k,0})/MSE(\check{g}_N^k)$ .

Результати зображені на рисунку 3.1.

Ці результати показують, що лінійні оцінки з мінімаксними ваговими коефіцієнтами мають менші середньоквадратичні похибки при обсягах вибірки до 500 спостережень. Для вибірок більшого обсягу більш ефективними виявляються адаптивні оцінки. Зрозуміло, що для інших розподілів та інших концентрацій поріг, після якого адаптивні оцінки починають переважати мінімаксні, може бути іншим. Але загальна картина залишається незмінною: при малих обсягах вибірок прості мінімаксні лінійні оцінки дають кращі результати, на великих вибірках вони поступаються адаптивним.

При використанні  $\hat{g}_N^k$  на роль пілотних оцінок, можуть виникати проблеми, пов'язані з тим, що ці оцінки самі не є моментами деякого ймовірного розподілу: наприклад, оцінка  $\hat{g}^{k,0}$  може бути від'ємною, хоча  $\overline{g}^{k,0}$  завжди додатній. В результаті матриця  $\Gamma_N^{-1}(\hat{\gamma}_N)$  може не бути додатньо-визначеною і навіть бути виродженою. Це, звичайно погіршує поведінку адаптивних оцінок на вибірках скінченного обсягу, хоча і не впливає на їх асимптотичні властивості.

Для того, щоб пом'якшити цей недолік, можна запропонувати на роль пілотних оцінки, аналогічні лінійним оцінкам 3.3, у яких зважена емпірична функція замінюється виправленню емпіричною функцією розподілу, запропонованою у п. 2.3. Такі оцінки ми розглянемо у наступному параграфі.

### 3.3 Виправлені оцінки для моментів

У цьому параграфі ми обмежуємось розглядом одновимірних спостережень, тобто  $\mathcal{X} = \mathbb{R}$  і  $\xi_{j:N}$  є випадковими величинами. У цьому випадку лінійна оцінка для функціонального момента  $\bar{g}^k$  набуває вигляду

$$\hat{g}_{k:N} := \int g(x) \hat{F}_N(dx, a) = \frac{1}{N} \sum_{j=1}^N a_{j:N} g(\xi_{j:N}) \quad (3.12)$$

де, як і раніше

$$\hat{F}_N(x, a) := \frac{1}{N} \sum_{j=1}^N a_{j:N} \mathbb{1}\{\xi_{j:N} < x\} \quad (3.13)$$

— зважена емпірична функція розподілу, побудована по спостереженнях  $\xi_{j:N}$ .

У п. 2.3 ми ввели виправлені емпіричні функції розподілу  $\tilde{F}_N^+(x, a)$  та подібні до них. На відміну від  $\hat{F}_N(x, a)$ ,  $\tilde{F}_N^+(x, a)$  самі є функціями розподілу деяких ймовірнісних мір, тому можна очікувати, що оцінки моментів, побудовані на основі цих функцій, будуть мати кращі властивості при фіксованому обсязі вибірки, ніж оцінки (3.12). Найпростішим варіантом такої оцінки є

$$\tilde{g}_{k:N}^+ := \int g(x) \tilde{F}_N^+(dx, a). \quad (3.14)$$

У даному параграфі нашою метою буде дослідити асимптотичну поведінку цих та подібних оцінок і показати, що за певних умов виправлення емпіричної функції розподілу не змінює асимптотику оцінок. При цьому ми будемо

спиратись на результати про асимптотику емпіричних процесів (нормованих відхилень емпіричних ф.р. від оцінюваних).

На роль виправленої з.е.ф.р. ми будемо використовувати функції визначені (2.30-2.34), тобто  $\hat{F}_N^+(x, a)$ ,  $\hat{F}_N^-(x, a)$ ,  $\tilde{F}_N^-(x, a)$ ,  $\tilde{F}_N^\pm(x, a)$ ,  $\hat{F}_N^\pm$ . Як і раніше,  $F_N^*(x, a)$  може бути будь-якою з цих функцій.

У п. 2.3 показано, що функції  $F_N^*(x, a)$  можна зобразити у вигляді

$$F_N^*(x, a) = \frac{1}{N} \sum_{j=1}^N b_{j:N}^* \mathbb{I}\{\xi_{j:N} < x\},$$

де  $b_{j:N}^*$  — деякі коефіцієнти, що залежать від вибірки  $\Xi_N$ . Відповідно, оцінки для моментів, побудовані на основі  $F_N^*(x, a)$ , можна записати як

$$g_{k:N}^* := \int g(x) F_N^*(dx, a) = \frac{1}{N} \sum_{j=1}^N b_{j:N}^* g(\xi_{j:N}).$$

Нас буде цікавити асимптотична нормальність таких оцінок, тобто слабка збіжність  $Y_{k:N}^* := \sqrt{N}(g_{k:N}^* - \bar{g})$  до нормального розподілу з нульовим математичним сподіванням.

Як показано у теоремі 3.1.2, якщо для всіх  $H_m$ ,  $m = 1, \dots, M$  є скінченними функціональні моменти  $\overline{g_m^2} = \int g^2(x) H_m(dx)$ , то  $Y_{k:N}^* := \sqrt{N}(\hat{g}_{k:N} - \bar{g}) \Rightarrow Y$ , де  $Y$  має нормальний розподіл з нульовим середнім і дисперсією  $\sigma^2 = \langle (a)^2 d \rangle$ ,  $d_{j:N} = \sum_{m=1}^M \overline{g_m^2} w_{j:N}^m - (\sum_{m=1}^M \overline{g_m} w_{j:N}^m)^2$ .

### Теорема 3.3.1 Нехай

1.  $\sup_{j:N} |a_{j:N}| < A < \infty$ .
  2. Для всіх  $m = 1, \dots, M$ ,  $\exists \langle w^k w^m(a)^2 \rangle$ .
  3. Для всіх  $m = 1, \dots, M$ ,  $H_m$  є неперервними функціями на  $\mathbb{R}$ .
  4. Для всіх  $m = 1, \dots, M$ ,  $\text{supp } H_m \subseteq \text{supp } H_k$ .
  5. Виконані умови незміщеності  $\hat{F}_N(x, a)$  як оцінки для  $H_k$ , тобто  $\langle a w^m \rangle_N = \mathbb{I}\{k = m\}$  для всіх  $m = 1, \dots, M$ .
  6.  $g$  — функція обмеженої варіації на  $\mathbb{R}$ .
- Тоді  $Y_{k:N}^* \Rightarrow Y$  для всіх розглядуваних  $g_{k:N}^*$ .

**Зауваження.** Насправді досить вимагати, щоб  $g$  була функцією обмеженої варіації на  $\text{supp } H_m$ ,  $m = 1, \dots, M$ .

*Доведення.* Помітимо, що

$$Y_{k:N}^* = \sqrt{N} \left( \int g(x) F_N^*(dx, a) - \int g(x) H(dx) \right)$$

$$= \int g(x) B_N^*(dx, a).$$

За теоремою 2.3.2, існують процеси  $\check{B}_N^*(x)$  та  $\check{B}_N(x)$ , такі, що розподіл  $\check{Y}_N^* := \int g(x) \check{B}_N^*(x)$  той же самий, що і у  $Y_{k;N}^*$ , а розподіл  $\check{Y}_N := \int g(x) \check{B}_N(x)$  той же, що у  $Y_N$ , причому  $\sup_x |\check{B}_N^*(x) - \check{B}_N(x)| \rightarrow 0$  за ймовірністю.

Отже

$$\begin{aligned} |\check{Y}_N^* - \check{Y}_N| &= \left| \int g(x) (\check{B}_N^*(dx) - \check{B}_N(dx)) \right| \\ &= \left| \int (\check{B}_N^*(x) - \check{B}_N(x)) g(x) dx \right| \leq \text{VAR}_x g(x) \cdot \sup_x |\check{B}_N^*(x) - \check{B}_N(x)| \rightarrow 0 \end{aligned}$$

за ймовірністю. Звідси випливає, що розподіл  $\check{Y}_N^*$  (а отже і  $Y_{k;N}^*$ ) збігається слабо до тієї ж границі, до якої збігається  $Y_{k;N}$ , тобто до  $Y$ .

Теорема доведена.

**Теорема 3.3.2** *Нехай*

1.  $\sup_{j,N} |a_{j;N}| < A < \infty$ .
2. Для всіх  $m = 1, \dots, M$ ,  $\exists \langle w^k w^m(a)^2 \rangle$ .
3. Для всіх  $m = 1, \dots, M$ ,  $H_m$  є неперервними функціями на  $\mathbb{R}$ .
4. Для всіх  $m = 1, \dots, M$ ,  $\text{supp } H_m \subseteq \text{supp } H_k$ .
5. Виконані умови незміщеності  $\hat{F}_N(x, a)$  як оцінки для  $H_k$ , тобто  $\langle a w^m \rangle_N = \mathbb{1}\{k = m\}$  для всіх  $m = 1, \dots, M$ .
6.  $g : \mathbb{R} \rightarrow \mathbb{R}$  — неперервна монотонна функція і для всіх  $m = 1, \dots, M$  та деяких  $0 < D, C < \infty$ ,  $\gamma > 0$

$$H_m(x) \leq \frac{D}{|g(x)|^{2+\gamma}}, \quad \forall x < -C, \quad (3.15)$$

$$1 - H_m(x) \leq \frac{D}{|g(x)|^{2+\gamma}}, \quad \forall x > C. \quad (3.16)$$

Тоді  $\hat{Y}_{k;N}^\pm \Rightarrow Y$  при  $N \rightarrow +\infty$ .

**Зауваження.** Оскільки будь-яку функцію  $g$ , що має обмежену варіацію на всіх скінченних інтервалах, можна зобразити у вигляді  $g(x) = g^+(x) - g^-(x)$ , де  $g^+$  та  $g^-$  — монотонні функції, то насправді теорему 3.3.2 можна переформулювати для функцій з обмеженою варіацією на скінченних інтервалах, але тоді умови (3.15-3.16) потрібно перевіряти окремо для  $g^+$  та  $g^-$ . Наприклад, для  $g(x) = x^2$ , ці умови перетворюються на  $H_m(x) = O(|x|^{-4-\gamma})$  при  $x \rightarrow -\infty$ ,  $1 - H_m(x) = O(x^{-4-\gamma})$  при  $x \rightarrow \infty$ .

Крім теореми 2.3.2 про близькість емпіричних процесів для виправлених та не виправлених з.е.ф.р. у рівномірній нормі для доведення теореми 3.3.2 нам буде потрібна характеристика поведінки  $B_N(x)$ ,  $B_N^+(x)$  та  $B_N^-(x)$  при  $x \rightarrow \infty$  та  $x \rightarrow -\infty$ .

Позначимо

$$\bar{H}(x) = \sum_{m=1}^M H_m(x). \quad (3.17)$$

**Теорема 3.3.3** *Нехай*

1.  $\sup_{j,N} |a_{j:N}| < A < \infty$ .
  2. Для всіх  $m = 1, \dots, M$ ,  $\exists \langle w^k w^m(a)^2 \rangle$ .
  3. Для всіх  $m = 1, \dots, M$ ,  $H_m$  є неперервними функціями на  $\mathbb{R}$ .
  4. Для всіх  $m = 1, \dots, M$ ,  $\text{supp } H_m \subseteq \text{supp } H_k$ .
  5. Виконані умови незміщеності  $\bar{F}_N(x, a)$  як оцінки для  $H_k$ , тобто  $\langle a w^m \rangle_N = \mathbb{1}\{k = m\}$  для всіх  $m = 1, \dots, M$ .
- Тоді, для довільних  $b$  і  $\delta$ ,  $0 < \delta < 1/2$ ,

$$\begin{aligned} \sup_N \mathbb{P} \left\{ \sup_{x < b} \frac{|B_N(x)|}{\bar{H}(x)^{1/2-\delta}} > \lambda \right\} &\rightarrow 0 \text{ при } \lambda \rightarrow \infty, \\ \sup_N \mathbb{P} \left\{ \sup_{x < b} \frac{|\hat{B}_N^+(x)|}{\bar{H}(x)^{1/2-\delta}} > \lambda \right\} &\rightarrow 0 \text{ при } \lambda \rightarrow \infty, \\ \sup_N \mathbb{P} \left\{ \sup_{x > b} \frac{|B_N(x)|}{(M - \bar{H}(x))^{1/2-\delta}} > \lambda \right\} &\rightarrow 0 \text{ при } \lambda \rightarrow \infty, \\ \sup_N \mathbb{P} \left\{ \sup_{x > b} \frac{|\hat{B}_N^-(x)|}{(M - \bar{H}(x))^{1/2-\delta}} > \lambda \right\} &\rightarrow 0 \text{ при } \lambda \rightarrow \infty. \end{aligned}$$

Доведення цієї теореми спирається на наступну лему.

**Лема 3.3.1** *В умовах теореми 3.3.3 існує таке  $C < \infty$ , не залежне від  $N$ , що для всіх  $\varepsilon > 0$ ,  $x \in \mathbb{R}$ ,*

$$\mathbb{P}\{\sup_{t < x} |B_N^+(t)| \geq \varepsilon\} \leq \mathbb{P}\{\sup_{t < x} |B_N(t)| \geq \varepsilon\} \leq C(\bar{H}^2(x)\varepsilon^{-4} + \bar{H}(x)\varepsilon^{-2}), \quad (3.18)$$

$$\begin{aligned} \mathbb{P}\{\sup_{t > x} |B_N^-(t)| \geq \varepsilon\} &\leq \mathbb{P}\{\sup_{t > x} |B_N(t)| \geq \varepsilon\} \\ &\leq C((M - \bar{H}(x))^2\varepsilon^{-4} + (M - \bar{H}(x))\varepsilon^{-2}), \end{aligned}$$

**Доведення.** Щоб довести другу нерівність у (3.18), застосуємо лему 7.4.2 на інтервалі  $(-\infty, x]$  з  $\gamma = 2$ ,  $\alpha = 1$ . Для цього оцінимо

$$J := \mathbf{E}(B_N(t) - B_N(t_1))^2(B_N(t_2) - B_N(t_1))^2.$$

Позначимо

$$\eta_j(x, y) := a_{j:N}(\mathbb{I}\{\xi_{j:N} \in [y, x]\} - \mathbf{P}\{\xi_{j:N} \in [y, x]\}).$$

Тоді

$$B_N(t) - B_N(s) = \frac{1}{\sqrt{N}} \sum_{j=1}^N \eta_j(s, t)$$

і

$$\begin{aligned} J &= \frac{1}{N^2} \mathbf{E} \sum_{j,k,l,m=1}^N \eta_j(t, t_1) \eta_k(t, t_1) \eta_l(t_2, t) \eta_m(t_2, t) \\ &\leq \frac{C}{N^2} \sum_{j \neq k} \{ \mathbf{E}(\eta_j(t, t_1))^2 (\eta_k(t_2, t))^2 + \mathbf{E} \eta_j(t, t_1) \eta_j(t_2, t) \eta_k(t, t_1) \eta_k(t_2, t) \} \\ &\quad + \frac{1}{N^2} \sum_{j=1}^N \mathbf{E}(\eta_j(t, t_1) \eta_j(t_2, t))^2 \end{aligned}$$

(ми скористались тим, що  $\mathbf{E} \eta_j = 0$  і  $\eta_k$  та  $\eta_m$  незалежні при  $k \neq m$ ).

Оскільки

$$\mathbf{E} \eta_j^2(t, t_1) \leq \mathbf{E}(a_{j:N})^2 \mathbb{I}\{\xi_{j:N} \in [t_1, t]\} \leq A^2 \bar{H}([t_1, t]) \leq A^2 \bar{H}([t_1, t_2]),$$

$$\begin{aligned} \mathbf{E} \eta_j(t, t_1) \eta_j(t_2, t) &= -(a_{j:N})^2 \mathbf{P}\{\xi_{j:N} \in [t_1, t]\} \mathbf{P}\{\xi_{j:N} \in [t, t_2]\} \leq A^2 \bar{H}^2([t_1, t_2]), \\ \mathbf{E}(a_{j:N})^4 (\mathbb{I}\{\xi_{j:N} \in [t_1, t]\} - \mathbf{P}\{\xi_{j:N} \in [t_1, t]\})^2 (\mathbb{I}\{\xi_{j:N} \in [t, t_2]\} - \mathbf{P}\{\xi_{j:N} \in [t, t_2]\})^2 \\ &\leq C A^4 \bar{H}^3([t_1, t_2]) \end{aligned}$$

і  $\bar{H}(\mathbb{R}) \leq M$ , отримуємо  $J \leq C(\bar{H}([t_1, t_2]))^2$ . Отже, за лемою 7.4.2,

$$\mathbf{P}\{\sup_{t \leq x} |B_N(t)| > \varepsilon\} \leq \mathbf{P}\{|B_N(x)| > \frac{\varepsilon}{2}\} + \frac{C}{\varepsilon^4} (\bar{H}(x))^2 \quad (3.19)$$

(оскільки  $B_N(-\infty) = \bar{H}(-\infty) = 0$ ). Оцінимо

$$\mathbf{P}\{|B_N(x)| < \varepsilon/2\} \leq \frac{\text{Var } B_N(x)}{\varepsilon^2} \leq \frac{A^2 \bar{H}(x)}{\varepsilon}, \quad (3.20)$$

оскільки  $\text{Var } B_N(x) \leq \frac{1}{N} \sum_{j=1}^N (a_{j:N})^2 \mathbf{P}\{\xi_{j:N} < x\}$ . З (3.19) та (3.20) отримуємо другу нерівність (3.18).

Перша нерівність у (3.18) випливає з того, що  $B_N^+(x) \geq B_N(x)$  для всіх  $x$ , і, в той же час, при  $t < x$ ,

$$\begin{aligned} B_N^+(t) &= \sqrt{N} \sup_{y < t} (\hat{F}_N(y, a) - H_k(t)) \\ &= \sqrt{N} \sup_{y < t} (B_N(y)/\sqrt{N} + H_k(y) - H_k(t)) \leq \sup_{y < x} B_N(y), \end{aligned}$$

оскільки  $H_k(y) - H_k(t) \leq 0$  при  $y \leq t$ . Друга пара нерівностей леми доводиться аналогічно.

**Доведення теореми 3.3.3.** Доведемо перше твердження теореми. Твердження 2-4 доводяться аналогічно. Позначимо

$$p_\lambda := \mathbf{P} \left\{ \sup_{x < b} \frac{|B_N(x)|}{\bar{H}(x)^{1/2-\delta}} > \lambda \right\},$$

$x_j$  — число, для якого  $\bar{H}(x_j) = 2^{-j}$ . Тоді

$$\mathbf{P} \left\{ \sup_{x_{j+1} \leq x \leq x_j} \frac{|B_N(x)|}{\bar{H}(x)^{1/2-\delta}} > \lambda \right\} \subseteq A_j$$

де

$$A_j := \{\forall x < x_j, B_N(x) \leq \lambda H^{1/2-\delta}(x_{j+1})\}$$

Застосовуючи лему 3.3.1 до події  $A_j$  з  $\varepsilon_j = \lambda \bar{H}^{1/2-\delta}(x_{j+1})$ , отримуємо

$$\begin{aligned} p_\lambda &\leq \sum_{j=1}^{\infty} \mathbf{P}\{A_j\} \leq \sum_{j=1}^{\infty} C(\bar{H}^2(x_j)\varepsilon^{-4} + \bar{H}(x_j)\varepsilon^{-2}) \\ &= C \sum_{j=1}^{\infty} \left( \frac{2^{2j}}{\lambda^4 2^{(-2+4\delta)j}} + \frac{2^j}{\lambda^2 2^{(-1+2\delta)j}} \right) \leq C(\lambda^{-4} + \lambda^{-2}) \rightarrow 0 \text{ при } \lambda \rightarrow +\infty. \end{aligned}$$

**Лема 3.3.2** В умовах теореми 3.3.3, якщо  $H_k(b) < 1/2$ ,  $H_k(c) > 1/2$ , то

$$\mathbf{P}\{\exists x < b : \hat{F}_N^\pm(x, a) \neq \hat{F}_N^+(x, a)\} \rightarrow 0, \quad N \rightarrow \infty,$$

$$\mathbf{P}\{\exists x > c : \hat{F}_N^\pm(x, a) \neq \hat{F}_N^-(x, a)\} \rightarrow 0, \quad N \rightarrow \infty,$$

**Доведення.** З теореми 2.3.2 випливає, що  $\sup_x |\hat{F}_N^+(x, a) - H_k(x)| \rightarrow 0$  і  $\sup_x |\hat{F}_N^-(x, a) - H_k(x)| \rightarrow 0$  при  $N \rightarrow \infty$  за ймовірністю. Отже, враховуючи монотонність  $H_k$  і умову леми, маємо  $\mathbf{P}\{\sup_{x < b} \hat{F}_N^+(x, a) > 1/2\} \rightarrow 0$ ,  $\mathbf{P}\{\inf_{x > c} \hat{F}_N^-(x, a) < 1/2\} \rightarrow 0$ , при  $N \rightarrow \infty$ .

Враховуючи (2.34), отримуємо твердження леми.

**Доведення теореми 3.3.2.** Доведемо теорему для випадку, коли  $g$  — монотонно зростаюча функція. Згідно з теоремою 2.3.2, існують такі випадкові процеси  $\check{B}_N^\pm$  та  $\check{B}_N$ , що  $\check{B}_N^\pm$  має той самий розподіл, що і  $\hat{B}_N^\pm$ , а  $\check{B}_N$  — той же, що і  $\hat{B}_N$ , причому  $\sup_x |\check{B}_N^\pm(x) - \check{B}_N(x)| \rightarrow 0$  при  $N \rightarrow \infty$  за ймовірністю. Помітимо, що  $\check{Y}_{k,N}^\pm = \int g(x) \check{B}_N^\pm(dx)$  має той же розподіл, що і  $\hat{Y}_{k,N}^\pm$ , а  $\check{Y}_{k,N} = \int g(x) \check{B}_N(dx)$  — той же, що і  $\hat{Y}_{k,N}$ . Тому для доведення теореми досить переконатись, що  $\check{Y}_{k,N}^\pm - \check{Y}_{k,N} \rightarrow 0$  за ймовірністю. Помітимо, що для довільного  $b > 0$ ,

$$J := |\check{Y}_{k,N}^\pm - \check{Y}_{k,N}| = \left| \int_{-\infty}^{+\infty} (\check{B}_N^\pm(x) - \check{B}_N(x))g(dx) \right| \leq J_1 + J_2 + J_3$$

де

$$J_1 = \int_{-b}^b |\check{B}_N^\pm(x) - \check{B}_N(x)|g(dx),$$

$$J_2 = \int_{-\infty}^{-b} |\check{B}_N^\pm(x) - \check{B}_N(x)|g(dx),$$

$$J_3 = \int_b^{\infty} |\check{B}_N^\pm(x) - \check{B}_N(x)|g(dx).$$

Ми доведемо, що для довільних фіксованих  $\alpha > 0$ ,  $\varepsilon > 0$  можна обрати числа  $b$  та  $N_0$  так, що

$$\sup_{N > N_0} \mathbf{P}\{J_2^N > \alpha\} < \varepsilon, \quad (3.21)$$

$$\sup_{N > N_0} \mathbf{P}\{J_3^N > \alpha\} < \varepsilon. \quad (3.22)$$

Звідси, враховуючи, що для будь-якого фіксованого  $b$ ,

$$J_1^N \leq (g(b) - g(-b) \sup_{|x| < b} |\check{B}_N^\pm(x) - \check{B}_N(x)|) \rightarrow 0$$

при  $N \rightarrow \infty$  за ймовірністю, отримуємо твердження теореми.



Отже, доведемо (3.21). (Доведення (3.22) аналогічне). Оберемо  $b$  так, щоб  $H_k(b) < 1/2$ . Тоді, враховуючи лему 3.3.2, досить довести, що при достатньо великих  $b$ ,  $\sup_N \mathbf{P}\{\tilde{J}_2^N > \alpha\} < \varepsilon$ , де  $\tilde{J}_2^N := J_{21}^N + J_{22}^N$ ,  $J_{21}^N := \int_{-\infty}^{-b} |B_N(x)|g(dx)$ ,  $J_{21}^N := \int_{-\infty}^{-b} |B_N^+(x)|g(dx)$ .

Оцінимо  $J_{21}^N$ . Фіксуємо довільне число  $r$  та  $0 < \delta < 1/2$  таке, що  $\gamma' := (2 + \gamma)(1/2 - \delta) > 1$ . За теоремою 3.3.3 існує таке  $\lambda$ , що для подій

$$A_N := \left\{ \sup_{t < r} \frac{|B_N(t)|}{\bar{H}^{1/2-\delta}(t)} > \lambda \right\}$$

виконано  $\sup_N \{A_N\} < \varepsilon$ . Якщо виконано  $\bar{A}_N$ , то  $\forall t < r$ ,

$$|B_N(t)| < \lambda \bar{H}^{1/2-\delta}(t) \leq \frac{C}{|g(t)|^{\gamma'}}$$

і, відповідно, при  $b < r$ ,  $J_{21}^N \leq \int_{-\infty}^{-b} \frac{C}{(g(t))^{\gamma'}} g(dt) < \infty$ . Тому можна обрати  $b$  достатньо великим, щоб  $J_{21}^N < \alpha/2$  при виконанні  $\bar{A}_N$ . Отже  $\sup_N \mathbf{P}\{J_{21}^N > \alpha/2\} \leq \sup_N \mathbf{P}\{A_N\} \leq \varepsilon$ .

Аналогічно оцінюється  $J_{22}^N$ . Таким чином, (3.21) доведено, що і завершує доведення теореми.

### 3.4 Оцінювання квантилів

Квантилі розподілів випадкових величин грають у статистиці роль не меншу, ніж моменти. Задача оцінки квантилів природно виникає при побудові довірчих інтервалів. Оцінки, що використовують емпіричні квантилі, виявляються найбільш стійкими по відношенню до забруднень вибірки. Використання квантилів лежить в основі так званої VaR-технології (Value At Risk technology) у фінансовій математиці.

У цьому розділі ми розглянемо задачу оцінки квантиля розподілу однієї компоненти суміші зі змінними концентраціями. Кажуть, що  $x$  є квантилем рівня  $\alpha$  функції розподілу  $H$  деякої випадкової величини  $\eta$ , якщо  $\mathbf{P}\{\eta < x\} = \alpha$ . Це позначають  $x = Q^\eta(\alpha) = Q^\eta(\alpha)$ .

Якщо  $H$  — неперервна функція розподілу, строго зростаюча в околі точки  $Q^\eta(\alpha)$ , то

$$Q^H(\alpha) = Q^\eta(\alpha) = H^{-1}(\alpha),$$

де  $H^{-1}$  — функція, обернена до  $H$ .

Якщо  $H(x)$  є константою:  $H(x) = \alpha$  на деякому інтервалі  $x \in (x_0, x_1]$ , то будь-яке число з цього інтервалу можна вважати квантилем рівня  $\alpha$  для  $H$ . Якщо  $H$  є розривною у деяких точках, то квантилі відповідного рівня у нашому розумінні не існуватимуть<sup>1</sup>. Надалі ми вважаємо, що функції розподілу всіх компонент суміші, яка розглядається є неперервними і оцінюваний квантиль є точкою росту для них усіх. Таким чином, у цьому випадку наше означення квантиля є однозначним.

На роль оцінки для квантиля  $Q^\eta(\alpha)$  за спостереженнями незалежних копій  $\eta$ , як правило, використовують  $Q^{\hat{F}_N}(\alpha)$ , де  $\hat{F}_N$  — відповідним чином згладжена емпірична функція розподілу, побудована по спостереженнях. У випадку спостережень з суміші зі змінними концентраціями природно замість  $\hat{F}_N$  використати зважену емпіричну функцію розподілу, теж відповідним чином підправлену. По-перше, оскільки зважена емпірична функція розподілу  $\hat{F}_N(x, a)$ , як правило, є немонотонною, її варто виправити одним із способів, розглянутих у п. 2.3. По-друге, після цього доцільно згладити стрибки виправленої з.е.ф.р. так, щоб відповідний квантиль для неї завжди існував. Процедура згладжування доцільно обрати так, щоб її результат не надто сильно відрізнявся від початкової функції.

Опишемо тепер більш детально запропоновану оцінку. Нехай  $\mathcal{X} = \mathbb{R}$ , спостереження являють собою вибірку  $\Xi_N = \{\xi_{i:N}, j = 1, \dots, N\}$  з суміші зі змінними концентраціями, розподіл якої задано 3.1. Ми будемо оцінку для квантиля рівня  $\alpha$  ( $0 < \alpha < 1$ ) функції розподілу  $k$ -тої компоненти, тобто для  $Q^{H^k}(\alpha)$ . Для цього розглянемо довільну виправлену зважену функцію розподілу, отриману за однією з формул (2.30-2.34). Знову через  $F_N^*(x, a)$  будемо позначати яку-небудь з цих функцій. (Як стане зрозуміло далі, досліджувана нами асимптотика отриманих оцінок не залежить від способу виправлення з.е.ф.р.).

Використовуючи алгоритм, описаний у п. 2.3, функцію  $F_N^*(x, a)$  можна зобразити у вигляді

$$F_N^*(x, a) = \frac{1}{N} \sum_{j=1}^N b_{j:N}^* \mathbb{1}\{\xi_{j:N} < x\},$$

де  $b_{j:N}^*$  — деякі коефіцієнти, що залежать від вибірки  $\Xi_N$ . Перенумеруємо точки стрибків  $F_N^*(x, a)$  у порядку зростання  $x_1, \dots, x_n$ . (Зрозуміло,

<sup>1</sup>Звичайно, можна усунути ці неприємні особливості, визначивши функцію, що задає квантилі, як узагальнену (у якому-небудь розумінні) обернену для функції розподілу. Однак варто відмітити, що практична користь від таких “узагальнених” квантилів мінімальна.

що  $x_j$  дорівнюють вибірковим значенням, розташованим у порядку зростання, тобто порядковим статистикам. Однак деякі порядкові статистики виявляються пропущеними, оскільки, за способом побудови  $F_N^*(x, a)$ , деякі  $b_{j:N}^* = 0$  і у цих точках стрибків  $F_N^*(x, a)$  немає. Тому  $n < N$ .) Значення коефіцієнта  $b_{j:N}^*$ , який відповідає  $x_i$ , позначимо  $b_i$ . Тоді

$$F_N^*(x, a) = \frac{1}{N} \sum_{i=1}^n b_i \mathbb{I}\{x_i < x\}, \quad (3.23)$$

Стрибок функції  $F_N^*(x, a)$  у точці  $x_i$  дорівнює  $b_i/N$ .

Згладимо функцію  $F_N^*(x, a)$  функцією  $\check{F}(x)$ , графік якої є ламаною, яка має злами у точках  $x_i$  і проходить у цих точках через середину стрибка  $F_N^*(x, a)$ . Точніше, позначимо  $y_i = F_N^*(x_i, a) \frac{1}{N} \left( \sum_{j < i} b_j + \frac{1}{2} b_i \right)$ , тоді

$$\check{F}(x) = \check{F}_N(x) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} (x - x_i) + y_i, \quad (3.24)$$

якщо  $x \in [x_i, x_{i+1}]$ . (За межами  $[x_1, x_n]$  функція  $\check{F}$  не визначена).

Оцінку  $\hat{Q}_N^k(\alpha)$  для квантиля  $Q^{H_k}(\alpha)$  визначимо як

$$\hat{Q}_N^k(\alpha) = \check{F}^{-1}(\alpha). \quad (3.25)$$

Зрозуміло, що при такому означенні можуть виникнути проблеми при знаходженні квантилів дуже високого  $\alpha \simeq 1$  або дуже низького  $\alpha \simeq 0$  рівнів, коли відповідна обернена функція виявляється невизначеною. В принципі, щоб усунути цей недолік, можна довизначити  $\check{F}(x)$  за межами інтервалу  $[x_1, x_n]$  будь-яким способом, так, щоб вона стала неперервною строго монотонною функцією розподілу. Тоді (3.25) визначатиме оцінку для будь-якого рівня  $\alpha$ . Спосіб довизначення  $\check{F}(x)$  не вплине на асимптотику  $\hat{Q}_N^k(\alpha)$  при фіксованому  $\alpha$  та  $N \rightarrow \infty$ , якщо у початкової зваженої емпіричної функції розподілу вагові коефіцієнти рівномірно обмежені, тобто

$$\sup_{j,N} |a_{j,N}| < C < \infty. \quad (3.26)$$

Однак слід відмітити, що для екстремальних квантилів (тобто для квантилів з рівнями  $\alpha \simeq 0$  або  $\alpha \simeq 1$ ) оцінювання доцільно проводити методами, що використовують техніку аналізу хвостів розподілу. У граничному випадку  $\alpha = 1$  це питання розглядається у наступному параграфі. Запропонований метод призначений для квантилів, що знаходяться посередині розподілу — таких, як медіана, кuartилі або децилі.

Неважко визначити оцінку  $\hat{Q}_N^k(\alpha)$  у явному вигляді. Для цього потрібно знайти таке  $i$ , при якому  $y_i < \alpha < y_{i+1}$  і покласти

$$\hat{Q}_N^k(\alpha) = \frac{\alpha - y_i}{y_{i+1} - y_i}(x_{i+1} - x_i) + x_i.$$

По суті, це формула для лінійної інтерполяції (так само, як і (3.24)). Однак для дослідження асимптотичної поведінки оцінок ми скористаємось означенням (3.25). При цьому нам буде корисною наступна лема.

**Лема 3.4.1** *Нехай  $F$  — неперервна функція розподілу на  $\mathbb{R}$ ,  $F_N^*$  — послідовність функцій розподілу вигляду (3.23), така, що  $\sup_{x \in \mathbb{R}} |F_N^*(x) - F(x)| \rightarrow \infty$ . Тоді  $\sup_{x \in \mathbb{R}} |\check{F}_N(x) - F(x)| \rightarrow \infty$ ,*

*(Функції  $\check{F}_N(x)$  довизначаються за межами інтервалів їх визначення довільним чином, так, щоб бути функціями розподілу на  $\mathbb{R}$ ).*

**Доведення.** Застосовуючи перетворення аргументу  $x \rightarrow 2 \tan(x)/\pi$  можна перевести всі розглядувані функції у функції розподілу, зосереджені на  $[-1, 1]$ , причому рівномірна норма не змінюється. Тому будемо одразу вважати, що всі розглядувані функції зосереджені на  $[-1, 1]$  і  $\sup$  береться по  $x \in [-1, 1]$ .

Зафіксуємо довільне  $\varepsilon > 0$ . Оскільки  $F$  є неперервною на  $[-1, 1]$ , то можна побудувати скінченний набір відкритих інтервалів  $U_1, \dots, U_K$ , який покриває  $[-1, 1]$  і для всіх  $k = 1, \dots, K$ ,  $\sup_{x, y \in U_k} |F(x) - F(y)| < \varepsilon/3$ . Враховуючи рівномірну збіжність  $F_N^*$  до  $F$  отримуємо, що, при достатньо великих  $N$ , для всіх  $k = 1, \dots, K$ ,  $\sup_{x, y \in U_k} |F_N^*(x) - F_N^*(y)| < 2\varepsilon/3$ . Отже, для таких  $N$  величини стрибків  $b_i/N$  функцій  $F_N^*$  не можуть бути більшими, ніж  $2\varepsilon/3$ . Але  $\sup_{x \in [-1, 1]} |F_N^*(x) - \check{F}_N(x)| \leq \max_i b_i/N$  за побудовою  $\check{F}_N$ . Тому  $\sup_{x \in \mathbb{R}} |\check{F}_N(x) - F(x)| \leq \sup_{x \in [-1, 1]} |F_N^*(x) - \check{F}_N(x)| + \sup_{x \in [-1, 1]} |\check{F}_N(x) - F(x)| \leq 2\varepsilon/3 + \varepsilon/3 = \varepsilon$  для достатньо великих  $N$ .

Внаслідок довільності  $\varepsilon$  отримуємо твердження леми.

**Теорема 3.4.1** *Нехай  $H_k$  — неперервна ф.р., для вагових коефіцієнтів  $a$  виконана умова незміщеності (2.5):*

$$\langle aw^m \rangle_N = \mathbb{1}\{m = k\} \text{ для всіх } m = 1, \dots, M;$$

$$i \sup_{j, N} |a_{j, N}| < \infty.$$

$$\text{Тоді } \sup_{x \in \mathbb{R}} |\check{F}_N(x) - H_k(x)| \rightarrow 0 \text{ м.н. при } N \rightarrow \infty.$$

**Доведення.** Враховуючи (2.21) отримуємо, що в умовах теореми

$$\sup_{x \in \mathbb{R}} |\hat{F}_N(x, a) - H_k(x)| \rightarrow 0$$

м.н. З урахуванням (2.36-2.35) звідси випливає, що  $\sup_{x \in \mathbb{R}} |F_N^*(x, a) - H_k(x)|$  прямує до 0, а, отже, за лемою 3.4.1,  $\sup_{x \in \mathbb{R}} |\check{F}_N(x) - H_k(x)| \rightarrow 0$ .

Теорема доведена.

З цієї теореми одразу отримуємо консистентність  $\hat{Q}_N^k(\alpha)$  як оцінки  $Q^{H_k}(\alpha)$  у випадку, коли  $H_k$  — неперервна функція.

**Теорема 3.4.2** *Нехай  $H_k$  — неперервна ф.р., для вагових коефіцієнтів  $a$  виконана умова незміщеності (2.5) і  $\sup_{j,N} |a_{j:N}| < \infty$ . Тоді  $\hat{Q}_N^k(\alpha) \rightarrow Q^{H_k}(\alpha)$  м.н. при  $N \rightarrow \infty$ .*

**Доведення** випливає з теореми 3.4.1, оскільки відображення  $F \rightarrow Q^F(\alpha)$  є неперервним у просторі неперервних функцій розподілу з рівномірною метрикою.

Тепер доведемо асимптотичну нормальність  $\hat{Q}_N^k(\alpha)$ .

**Теорема 3.4.3** *Нехай*

1. Для деякого  $A < \infty$   $\sup_{j,N} |a_{j:N}| < A$ .
2. Для всіх  $l, m = 1, \dots, M$  існують границі

$$\langle w^l w^m(a)^2 \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N w_{j:N}^l w_{j:N}^m (a_{j:N})^2.$$

3.  $H_m$  є неперервними функціями на  $\mathbb{R}$  при всіх  $m = 1, \dots, M$ .

4. Виконана умова незміщеності (2.5).

5. Існує такий окіл  $I$  точки  $Q^{H_k}(\alpha)$ , що для всіх  $m = 1, \dots, M$ , всі точки  $I$  є точками росту  $H_m$ .

6. На  $I$  у  $H_k(x)$  існує неперервна похідна  $h_k(x)$  і  $h_k(Q^{H_k}(\alpha)) \neq 0$ .

Тоді розподіл  $\sqrt{N}(\hat{Q}_N^k(\alpha) - Q^{H_k}(\alpha))$  слабо збігається при  $N \rightarrow \infty$  до гауссового розподілу з нульовим середнім і дисперсією, рівною

$$s^2 = \frac{1}{(h_k(Q^{H_k}(\alpha)))^2} \left( \sum_{m=1}^M \langle w^m(a)^2 \rangle H_m(Q^{H_k}(\alpha)) - \sum_{i,m=1}^M \langle w^m w^i(a)^2 \rangle H_m(Q^{H_k}(\alpha)) H_i(Q^{H_k}(\alpha)) \right)$$

**Доведення.** Позначимо

$$B_N^*(x) = B_N(x, a) := \sqrt{N}(F_N^*(x, a) - H_k(x)).$$

За теоремою 2.3.2,  $B_N \xrightarrow{\text{Sk}} B$  де  $B$  — неперервний м.н. гауссів процес, описаний у цій теоремі. Покажемо, що процес

$$\check{B}_N(x) = B_N(x, a) := \sqrt{N}(\check{F}_N(x, a) - H_k(x))$$

також збігається за Скороходом до  $B(x)$  на  $I$ .

Для цього помітимо, що за побудовою  $F_N^*(x, a)$

$$0 < b_j^* \leq \sup_{j,N} |a_{j:N}| < A,$$

а за побудовою  $\check{F}_N(x, a)$

$$\sup_{x \in [x_i, x_{i+1}]} |\check{F}_N(x, a) - F_N^*(x, a)| \leq \frac{1}{2N} \max(b_i^*, b_{i+1}^*).$$

Отже

$$\sup_{x \in \mathbb{R}} |\check{F}_N(x, a) - F_N^*(x, a)| \leq \frac{A}{2N}.$$

Тому  $|\check{B}_N(x) - B_N^*(x)| \leq \frac{A}{2\sqrt{N}}$  і  $\check{B}_N \xrightarrow{\text{Sk}} B$ . Оскільки нас цікавить лише слабка збіжність розподілів, можна ототожнити  $\check{B}_N$  з їх копіями, побудованими на одному ймовірнісному просторі і вважати, що

$$\sup_{x \in I} |\check{B}_N(x) - B(x)| \rightarrow 0, \text{ при } N \rightarrow \infty. \quad (3.27)$$

Для довільного  $x \in \mathbb{R}$  позначимо

$$A_N = \{\sqrt{N}(\hat{Q}_N^k(\alpha) - Q^{H_k}(\alpha)) < x\}.$$

Покажемо, що  $P(A_N) \rightarrow P\{\eta < x\}$ , де  $\eta \sim N(0, s^2)$ . Це є твердження теорему.

Отже

$$\begin{aligned} A_N &= \{\sqrt{N}(\check{F}_N^{-1}(\alpha) - H_k^{-1}(\alpha)) < x\} \\ &= \{\check{F}_N^{-1}(\alpha) < H_k^{-1}(\alpha) + x/\sqrt{N}\} \\ &= \{\alpha < \check{F}_N(H_k^{-1}(\alpha) + x/\sqrt{N})\} \end{aligned}$$

$$\begin{aligned}
&= \left\{ \alpha < H_k(H_k^{-1}(\alpha) + x/\sqrt{N}) + \frac{\check{B}_N(H_k^{-1}(\alpha) + x/\sqrt{N})}{\sqrt{N}} \right\} \\
&= \{-\check{B}_N(H_k^{-1}(\alpha) + x/\sqrt{N}) < \sqrt{N}(H_k(H_k^{-1}(\alpha) + x/\sqrt{N}) - \alpha)\}
\end{aligned}$$

Звідси, враховуючи збіжність  $\check{B}_N$  до  $B$ , неперервність  $B$ , і той факт, що

$$\Delta(H_k(H_k^{-1}(\alpha) + x\Delta) - \alpha)/\Delta \rightarrow h_k(H_k^{-1}(\alpha))x,$$

отримуємо, що

$$\mathbf{P}(A_N) \rightarrow \mathbf{P}\{B(H_k^{-1}(\alpha)) < h_k(H_k^{-1}(\alpha))x\}.$$

Оскільки  $B(H_k^{-1}(\alpha))/h_k(H_k^{-1}(\alpha)) \sim N(0, s^2)$ , отримуємо твердження теореми.

Теорема доведена.

### 3.5 Оцінка екстремальних точок розподілів КОМПОНЕНТ

Нехай  $\xi$  — деяка дійснозначна випадкова величина з неперервною функцією розподілу  $H$ . Позначимо  $m_\xi = m_H = \text{essinf}(\xi) = \sup\{x : \mathbf{P}\{\xi < x\} = H(x) = 0\}$ ,  $M_\xi = M_H = \text{ess sup}(\xi) = \inf\{x : \mathbf{P}\{\xi \geq x\} = 1 - H(x) = 0\}$ . Значення  $m_\xi$  та  $M_\xi$  називають відповідно нижньою та верхньою екстремальними точками розподілу в.в.  $\xi$ . У даному параграфі ми розглянемо задачу оцінювання нижньої екстремальної точки (верхня оцінюється аналогічно) по спостереженнях, що являють собою вибірку з суміші зі змінними концентраціями.

Задача оцінювання екстремальних точок у випадку вибірок з незалежних, однаково розподілених (з ф.р.  $H$ ) випадкових величин  $\xi_1, \dots, \xi_N$  добре досліджена [2], с.203–207. Нехай, скажімо,  $m_\xi > -\infty$  і для деяких  $c > 0$ ,  $\delta > 0$ ,

$$\mathbf{P}\{\xi \in [m_\xi, x]\} \geq c(x - m_\xi) \text{ для всіх } x \in (m_\xi, m_\xi + \delta). \quad (3.28)$$

Розглянемо оцінку

$$\tilde{m}_N = \min\{\xi_1, \dots, \xi_N\}. \quad (3.29)$$

Легко бачити, що для будь-якого  $\varepsilon_N$ , такого, що  $\varepsilon_N \rightarrow 0$  при  $N \rightarrow \infty$ ,

$$\mathbf{P}\{\tilde{m}_N - m_\xi > \varepsilon_N\} = \prod_{j=1}^N (1 - \mathbf{P}\{\xi_j \in [m_\xi, m_\xi + \varepsilon_N]\}) \leq (1 - c\varepsilon_N)^N \sim e^{-cN\varepsilon_N},$$

отже, для довільної числової послідовності  $A_N \rightarrow \infty$ ,  $N \rightarrow \infty$ , поклавши

$$\varepsilon_N = \frac{A_N \ln N}{N}, \quad (3.30)$$

отримуємо

$$|\tilde{m}_N - m_\xi| \leq \varepsilon_N \text{ м.н. при достатньо великих } N. \quad (3.31)$$

Збіжність зі швидкістю  $\varepsilon_N$  свідчить про нерегулярність задачі оцінювання  $m_\xi$ , оскільки для регулярних задач, з якими ми мали справу досі, характерною є швидкість  $1/\sqrt{N}$  (точніше  $\ln \ln N/\sqrt{N}$ , якщо йдеться про збіжність майже напевне). І справді, умова (3.28) виконується, якщо, наприклад, існує щільність розподілу  $h(x) = dH(x)/dx$ , відділена від 0 на інтервалі  $[m_\xi, m_\xi + \delta]$ . Оскільки при  $x < m_\xi$ ,  $h(x) = 0$ , то в точці  $m_\xi$  щільність має стрибок. Як відомо ([10], розділ 5), у задачах оцінювання по однорідних вибірках параметрів, пов'язаних з стрибками щільності, справедливі нерівності, аналогічні (3.30-3.31), а слабка збіжність нормованих оцінок має місце при нормуванні порядку  $1/N$ . Ми отримаємо для сумішей зі змінними концентраціями оцінки, які матимуть швидкість збіжності, аналогічну (3.30-3.31).

Наша мета полягає в тому, щоб по спостереженням  $\xi_{1:N}, \dots, \xi_{N:N}$  оцінити нижню екстремальну точку першої (скажімо) компоненти суміші —  $m_{H_1}$ . Якщо для всіх  $k = 2, \dots, M$   $m_{H_1} \leq m_{H_k}$ , то оцінка  $\tilde{m}_N$  буде консистентною для  $m_{H_1}$  і збіжність зі швидкістю (3.30) буде мати місце. Але, якщо, при деякому  $k$ ,  $m_{H_1} > m_{H_k}$ , то спостереження, що відповідають  $k$ -тій компоненті, будуть заважати оцінці. Як у такому випадку оцінити нижню екстремальну точку першої компоненти?

Використовуючи означення  $m_{H_1}$ , можна запропонувати для нього оцінку  $m_N^* = \sup\{x : \hat{H}_1^N(x) < C_N\}$ , де  $\hat{H}_1^N(x)$  — зважена емпірична функція розподілу з мінімаксними коефіцієнтами для оцінки ф.р. першої компоненти,  $C_N$  — деяке число, близьке до 0. (Точніше, слід покласти  $C_N \rightarrow 0$  при  $N \rightarrow \infty$ .) Але така оцінка не може мати швидкості збіжності порядку (3.30). Дійсно, згідно з теоремою 2.2.5, процес  $\sqrt{N}(\hat{H}_1^N(x) - H_1(x))$  слабо збігається до невідродженого гаусівського процесу у рівномірній метриці при  $N \rightarrow \infty$ . Тому, щоб забезпечити виконання  $\hat{H}_1^N(x) < C_N$  при  $x < m_{H_1}$  м.н., слід обрати рівень  $C_N > 1/\sqrt{N}$ . Але в цьому випадку, щоб забезпечити  $H_1(x) > C_N$  за умови, що щільність розподілу першої компоненти обмежена, значення  $x$  має бути більшим, ніж  $m_{H_1} + \alpha/\sqrt{N}$  для деякого додатного  $\alpha$ . Тобто швидкість збіжності оцінки  $m_N^*$  не може бути кращою



ніж  $1/\sqrt{N}$ . Чи можна побудувати оцінку, яка для сумішей зі змінними концентраціями мала б швидкість збіжності порядку (3.30)?

Відповідь на це запитання позитивна, в усякому випадку, для достатньо гладеньких розподілів компонент суміші. Спочатку ми побудуємо оцінку, а потім доведемо, що вона має гарну швидкість збіжності при правильному виборі параметрів.

Ідея оцінки полягає в тому, щоб шукати стрибок щільності, який має розподіл  $H_1$  у точці  $m_{H_1}$ , за допомогою непараметричної оцінки щільності. Ми застосуємо гістограмну оцінку щільності, що використовує  $\hat{H}_1^N$  як базову оцінку функції розподілу. Зауважимо, що швидкість збіжності оцінок щільності, як правило, гірша, ніж у оцінок функцій розподілу, на яких вони ґрунтуються. (В нашому випадку це теж так, див. п. 4.4). Але стрибок щільності є настільки помітним явищем, що гістограма реагує на нього швидше, ніж регулярна оцінка  $\hat{H}_1^N$ .

Отже, нехай нам відомо, що оцінювана екстремальна точка  $m_{H_1}$  належить деякому інтервалу  $(x_L, x_R)$ ,  $-\infty < x_L < x_R < \infty$ . Розіб'ємо цей інтервал на  $K = K_N$  підінтервалів однакової довжини  $\Delta_N = (x_R - x_L)/K_N$  точками  $t_k = x_L + \Delta_N k$ ,  $k = 0, 1, \dots, K$ . Обчислимо

$$Z_k^N = N(\hat{H}_1^N(t_{k+1}) - \hat{H}_1^N(t_k)) = \sum_{j=1}^N a_{j:N}^1 \mathbb{I}\{\xi_{j:N} \in [t_k, t_{k+1})\}.$$

Нехай  $T_N$  — деякий “поріг”, тобто додатне число. Як оцінку для  $m_{H_1}$  виберемо

$$\hat{m}_N = \inf\{t_k : Z_k^N > T_N\}. \quad (3.32)$$

Інакше кажучи,  $Z_k^N$  — це висота стовпчика гістограми<sup>2</sup>, побудованого над інтервалом  $(t_k, t_{k+1}]$ . Ми знаходимо крайній стовпчик ліворуч, який перевищує поріг, і лівий його кінець приймаємо як оцінку для нижньої екстремальної точки. При правильному виборі порогу та кількості інтервалів розбиття така оцінка має потрібну нам швидкість збіжності (3.30).

Зауважимо, що вимога апіорного задання інтервалу  $(x_L, x_R)$ , якому належить оцінювана екстремальна точка, не є суттєвою. Дійсно, використовуючи “негарну” оцінку  $m_N^*$ , можна спочатку знайти інтервал, якому  $m_{H_1}$  належить з великою ймовірністю, а потім застосувати оцінку  $\hat{m}_N$  на цьому інтервалі.

<sup>2</sup>Тут ми маємо на увазі гістограму абсолютних частот, про гістограму відносних частот, як оцінку щільності, див. 4.4.

**Теорема 3.5.1** *Нехай виконуються наступні умови*

1. Для деяких відомих  $-\infty < x_L, x_R < \infty, m_{H_1} \in (x_L, x_R)$
2. Для деяких  $\alpha > 0, \delta > 0$  при всіх  $x_1, x_2 \in (m_{H_1}, m_{H_1} + \delta)$  виконується нерівність

$$|H_1(x_1) - H_1(x_2)| \geq \alpha |x_1 - x_2|.$$

3. Для всіх  $k = 1, \dots, M$  при деякому  $L < \infty$ , для всіх  $x_1, x_2 \in \mathbb{R}$  виконується нерівність

$$|H_k(x_1) - H_k(x_2)| \leq L |x_1 - x_2|.$$

4. Вагові коефіцієнти  $a_{j:N}^1$  задовольняють умову незміщеності (2.5) і для деякого  $D < \infty, |a_{j:N}^1| < D$  при всіх  $j = 1, \dots, N, N \in \mathbb{N}$ .

5. Оцінка  $\hat{m}_N$  визначена (3.32) з  $T_N = R_N \ln N, K_N = \left\lfloor \frac{N}{S_N \ln N} \right\rfloor$ , де  $[x]$  — ціла частина числа  $x$ ;  $S_N, R_N$  — числові послідовності, такі, що при  $N \rightarrow \infty, S_N \rightarrow \infty, R_N \rightarrow \infty, K_N \rightarrow \infty, R_N/S_N \rightarrow 0, R_N^2/S_N \rightarrow \infty$ .

Тоді нерівність

$$|\hat{m}_N - m_{H_1}| \leq \frac{2(x_R - x_L)}{K_N}$$

виконується м.н. при достатньо великих  $N$ .

**Зауваження.1.** Умова 2) виконується, якщо розподіл першої компоненти має щільність, відділену від 0 на інтервалі  $(m_{H_1}, m_{H_1} + \delta)$ .

2. Умова 4) виконується, якщо  $\liminf_{N \rightarrow \infty} \Gamma_N > 0$  і  $a_{j:N}^1$  — мінімаксні вагові коефіцієнти, визначені за (2.10).

3. Для довільного  $S_N \rightarrow \infty$  обравши  $R_N = S_N^{3/4}$ , отримуємо  $R_N/S_N \rightarrow 0, R_N^2/S_N \rightarrow \infty$ , отже в умові 5)  $S_N$  може прямувати до  $\infty$  як завгодно повільно.

**Доведення.** Позначимо

$$\chi_{jN}^k = a_{j:N}^1 (\mathbb{I}\{\xi_{j:N} \in [t_k, t_{k+1})\} - \mathbb{P}\{\xi_{j:N} \in [t_k, t_{k+1})\}),$$

$Y_k^N = \sum_{j=1}^N \chi_{jN}^k$ . Якщо  $K_N$  та  $T_N$  задовольняють умові 5) теореми, то

$$\sup_{1 \leq k \leq K_N} |Y_k^N| \leq T_N \quad (3.33)$$

м.н. при достатньо великих  $N$ .

Дійсно,

$$\mathbb{P}\{\sup_k Y_k^N > T_N\} \leq \sum_{k=1}^{K_N} \mathbb{P}\{Y_k^N > T_N\}. \quad (3.34)$$

Внаслідок умови 4),  $|\chi_{jN}^k| < D$  і  $\chi_{jN}^k$  є незалежними випадковими величинами. Отже виконується нерівність Прохорова (див. теорему 7.3.1):

$$\mathbb{P}\{Y_k^N > T_N\} \leq \exp\left(-\frac{T_N}{2D} \operatorname{arcsch}\left(\frac{DT_N}{2B_N}\right)\right),$$

де

$$\begin{aligned} B_N &= \sum_{j=1}^N \mathbb{E}(\chi_{jN}^k)^2 = \sum_{j=1}^N (a_{j:N}^1)^2 (\mathbb{P}\{\xi_{j:N} \in [t_k, t_{k+1})\} - (\mathbb{P}\{\xi_{j:N} \in [t_k, t_{k+1})\})^2) \\ &\leq D^2 \sum_{j=1}^N \mathbb{P}\{\xi_{j:N} \in [t_k, t_{k+1})\} \leq D^2 L N \Delta_N \leq D^2 S_N \ln N. \end{aligned}$$

Таким чином, для деякої константи  $C < \infty$ ,

$$\begin{aligned} \mathbb{P}\{Y_k^N > T_N\} &\leq \exp\left(-\frac{R_N \ln N}{2D} \operatorname{arcsch}\left(\frac{R_N}{DS_N}\right)\right) \\ &\leq C \exp\left(-\frac{R_N^2}{2D^2 S_N} \ln N\right) \end{aligned}$$

при достатньо великих  $N$  (оскільки  $R_N/S_N \rightarrow 0$  і  $\operatorname{arcsch}(z) \sim z$  при  $z \rightarrow 0$ ). Отже, продовжуючи (3.34), маємо, для деякого  $C < \infty$ ,

$$\mathbb{P}\{\sup_k Y_k^N > T_N\} \leq CN \cdot N^{-R_N^2/(2D^2 S_N)}.$$

Звідси, враховуючи  $R_N^2/S_N \rightarrow \infty$ , маємо

$$\sum_{N=1}^{\infty} \mathbb{P}\{\sup_k Y_k^N > T_N\} < \infty.$$

Отже, за лемою Бореля-Кантеллі,  $\sup_k Y_k^N < T_N$  м.н. при достатньо великих  $N$ . Аналогічно  $\sup_k (-Y_k^N) < T_N$  м.н. Нерівність (3.33) доведено.

Легко бачити, що  $Z_k^N = Y_k^N + \bar{Z}_k^N$ , де

$$\bar{Z}_k^N = \mathbb{E} Z_k^N = N \mathbb{E}(\hat{H}_1^N(t_{k+1}) - \hat{H}_1^N(t_k)).$$

Оскільки  $\hat{H}_1^N$  є незміщеною оцінкою  $H_1$ , то при  $t_{k+1} < m_{H_1}$ ,  $\bar{Z}_k^N = 0$ . Для значення  $k^*$ , такого, що  $t_{k^*-1} \leq m_{H_1} \leq t_{k^*}$ , внаслідок умови 2) отримуємо, при достатньо великих  $N$ ,

$$\bar{Z}_{k^*} \geq N\alpha|t_{k^*+1} - t_{k^*}| \geq \alpha S_N \ln N. \quad (3.35)$$

Оскільки  $R_N/S_N \rightarrow 0$ , то з (3.33) та (3.35) отримуємо при достатньо великих  $N$ ,

$$Z_{k^*}^N \geq \bar{Z}_{k^*}^N - |Y_{k^*}^N| \geq \alpha S_N \ln N - R_N \ln N \geq R_N \ln N = T_N.$$

Отже, при достатньо великих  $N$ ,  $\hat{m}_N$  дорівнює або  $t_{k^*}$ , або  $t_{k^*-1}$ . Таким чином,  $|\hat{m}_N - m_{H_1}| \leq 2\Delta_N$ .

Теорему доведено.

## Розділ 4

# Оцінювання щільностей розподілів компонент

Оцінки, що були розглянуті нами досі, мали “параметричну” швидкість збіжності порядку  $1/\sqrt{N}$  (де  $N$  — обсяг вибірки). Оцінювання щільності є типовим прикладом задач непараметричного оцінювання, для яких ця швидкість збіжності є недосяжною. У цьому розділі розглянуті оцінки щільності компонент на основі вибірки зі змінними концентраціями. Основна увага приділена зваженим ядерним оцінкам. Ми покажемо, що порядок збіжності цих оцінок є таким самим, як у випадку оцінювання за вибіркою з однаково розподілених випадкових величин.

Отримані оцінки будуть використані далі при побудові класифікаторів у розділі 6.

## 4.1 Ядерні оцінки щільності

Мабуть найбільш популярними непараметричними оцінками щільності розподілу по вибірках з незалежних, однаково розподілених одновимірних спостережень є гістограми. Друге (а може і перше) місце належить ядерним оцінкам щільності. У випадку, коли простір спостережень  $\mathcal{X} = \mathbb{R}^d$ , причому його вимірність  $d > 1$ , але не дуже велика ( $d = 2, 3$ ), ядерні оцінки безумовно займають перше місце по популярності. Це пояснюється як порівняною простотою їх обчислення, так і хорошими асимптотичними властивостями: можна показати, що при виконанні деяких, досить широких умов на оцінювану щільність та на параметри ядерних оцінок, ці оцінки є, у певному розумінні, асимптотично ефективними, точніше на них досягається локально-мінімаксна межа середньоквадратичного ризику (див. п. 5, розділу 4 у [10]).

У цьому параграфі ми введемо ядерні оцінки щільності для спостережень з суміші зі змінними концентраціями у випадку, коли спостереження належать  $\mathcal{X} = \mathbb{R}^d$  та доведемо теорему про консистентність цих оцінок у просторі  $L_1(\mathbb{R}^d)$ . Вибір саме цього простору не є випадковим — на збіжності оцінок щільності у  $L_1(\mathbb{R}^d)$  базується доведення консистентності емпірично-баєсових класифікаторів, що використовують ці оцінки (див. розділ 6).

Як і раніше, будемо вважати, що незалежні спостереження  $\xi_{j:N}$  описуються моделлю

$$P\{\xi_j \leq x\} = \sum_{k=1}^M w_{j:N}^k H_k(x), \quad (4.1)$$

де  $H_m$  — розподіл  $m$ -тої компоненти,  $w_{j:N}^m$  — її концентрація у суміші. Будемо вважати, що розподіли усіх компонент суміші мають щільності відносно  $\lambda$  — міри Лебега на  $\mathbb{R}^d$ . Позначимо ці щільності  $h_i(x) = \frac{\partial H_i}{\partial \lambda}(x)$ . Інакше кажучи, для будь-якої вимірної множини  $A \subseteq \mathbb{R}^d$ ,

$$H_i(A) = \int_A h_i(x) dx$$

(тут  $dx = \lambda(dx)$  позначає інтегрування по мірі Лебега у  $\mathbb{R}^d$ , тобто, фактично, цей інтеграл є  $d$ -кратним інтегралом).

Функцію  $K : \mathbb{R}^d \rightarrow [0, +\infty]$  будемо називати ядром, якщо  $\int_{\mathbb{R}^d} K(x) dx = 1$ . Можна сказати, що ядро — це щільність деякої ймовірнісної міри на  $\mathbb{R}^d$ .

Отже, нехай  $K$  – ядро на  $\mathbb{R}^d$ , а  $s = s_N \rightarrow 0$ ,  $N \rightarrow \infty$  деяка послідовність додатних чисел. Ядерною оцінкою з ядром  $K$  називають

$$\begin{aligned}\hat{h}_i^N(x) &= s^{-d} \frac{1}{N} \sum_{j=1}^N a_{j:N}^i K\left(\frac{x - \xi_j}{s}\right) \\ &= s^{-d} \int K\left(\frac{x - y}{s}\right) \hat{H}_i^N(dy),\end{aligned}\quad (4.2)$$

де  $a^i$  – мінімаксні вагові коефіцієнти, визначені (2.10), а  $\hat{H}_i^N$  – зважена емпірична міра з цими коефіцієнтами, тобто мінімаксна оцінка  $H_i$ .

Число  $s_N$  називають “параметром згладжування” (smoothing parameter) оцінки  $\hat{h}_i^N(x)$ . Для вибору параметра згладжування існують дві протилежні вимоги. По-перше, він повинен бути достатньо малим, оскільки чим більше  $s_N$ , тим більш гладенькою виходить оцінка, і при занадто великих  $s_N$  вона перетворюється на “майже константу”. По-друге, він повинен бути достатньо великим, оскільки при малих значеннях параметру згладжування оцінка має велику статистичну розкиданість.

Наступна теорема задає умови на параметр згладжування, при яких ядерна оцінка буде консистентною у  $L_1(\mathbb{R}^d)$ .

**Теорема 4.1.1** *Нехай*

1. *Для деякої константи  $C > 0$ ,  $\det \Gamma_N > C$ .*
2.  *$s_N \rightarrow 0$ ,  $\sqrt{\frac{\ln N}{N}} s_N^{-d} \rightarrow 0$  при  $N \rightarrow \infty$ .*

*Тоді*

$$\int_{\mathbb{R}^d} |\hat{h}_i^N(x) - h_i(x)| dx \rightarrow 0 \text{ м.н.}$$

Для доведення теореми нам буде потрібна одна лема.

Для  $x \in \mathbb{R}^d$ ,  $A, B \subseteq \mathbb{R}^d$ ,  $s \in \mathbb{R}$  будемо позначати

$$x + A = \{x + a : a \in A\},$$

$$A + B = \{a + b : a \in A, b \in B\},$$

$$sA = \{sa : a \in A\}.$$

**Лема 4.1.1** *Для довільної скінченної міри  $\nu$*

$$\int_{x \in B} \nu(x + A) dx = \int_{z \in A} \nu(z + B) dz.$$

Доведення. Якщо  $\nu$  має щільність  $h$  відносно  $\lambda$ , то

$$\begin{aligned} \int_{x \in B} \nu(x + A) dx &= \int \int \mathbb{I}\{x \in B\} \mathbb{I}\{y - x \in A\} \nu(dy) dx = \\ &= \int \int \mathbb{I}\{x \in B\} \mathbb{I}\{z \in A\} h(x + z) dz dx = \int \mathbb{I}\{z \in A\} \nu(z + B) dz. \end{aligned}$$

Перехід до  $\nu$  загального вигляду можна здійснити, апроксимуючи  $\nu$  абсолютно неперервними мірами.

Доведення теореми. Позначимо  $\varepsilon_N = \sqrt{\frac{\ln N}{N}}$ . З умови 1 випливає, що  $\sup_{j,N} |a_{j,N}^i| < C' < \infty$ . Тому, враховуючи, що клас всіх прямокутників у  $\mathbb{R}^d$  є класом Валника-Червоненкіса, за теоремою 2.2.4 маємо, що, для всіх прямокутників<sup>1</sup>  $B \subseteq \mathbb{R}^d$ ,

$$|\tilde{H}_i^N(B) - H_i(B)| \leq \Lambda \varepsilon_N \text{ м.н.}$$

$$\text{і } |\hat{H}_i^N| = \text{VAR} \hat{H}_i^N \leq C < \infty \text{ м.н.}^2$$

Фіксуємо деяке  $\delta > 0$ . Виберемо нове ядро вигляду

$$K^*(x) = \sum_{k=1}^n d_k \mathbb{I}\{x \in D_k\},$$

де  $D_i$  – обмежені прямокутники,  $0 \leq d_i \leq H < \infty$ , так, щоб  $\int |K(x) - K^*(x)| dx < \delta$ . Позначимо  $g_s(x) = s^{-d} \int K(\frac{x-y}{s}) H_i(dy)$ ,  $g_s^*$  та  $h_N^*$  визначаються аналогічно  $g_s$  і  $\hat{h}_i^N$  відповідно, з заміною  $K$  на  $K^*$ . Згідно [9],

$$\int |h_i(x) - g_s(x)| dx \rightarrow 0 \text{ при } s \rightarrow 0.$$

Оцінимо

$$J = \int |\hat{h}_i^N(x) - g_s(x)| dx \leq J_1 + J_2 + J_3,$$

де

$$J_1 = \int s^{-d} \int |K^*((x-y)/s) - K((x-y)/s)| \times |\hat{H}_i^N|(dy) dx \leq C\delta,$$

<sup>1</sup>Тут і далі ми маємо на увазі прямокутники (паралелограми, бруски) з ребрами, паралельними осям координат.

<sup>2</sup> $\text{VAR} \mu$  позначає повну варіацію заряду  $\mu$



$$J_2 = \int s^{-d} \int |K^*((x-y)/s) - K((x-y)/s)| h_i(y) dy dx \leq \delta,$$

$$J_3 = \int |h_N^*(x) - g_s^*(x)| dx \leq H s^{-d} \sum_{k=1}^n \int |H_i(x + sD_k) - \hat{H}_i^N(x + sD_k)| dx.$$

Залишилось тільки довести, що

$$J' = s^{-d} \int |H_i(x + sD_k) - \hat{H}_i^N(x + sD_k)| dx \rightarrow 0 \text{ м.н. при } N \rightarrow \infty.$$

Позначимо  $B \subset \mathbb{R}^d$  обмежений прямокутник, такий, що  $H_i(D_k + \mathbb{R}^d \setminus B) \leq \delta'$ . Тоді  $J' \leq J'_1 + J'_2 + J'_3$ , де

$$J'_1 = s^{-d} \int_B |H_i(x + sD_k) - \hat{H}_i^N(x + sD_k)| dx = s^{-d} \int_B C \varepsilon_N dx \leq \Lambda' \varepsilon_N s^{-d},$$

$$J'_2 = s^{-d} \int_{\mathbb{R}^d \setminus B} H_i(x + sD_k) dx = s^{-d} \int_{sA_i} H_i(\mathbb{R}^d \setminus B + sD_k) dx \leq \delta' \lambda(D_k)$$

якщо  $s < 1$  (за леммою 4.1.1),

$$J'_3 = s^{-d} \int_{\mathbb{R}^d \setminus B} |\hat{H}_i^N(x + sD_k)| dx \leq \delta' \lambda(D_k) + \Lambda' \varepsilon_N \lambda(D_k).$$

Об'єднуючи всі ці оцінки, маємо  $J' \leq C(\delta + \delta') + \Lambda' \varepsilon_N s^{-d}$  та  $\Lambda' \varepsilon_N s^{-d} \rightarrow 0$  при  $N \rightarrow \infty$  за умовою теореми. Внаслідок довільності  $\delta$  та  $\delta'$  отримуємо  $J \rightarrow 0$  м.н.

Теорему доведено.

Доведення теореми не дає жодних натяків на те, якою може бути швидкість збіжності ядерних оцінок до оцінюваної щільності. У [9] показано, що навіть для незалежних, однаково розподілених спостережень, ядерні оцінки щільності можуть збігатись як завгодно повільно, якщо оцінювана щільність не є гладенькою. Накладаючи різні умови на гладкість оцінюваних щільностей та на збіжність до 0 параметра згладжування можна отримувати відповідні швидкості збіжності ядерних оцінок. Приклади дослідження збіжності ядерних оцінок ми розглянемо далі у цьому розділі.

## 4.2 Асимптотична нормальність ядерних оцінок

У цьому параграфі ми розглянемо одновимірні спостереження  $\mathcal{X} = \mathbb{R}$  і покажемо, що ядерна оцінка щільності, розглянута у попередньому параграфі, є асимптотично нормальною, але зміст цього твердження буде дещо відмінним від того, як ми розуміли асимптотичну нормальність для емпіричних функцій розподілу або зважених емпіричних моментів.

По-перше, ми будемо досліджувати відхилення оцінки (ядерної) не від оцінюваної величини (щільності розподілу), а від математичного сподівання самої оцінки. У випадку зважених емпіричних функцій розподілу або лінійних оцінок моментів така відмінність була б несуттєвою: оскільки ми розглядали лише незміщені оцінки, їх математичне сподівання дорівнювало оцінюваним характеристикам. Для виправлених з.е.ф.р., адаптивних оцінок розподілів та моментів це вже не так — вони не є незміщеними. Однак відповідні теореми про асимптотичну нормальність цих оцінок показують, що асимптотично це зміщення є несуттєвим, порівняно зі статистичним розкидом оцінки навколо свого середнього значення.

У випадку оцінок щільності це не так. Виявляється, що для того, щоб досягти оптимальної швидкості збіжності цих оцінок, потрібно обирати параметр згладжування так, щоб забезпечити приблизно однаковий внесок у загальне відхилення оцінки статистичних відхилень від середнього з одного боку та зміщення середнього від оцінюваної щільності з іншого боку. Теорема про асимптотичну нормальність дозволяє досить точно оцінити розмір статистичного відхилення.

По-друге, нормування відхилення оцінки від її математичного сподівання буде нормуватись не  $\sqrt{N}$ , а  $\sqrt{Ns_N}$ , де  $s_N$  — параметр згладжування ядерної оцінки. Оскільки для розумних оцінок  $s_N \rightarrow 0$ , це означає, що отримана швидкість збіжності буде повільнішою, ніж  $\sqrt{N}$  — традиційна швидкість збіжності оцінок у параметричній статистці. Відомо, що у випадку оцінювання по незалежних однаково розподілених випадкових величинах, нормування ядерних оцінок для отримання асимптотичної нормальності повинно бути саме  $\sqrt{Ns_N}$ . Це характеризує “нерегулярність” задачі оцінювання щільності порівняно з параметричними задачами та з задачами непараметричного оцінювання моментів або функцій розподілу. Таким чином, у цьому розумінні задача оцінки щільностей компонент суміші зі змінними концентраціями є так само нерегулярною (не більше і не менше) як і ця задача у випадку однаково розподілених спостережень.

**Теорема 4.2.1** *Нехай виконуються наступні умови*

а) щільність  $h_i(x)$  неперервна і обмежена:  $h_i(x) < c$  для усіх  $i = 1 \div M$  і деякої константи  $c$ ;

б)

$$d^2 = \int_{-\infty}^{\infty} K^2(t) dt < \infty;$$

в)

$$\sigma_i^2(x) = \lim_{N \rightarrow \infty} \sum_{k=1}^M \langle (a^i)^2 (w^k)^2 \rangle_N h_k(x) < \infty;$$

г)  $\det \Gamma_N > g$  для деякого  $g > 0$  і всіх  $N \in \mathbb{N}$ ;

д)  $s_N \rightarrow 0$ ,  $Ns_N \rightarrow \infty$  при  $N \rightarrow \infty$ .

Тоді

$$\hat{h}_i^N(x) = h_i^N(x) + \frac{1}{\sqrt{Ns_N}} \zeta_N^i, \quad (4.3)$$

де  $h_i^N(x) = \mathbb{E} \hat{h}_i^N(x)$  є не випадковою функцією,  $h_i^N(x) \rightarrow h_i(x)$  при  $s_N \rightarrow 0$ , а випадкові величини  $\zeta_N^i$  є асимптотично нормальними з параметрами  $N(0, d^2 \sigma_i^2(x))$ .

**Доведення.** Спочатку розглянемо функцію  $h_i^N(x)$ .

$$\begin{aligned} h_i^N(x) &= \mathbb{E} \hat{h}_i^N(x) = \frac{1}{Ns_N} \sum_{j=1}^N a_{j:N}^i \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) = \\ &= \frac{1}{Ns_N} \sum_{j=1}^N a_{j:N}^i \sum_{k=1}^M w_{j:N}^k \int_{-\infty}^{+\infty} K \left( \frac{x - y}{s_N} \right) h_k(y) dy. \end{aligned}$$

Використавши умову незміщеності (2.5) і зробивши заміну змінної в інтегралі, одержуємо

$$h_i^N(x) = \int_{-\infty}^{+\infty} K(z) h_i(x - s_N z) dz \rightarrow h_i(x)$$

при  $s_N \rightarrow 0$ . Введемо у розгляд послідовність

$$X_{j,N}^i = \frac{1}{\sqrt{Ns_N}} a_{j:N}^i \left( K \left( \frac{x - \xi_{j:N}}{s_N} \right) - \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) \right).$$

Тоді

$$\hat{h}_i^N(x) - h_i^N(x) = \frac{1}{\sqrt{Ns_N}} \sum_{j=1}^N X_{j,N}^i; \quad \mathbb{E} X_{j,N}^i = 0.$$

Оскільки спостереження  $\xi_{j:N}$  незалежні,  $\{X_{j,N}^i\}_{j=1}^N$  являють собою послідовність незалежних випадкових величин. Для доведення теореми 4.2.1 застосуємо центральну граничну теорему з умовою Ліндеберга (теорема 7.3.8) до послідовності  $\{X_{j,N}^i\}_{j=1}^N$ . Спершу перевіримо виконання умови (7.6).

$$\begin{aligned} \text{Var } \zeta_N^i &= \sum_{j=1}^N \text{Var } X_{j,N}^i \\ &= \sum_{j=1}^N \frac{(a_{j:N}^i)^2}{Ns_N} \left( \mathbb{E} K^2 \left( \frac{x - \xi_{j:N}}{s_N} \right) - \left( \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) \right)^2 \right) \\ &= S_1^N - S_2^N \end{aligned} \quad (4.4)$$

де

$$S_1^N = \sum_{j=1}^N \frac{(a_{j:N}^i)^2}{Ns_N} \mathbb{E} K^2 \left( \frac{x - \xi_{j:N}}{s_N} \right)$$

і

$$S_2^N = - \sum_{j=1}^N \frac{(a_{j:N}^i)^2}{Ns_N} \left( \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) \right)^2$$

Розглянемо кожний доданок правої частини (4.4) окремо.

$$\begin{aligned} \frac{1}{s_N} \left( \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) \right)^2 &= \frac{1}{s_N} \left( \sum_{k=1}^M w_{j:N}^k \int_{-\infty}^{\infty} K \left( \frac{x-y}{s_N} \right) h_k(y) dy \right)^2 = \\ &= s_N \left( \sum_{k=1}^M w_{j:N}^k \int_{-\infty}^{\infty} K(z) h_k(x - s_N z) dz \right)^2. \end{aligned} \quad (4.5)$$

Величини  $w_{j:N}^k, a_{j:N}^i, h_k(x)$  є обмеженими;  $K(z)$  — щільність на  $\mathbb{R}$ , тому очевидно, що

$$S_2^N < s_N C_1,$$

де  $C_1$  — деяка константа. Отже,

$$\lim_{N \rightarrow \infty} S_2^N = 0.$$

Оцінимо  $S_1^N$ .

$$\begin{aligned} \frac{1}{s_N} \mathbb{E} K^2 \left( \frac{x - \xi_{j:N}}{s_N} \right) &= \frac{1}{s_N} \sum_{k=1}^M w_{j:N}^k \int_{-\infty}^{\infty} K^2 \left( \frac{x - y}{s_N} \right) h_k(y) dy = \\ &= \sum_{k=1}^M w_{j:N}^k \int_{-\infty}^{\infty} K^2(z) h_k(x - s_N z) dz. \end{aligned}$$

Отже,

$$\begin{aligned} S_1^N &= \sum_{j=1}^N \frac{(a_{j:N}^i)^2}{N} \sum_{k=1}^M w_{j:N}^k \int_{-\infty}^{\infty} K^2(z) (h_k(x - z s_N) - h_k(x) + h_k(x)) dz = \\ &= d^2 \langle (a^i)^2 w^k \rangle h_k(x) + \varepsilon_N, \end{aligned}$$

де

$$\varepsilon_N \leq c_2 \int_{-\infty}^{\infty} K^2(z) \sup_{1 \leq k \leq M} |h_k(x - z s_N) - h_k(x)| dz;$$

$c_2$  - деяка константа.  $\varepsilon_N \rightarrow 0$  при  $N \rightarrow \infty$ , отже,

$$\lim_{N \rightarrow \infty} S_1^N = d^2 \sigma_i(x).$$

Умова (7.6) теореми 7.3.8 перевірена. Розглянемо виконання умови Ліндеберга (7.7). З нерівності

$$(x - y)^2 \leq 2(x^2 + y^2)$$

випливає

$$(X_{j,N}^i)^2 \leq \frac{2(a_{j:N}^i)^2}{N s_N} \left( K^2 \left( \frac{x - \xi_{j:N}}{s_N} \right) + \left( \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) \right)^2 \right).$$

Тому

$$\begin{aligned} B &= \sum_{j=1}^N \mathbb{E} (X_{j,N}^i)^2 \mathbb{I}(|X_{j,N}^i| > \tau) \leq \\ &\leq \frac{2}{N s_N} \sum_{j=1}^N (a_{j:N}^i)^2 \mathbb{E} K^2 \left( \frac{x - \xi_{j:N}}{s_N} \right) \mathbb{I}\{|X_{j,N}^i| > \tau\} + \end{aligned}$$

$$\begin{aligned}
& + \frac{2}{N s_N} \sum_{j=1}^N (a_{j:N}^i)^2 \left( \mathbb{E} K^2 \left( \frac{x - \xi_{j:N}}{s_N} \right) \right)^2 P\{|X_{j:N}^i| > \tau\} = \\
& = S_3 + S_4.
\end{aligned}$$

З того, що  $a_{j:N}^i$  обмежені і виконується співвідношення (4.5), випливає, що  $S_4 \rightarrow 0$  при  $N \rightarrow \infty$ . Розглянемо доданок  $S_3$ .

$$\begin{aligned}
S_3 &= \frac{2}{N s_N} \sum_{j=1}^N (a_{j:N}^i)^2 \mathbb{E} K^2 \left( \frac{x - \xi_{j:N}}{s_N} \right) \\
& \times \mathbb{I}\left\{ \left| K \left( \frac{x - \xi_{j:N}}{s_N} \right) - \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) \right| > \frac{\tau \sqrt{N s_N}}{a_{j:N}^i} \right\} = \\
&= \frac{2}{N} \sum_{j=1}^N (a_{j:N}^i)^2 \sum_{k=1}^M w_{j:N}^k \int_{B_{N,\tau}} K^2(z) h_k(x - z s_N) dz, \tag{4.6}
\end{aligned}$$

де

$$B_{N,\tau} = \left\{ z : \left| K(z) - \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) \right| > \frac{\tau \sqrt{N s_N}}{a_{j:N}^i} \right\}.$$

Щільності  $h_k(x) < c$  для усіх  $k = 1 \div M$ , отже,

$$\int_{B_{N,\tau}} K^2(z) h_k(x - z s_N) dz \leq c \int_{B_{N,\tau}} K^2(z) dz. \tag{4.7}$$

Враховуючи, що

$$\mathbb{E} K \left( \frac{x - \xi_j^N}{s_N} \right) = O(s_N)$$

при  $N \rightarrow \infty$ , і завдяки умові (б) теореми, можемо записати

$$\int_{A_{N,\tau}} K^2(z) dz \rightarrow 0,$$

де  $A_{N,\tau} = \{z : K(z) > C_N\}$ , а  $C_N \rightarrow \pm\infty$ . Отже,  $S_3 \rightarrow 0$  при  $N \rightarrow \infty$ , для довільної сталої  $\tau$ , умова Ліндеберга виконується.

Теорема 4.2.1 доведена.

### 4.3 Вибір параметра згладжування

Виникає питання про оптимальний вибір  $s_N$  і  $K(z)$ . Для того, щоб оцінити якість оцінки  $\hat{h}_i^N(x)$  “в середньому” і вибрати найкраще значення  $s_N$ , традиційно використаємо  $L_2$ -підхід. Ми будемо мінімізувати по  $s_N$  проінтегровану середньоквадратичну похибку (mean integrated squared error, MISE)

$$J_N = \int_{-\infty}^{\infty} \mathbb{E} \left( \hat{h}_i^N(x) - h_i(x) \right)^2 dx. \quad (11)$$

Припустимо, що виконуються умови

а)  $h_i(x) > 0$  лише на скінченному інтервалі;

б) існує  $(h_i(x))'' = \bar{h}_i(x)$  друга похідна щільності  $i$ -тої компоненти суміші

та  $\phi_i = \int_{-\infty}^{\infty} (\bar{h}_i(x))^2 dx < \infty$ ;

в)  $\int_{-\infty}^{\infty} zK(z)dz = 0$ ;

г)  $D^2 = \int_{-\infty}^{\infty} z^2 K(z)dz < \infty$ ;

д)  $A_N = \frac{1}{N} \sum_{j=1}^N (a_{j:N}^i)^2 \rightarrow A$  при  $N \rightarrow \infty$ , де  $A$  - деяка константа.

**Теорема 4.3.1** В умовах (а) - (д) виконується співвідношення

$$\liminf_{N \rightarrow \infty} N^{\frac{4}{5}} J_N \geq \frac{5}{4} (DA d^2)^{\frac{4}{5}},$$

причому рівність досягається при

$$s_N = \left( \frac{Ad^2}{ND^4\phi_i} \right)^{\frac{1}{5}}. \quad (4.8)$$

**Доведення.** Перетворимо  $J_N$  таким чином

$$J_N = \int_{-\infty}^{\infty} E \left( \hat{h}_i^N(x) - h_i^N(x) \right)^2 dx + \int_{-\infty}^{\infty} (h_i^N(x) - h_i(x))^2 dx = B_1^N + B_2^N.$$

Відомо, що

$$B_2^N = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} K(z) h_i(x - zs_N) dz - h_i(x) \right)^2 dx = \left( \frac{D^2 s_N^2}{2} \right)^2 \phi_i + o(s_N^4). \quad (4.9)$$

(Доведення цього факту є, наприклад, у [27], с. 132.) Оцінимо  $B_1^N$ .

$$\mathbb{E} \left( \hat{h}_i^N(x) - h_i^N(x) \right)^2 =$$

$$= \sum_{j=1}^N \frac{(a_{j:N}^i)^2}{Ns_N} \left( \mathbb{E} K^2 \left( \frac{x - \xi_{j:N}}{s_N} \right) - \left( \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) \right)^2 \right) \quad (4.10)$$

Знайдемо інтеграл від кожного доданку (4.10) окремо.

$$\begin{aligned} & \frac{1}{s_N} \int_{-\infty}^{\infty} \mathbb{E} K^2 \left( \frac{x - \xi_{j:N}}{s_N} \right) dx \\ &= \sum_{k=1}^M w_{j:N}^k \int_{-\infty}^{\infty} K^2(z) \left[ \int_{-\infty}^{\infty} h_k(x - zs_N) dx \right] dz = d^2. \end{aligned} \quad (4.11)$$

Розглянемо другий доданок у (4.10).

$$\frac{1}{s_N} \left( \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) \right)^2 = s_N \left( \sum_{k=1}^M w_{j:N}^k \int_{-\infty}^{\infty} K(z) h_k(x - zs_N) dz \right)^2.$$

Оскільки  $h_k(x)$  обмежена для усіх  $k$ ,  $h_k(x) > 0$  лише на скінченному інтервалі, то

$$\frac{1}{s_N} \int_{-\infty}^{\infty} \left( \mathbb{E} K \left( \frac{x - \xi_{j:N}}{s_N} \right) \right)^2 dx = O(s_N) \quad (4.12)$$

Враховуючи (4.10-4.12), отримуємо

$$B_1^N = \frac{1}{Ns_N} \frac{1}{N} \sum_{j=1}^N (a_{j:N}^i)^2 (d^2 + O(s_N)) \simeq \frac{Ad^2}{Ns_N} + O\left(\frac{A}{N}\right) \quad (4.13)$$

Підставимо (4.9) і (4.13) у вираз  $J_N$ .

$$J_N = \frac{D^4 \phi_i}{4} s_N^4 + \frac{Ad^2}{Ns_N} + o((s_N)^4) \quad (4.14)$$

Мінімум головної частини (4.14) досягається, коли  $s_N$  задається формулою (4.8). Підставляючи (4.8) у (4.8), одержуємо твердження теореми 4.3.1.

Ми одержали вираз для  $J_N$ , що характеризує збіжність у середньоквадратичному ядерної оцінки щільності компоненти суміші до щільності цієї компоненти. Але  $J_N$  містить параметр  $\phi_i$ , який залежить від щільності  $h_i(x)$ . Його потрібно оцінити за даними спостережень  $\{\xi_{j:N}\}_{j=1}^N$ . Нехай  $\{\kappa_N\}$  - послідовність додатних чисел, таких, що  $\kappa_N \rightarrow 0$  при  $N \rightarrow \infty$ , причому  $\kappa_N = O(N^{-\beta})$ , де  $0 < \beta < \frac{1}{10}$ . За оцінку  $\phi_i$  приймемо

$$\hat{\phi}_i^N = \int_{-\infty}^{\infty} \left( \left( \hat{h}_i^N(x, \kappa_N) \right)'' \right)^2 dx,$$



де

$$\hat{h}_i^N(x, \kappa) = \hat{h}_i^N(x) = \frac{1}{N\kappa} \sum_{j=1}^N a_{j:N}^i K\left(\frac{x - \xi_j}{\kappa}\right)$$

— ядерна оцінка щільності з параметром згладжування  $\kappa$ . Легко бачити, що

$$\hat{\phi}_i^N = \frac{1}{N^2 \kappa_N^6} \int_{-\infty}^{\infty} \left( \sum_{j=1}^N a_{j:N}^i \bar{K}\left(\frac{x - \xi_{j:N}}{\kappa_N}\right) \right)^2 dx, \quad (4.15)$$

тут  $\bar{K}(z) = K''(z)$  - друга похідна  $K(z)$ . Припустимо, що виконуються наступні умови

а)  $\bar{h}_i(x)$  є неперервною обмеженою функцією, що належить  $L_2$ ;

б)  $K(x)$  - симетричне ядро на  $\mathbb{R}$ ;  $K'(x) \rightarrow 0$  при  $x \rightarrow \pm\infty$ ;

в)

$$\int_{-\infty}^{\infty} x K(x) dx = 0; \quad \int_{-\infty}^{\infty} x^2 |\bar{K}(x)| dx < \infty.$$

**Теорема 4.3.2** В умовах (а) - (в) для довільного цілого  $m > 0$  оцінка (4.15) задовольняє співвідношення

$$E \left( \hat{\phi}_i - \phi_i \right)^{2m} = O \left( \left( N^{-(1-10\beta)} \right)^m \right) \quad (4.16)$$

**Доведення.** Позначимо

$$\mu_N^i(x) = \int_{-\infty}^{\infty} K(z) \bar{h}_i(x - z\kappa_N) dz$$

Скористаємося наступною лемою [6, с. 133].

**Лема 4.3.1** У припущеннях (а) - (в) виконується

$$\lim_{N \rightarrow \infty} \int_{-\infty}^{\infty} (\mu_N^i(x))^2 dx = \int_{-\infty}^{\infty} (\bar{h}_i(x))^2 dx = \phi_i.$$

Отже, нам залишається довести замість (4.16) твердження

$$\alpha = E \left( \hat{\phi}_i - \int_{-\infty}^{\infty} (\mu_N^i(x))^2 dx \right)^{2m} = O \left( N^{-(1-10\beta)m} \right). \quad (4.17)$$

Введемо позначення

$$K^*(u) = \int_{-\infty}^{\infty} \bar{K}(v) \bar{K}(u-v) dv; \quad T_N(u) = \frac{1}{\kappa_N} \int_{-\infty}^{\infty} K^*\left(\frac{u-v}{\kappa_N}\right) h_i(v) dv$$

Неважко пересвідчитися, що

$$\begin{aligned} \int_{-\infty}^{\infty} (\mu_N^i(x))^2 dx &= \frac{1}{\kappa_N^4} \int_{-\infty}^{\infty} h_i(u) T_N(u) du = \\ &= \frac{1}{N \kappa_N^4} \mathbb{E} \sum_{j=1}^N a_{j:N}^i T_N(\xi_{j:N}). \end{aligned} \quad (4.18)$$

Окрім того,

$$\begin{aligned} \hat{\phi}_i &= \frac{1}{N^2 \kappa_N^6} \int_{-\infty}^{\infty} \left( \sum_{j=1}^N a_{j:N}^i \bar{K}\left(\frac{x - \xi_{j:N}}{\kappa_N}\right) \right)^2 dx = \\ &= \frac{1}{N^2 \kappa_N^5} \sum_{j=1}^N \sum_{l=1}^N a_{j:N}^i a_{l:N}^i K^*\left(\frac{\xi_{l:N} - \xi_{j:N}}{\kappa_N}\right). \end{aligned} \quad (4.19)$$

Підставимо (4.18), (4.19) у ліву частину (4.17):

$$\begin{aligned} \alpha &= \mathbb{E} \left( \frac{1}{N^2 \kappa_N^5} \sum_{j=1}^N \sum_{l=1}^N a_{j:N}^i a_{l:N}^i K^*\left(\frac{\xi_{l:N} - \xi_{j:N}}{\kappa_N}\right) \right. \\ &\quad \left. - \frac{1}{N \kappa_N^4} \mathbb{E} \sum_{j=1}^N a_{j:N}^i T_N(\xi_{j:N}) \right)^{2m} = \\ &= \frac{1}{\kappa_N^{8m}} \mathbb{E} \left( \frac{1}{N^2} \sum_{j=1}^N \sum_{l=1}^N a_{j:N}^i a_{l:N}^i \frac{1}{\kappa_N} K^*\left(\frac{\xi_{l:N} - \xi_{j:N}}{\kappa_N}\right) - \right. \\ &\quad \left. - \frac{1}{N} \sum_{j=1}^N a_{j:N}^i T_N(\xi_{j:N}) + \frac{1}{N} \sum_{j=1}^N a_{j:N}^i T_N(\xi_{j:N}) - \frac{1}{N} \sum_{j=1}^N a_{j:N}^i \mathbb{E} T_N(\xi_{j:N}) \right)^{2m} \end{aligned}$$

Звідси, відповідно нерівності

$$\left| \sum_{k=1}^m z_k \right|^p \leq m^{p-1} \sum_{k=1}^m |z_k|^p,$$

де  $p > 0$  — ціле число, знаходимо

$$\alpha \leq 2^{2m-1} \kappa_N^{-8m} (E_1^N + E_2^N), \quad (4.20)$$

де

$$E_1^N = \mathbb{E} \left( \frac{1}{N^2} \sum_{j=1}^N \sum_{l=1}^N a_{j:N}^i a_{l:N}^i \frac{1}{\kappa_N} K^* \left( \frac{\xi_{l:N} - \xi_{j:N}}{\kappa_N} \right) - \frac{1}{N} \sum_{j=1}^N a_{j:N}^i T_N(\xi_{j:N}) \right)^{2m};$$

$$E_2^N = \frac{1}{N^{2m}} \mathbb{E} \left( \sum_{j=1}^N a_{j:N}^i (T_N(\xi_{j:N}) - \mathbb{E} T_N(\xi_{j:N})) \right)^{2m}.$$

Для оцінки  $E_2^N$  скористаємося наступним твердженням.

**Теорема 4.3.3** [28], с. 89. Нехай  $x_1, x_2, \dots, x_N$  — незалежні випадкові величини;  $\mathbb{E} x_k = 0$  для усіх  $k = 1 \div n$ . Тоді

$$\mathbb{E} |S_N|^p \leq C(p) n^{\frac{p}{2}-1} \sum_{k=1}^n \mathbb{E} |x_k|^p;$$

тут  $S_n = \sum_{i=1}^n x_i$ ;  $p \geq 2$ ;  $C(p)$  — константа, що залежить лише від  $p$ .

Отже,

$$E_2^N \leq \frac{c}{N^{2m}} \sum_{j=1}^N (a_{j:N}^i)^{2m} \mathbb{E} (T_N(\xi_{j:N}) - \mathbb{E} T_N(\xi_{j:N}))^{2m} N^{m-1}.$$

(Тут і далі через  $c, c_1, c_2, \dots$  позначаємо різні константи.) З того, що

$$T_N(x) \leq \frac{c_1}{\kappa_N}, \quad (4.21)$$

випливає

$$E_2^N = \frac{cc_2}{N^{m+1}(\kappa_N)^{2m}} \sum_{j=1}^N (a_{j:N}^i)^{2m} \leq \frac{c_3}{(N\kappa_N^2)^m};$$

$$E_2^N = O \left( \frac{1}{(N\kappa_N^2)^m} \right). \quad (4.22)$$

Тепер розглянемо  $E_1^N$ . Функція  $K^*(u)$  обмежена і виконується (4.21), тому можна окремо оцінити доданок з  $E_1^N$ , для якого  $l = j$ .

$$E_1^N \leq \frac{c_4}{(N\kappa_N)^{2m}} + \mathbb{E} \left( \frac{1}{N} \sum_{j=1}^N a_{j:N}^i \left( \frac{1}{N} \sum_{l=1; l \neq j}^N a_{l:N}^i \frac{1}{\kappa_N} K^* \left( \frac{\xi_{l:N} - \xi_{j:N}}{\kappa_N} \right) - T_N(\xi_{j:N}) \right) \right)^{2m}.$$

Скористаємося нерівністю Єнсена:

$$\begin{aligned} E_1^N &\leq \frac{c_4}{(N\kappa_N)^{2m}} \\ &+ \frac{1}{N} \sum_{j=1}^N (a_{j:N}^i)^{2m} \mathbb{E} \left( \frac{1}{N} \sum_{l=1; l \neq j}^N a_{l:N}^i \frac{1}{\kappa_N} K^* \left( \frac{\xi_{l:N} - \xi_{j:N}}{\kappa_N} \right) - T_N(\xi_{j:N}) \right)^{2m} \\ &= \frac{c_4}{(N\kappa_N)^{2m}} + E_3^N \end{aligned} \quad (4.23)$$

де

$$\begin{aligned} E_3^N &= \frac{1}{N} \sum_{j=1}^N (a_{j:N}^i)^{2m} \mathbb{E} (\gamma(\xi_{j:N})), \\ \gamma(y) &= \mathbb{E} \left[ \left( \frac{1}{N} \sum_{l=1; l \neq j}^N a_{l:N}^i \frac{1}{\kappa_N} K^* \left( \frac{\xi_{l:N} - \xi_{j:N}}{\kappa_N} \right) - T_N(\xi_{j:N}) \right)^{2m} \middle| \xi_{j:N} = y \right] \\ &= E \left( \frac{1}{N} \sum_{l=1; l \neq j}^N a_{l:N}^i \frac{1}{\kappa_N} K^* \left( \frac{\xi_{l:N} - y}{\kappa_N} \right) - T_N(y) \right)^{2m}. \end{aligned}$$

Таким чином, нам необхідно оцінити  $E_3^N$ . Представимо

$$T_N(y) = \frac{1}{N} \sum_{l=1}^N a_{l:N}^i \frac{1}{\kappa_N} \sum_{k=1}^M w_{l:N}^k \int_{-\infty}^{\infty} K^* \left( \frac{u - y}{\kappa_N} \right) h_k(u) du$$

і позначимо

$$V_N(\xi_{l:N}) = a_{l:N}^i \frac{1}{\kappa_N} \left( K^* \left( \frac{\xi_{l:N} - y}{\kappa_N} \right) - \sum_{k=1}^M w_{l:N}^k \int_{-\infty}^{\infty} K^* \left( \frac{u - y}{\kappa_N} \right) h_k(u) du \right)$$

Тоді

$$\begin{aligned} \gamma(y) \leq & c_5 \left( \mathbb{E} \left( \frac{1}{N} \sum_{l=1; l \neq j} a_{l:N}^i \frac{1}{\kappa_N} (K^* \left( \frac{\xi_{l:N} - y}{\kappa_N} \right)) \right. \right. \\ & \left. \left. - \sum_{k=1}^M w_{l:N}^k \int_{-\infty}^{\infty} K^* \left( \frac{u - y}{\kappa_N} \right) h_k(u) du \right)^{2m} \right. \\ & \left. + \left( \frac{1}{N} a_{j:N}^i \frac{1}{\kappa_N} \sum_{k=1}^M w_{j:N}^k \int_{-\infty}^{\infty} K^* \left( \frac{u - y}{\kappa_N} \right) h_k(u) du \right)^{2m} \right) \end{aligned} \quad (4.24)$$

Другий доданок у (4.24) не перевищує  $(c_6/(N\kappa_N))^{2m}$ . Таким чином,

$$\gamma(y) \leq c_5 \mathbb{E} \left( \frac{1}{N} \sum_{l=1; l \neq j}^N V_N(\xi_{l:N}) \right)^{2m} + \frac{c_6}{(N\kappa_N)^{2m}}.$$

$V_N(\xi_{l:N})$  являє собою послідовність незалежних випадкових величин, причому  $\mathbb{E} V_N(\xi_{l:N}) = 0$  для усіх  $l$ . Застосуємо до цієї послідовності теорему 4.3.3:

$$\begin{aligned} \mathbb{E} \left( \sum_{l=1; l \neq j}^N V_N(\xi_{l:N}) \right)^{2m} & \leq c_7 N^{m-1} \sum_{l=1; l \neq j}^N (\mathbb{E} V_N(\xi_{l:N}))^{2m} \\ & \leq c_8 N^{m-1} \frac{N-1}{\kappa_N^{2m}} < c_8 \frac{N^m}{\kappa_N^{2m}} \end{aligned}$$

Отже, для довільного  $y$

$$\gamma(y) \leq \frac{c_6}{(N\kappa_N)^{2m}} + \frac{c_8}{(N\kappa_N^2)^m}$$

Підсумовуючи оцінки (4.23) і (4.24), одержуємо

$$E_1^N \leq \frac{c_4}{(N\kappa_N)^{2m}} + \frac{c_6}{(N\kappa_N)^{2m}} + \frac{c_8}{(N\kappa_N^2)^m} = O \left( \left( \frac{1}{N\kappa_N^2} \right)^m \right) \quad (4.25)$$

Підставляючи (4.22) і (4.25) у (4.20), отримуємо (4.17).

Теорема доведена.

Відмітимо, що асимптотика вектора ядерних оцінок щільностей компонент суміші зі змінними концентраціями, а також асимптотичні властивості похідних від цих щільностей досліджені в роботі [13]. Зокрема, в

подальшому розгляді з цього джерела нам будуть потрібні наступні результати, які ми наводимо без доведення.

Позначимо похідну від ядерної оцінки щільності

$$\hat{h}'_{k,N}(x) = \frac{1}{N s_N^2} \sum_{j=1}^N a_{j:N}^k K' \left( \frac{x - \xi_{j:N}}{s_N} \right), \quad 1 \leq k \leq M.$$

**Теорема 4.3.4** *Нехай виконуються наступні умови.*

- (i) щільності  $h_k(x)$  є неперервними та обмеженими деякою сталою:  $\exists c > 0 : h_k(x) < c, 1 \leq k \leq M$ ;
- (ii) існують границі

$$\sigma_k^2(x) = \lim_{N \rightarrow \infty} \sum_{r=1}^M \langle (a^k)^2 w^r \rangle_N h_r(x) < \infty; \quad 1 \leq k \leq M;$$

- (iii) існують  $h'_k(x)$ , які є обмеженими деякою сталою  $\exists c_1 > 0 : h'_k(x) < c_1, 1 \leq k \leq M$ ;
- (iv)  $d_2^2 = \int_{-\infty}^{\infty} (K'(z))^2 dz < \infty$ .

Тоді

$$\hat{h}'_{k,N}(x) = \bar{h}'_{k,N}(x) + \frac{1}{\sqrt{N s_N^2}} \zeta_N^k, \quad (4.26)$$

де  $\bar{h}'_{k,N}(x) = E \hat{h}'_{k,N}(x)$  є не випадковими функціями,  $\bar{h}'_{k,N}(x) \rightarrow f'_k(x)$  при  $s_N \rightarrow 0$ , а випадкові величини  $\zeta_N^k$  є асимптотично нормальними з параметрами  $N(0, d_2^2 \sigma_k^2(x))$ .

Позначимо  $\vec{\zeta}_N = N^{2/5} (\hat{h}_i^N(x) - h_i(x))_{i=1}^d$ .

**Теорема 4.3.5** *Нехай виконуються наступні умови*

- (i) Функції  $h_k(x)$  та їх другі похідні  $h_k''(x)$  є неперервними та обмеженими для  $x \in \mathbb{R}$ .
- (ii) Існують границі  $\sigma_{km}^2(x) = \lim_{N \rightarrow \infty} \sum_{r=1}^M \langle a^k a^m w^r \rangle_N h_r(x)$ .
- (iii)  $d^2 = \int_{-\infty}^{\infty} K^2(z) dz < \infty$ ,  $\int_{-\infty}^{\infty} z K(z) dz = 0$ ,  $D^2 = \int_{-\infty}^{\infty} z^2 K(z) dz < \infty$ .
- (iv)  $s_N = c N^{-1/5}$  де  $0 < c < \infty$  — деяка константа.

Тоді  $\zeta_N \Rightarrow \zeta$  при  $N \rightarrow \infty$ , де  $\zeta$  — гауссів випадковий вектор з математичним сподіванням  $D^2 c^{2/5} \vec{h}''(x)/2$  і коваріаційною матрицею  $\frac{d^2}{c^{1/5}} G(x)$ ,  $\vec{h}''(x) = (h_1''(x), \dots, h_M''(x))^T$ ,  $G(x) = (\sigma_{km}^2(x))_{k,m=1}^M$ .

## 4.4 Неядерні оцінки щільностей розподілів

Розглянуті у попередніх параграфах ядерні оцінки щільностей компонент суміші мають ряд хороших властивостей: вони порівняно просто обчислюються при помірних обсягах вибірки, забезпечують оптимальну швидкість збіжності до оцінюваної щільності, їх легко використовувати для багатовимірних даних. Однак у деяких задачах виявляються більш зручними інші підходи до оцінювання щільностей. У цьому параграфі представлено короткий огляд таких підходів. Оскільки навіть для непараметричного оцінювання щільностей за однаково розподіленими спостереженнями існує велетенський обсяг літератури, що включає у себе чимало монографій, наш огляд ніяк не може претендувати на вичерпність.

**Гістограмні оцінки.** Гістограми абсолютних та відносних частот є настільки поширеним і загальновизнаним засобом візуального аналізу даних, що багато дослідників, які використовують статистичні методи на практиці, сприймають щільність розподілу як “граничний” або “теоретичний” аналог гістограми. Тому розробка версій гістограмних оцінок для вибірки з суміші зі змінними концентраціями потрібна хоча б для того, щоб дати таким дослідникам можливість порівнювати гістограми, отримані за однорідними даними, з результатами аналізу сумішей.

Нагадаємо спочатку техніку побудови гістограм для однорідної вибірки. Нехай вибірка складається з  $\zeta_1, \dots, \zeta_N$  — одновимірних спостережень з одним і тим самим розподілом. Для побудови гістограми потрібно обрати скінченний інтервал, на якому зосереджений цей розподіл —  $A = [b, c]$  та кількість  $K$  маленьких “підінтервалів”, на які буде розбито цей інтервал. У класичному варіанті для побудови гістограми використовують рівномірне розбиття  $A$  на підінтервали  $A_k = [t_{k-1}, t_k)$ ,  $k = 1, \dots, K-1$ ,  $A_K = [t_{K-1}, t_K]$  де  $t_k = b + k\delta$ ,  $\delta = (c - b)/K$  — ширина підінтервалів.

Після цього підраховують “абсолютні частоти”  $n_k$  інтервалів  $A_k$  у вибірці:  $n_k = \sum_{j=1}^N \mathbb{I}\{\zeta_j \in A_k\}$ . Інакше кажучи,  $n_k$  — це кількість елементів вибірки, які потрапили у  $k$ -тий підінтервал. Гістограма абсолютних частот складається зі стовпчиків, основами яких є  $A_k$ , а висоти дорівнюють  $n_k$ . У гістограмі абсолютних частот висоту  $k$ -того стовпчика обирають рівною  $f_k = n_k/(N\delta)$ . Оскільки відносні частоти (або просто “частоти”) інтервалів визначаються як  $\nu_k = n_k/N$ , то  $f_k = \nu_k/\delta$ .

Очевидно, що за формою гістограми абсолютних та відносних частот цілком однакові і відрізняються лише масштабом вертикальної осі. Оцінкою для щільності розподілу є гістограма відносних частот, але гістогра-

ми абсолютних частот бувають зручнішими при використанні, наприклад, для аналізу викидів. Надалі під гістограмою ми будемо розуміти саме гістограму відносних частот.

Точніше, формально гістограму відносних частот можна визначити як графік функції

$$\hat{f}(x) = \begin{cases} f_k & \text{якщо } x \in A_k \\ 0 & \text{у іншому випадку} \end{cases} = \sum_{k=1}^K f_k \mathbb{I}\{x \in A_k\}.$$

Саму цю функцію також інколи називають гістограмою або гістограмною оцінкою щільності розподілу [9]. При правильному виборі параметрів (інтервалу-носія гістограми  $A$  та кількості підінтервалів  $K$ ) в залежності від обсягу вибірки, гістограма є консистентною оцінкою щільності розподілу  $f$  за вибіркою з незалежних однаково розподілених випадкових величин.

Щоб створити аналог гістограми для оцінювання щільності  $m$ -тої компоненти по спостереженнях  $\xi_{j:N}$ ,  $j = 1, \dots, N$  з суміші зі змінними концентраціями, досить задати аналоги відносних частот  $\nu_k$ . Оскільки  $\nu_k$  є насправді оцінками для ймовірностей  $P\{\zeta_j \in A_k\} = \int_{A_k} f(x)dx$ , то у випадку суміші їх можна замінити на  $\hat{H}_m^N(A_k) = \hat{H}_m^N(t_k) - \hat{H}_m^N(t_{k-1})$ , де  $\hat{H}_m^N$  — оцінка для функції розподілу  $m$ -тої компоненти  $H_m$  по спостереженнях з суміші. На роль  $\hat{H}_m^N$  можна обрати

- (i) мінімаксну оцінку  $\hat{H}_m^N$  з п. 2.1;
- (ii) яку-небудь версію виправленої зваженої емпіричної функції розподілу  $F^*(x, a^m)$  з розглянутих у п. 2.3;
- (iii) оцінку, отриману за допомогою групування (дискретизації) даних і застосування асимптотично ефективної техніки оцінювання з п. 2.4.

У варіанті (i) отримуємо наступну оцінку щільності  $h_m$ :

$$\tilde{h}_m^N = \sum_{k=1}^K \left( \hat{H}_m^N(A_k) \mathbb{I}\{x \in A_k\} \right) = \sum_{k=1}^K \sum_{j=1}^N a_{j:N}^m \mathbb{I}\{x \in A_k, \xi_{j:N} \in A_k\}, \quad (4.27)$$

де  $a_{j:N}^m$  — мінімаксні вагові коефіцієнти з (2.10).

Варіант (ii) відрізняється від (i) лише тим, що замість вагових коефіцієнтів мінімаксної оцінки у (4.27) використовуються коефіцієнти  $b_j^*$  відповідної виправленої зваженої емпіричної ф.р. описані у п. 2.3.

Нарешті, при використанні підходу (iii), ми “групуємо” спостереження, тобто заміняємо кожне  $\xi_{j:N}$  величиною  $\zeta_{j:N}$ , яка дорівнює номеру того



підінтервалу, на який потрапляє  $\xi_{j:N}$ :  $\zeta_{j:N} = \sum_{k=1}^K k \mathbb{1}\{\xi_{j:N} \in A_k\}$ . Якщо інтервали  $A_k$  — не випадкові (тобто вибрані незалежно від спостережень), то розподіл  $\zeta_{j:N}$  описується формулою (2.49):

$$\mathbb{P}\{\zeta_{j:N} = k\} = \sum_{m=1}^M w_j^m H_{(k,m)}, \quad (4.28)$$

де<sup>3</sup>  $H_{(k,m)} = H_m(A_k) = H_m(t_k) - H_m(t_{k-1})$ . Оцінки  $\check{H}_{(k,m)}^N$  для  $H_{(k,m)}$ , отримані у п. 2.4, підставляються у формулу для гістограми замість відносних частот. Оскільки ці оцінки також мають вигляд сум вагових коефіцієнтів, помножених на індикатори, загальний вигляд оцінки залишається подібним до (4.27), однак є і суттєва відмінність: при розрахунку  $\hat{H}_m^N(A_k)$  підсумовуються лише вагові коефіцієнти  $a_{j:N}^m$ , що відповідають спостереженням, які потрапили на підінтервал  $A_k$ . У формулі для  $\check{H}_{(k,m)}^N$  використовуються коефіцієнти, пов'язані з усіма спостереженнями і усіма підінтервалами, на які розбивається носій гістограми  $A$ . Врахування цієї інформації необхідне для отримання ефективної оцінки.

Нажаль, як показують дослідження на модельованих даних, переваги адаптивних оцінок стають помітними лише при достатньо великих обсягах вибірки (потрібні сотні спостережень на один підінтервал розбиття), що робить їх незручними для використання у гістограмних оцінках, де ми, як правило, намагаємось зробити підінтервали якомога вузкими, щоб помітити особливості оцінюваної щільності.

Точніше, для вибору ширини підінтервалів  $\delta$  (або, відповідно, їх кількості  $K$ ) ми маємо дві протилежні рекомендації: з одного боку,  $\delta$  має бути великим, щоб на кожен підінтервал  $A_k$  потрапляла достатня кількість спостережень для оцінки  $\mathbb{P}\{A_k\}$ , з іншого —  $\delta$  повинно бути досить малим, щоб оцінювана щільність не дуже сильно змінювалась всередині підінтервалів. Фактично, у гістограмних оцінках  $\delta$  (точніше  $(c - b)/K$ , оскільки вибирають саме  $K$ ) відіграє роль параметра згладжування аналогічного параметру  $s$  для ядерних оцінок.

Ми не будемо розглядати тут загальні теореми про вибір кількості підінтервалів в залежності від обсягу вибірки та швидкості збіжності оцінки. Прикладом тверджень такого роду є лема 5.2.3 у наступному розділі.

Значення  $K$  можна намагатись обирати адаптивно, виходячи з деякого функціонала якості оцінки, методами, подібними розглянутому у п. 4.3.

<sup>3</sup>Оскільки ми оцінюємо щільність розподілу  $H_m$ , то його слід вважати неперервним, тому  $H_m([t_{k-1}, t_k]) = H_m([t_{k-1}, t_k))$ , тобто з відкритими/замкненими кінцями інтервалів проблем не виникає.

Однак можливий і інший підхід, при якому ширина кожного інтервалу розбиття регулюється окремо, щоб забезпечити адаптацію до локальних особливостей оцінюваної щільності.

Простіший приклад такої адаптації полягає в тому, щоб розбивати досліджуваний інтервал (емпіричними) квантилями розподілу оцінюваної компоненти. На роль таких емпіричних квантилів можна використати оцінки  $\hat{Q}_N^m(\alpha)$ , визначені (3.25). Таким чином, алгоритм побудови оцінки наступний:

1. Вибираємо кількість інтервалів  $K$  та носій оцінки  $[b, c]$ .
2. Задаємо точки розбиття інтервалу  $[b, c]$ :  $t_0 = b$ ,  $t_K = c$ ,  $t_k = Q_m^N(k/K)$  для  $k = 1, \dots, K-1$  та інтервали розбиття  $A_k = [t_{k-1}, t_k)$ ,  $k = 1, \dots, K-1$ ,  $A_K = [t_{K-1}, t_K]$ .
3. Визначаємо

$$f_k = \frac{1}{K(t_k - t_{k-1})}$$

4. Задаємо оцінку для  $h_m$  формулою

$$\hat{h}_m^N(x) = \sum_{k=1}^K f_k \mathbb{1}\{x \in A_k\}.$$

Тут коментарів вимагає лише пункт 3: насправді величина  $f_k$  повинна бути оцінкою відношення ймовірності того, що випадкова величина потрапить на інтервал  $A_k$ , до довжини цього інтервалу. Але якщо інтервал утворений квантилями рівнів  $(k-1)/K$  та  $k/K$ , то ймовірність потрапити на нього дорівнює  $1/K$ .

Для випадку незалежних однаково розподілених спостережень аналогічна оцінка зветься поліграмою [68]. Вона не набула великого поширення через те, що оцінки квантилів є, як правило, менш точними, ніж оцінки ймовірностей і, отже, поліграма часто поступається у точності оцінювання звичайній гістограмі.

Інший варіант локальної адаптації полягає в тому, щоб проводити розбиття носія на підінтервали послідовно. На першому кроці розбиття складається з одного елемента — самого носія. На кожному наступному кроці перебирають всі елементи і намагаються розбити їх навпіл. На кожній половинці оцінюють досліджувану щільність, вважаючи її (майже) константою, тобто для н.о.р. спостережень використовують відношення частоти до довжини інтервалу, а для суміші — один з підходів (i)-(iii). Потім порівнюють отримані оцінки по двох половинах. Якщо вони не дуже

відрізняються одна від одної, елемент залишають цілим і більше не змінюють. Якщо оцінки сильно відрізняються — обидві половинки включають у розбиття і вони можуть бути розбиті далі на наступних кроках.

Процедура зупиняється коли всі елементи виявляються “стійкими” до розбиття або настільки вузькими, що їх немає рації розбивати.

Така процедура забезпечує локальну адаптацію, але її доцільно застосовувати (як і інші адаптивні процедури) лише при наявності великого обсягу даних. Неважко розробити аналогічну процедуру послідовного розбиття і для поліграмних оцінок.

При використанні всіх варіантів як поліграми, так і гістограми важливим є питання про вибір носія  $[b, c]$ . Для незалежних, однаково розподілених спостережень, як правило,  $b$  обирають рівним найменшому значенню у вибірці (або трошки менше), а  $c$  — найбільшому (або трошки більше). Це робить стандартну гістограму дуже чутливою до викидів. У випадку спостережень з суміші такий вибір носія створює додаткову проблему, оскільки найменше та найбільше значення у вибірці зовсім не обов’язково належать тій компоненті вибірки, для якої оцінюється щільність. Тому доцільно більш-менш точно оцінити межі справжнього носія щільності аналізованої компоненти, наприклад, використовуючи техніку, описану у п. 3.5. Отримані оцінки можна використовувати як межі носія гістограми.

А що робити, коли оцінювана щільність має необмежений носій? Гістограмні оцінки не призначені для таких випадків. Втім, і ядерні оцінки не дають цілком адекватних результатів при застосуванні до щільностей з необмеженим розподілом, особливо, якщо це щільності розподілів, що породжують викиди (тобто до розподілів з важкими хвостами). Можливий обхідний маневр у цьому випадку полягає в тому, щоб використати попередню трансформацію вибірки за допомогою монотонного перетворення, яке відображає  $\mathbb{R}$  на інтервал  $[0, 1]$ , оцінити щільність на цьому інтервалі, а потім “розгорнути” оцінку назад оберненим перетворенням. Для н.о.р. спостережень така процедура описана у [9]. При цьому слід мати на увазі, що неакуратне використання таких перетворень без урахування “хвостової” поведінки оцінюваної щільності може привести до того, що щільність відображених на  $[0, 1]$  стане необмеженою — прямуватиме до  $+\infty$  в околі 0 або 1. Звичайні оцінки щільностей зовсім не пристосовані до таких ситуацій.

Нажаль, техніка аналізу хвостів розподілу по спостереженнях з суміші іще зовсім не розроблена. Невідомо навіть, чи існують методи, які дозволяли б оцінювати екстремальні індекси окремо для кожної компоненти

суміші.

**Проекційні оцінки.** Інший підхід до оцінювання невідомих щільностей полягає в тому, щоб розглядати їх як елементи гільбертового простору  $L_2$  і оцінювати коефіцієнти у розкладі аналізованої функції за деяким базисом у  $L_2$ . Оцінки такого роду мають назву проекційних, оскільки, по суті, оцінюється не сама щільність, а її проекція на деякий підпростір у  $L_2$ .

Отже, нехай потрібно оцінити щільність  $h_m = \frac{\partial}{\partial x} H_m$   $m$ -тої компоненти суміші за спостереженнями  $\xi_{j:N}$ , що описуються стандартною моделлю 4.1. Припустимо, що розподіл  $H_m$  зосереджений на деякому інтервалі  $S \subseteq \mathbb{R}$  (скінченному або нескінченному) і  $h_m \in L_2(S)$ , де  $L_2(S)$  — простір функцій, інтегрованих з квадратом відносно міри Лебега на  $S$  зі скалярним добутком

$$\langle a, b \rangle_{L_2} = \int_S a(x)b(x)dx.$$

Як відомо,  $L_2(S)$  є сепарабельним гільбертовим простором. Виберемо і зафіксуємо ортонормований базис у цьому просторі:  $p_1, p_2, \dots$ . Функція  $h_m$  розкладається у ряд за цим базисом, який збігається у нормі  $L_2$ :

$$h_m(x) = \sum_{i=1}^{\infty} c_i p_i(x),$$

де  $c_i$  — узагальнені коефіцієнти Фур'є функції  $h_m$ , які можна обчислити за формулою

$$c_i = \int_S p_i(x)h_m(x)dx = \int p_i(x)H_m(dx) = \mathbb{E} p_i(\eta_m),$$

$\eta_m$  — випадкова величина з розподілом  $H_m$ .

Таким чином,  $c_i$  є функціональними моментами розподілу  $H_m$ , що відповідають функціям  $p_i$ . Для їх оцінки можна використовувати методи, розроблені у розділі 3. Наприклад, використовуючи лінійну оцінку 3.3, отримуємо

$$\hat{c}_i^N = \frac{1}{N} \sum_{j=1}^N a_{j:N}^m p_i(\xi_{j:N}), \quad (4.29)$$

де, як і раніше,  $a_{j:N}^m$  — мінімаксні вагові коефіцієнти для оцінки  $H_m$ . Проекційна оцінка для  $h_m$  має вигляд

$$\hat{h}_m^N(x) = \sum_{i=1}^L \hat{c}_i p_i(x) \quad (4.30)$$

$$= \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^L p_i(\xi_{j:N}) p_i(x) = \frac{1}{N} \sum_{j=1}^N K_L(\xi_{j:N}, x),$$

де  $L$  — деяке ціле число, що є параметром алгоритму,

$$K_L(x, y) = \sum_{i=1}^L \hat{p}_i(x) p_i(y).$$

Функцію  $K_L$  називають ядром проекційної оцінки. Вона відіграє приблизно таку ж роль, яку “масштабоване” ядро

$$K_h(x, y) = \frac{1}{h} K\left(\frac{x - y}{h}\right)$$

грає для ядерної оцінки.

Відповідно, задача вибору базису  $p_i, i = 1, 2, \dots$  для проекційної оцінки виявляється подібною до вибору ядра для ядерної оцінки, а “ядерному” параметру згладжування  $h$  відповідає “проекційний” параметр  $L$ .

Оцінка  $\hat{h}_m^N$  оцінює скоріше не  $h_m(x)$ , а  $\sum_{i=1}^L c_i p_i(x)$ , тобто проекцію  $h_m(x)$  на простір, натягнутий на перші  $L$  базисних функцій. Тому параметр  $L$  називають вимірністю простору проекції. Вибір скінченного  $L$  обумовлений міркуваннями як практичного, так і теоретичного характеру.

З практичної точки зору, нескінченну суму  $\sum_{i=1}^{\infty} \hat{c}_i p_i(x)$  просто неможливо обчислити і тому її замінюють скінченною. Крім того, якщо обрати  $L$  не дуже великим (помітно меншим, ніж  $N$ ), то проекційні оцінки виявляються зручнішими для користування, ніж ядерні. Дійсно, якщо потрібно обчислювати одну і ту ж саму оцінку  $\hat{h}_m^N(x)$  для багатьох різних значень  $x$  (наприклад, при побудові її графіку або при використанні у задачах класифікації), то досить один раз обчислити  $L$  коефіцієнтів  $\hat{c}_i$ , виконавши для цього  $\sim LN$  операцій, а потім підставляти ці коефіцієнти у формулу (4.30) стільки разів, скільки це потрібно для різних  $x$ . При підрахунку ядерної оцінки за (4.2) для нових  $x$  оцінку потрібно переобчислювати за даними наново, кожного разу виконуючи  $\sim N$  операцій.

З точки зору точності оцінювання на  $L$  накладаються такі ж суперечливі вимоги, як на параметр згладжування  $h$  ядерної оцінки та кількість підінтервалів  $K$  гістограми або поліграми. З одного боку, чим більшим є  $L$ , тим ближчою є оцінювана проекція до справжньої щільності і, отже, тим точнішою буде оцінка. З іншого боку, при зростанні  $L$  ми змушені оцінювати більше коефіцієнтів  $\hat{c}_i$ . Оскільки кожен коефіцієнт оцінюється з похибкою, то і сумарна похибка оцінки зростає із зростанням  $L$ .

Таким чином, при зростанні обсягу вибірки  $N \rightarrow \infty$ ,  $L$  також повинно зростати, але помітно повільніше, ніж  $N$ . Можливий адаптивний вибір  $L$  на основі попереднього оцінювання ступеня гладкості щільності  $h_m$ . Цікавий варіант адаптивного проекційного оцінювання полягає в тому, щоб  $L$  обрати порівняно великим, але значення  $\hat{c}_i$ , які виявляються малими за абсолютною величиною, покласти рівними 0. Формально, задається деякий поріг  $M_N$  і оцінка визначається як

$$\tilde{h}_m^N(x) = \sum_{i=1}^L \hat{c}_i \mathbb{I}\{|\hat{c}_i| < M_N\} p_i(x). \quad (4.31)$$

Можна показати, що такі оцінки при правильному виборі порогу  $M_N$  та вимірності простору проекції  $L = L_N$  мають швидкість збіжності, близьку до оптимальної.

На відміну від ядерних оцінок, у яких асимптотичну теорію можна побудувати для широкого класу ядер одразу, поведінка проекційних оцінок дуже сильно залежить від властивостей обраного базису проекції.

Останнім часом у статистиці великого поширення набули алгоритми, які спираються на використання вейвлет-базисів (див., наприклад, [14]). Вейвлети зручні для апроксимації функцій (у статистиці звичайно — функцій регресії або щільностей розподілу) тим, що вони дозволяють використовувати локальні особливості функції у різних частинах області її визначення. Там, де функція є гладенькою, для її апроксимації використовується менше елементів вейвлет-базису, там, де вона має особливості — більше.

Оцінкам щільностей розподілів компонент суміші зі змінними концентраціями присвячені роботи Д.І.Похилька [29]-[31].

**Оцінки методу найближчого сусіда.** Розглянуті вище оцінки призначені для використання, як правило, у одновимірному просторі даних  $\mathcal{X}$ . Хоча, в принципі, існують версії “багатовимірних” гістограм та проекційних оцінок, але на практиці їх використовують рідко. Для багатовимірних даних, або навіть для даних з довільних метричних просторів, найбільш поширеною є оцінка методу  $k$  найближчих сусідів. Такі оцінки зручні, зокрема тим, що дозволяють оцінювати щільність не тільки відносно міри Лебега, а відносно довільної міри на  $\mathcal{X}$ . Це, наприклад, дозволяє використовувати такі оцінки для побудови емпірично-баєсових класифікаторів у дуже загальних ситуаціях (див. п. 6.2).

Ми опишемо версію оцінки методу  $k$  найближчих сусідів, призначену для аналізу сумішей зі змінними концентраціями. Нехай спостереження

$\xi_{j:N}$  є елементами сепарабельного метричного простору  $\mathcal{X}$  з метрикою  $\rho$ . Розподіл  $\xi_{j:N}$  описується моделлю суміші зі змінними концентраціями:

$$\mathbb{P}\{\xi_{j:N} \in A\} = \sum_{m=1}^M w_{j:N}^m H_m(A)$$

для будь-якої множини  $A$  з борелевої  $\sigma$ -алгебри на просторі  $(\mathcal{X}, \rho)$ . Нехай на цьому просторі задана деяка міра  $\nu$  і ми хочемо побудувати оцінку для щільності розподілу  $m$ -тої компоненти відносно цієї міри, тобто для похідної Радона  $h_m = \frac{\partial H_m}{\partial \nu}$ .

Для довільного  $x \in \mathcal{X}$  розглянемо набір відстаней від  $x$  до всіх елементів вибірки:  $z_j = \rho(x, \xi_{j:N})$ ,  $j = 1, \dots, N$ . Впорядкуємо цей набір у порядку зростання:  $z_{[1]} \leq z_{[2]} \leq z_{[3]} \leq \dots \leq z_{[N]}$ . Зрозуміло, що  $z_{[1]}$  — це відстань від  $x$  до його найближчого сусіда серед елементів вибірки,  $z_{[2]}$  — відстань до наступного (другого) сусіда і так далі. (Якщо серед  $z_j$  є однакові, тобто відстані від  $x$  до деяких елементів вибірки є рівними між собою, то порядок нумерації таких сусідів по ступеню близькості є для нас несуттєвим). Позначимо  $r(x, k) = z_{[k]}$  — відстань до  $k$ -того найближчого сусіда.

Позначимо  $B(x, r) = \{y \in \mathcal{X} : \rho(x, y) \leq r\}$  — замкнена куля в  $\mathcal{X}$  радіуса  $r$  з центром у  $x$ . Як оцінку для  $h_m(x)$  використаємо

$$\hat{h}_m^N(x) = \frac{\hat{H}_m^N(B(x, r(x, k)))}{\nu(B(x, r(x, k)))}, \quad (4.32)$$

де  $\hat{H}_m^N(A)$  — деяка оцінка для  $H_m$  за спостереженнями. Наприклад, це може бути зважена емпірична функція розподілу з мінімаксними ваговими коефіцієнтами. Параметр  $k = k_N$  (кількість найближчих сусідів) відіграє роль, аналогічну параметру згладжування у ядерних оцінках (з точністю до навпаки): обираючи  $k_N \rightarrow \infty$  так, щоб послідовність  $k_N$  зростала не дуже швидко, можна отримати консистентну оцінку.

Ідею оцінки (4.32) особливо легко зрозуміти якщо  $\mathcal{X} = \mathbb{R}^d$ , а  $\nu \in$  мірою Лебега. У цьому випадку, якщо  $h_m$  існує, то при  $N \rightarrow \infty$ ,  $k_N/N \rightarrow 0$  буде мати місце збіжність  $r(x, k_N) \rightarrow 0$ . Отже, куля  $B(x, r(x, k_N))$  буде стягуватися до  $x$  і  $\frac{H_m(B(x, r(x, k_N)))}{\nu(B(x, r(x, k_N)))} \rightarrow h_m(x)$ . Якщо  $k_N \rightarrow \infty$  при  $N \rightarrow \infty$ , то кількість елементів вибірки, які попадають до  $B(x, r(x, k))$ , прямує до нескінченності, тому можна сподіватись, що  $\hat{H}_m^N(B(x, r(x, k))) \sim H_m(B(x, r(x, k)))$  (обидві ці величини за даних умов прямують до 0). Звідси і отримуємо консистентність оцінки.

Але оцінка (4.32) виявляється консистентною і у значно більш широким умовам, наприклад, якщо  $H_m$  та  $\nu$  — дискретні міри. Ми не будемо тут формулювати загальні результати для оцінок методу  $k$  найближчих сусідів. У розділі 6 ці оцінки використовуються для побудови емпірично-баєсових класифікаторів. Там наведені і результати щодо цих оцінок, які дозволяють довести консистентність відповідних класифікаторів.

Іще один напрямок розробки оцінок щільності — застосування певних модифікацій **методу найбільшої вірогідності**. Як відомо, оцінки найбільшої вірогідності для щільностей розподілу в загальному випадку не існують. Тому застосовують різні техніки виправлення цього методу, які дозволяють будувати консистентні оцінки — або з використанням штрафних функцій, або з обмеженням класу, по якому шукають точку максимуму функції вірогідності. Застосування останнього підходу (метод максимуму відсіяної вірогідності) ми розглянемо у наступному розділі в контексті аналізу даних з домішкою.



## Розділ 5

# Аналіз спостережень з домішкою

### 5.1 Оцінки щільності по спостереженнях з домішкою

У медико-біологічній статистиці часто виникає задача аналізу даних з домішками. Опишемо модельний приклад таких даних.

Нехай досліджуються  $N$  пацієнтів, яким поставлено попередній діагноз, скажімо “атипічна пневмонія”(АП). У кожного з досліджуваних вимірюється певна фізіологічна характеристика  $\xi$  (наприклад — кількість еритроцитів у крові). В результаті отримано набір  $\Xi_N = (\xi_1, \dots, \xi_N)$ , де  $\xi_j$  — значення  $\xi$  у  $j$ -того пацієнта. Припускається, що розподіл  $\xi$  у хворих на АП не такий, як у пацієнтів, що не є хворими на АП. Позначимо щільність розподілу  $\xi$  у хворих на АП через  $h_1$ , у не хворих —  $h_2$ . Якби було точно відомо, що всі пацієнти з досліджуваної вибірки хворі на АП, то можна було б вважати, що  $\Xi_N$  — проста вибірка з щільністю спостережень  $h_1$ . Нажаль, АП важко відрізнити від інших захворювань (звичайна пневмонія, грип, застуда) тому, скоріше за все, в  $\Xi_N$  є домішка  $\xi_j$  з щільністю  $h_2$ . Часто за симптомами можна вказати (оцінити) ймовірність  $w_j$  того, що  $j$ -тий пацієнт є хворим саме АП. У такому випадку щільність  $\xi_j$  являє собою суміш щільностей  $h_i$ :

$$w_j h_1(x) + (1 - w_j) h_2(x).$$

Як за такими даними оцінити  $h_1$ ?

Цей параграф присвячено задачі оцінювання щільностей за даними з домішкою у випадку, коли щільність основної компоненти  $h_1$  повністю

невідома, тобто її потрібно оцінювати непараметрично. Щодо щільності домішки,  $h_2$ , можливі різні припущення:

- (1)  $h_2$  може бути так само невідома, як і  $h_1$  (непараметричний випадок);
- (2)  $h_2$  може бути повністю відома, наприклад, оцінена з достатньою точністю за спостереженнями  $\xi$  у осіб, які напевне не є хворими на АП (детермінований випадок);
- (3)  $h_2$  може бути відома з точністю до параметру (параметрична модель домішки).

Третій варіант виникає, коли ми припускаємо, що пацієнти “не хворі на АП” у нашій вибірці мають розподіл  $\xi$  такого ж типу, як і здорові (скажімо — гауссів) але, можливо, з іншими параметрами (наприклад, з іншим середнім) оскільки ці особи все ж є хворими якимось (неідентифікованими) хворобами.

Ми розглянемо оцінки гістограмного типу для щільностей розподілу у всіх трьох випадках, побудовані методом відсіяної найбільшої вірогідності (sieve maximum likelihood)<sup>1</sup>. Для третього випадку побудуємо також оцінки невідомих параметрів розподілу домішки. Доведемо консистентність цих оцінок і оцінимо швидкість їх збіжності.

Як і раніше, вибірку  $\Xi_N$  будемо вважати елементом схеми серій:  $\Xi_N = (\xi_{1:N}, \dots, \xi_{N:N})$  де  $\xi_{j:N}$  — незалежні при фіксованому  $N$  випадкові величини і

$$\Pr\{\xi_{j:N} < x\} = w_{j:N}H_1(x) + (1 - w_{j:N})H_2(x),$$

де  $H_1$  — функція розподілу (ф.р.) основної компоненти,  $H_2$  — ф.р. домішки,  $w_{j:N}$  — концентрація основної компоненти у суміші в момент  $j$ -того спостереження, тобто ймовірність того, що  $j$ -те спостереження вибрано з основної компоненти.

Надалі припускаємо, що розподіли  $H_i$  зосереджені на скінченному інтервалі, який (без обмеження загальності) вважаємо рівним  $[0, 1]$ . Вважаємо, що існують щільності  $h_i$  розподілів  $H_i$  відносно міри Лебега, причому виконана умова

(А) існують числа  $0 < u < U < \infty$ , такі, що  $u \leq h_i(x) \leq U$  для всіх  $x \in [0, 1]$ .

Для побудови оцінок скористаємося методом максимуму відсіяної вірогідності. Будемо шукати оцінку  $\hat{h}_i^N(x)$  для  $h_i(x)$  лише серед східчастих

<sup>1</sup>Про метод відсіяної найбільшої вірогідності див., наприклад, [65].

функцій вигляду

$$g_k(x) = \sum_{k=1}^{K_N} g_k \mathbb{I}\{x \in A_k\}, \quad (5.1)$$

де  $K_N$  — кількість підінтервалів розбиття  $A_k$ ,  $A_k = [t_{k-1}, t_k)$  для  $k = 1, \dots, K_N - 1$ ,  $A_{K_N} = [t_{K_N-1}, t_{K_N}]$ ,  $t_k = k/K_N$ ,  $g_i$  — довільні додатні числа, такі, що  $\frac{1}{K_N} \sum_{k=1}^{K_N} g_k = 1$ . Нехай  $0 < u_N < U_N < \infty$  деякі дійсні числа. Множину всіх функцій вигляду (5.1), для яких  $u_N \leq g_k \leq U_N$ ,  $k = 1, \dots, K_N$ , позначимо  $\mathcal{H}_N$ . Нехай  $\chi_{kj} = \mathbb{I}\{\xi_{j:N} \in A_k\}$ . Якщо припустити, що справжні щільності компонент  $h_i = g_i(x)$  де  $g_i(x) = \sum_{k=1}^{K_N} g_{ki} \mathbb{I}\{x \in A_k\}$ , то логарифмічна функція вірогідності, підрахована за вибіркою,  $\Xi_N$  має вигляд:

$$\begin{aligned} l(g_1, g_2) &= \sum_{j=1}^N \ln \left( w_{j:N} \sum_{k=1}^{K_N} g_{k1} \chi_{kj} + (1 - w_{j:N}) \sum_{k=1}^{K_N} g_{k2} \chi_{kj} \right) \\ &= \sum_{j=1}^N \sum_{k=1}^{K_N} \chi_{kj} \ln(w_{j:N} g_{k1} + (1 - w_{j:N}) g_{k2}). \end{aligned}$$

Якщо і  $h_1$  і  $h_2$  повністю невідомі, на роль оцінки для пари  $(h_1, h_2)$  оберемо пару  $(\hat{h}_1^N, \hat{h}_2^N) \in \mathcal{H}_N \times \mathcal{H}_N = \mathcal{H}_N^2$ , яка максимізує функцію  $l$  на  $\mathcal{H}_N^2$ :

$$(\hat{h}_1^N, \hat{h}_2^N) = \underset{(g_1, g_2) \in \mathcal{H}_N^2}{\operatorname{argmax}} l(g_1, g_2). \quad (5.2)$$

(Якщо  $\operatorname{argmax}$  досягається на кількох парах функцій, оцінкою може бути будь-яка з них.)

Для випадку однорідних даних (без домішок) метод відсіяної вірогідності приводить до гістограмних оцінок щільності. По аналогії будемо називати гістограмами всі оцінки щільностей, які мають вигляд (5.1).

Якщо щільність другої компоненти,  $h_2$  відома з точністю до параметру  $\vartheta$ , тобто  $h_2(x) = h_2(x; \vartheta)$ , де  $\vartheta \in \Theta$  — невідомий параметр, позначимо

$$\bar{h}_2(x, \vartheta) = \sum_{k=1}^{K_N} \bar{h}_{k2}(\vartheta) \mathbb{I}\{x \in A_k\},$$

де

$$\bar{h}_{k2} = \bar{h}_{k2}(\vartheta) = \frac{1}{s_N} \int_{t_{k-1}}^{t_k} h_2(x, \vartheta) dx.$$

Позначимо  $\Theta_N = \{\vartheta \in \Theta : \bar{h}_2(\cdot, \vartheta) \in \mathcal{H}_N\}$ . На роль оцінки для пари невідомих параметрів розподілу даних  $(h_1, \vartheta)$  використаємо

$$(\hat{h}_1^N, \hat{\vartheta}^N) = \operatorname{argmax}_{g_1 \in \mathcal{H}_N, \tau \in \Theta_N} l(g_1, \bar{h}_2(\cdot, \tau)), \quad (5.3)$$

тобто максимум шукається лише по тих “гістограмах”, які для другої компоненти можна отримати усередненням щільності, що відповідає обраній параметричній моделі.

Якщо  $h_2$  повністю відома, можна скористатись (5.3) з множиною можливих значень параметра  $\Theta$ , що складається з одного елемента. Тобто в цьому випадку

$$\hat{h}_1 = \operatorname{argmax}_{g_1 \in \mathcal{H}_N} l(g_1, \bar{h}_2(\cdot)).$$

**Основні теореми.** Позначимо  $s_N = 1/K_N$  — довжина підінтервалів розбиття  $A_k$ ,  $\bar{w}_N = \frac{1}{N} \sum_{j=1}^N w_{j:N}$  — “середня концентрація” основної компоненти у суміші.

Наступна теорема дає умови консистентності гістограмних оцінок у непараметричному випадку в просторі  $L_2[0, 1]$ , тобто у просторі функцій на  $[0, 1]$  з нормою

$$\|a\|_2 = \left( \int_0^1 a^2(x) dx \right)^{1/2}.$$

**Теорема 5.1.1** *Нехай виконана умова (A) і*

- 1) *Існує  $c > 0$ , таке, що  $\frac{1}{N} \sum_{j=1}^N (w_{j:N} - \bar{w}_N)^2 > c$  для всіх  $N$ ;*
- 2)  *$s_N \rightarrow 0$ ,  $s_N N \rightarrow \infty$  при  $N \rightarrow \infty$ ;*
- 3)  *$U_N \rightarrow \infty$ ,  $u_N \rightarrow 0$ ,*

$$\frac{U_N^5}{N s_N u_N^4} \rightarrow 0, \quad \frac{U_N \ln^2 u_N}{s_N N} \rightarrow 0 \text{ при } N \rightarrow \infty.$$

*Тоді для оцінок  $\hat{h}_i^N$ ,  $i = 1, 2$ , визначених (5.2),*

$$\|\hat{h}_i^N - h_i\|_2 \rightarrow 0$$

*за ймовірністю при  $N \rightarrow \infty$ .*

**Зауваження.** Умова 1 теореми еквівалентна умові  $\det \Gamma_N > c > 0$ .

У випадку параметричної моделі домішки позначимо

$$\omega(s) = \sup_{|x-y| \leq s} \sup_{\tau \in \Theta} |h_2(x, \tau) - h_2(y, \tau)|$$

— рівномірний модуль неперервності  $h_2$  при всіх можливих значеннях невідомого параметру.

Умови консистентності в цьому випадку дає наступна теорема

**Теорема 5.1.2** *Припустимо, що виконані умова (A) і умови 1)-3) теореми 5.1.1. Нехай крім того виконано:*

1)  $\omega(s) \rightarrow 0$  при  $s \rightarrow 0$ ;

2)  $\Theta$  є компактом у напівметриці  $\rho(\alpha, \tau) = \|h_2(\cdot, \alpha) - h_2(\cdot, \tau)\|_2$ ;

3) При  $\tau \neq \vartheta$ ,  $\rho(\tau, \vartheta) > 0$ .

Тоді для оцінок  $\hat{h}_1^N$  та  $\hat{\vartheta}^N$ , визначених (5.3), має місце збіжність  $\|\hat{h}_1^N - h^1\|_2 \rightarrow 0$  і  $\rho(\hat{\vartheta}^N, \vartheta) \rightarrow 0$  за ймовірністю при  $N \rightarrow \infty$ .

Накладаючи додаткові умови на щільності  $h_i$ , можна отримати швидкості збіжності оцінок. Наприклад, нехай

$$\omega_i(s) = \sup_{|x-y| \leq s} |h_i(x) - h_i(y)|.$$

Розглянемо непараметричний випадок. Літерою  $C$  будемо позначати довільні скінченні додатні константи, можливо, різні.

**Теорема 5.1.3** *Нехай виконані умова (A) і умова 1) теореми (5.1.1). Додатково припустимо, що:*

1) для деякого  $\beta > 0$ ,  $\omega_i(s) \leq Cs^\beta$ ;

2) для деякого  $\alpha > 0$ ,  $s_N = CN^{-\alpha}$ ;

3)  $U_N = C \ln N$ ,  $u_N = C / \ln N$ .

Тоді для оцінок, визначених (5.2), для будь-якого  $\gamma$ , такого, що  $0 < \gamma < \alpha\beta$ ,  $\gamma < (1 - \alpha)/4$ , знайдуться константи  $C_1, C_2 > 0$ , такі, що для всіх  $\lambda > 0$ , при достатньо великих  $N$ ,

$$\mathbf{P}\{\|\hat{h}_i^N - h_i\|_2 > \lambda N^{-\gamma}\} \leq \frac{C_1 \ln N}{N^{1-\alpha}(\lambda N^{-\gamma} - C_2 N^{-\alpha\beta})^4}$$

**Доведення теорем.** Для  $g_{k1}, g_{k2} > 0$  позначимо

$$\hat{J}_k(g_{k1}, g_{k2}) = \frac{1}{N} \sum_{j=1}^N \chi_{kj} \ln(w_{j:N} g_{k1} + (1 - w_{j:N}) g_{k2}).$$

Тоді

$$\frac{1}{N} l(g_1, g_2) = \hat{J}(g_1, g_2) := \sum_{k=1}^{K_N} \hat{J}_k(g_{k1}, g_{k2}).$$

(Надалі ми східчасту функцію  $\sum_{k=1}^{K_N} g_{ki} \mathbb{I}\{x \in A_k\}$  і набір чисел  $(g_{ki}, k = 1, \dots, K_N)$  позначаємо одним символом  $g_i$ .)

Нехай  $\eta_i$  — випадкові величини з розподілом  $H_i$ . Позначимо

$$\begin{aligned} \bar{h}_{ki} &= \frac{1}{s_N} \mathbb{P}\{\eta_i \in A_k\} = \frac{1}{s_N} \int_{t_{k-1}}^{t_k} h_i(x) dx, \\ \bar{h}_i(x) &= \sum_{k=1}^{K_N} \bar{h}_{ki} \mathbb{I}\{x \in A_k\}, \\ J_k(g_{k1}, g_{k2}) &= \mathbb{E} \hat{J}_k(g_{k1}, g_{k2}) \\ &= \frac{1}{N} \sum_{j=1}^N s_N (w_{j:N} \bar{h}_{k1} + (1 - w_{j:N}) \bar{h}_{k2}) \cdot \ln(w_{j:N} g_{k1} + (1 - w_{j:N}) g_{k2}), \\ J(g_1, g_2) &= \sum_{k=1}^{K_N} J_k(g_{k1}, g_{k2}). \end{aligned}$$

**Лема 5.1.1** При виконанні умови (A) для всіх  $\varepsilon > 0$

$$\begin{aligned} &\mathbb{P} \left\{ \sup_{g_1, g_2 \in \mathcal{H}_N} |\hat{J}(g_1, g_2) - J(g_1, g_2)| > \varepsilon \right\} \\ &\leq \frac{4U}{s_N N \varepsilon^2} \left( \frac{U_N^4}{u_N^4} + 2 \frac{U_N^2}{u_N^2} + \max(\ln^2(u_N), \ln^2(U_N)) \right). \end{aligned}$$

**Доведення.** Позначимо  $f(g_{k1}, g_{k2}) = \hat{J}_k(g_{k1}, g_{k2}) - J_k(g_{k1}, g_{k2})$ . Скористаємося рівністю (7.2):

$$\begin{aligned} f(x_1, x_2) &= \int_{u_N}^{x_2} \int_{u_N}^{x_1} \frac{\partial^2}{\partial t_1 \partial t_2} f(t_1, t_2) dt_1 dt_2 \\ &\quad + \int_{u_N}^{x_2} \frac{\partial}{\partial t_2} f(u_N, t_2) dt_2 + \int_{u_N}^{x_1} \frac{\partial}{\partial t_1} f(t_1, u_N) dt_1 + f(u_N, u_N) \end{aligned} \quad (5.4)$$

Застосовуючи нерівність Коші-Буняківського до перших трьох доданків (5.4), отримуємо:

$$\sup_{x_1, x_2 \in [u_N, U_N]} |f(x_1, x_2)| \leq \left( \int_{u_N}^{U_N} \int_{u_N}^{U_N} \left( \frac{\partial^2}{\partial t_1 \partial t_2} f(t_1, t_2) \right)^2 dt_1 dt_2 (U_N - u_N)^2 \right)^{1/2}$$

$$\begin{aligned}
& + \left( \int_{u_N}^{U_N} \left( \frac{\partial}{\partial t_2} f(u_N, t_2) \right)^2 dt_2 (U_N - u_n) \right)^{1/2} \\
& + \left( \int_{u_N}^{U_N} \left( \frac{\partial}{\partial t_1} f(t_1, u_N) \right)^2 dt_1 (U_N - u_n) \right)^{1/2} + f(u_N, u_N).
\end{aligned}$$

Оцінимо

$$\mathbb{E} \sup_{x_1, x_2 \in [u_N, U_N]} (f(x_1, x_2))^2 \leq 4(R_1(U_N - u_n)^2 + R_2(U_N - u_N) + R_3(U_N - u_N) + R_4),$$

де

$$\begin{aligned}
R_1 &= \int_{u_N}^{U_N} \int_{u_N}^{U_N} \left( \frac{\partial^2}{\partial t_1 \partial t_2} f(t_1, t_2) \right)^2 dt_1 dt_2, \quad R_2 = \int_{u_N}^{U_N} \left( \frac{\partial}{\partial t_2} f(u_N, t_2) \right)^2 dt_2, \\
R_3 &= \int_{u_N}^{U_N} \left( \frac{\partial}{\partial t_1} f(t_1, u_N) \right)^2 dt_1, \quad R_4 = f^2(u_N, u_N).
\end{aligned}$$

Розглянемо кожен доданок окремо. Маємо

$$\begin{aligned}
R_1 &= \frac{1}{N^2} \int_{u_N}^{U_N} \int_{u_N}^{U_N} \sum_{j,l=1}^N \frac{w_{j:N}(1 - w_{j:N})w_{l:N}(1 - w_{l:N})}{(w_{j:N}t_1 + (1 - w_{j:N})t_2)^2(w_{l:N}t_1 + (1 - w_{l:N})t_2)^2} \times \\
& \times \mathbb{E}(\chi_{jk} - \mathbb{E} \chi_{jk})(\chi_{lk} - \mathbb{E} \chi_{lk}) dt_1 dt_2.
\end{aligned}$$

При  $l \neq j$ ,  $\mathbb{E}(\chi_{jk} - \mathbb{E} \chi_{jk})(\chi_{lk} - \mathbb{E} \chi_{lk}) = 0$ , а при  $l = j$ ,  $\mathbb{E}(\chi_{jk} - \mathbb{E} \chi_{jk})(\chi_{lk} - \mathbb{E} \chi_{lk}) \leq \mathbb{E} \chi_{j,k}$ . Тому

$$\begin{aligned}
R_1 &\leq \frac{1}{N^2} \sum_{j=1}^N \int_{u_N}^{U_N} \int_{u_N}^{U_N} \frac{1}{u_N^4} s_N(w_{j:N} \bar{h}_{k1} + (1 - w_{j:N}) \bar{h}_{k2}) dt_1 dt_2 \\
&\leq \frac{s_N(U_N - u_n)^2 U_N}{N u_N^4}.
\end{aligned}$$

Аналогічно для  $R_2$  та  $R_3$  маємо

$$\begin{aligned}
R_2 &\leq \int_{u_N}^{U_N} \sum_{j=1}^N \frac{w_{j:N}^2}{(w_{j:N}t_1 + (1 - w_{j:N})u_N)^2} \mathbb{E}(\chi_{jk} - \mathbb{E} \chi_{jk})^2 dt_1 \\
&\leq \frac{s_N(U_N - u_N)U}{N u_N^2},
\end{aligned}$$

і, так само,

$$R_3 \leq \frac{s_N(U_N - u_N)U}{Nu_N^2}.$$

Для  $R_4$  маємо

$$\begin{aligned} R_4 &\leq \frac{1}{N^2} \sum_{j=1}^N \ln^2(u_N) \mathbb{E}(\chi_{j:N} - \mathbb{E} \chi_{j:N})^2 \\ &\leq \frac{1}{N} \ln^2(u_N) s_N U. \end{aligned}$$

Отже

$$\mathbb{E} \sup_{t_1, t_2 \in [u_N, U_N]} |\hat{J}_k(t_1, t_2) - J_k(t_1, t_2)|^2 \leq \frac{s_N D}{N},$$

де

$$D = 4U \left( \frac{(U_N - u_N)^4}{u_N^4} + 2 \frac{(U_N - u_N)^2}{u_N^2} + \ln^2(u_N) \right).$$

Тому

$$\begin{aligned} &\left( \mathbb{E} \sup_{g_1, g_2 \in \mathcal{H}_N} |\hat{J}(t_1, t_2) - J(t_1, t_2)|^2 \right)^{1/2} \\ &\leq \left( \mathbb{E} \left( \sum_{k=1}^{K_N} \sup_{g_{1k}, g_{2k} \in [u_N, U_N]} |\hat{J}_k(g_{1k}, g_{2k}) - J_k(g_{1k}, g_{2k})| \right)^2 \right)^{1/2} \\ &\leq \left( K_N \sum_{k=1}^{K_N} \mathbb{E} \sup_{g_{1k}, g_{2k} \in [u_N, U_N]} (\hat{J}_k(g_{1k}, g_{2k}) - J_k(g_{1k}, g_{2k}))^2 \right)^{1/2} \\ &\leq \sqrt{K_N^2 \frac{s_N D}{N}} = \sqrt{\frac{D}{s_N N}}. \end{aligned}$$

Звідси, використовуючи нерівність Чебишова, отримуємо

$$\mathbb{P} \left\{ \sup_{g_1, g_2 \in \mathcal{H}_N} |\hat{J}(t_1, t_2) - J(t_1, t_2)| \geq \varepsilon \right\} \leq \frac{D}{s_N N \varepsilon^2}.$$

Лема доведена.

**Доведення теореми 5.1.1.** Позначимо

$$B_N(\varepsilon) = \left\{ \sup_{g_1, g_2 \in \mathcal{H}_N} |\hat{J}(g_1, g_2) - J(g_1, g_2)| > \varepsilon \right\}.$$



Згідно з лемою 5.1.1,  $\mathbf{P}\{B_N(\varepsilon)\} \rightarrow 0$  при  $N \rightarrow \infty$  для всіх  $\varepsilon > 0$ . Нехай виконана подія  $\bar{B}_N(\varepsilon)$ , протилежна до  $B_N(\varepsilon)$ . Тоді  $|\hat{J}(\hat{h}_1^N, \hat{h}_2^N) - J(\hat{h}_1^N, \hat{h}_2^N)| < \varepsilon$  і  $|\hat{J}(\bar{h}_1, \bar{h}_2) - J(\bar{h}_1, \bar{h}_2)| < \varepsilon$ . Отже

$$J(\hat{h}_1^N, \hat{h}_2^N) \geq \hat{J}(\hat{h}_1^N, \hat{h}_2^N) - \varepsilon \geq \hat{J}(\bar{h}_1, \bar{h}_2) - \varepsilon \geq J(\bar{h}_1, \bar{h}_2) - 2\varepsilon.$$

Тому

$$0 \geq J(\hat{h}_1^N, \hat{h}_2^N) - J(\bar{h}_1, \bar{h}_2) \geq -2\varepsilon.$$

Позначимо

$$\rho_{1N}^j := \int_0^1 \ln \left( \frac{w_{j:N} \hat{h}_1^N(x) + (1 - w_{j:N}) \hat{h}_2^N(x)}{w_{j:N} \bar{h}_1(x) + (1 - w_{j:N}) \bar{h}_2(x)} \right) (w_{j:N} \bar{h}_1(x) + (1 - w_{j:N}) \bar{h}_2(x)) dx$$

Використовуючи означення  $J(h_1, h_2)$ , отримуємо

$$\rho_{1N} := \left| \frac{1}{N} \sum_{j=1}^N \rho_{1N}^j \right| \leq 2\varepsilon.$$

Кожен доданок  $\rho_{1N}^j$  у цій сумі являє собою дивергенцію Кульбака-Ляйблера між щільностями  $\hat{f}_j(x) := w_{j:N} \hat{h}_1^N(x) + (1 - w_{j:N}) \hat{h}_2^N(x)$  та  $f_j(x) := w_{j:N} \bar{h}_1(x) + (1 - w_{j:N}) \bar{h}_2(x)$ . Використовуючи нерівність між дивергенцією Кульбака-Ляйблера та відстанню Хелінгера з [2] с. 195, для

$$\rho_{2N}^j := \int_0^1 \left( \sqrt{\hat{f}_j(x)} - \sqrt{f_j(x)} \right)^2 dx$$

отримуємо  $\rho_{2N}^j \leq \rho_{1N}^j$  і

$$\rho_{2N} := \frac{1}{N} \sum_{j=1}^N \rho_{2N}^j \leq \rho_{1N} \leq 2\varepsilon.$$

Нехай

$$\rho_{3N}^j := \int_0^1 \left( \hat{f}_j(x) - f_j(x) \right)^2 dx.$$

Оскільки для всіх додатних  $a, b$ ,

$$|\sqrt{a} - \sqrt{b}| = |a - b| / (\sqrt{a} + \sqrt{b}) \geq |a - b| / 2 \max(\sqrt{a}, \sqrt{b}),$$

то

$$\rho_{3N} := \frac{1}{N} \sum_{j=1}^N \rho_{3N}^j \leq 2 \max(\sqrt{U_N} \sqrt{U}) \rho_{2N} \leq 4 \sqrt{U_N} \varepsilon$$

при достатньо великих  $N$ .

Позначимо  $b_{11} = N^{-1} \sum_{j=1}^N w_{j:N}^2$ ,  $b_{12} = b_{21} = N^{-1} \sum_{j=1}^N w_{j:N}(1 - w_{j:N})$ ,  $b_{22} = N^{-1} \sum_{j=1}^N (1 - w_{j:N})^2$ ,  $B = (b_{ik})_{i,k=1}^2$ . Маємо

$$\rho_{3N} = b_{11} \|\hat{h}_1^N - \bar{h}_1\|_2^2 + 2b_{12} \|\hat{h}_1^N - \bar{h}_1\|_2 \cdot \|\hat{h}_2^N - \bar{h}_2\|_2 + b_{22} \|\hat{h}_2^N - \bar{h}_2\|_2^2.$$

Тому

$$\rho_{3N} \geq \lambda_{\min}(\|\hat{h}_1 - \bar{h}_1\|_2^2 + \|\hat{h}_2 - \bar{h}_2\|_2^2),$$

де  $\lambda_{\min}$  — найменше,  $\lambda_{\max}$  — найбільше власне число матриці  $B$ . Оскільки всі елементи матриці  $B$  не перевищують 1, то  $\lambda_{\max} \leq 2$ . За умовою 1) теореми  $\lambda_{\min} \lambda_{\max} = \det B \geq c$ , тому  $\lambda_{\min} > c/2$ . Отже,

$$\|\bar{h}_i - \hat{h}_i^N\|_2^2 \leq \frac{8\sqrt{U_N}\varepsilon}{c}.$$

Покладемо  $\varepsilon = cz/(8\sqrt{U_N})$ , де  $z$  — довільне додатне число. Використовуючи лему 5.1.1, отримуємо

$$\mathbb{P}\{\|\bar{h}_i - \hat{h}_i^N\|_2^2 \geq z\} \leq \frac{36UU_N}{cs_N N z^2} \left( \frac{U_N^4}{u_N^4} + 2 \frac{U_N^2}{u_N^2} + \ln^2(u_N) \right) \quad (5.5)$$

За другою і третьою умовами теореми, права частина прямує до 0 при  $N \rightarrow \infty$ . За теоремою 6 з [9],  $\int_0^1 |\bar{h}_i(x) - h_i(x)| dx \rightarrow 0$  при  $s_N \rightarrow 0$ . За умовою (A),  $|h_i(x)| \leq U$ , тому  $|\bar{h}_i(x)| \leq U$ . Отже

$$\|\bar{h}_i - h_i\|_2^2 \leq 2U \int_0^1 |\bar{h}_i(x) - h_i(x)| dx \rightarrow 0,$$

при  $N \rightarrow \infty$ . Враховуючи (5.5), отримуємо твердження теореми.

**Доведення теореми 5.1.2.** Так само, як у теоремі 5.1.1, використовуючи лему 5.1.1, отримуємо  $\|\hat{h}_1^N - \bar{h}_1\|_2 \rightarrow 0$  і  $\|\bar{h}_2(\cdot, \hat{v}^N) - \hat{h}_2(\cdot, \vartheta)\|_2 \rightarrow 0$  при  $N \rightarrow \infty$ . Для довільного  $x \in [0, 1]$  виберемо  $A_i$ , для якого  $x \in A_i$ . Тоді для будь-якого  $\tau \in \Theta$ ,

$$|\bar{h}_2(x, \tau) - h_2(x, \tau)| \leq \frac{1}{s_N} \int_{A_i} |h_2(y, \tau) - h_2(x, \tau)| dy \leq \omega(s_N).$$

Отже,  $\|\bar{h}_2(\cdot, \tau) - h_2(\cdot, \tau)\|_2 \leq \omega(s_N) \rightarrow 0$  при  $N \rightarrow \infty$ . Тому  $\|h_2(\cdot, \hat{\vartheta}^N) - h_2(\cdot, \vartheta)\|_2 \rightarrow 0$  за ймовірністю. Враховуючи умови 2) і 3) теореми, отримуємо  $\hat{\vartheta}^N \rightarrow \vartheta$  за ймовірністю при  $N \rightarrow \infty$ .

Теорема доведена.

**Доведення теореми 5.1.3.** Так само, як у доведенні теореми 5.1.2, отримуємо  $\|\hat{h}_i - h_i\|_2 \leq \omega(s_N)$ . Отже

$$\begin{aligned} \mathbb{P}\{\|\hat{h}_i^N - h_i\|_2 \geq \lambda N^{-\gamma}\} &\leq \mathbb{P}\{\|\hat{h}_i^N - h_i\|_2 + \|\bar{h}_i - h_i\|_2 \geq \lambda N^{-\gamma}\} \\ &\leq \mathbb{P}\{\|\hat{h}_i^N - h_i\|_2 + \omega(s_N) \geq \lambda N^{-\gamma}\} \leq \mathbb{P}\{\|\hat{h}_i^N - h_i\|_2 \geq \lambda N^{-\gamma} - CN^{-\alpha\beta}\}. \end{aligned}$$

Використовуючи (5.5), отримуємо твердження теореми.

## 5.2 Адаптивні оцінки для параметрів

У цьому параграфі продовжується розгляд задач оцінювання по спостереженнях з домішкою, розпочатий у п. 5.1. Тепер ми дослідимо випадок, коли розподіл основної компоненти суміші заданий параметрично, а розподіл домішки — повністю невідомий. Задача полягає в оцінці невідомого параметру основної компоненти. Концентрації компонент у суміші змінюються від спостереження до спостереження і вважаються відомими.

Для оцінювання ми використаємо узагальнений метод моментів, в якому замість звичайних емпіричних моментів використано зважені функціональні моменти, які вивчалися у 3.1. Будуть доведені консистентність та асимптотична нормальність таких оцінок. Оскільки при побудові моментної оцінки пробну функцію можна задавати значною мірою довільно, виникає питання про оптимальний вибір цієї функції. Ми знайдемо пробну функцію, що забезпечує найменший коефіцієнт розсіювання оцінки. Нажаль, вона залежить від невідомого параметру та щільності розподілу домішки (теж невідомої).

Тому можна використати адаптивний підхід, подібний до розглянутого у п. 3.2 для оцінок моментів, а у п. — 2.4 — для оцінок розподілу. На першому кроці оцінюються параметр — за допомогою грубої пілотної оцінки (скажімо, неоптимальної моментної оцінки) і щільність розподілу домішки. На другому — ці оцінки підставляються у формулу для оптимальної пробної функції і отримана (випадкова) функція використовується у адаптивній моментній оцінці. Оскільки моментне рівняння для такої пробної функції, як правило, не розв'язується аналітично, справжній теоретичний момент у ньому замінюється диференціалом в околі пілотної оцінки.

Відомо, що у випадку однорідних спостережень без домішки, ця адаптивна схема приводить до наближених оцінок методу найбільшої вірогідності (оцінок Вальда), які у регулярних задачах є асимптотично ефективними (див. [2], п.2.26). Загальна теорія адаптивного оцінювання [39, 67] дає умови, за яких можна сподіватись асимптотичної ефективності адаптивних оцінок незалежно від того, наскільки ефективними є оцінки “першого кроку”. Нажаль, ці умови не виконуються у моделі спостережень з домішкою. Однак ми покажемо (у теоремі 5.2.3) що, за певних умов, адаптивна оцінка має той же коефіцієнт розсіювання, що і моментна оцінка з оптимальною пробною функцією. Умови теореми 5.2.3 дещо схожі на умови ефективності оцінок методу штрафних функцій з [65]. Вони виконуються, наприклад, при оцінюванні параметру зрізаного експоненційного розподілу при достатньо гладенькій щільності розподілу домішки.

**Постановка задачі.** Ми будемо розглядати спостереження з домішкою, тобто дані вигляду  $\Xi_N = \{\xi_{j:N}, j = 1, \dots, N\}$ , де, при фіксованому  $N$ ,  $\xi_{j:N}$  — незалежні між собою випадкові величини з функцією розподілу

$$\mathbf{P}\{\xi_{j:N} < x\} = w_{j:N}H_1(x, \vartheta) + (1 - w_{j:N})H_2(x),$$

де  $w_{j:N}$  — концентрація основної компоненти у суміші під час  $j$ -того спостереження,  $H_1(x, \vartheta)$  — функція розподілу основної компоненти,  $\vartheta \in \Theta \in \mathbb{R}$  — невідомий параметр,  $H_2$  — функція розподілу домішки (вважається повністю невідомою). Задача полягає в тому, щоб оцінити  $\vartheta$  за спостереженнями  $\Xi_N$ .

Для зручності позначень введемо випадкові величини  $\eta_1$  з ф.р.  $H_1(\cdot, \vartheta)$  і  $\eta_2$  з ф.р.  $H_2$ .

Ми будемо припускати, що

$$\Delta_N = \langle (w)^2 \rangle_N - (\langle w \rangle_N)^2 > c > 0 \quad (5.6)$$

для деякого  $c$  і всіх  $N$ . Для двоконпонентних сумішей  $\Delta_N = \det \Gamma_N$ , тому ця умова еквівалентна тому, що концентрації  $w_{j:N}$  не вироджуються у константу при  $N \rightarrow \infty$ .

Для оцінювання  $\vartheta$  можна застосувати відповідним чином модифікований метод моментів. Для цього задамо "пробну функцію"  $g : \mathbb{R} \rightarrow \mathbb{R}$  і розглянемо

$$G_g(t) := G(t) := \mathbf{E}_t g(\eta_1) = \int g(x) H_1(dx, t)$$

— функціональний момент основної компоненти, що відповідає значенню невідомого параметру  $\vartheta = t$ . Згідно з п. 3.1, хорошою оцінкою для  $G(\vartheta)$  є

зважений вибіркового функціональний момент

$$\hat{g}_N = \frac{1}{N} \sum_{j=1}^N a_{j:N} g(\xi_{j:N}),$$

де

$$a_{j:N} = a_{j:N}^1 = \frac{1}{\Delta_N} [(1 - \langle w \rangle_N) w_{j:N} + \langle (w)^2 \rangle_N - \langle w \rangle_N] \quad (5.7)$$

— мінімакський набір вагових коефіцієнтів для оцінювання розподілу основної компоненти (див. (2.10)).

Прирівнюючи теоретичний момент з невідомим значенням параметру  $t$  до емпіричного, отримуємо оцінку методу моментів:

$$\hat{\vartheta}_N(g) := \hat{\vartheta}_N := G^{-1}(\hat{g}_N), \quad (5.8)$$

де  $G^{-1}$  — функція, обернена до  $G$ . (Зрозуміло, що існування  $G^{-1}$  необхідне для того, щоб можна було визначити оцінку методу моментів з даною пробною функцією  $g$ ).

**Приклади. 1.** Нехай  $H_1$  — експоненційний розподіл з щільністю розподілу  $h_1(x, \vartheta) = \vartheta e^{-\vartheta x} \mathbb{I}\{x > 0\}$ ,  $\vartheta \in \Theta = (0, +\infty)$ . Тоді  $E_t \eta_1 = \frac{1}{t}$ , на роль пробної функції можна взяти  $g(x) = x$ , а відповідною оцінкою буде

$$\hat{\vartheta}_N = \left( \frac{1}{N} \sum_{j=1}^N a_{j:N} \xi_{j:N} \right)^{-1}.$$

**2.** Нехай  $H_1$  — “зрізаний експоненційний” розподіл з щільністю

$$h_1(x, \vartheta) = \frac{\vartheta e^{-\vartheta x}}{1 - e^{-\vartheta T}} \mathbb{I}\{x \in (0, T)\}, \quad (5.9)$$

$\vartheta \in \Theta = (0, +\infty)$  — невідомий параметр,  $T$  вважаємо відомим. Вочевидь, при такому розподілі основної компоненти можна вважати, що розподіл домішки теж зосереджений на  $[0, T]$  (оскільки всі спостереження за межами цього інтервалу запевне не належать основній компоненті і їх можна просто відкинути).

Виберемо пробну функцію  $g(x) = x$ . Тоді

$$G(\vartheta) = \frac{1 - e^{\vartheta T} + \vartheta T}{\vartheta(1 - e^{\vartheta T})}.$$

Ця функція є монотонно спадною і неперервною, тому у неї існує обернена і оцінку можна задавати (5.8).

**Асимптотика моментних оцінок.** Дослідимо поведінку моментних оцінок при зростанні обсягу вибірки.

**Теорема 5.2.1 (консистентність)** *Нехай*

- (i) існують  $E|g(\eta_i)| < \infty$  для  $i = 1, 2$ .
  - (ii) Виконана умова (5.6).
  - (iii) Функція  $G^{-1}$  існує і є неперервною в точці  $G(\vartheta)$ .
- Тоді  $\hat{\vartheta}_N \rightarrow \vartheta$  за ймовірністю.

**Доведення.** За теоремою 3.1.1, з умов (i)-(ii) випливає, що  $\hat{g}_N \rightarrow E g(\eta_1) = G(\vartheta)$ . Звідси, враховуючи неперервність  $G^{-1}$ , отримуємо твердження теореми.

**Теорема 5.2.2** *Нехай виконані наступні умови.*

- (i) Існують  $E(g(\eta_i))^2 < \infty$  для  $i = 1, 2$ .
- (ii) Виконана умова (5.6).
- (iii) Існує неперервна похідна  $G'(t) = \frac{dG(t)}{dt}$ , рівномірно відокремлена від 0 для всіх  $t \in \Theta$  (тобто або  $G'(t) > c > 0$  для всіх  $t \in \Theta$ , або  $G'(t) < c < 0$  для всіх  $t \in \Theta$ ).

Тоді розподіл  $\sqrt{N}(\hat{\vartheta}_N - \vartheta)/s_{\vartheta,N}$  слабо збігається до стандартного нормального розподілу.

Тут  $s_{\vartheta,N}^2 := \sigma_{\vartheta,N}^2 / (G'(\vartheta))^2$ ,

$$\begin{aligned} \sigma_{\vartheta,N}^2 := \sigma_{\vartheta,N}^2(g) := & \langle (a)^2 w \rangle_N E(g(\eta_1))^2 + \langle (a)^2 (1-w) \rangle_N E(g(\eta_2))^2 \\ & - [\langle (a)^2 (w)^2 \rangle_N (E g(\eta_1))^2 + 2 \langle (a)^2 w(1-w) \rangle_N E g(\eta_1) E g(\eta_2) \\ & + \langle (a)^2 (1-w)^2 \rangle_N (E g(\eta_2))^2]. \end{aligned} \quad (5.10)$$

**Наслідок 5.2.1** *Якщо виконані умови (i) та (iii) теореми 5.2.2, і умова (ii') Існують границі  $\langle (w)^k \rangle$  для  $k = 1, 2, 3, 4$ , і*

$$\Delta = \langle (w)^2 \rangle - (\langle w \rangle)^2 \neq 0,$$

то  $\sqrt{N}(\hat{\vartheta}_N - \vartheta)$  слабо збігається до нормального розподілу з нульовим середнім і дисперсією  $s_{\vartheta}^2(g) := s_{\vartheta}^2 := \sigma_{\vartheta}^2 / (G'(\vartheta))^2$ , де  $\sigma_{\vartheta}^2$  визначається за формулою (5.10) з заміною  $\langle \cdot \rangle_N$  на  $\langle \cdot \rangle$ .

**Зауваження.** Всі коефіцієнти у формулі (5.10), які залежать від  $a$ , виражаються через  $\langle (w)^k \rangle$ ,  $k = 1, 2, 3, 4$ , за допомогою (5.7).

**Доведення теореми 5.2.2.** Зауважимо, що з умови (iii) випливає виконання умови (iii) теореми 5.2.1. Тому оцінка  $\hat{\vartheta}_N$  є консистентною. При виконанні умов (i) та (ii) за теоремою 3.1.2 розподіли випадкових величин  $\sqrt{N}(\hat{g}_N - G(\vartheta))/\sigma_{\vartheta,N}$  слабо збігаються до стандартного нормального розподілу. Використовуючи теорему 7.3.9, з урахуванням (iii), отримуємо твердження теореми.

**Доведення наслідку 5.2.1.** З (ii') випливає виконання умови (5.6) а також збіжність  $\sigma_{\vartheta,N}^2 \rightarrow \sigma_{\vartheta}^2$ . Використовуючи теорему Слуцького 7.3.10, отримуємо твердження наслідку.

**Оптимальна пробна функція.** Згідно з наслідком 5.2.1, найкращою серед всіх пробних функцій  $g$  буде та, для якої коефіцієнт розсіювання (гранична дисперсія) оцінки  $s_{\vartheta}^2(g)$  буде найменшою. Позначимо цю функцію  $g^*$ . Визначимо, який вигляд матиме  $g^*$ , в припущенні, що  $\vartheta$  та  $H_2$  — відомі. Будемо також припускати, що існують щільності розподілів  $h_1^{\vartheta}(x) = \frac{\partial H_1(x, \vartheta)}{x}$ ,  $h_2(x) = \frac{\partial H_2(x)}{x}$ . Додаткові умови на ці функції ми будемо накладати далі в ході побудови  $g^*$ .

Помітимо, що при переході від пробної функції  $g^0(x)$  до

$$g(x) = \alpha g^0(x) + \beta, \quad (5.11)$$

(де  $\alpha$  та  $\beta$  — довільні дійсні числа) моментна оцінка, по суті, не змінюється, оскільки зміни  $\hat{g}_N$  компенсуються відповідними змінами  $G^{-1}$ . Тому  $s_{\vartheta}^2(g) = s_{\vartheta}^2(g^0)$  (це легко перевірити і безпосередньо за формулою (5.10)).

Отже, можна шукати оптимальну пробну функцію лише в класі певним чином нормованих та центрованих функцій  $g$ . Виберемо  $\alpha$  та  $\beta$  так, щоб виконувались рівності:

$$G(\vartheta) = \int g(x) h_1^{\vartheta}(x) dx = 0 \quad (\text{тобто } \mathbb{E} g(\eta_1) = 0) \quad (5.12)$$

та

$$G'(\vartheta) = \frac{\partial}{\partial \vartheta} \int g(x) h_1^{\vartheta}(x) dx = \int g(x) \dot{h}_1^{\vartheta}(x) dx = 1. \quad (5.13)$$

(Тут ми позначили  $\dot{h}_1^{\vartheta}(x) = \frac{\partial}{\partial \vartheta} h_1^{\vartheta}(x)$ ). Існування цієї похідної та можливість перестановки інтегрування і диференціювання долучимо до умов на розподіли компонент).

Точніше кажучи, для того, щоб лінійним перетворенням (5.11) з будь-якої функції  $g^0$  (не рівної константі) можна було отримати  $g$ , що задовольняє (5.12)-(5.13), потрібно, щоб функції  $1$ ,  $h_1^{\vartheta}$ ,  $\dot{h}_1^{\vartheta}$  не були компланарними:

ні при яких  $\alpha, \beta \in \mathbb{R}$  не повинно виконуватись  $h_1^\vartheta - \alpha \dot{h}_1^\vartheta - \beta = 0$  майже всюди.

При такому обмеженні на функції  $g$  наша задача зводиться до мінімізації функціоналу

$$\begin{aligned} \sigma_\vartheta^2(g) = & \langle (a)^2 w \rangle \int (g(x))^2 h_1^\vartheta(x) dx + \langle (a)^2 (1-w) \rangle \int (g(x))^2 h_2(x) dx \\ & - \langle (a)^2 (1-w)^2 \rangle \left( \int g(x) h_2(x) dx \right)^2 \end{aligned} \quad (5.14)$$

за умов

$$\begin{cases} \int g(x) h_1^\vartheta(x) dx = 0 \\ \int g(x) \dot{h}_1^\vartheta(x) dx = 1 \end{cases} \quad (5.15)$$

Помітимо, що  $\sigma_\vartheta^2(g)$  являє собою невід'ємно визначену квадратичну форму в лінійному просторі відповідних функцій  $g$  (невід'ємна визначеність випливає з того, що  $\sigma_\vartheta^2(g)$  є граничною дисперсією  $\hat{g}_N$ ). Умови (5.15) виділяють афінний підпростір цього простору. Отже, мінімум  $\sigma_\vartheta^2(g)$  завжди досягається, хоча може бути не єдиним. Цей мінімум можна шукати методом множників Лагранжа.

Позначимо

$$I = \int g(x) h_2(x) dx, \quad (5.16)$$

$$z(x) = \langle (a)^2 w \rangle h_1^\vartheta(x) + \langle (a)^2 (1-w) \rangle h_2(x),$$

$$\gamma = \langle (a)^2 (1-w)^2 \rangle.$$

Функція Лагранжа має вигляд

$$\mathcal{L}(g) = \int (g(x))^2 z(x) dx - \gamma \left( \int g(x) h_2(x) dx \right)^2 + \lambda_1 \int g(x) h_1^\vartheta(x) dx + \lambda_2 \int g(x) \dot{h}_1^\vartheta(x) dx,$$

де  $\lambda_1, \lambda_2$  — множники Лагранжа. Диференціал  $\mathcal{L}$ :

$$\delta \mathcal{L}(g) = \int (2g(x)z(x) + \lambda_1 h_1^\vartheta(x) + \lambda_2 \dot{h}_1^\vartheta(x) - 2\gamma I h_2(x)) \delta(x) dx.$$

У стаціонарній точці  $\delta \mathcal{L}(g) = 0$  для всіх можливих приростів аргументу  $\delta(x)$ . Отже, екстремум досягається у точці

$$g^*(x; \vartheta, h_2) := g^*(x) := \frac{\lambda_1 h_1^\vartheta(x) + \lambda_2 \dot{h}_1^\vartheta(x) + \gamma I h_2(x)}{\langle (a)^2 w \rangle h_1^\vartheta(x) + \langle (a)^2 (1-w) \rangle h_2(x)}, \quad (5.17)$$



де  $\lambda_1$ ,  $\lambda_2$  та  $I$  знаходяться з системи рівнянь, які відповідають умовам (5.15)-(5.16):

$$\begin{cases} \lambda_1 \int \frac{(h_1^\vartheta(x))^2}{z(x)} dx + \lambda_2 \int \frac{\dot{h}_1^\vartheta(x) h_1^\vartheta(x)}{z(x)} dx + I\gamma \int \frac{h_2(x) h_1^\vartheta(x)}{z(x)} dx = 0 \\ \lambda_1 \int \frac{h_1^\vartheta(x) \dot{h}_1^\vartheta(x)}{z(x)} dx + \lambda_2 \int \frac{(\dot{h}_1^\vartheta(x))^2}{z(x)} dx + I\gamma \int \frac{h_2(x) \dot{h}_1^\vartheta(x)}{z(x)} dx = 1 \\ \lambda_1 \int \frac{h_1^\vartheta(x) h_2(x)}{z(x)} dx + \lambda_2 \int \frac{\dot{h}_1^\vartheta(x) h_2(x)}{z(x)} dx + I\gamma \int \frac{(h_2(x))^2}{z(x)} dx = I \end{cases} \quad (5.18)$$

Функція  $g^*$ , якщо вона задовольняє умови теореми 5.2.2, є оптимальною пробною функцією для оцінки  $\vartheta$  методом моментів.

**Адаптивна оцінка.** Зрозуміло, що не знаючи справжніх  $\vartheta$  та  $h_2$ , неможливо безпосередньо використати оптимальну пробну функцію (5.17) для оцінювання. Вихід може полягати у застосуванні адаптивного підходу.

Для побудови адаптивної оцінки ми на першому кроці оцінюємо  $\vartheta$  та  $h_2$  за допомогою грубих “пілотних” оцінок  $\tilde{\vartheta}_N$  та  $\tilde{h}_{2,N}$  і підставляємо значення цих оцінок в (5.17). Отримана функція  $g_N^*(x) := g^*(x; \tilde{\vartheta}_N, \tilde{h}_{2,N})$  використовується як пробна функція для моментної оцінки:  $\vartheta_N^* = G_{g_N^*}^{-1}(\hat{g}_N^*)$ , де  $\hat{g}_N^* = \frac{1}{N} \sum_{j=1}^N a_{j:N} g_N^*(\xi_{j:N})$ .

При цьому виникають два ускладнення.

По-перше, зовсім не очевидно, чи буде існувати функція  $G_{g_N^*}^{-1}$  для обраної нами пробної функції  $g_N^*$ , а коли так, то як її обчислювати.

По-друге, оскільки  $g_N^*$  є випадковою функцією, залежною від даних, теорему 5.2.2 не можна застосовувати до відповідної оцінки. Тому асимптотична поведінка оцінки вимагає додаткового аналізу.

Перше ускладнення можна обійти, розв’язуючи рівняння

$$G_{g_N^*}(t) = \hat{g}_N^* \quad (5.19)$$

наближено методом Ньютона з початковим наближенням  $\tilde{\vartheta}_N$ . Дійсно, розкладаючи  $G_{g_N^*}(t)$  у ряд в околі  $\tilde{\vartheta}_N$ , отримуємо, що рівняння (5.19) приблизно еквівалентне

$$G_{g_N^*}(\tilde{\vartheta}_N) + G'_{g_N^*}(\tilde{\vartheta}_N)(t - \tilde{\vartheta}_N) = \hat{g}_N^*. \quad (5.20)$$

Оскільки функція  $g_N^*$  задовольняє умови (5.12)-(5.13) з  $\vartheta = \tilde{\vartheta}_N$ , то  $G_{g_N^*}(\tilde{\vartheta}_N) = 0$  і  $G'_{g_N^*}(\tilde{\vartheta}_N) = 1$ . Отже розв’язок (5.20) має вигляд

$$\tilde{\vartheta}_N = \tilde{\vartheta}_N + \hat{g}_N^*. \quad (5.21)$$

Цю оцінку ми і будемо назвати адаптивною оцінкою методу моментів для  $\vartheta$ .

Визначимо тепер умови, за яких  $\check{\vartheta}_N$  буде асимптотично нормальною з оптимальним коефіцієнтом розсіювання  $s_{\check{\vartheta}}^2(g^*)$ . Відповідне твердження ми сформулюємо для трохи більш загальної ситуації адаптивного оцінювання, ніж розглянута вище.

Нехай  $\mathcal{A}$  — деякий вимірний простір,  $g : \mathbb{R} \times \mathcal{A} \rightarrow \mathbb{R}$  — фіксована не випадкова вимірна функція. Визначимо  $\hat{g}(\alpha) = \frac{1}{N} \sum_{j=1}^N a_{j:N} g(\xi_{j:N}, \alpha)$ ,  $G(\vartheta, \alpha) = \mathbb{E}_{\vartheta} g(\eta_1, \alpha) = \int g(x, \alpha) h_1^{\vartheta}(x) dx$ .

Припустимо, що у нас є пілотна оцінка  $\tilde{\vartheta}_N$  для  $\vartheta$  і послідовність випадкових елементів  $\mathcal{A}$  (оцінок)  $\alpha_N$ , яка наближається до не випадкового елемента  $\alpha_{\infty}$ . При цьому  $\alpha_N \in \mathcal{A}_N$ , де  $\mathcal{A}_N$  — послідовність не випадкових підмножин  $\mathcal{A}$ . Розглянемо оцінку

$$\check{\vartheta}_N = \tilde{\vartheta}_N + \frac{\hat{g}_N(\alpha_N) - G(\tilde{\vartheta}_N, \alpha_N)}{G'(\tilde{\vartheta}_N, \alpha_N)} \quad (5.22)$$

де  $G'(t, \alpha) = \frac{\partial}{\partial t} G(t, \alpha)$ .

(Якщо  $\alpha_N = (\tilde{\vartheta}_N, \hat{h}_{2,N})$ , де  $\hat{h}_{2,N}$  — оцінка для  $h_2$ ,  $g(x, \alpha) = g^*(x, \vartheta_N, \hat{h}_{2,N})$ , то (5.22) перетворюється на (5.21)).

Позначимо через  $\mathcal{F}_N$  клас усіх множин вигляду  $\{x \in \mathbb{R} : g(x, \alpha) - g(x, \alpha_{\infty}) < C\}$  для всіх можливих  $\alpha \in \mathcal{A}_N$ ,  $C \in \mathbb{R}$ . Позначимо  $\nu_N(l) = \nu(l, \mathcal{F}_N)$  функцію зростання класу  $\mathcal{F}_N$ . (Означення функцій зростання див. у п. 2.2).

Нехай

$$\begin{aligned} \sigma_{\infty}^2 &:= \lim_{N \rightarrow \infty} \text{Var} \left( \frac{1}{\sqrt{N}} \sum_{j=1}^N a_{j:N} g(\xi_j, \alpha_{\infty}) \right) \\ &= \langle (a)^2 w \rangle \mathbb{E}(g(\eta_1, \alpha_{\infty}))^2 + \langle (a)^2 (1-w) \rangle \mathbb{E}(g(\eta_2, \alpha_{\infty}))^2 \\ &\quad - [\langle (a)^2 (w)^2 \rangle (\mathbb{E} g(\eta_1, \alpha_{\infty}))^2 + 2 \langle (a)^2 w(1-w) \rangle \mathbb{E} g(\eta_1, \alpha_{\infty}) \mathbb{E} g(\eta_2, \alpha_{\infty}) \\ &\quad + \langle (a)^2 (1-w)^2 \rangle (\mathbb{E} g(\eta_2, \alpha_{\infty}))^2]. \end{aligned}$$

**Теорема 5.2.3** *Нехай виконані наступні умови.*

- (i) *Існують  $\mathbb{E}(g(\eta_i, \alpha_{\infty}))^2 < \infty$  для  $i = 1, 2$ .*
- (ii) *Існують границі  $\langle (w)^k \rangle$  для  $k = 1, 2, 3, 4$ , і*

$$\Delta = \langle (w)^2 \rangle - (\langle w \rangle)^2 \neq 0,$$

- (iii)  *$\sup_N \mathbb{P}\{\sqrt{N}(\tilde{\vartheta}_N - \vartheta) > c\} \rightarrow 0$  при  $c \rightarrow \infty$*

(iv)  $G'(\vartheta, \alpha_\infty) \neq 0$ ,  $G'(t, \alpha)$  є неперервною по  $t$  в деякому околі  $\vartheta$  для всіх  $\alpha \in \mathcal{A}$  і  $G'(t_N, \alpha_N) \rightarrow G'(\vartheta, \alpha_\infty)$  при  $N \rightarrow \infty$  за ймовірністю, для всіх випадкових послідовностей  $t_N$ , таких, що  $t_N \rightarrow \vartheta$  за ймовірністю.

(v) Для деякої не випадкової послідовності  $\delta_N \rightarrow 0$ ,

$$\frac{1}{\delta_N} \sup_{x \in \mathbb{R}} |g(x, \alpha_N) - g(x, \alpha_\infty)| \rightarrow 0$$

при  $N \rightarrow \infty$  за ймовірністю.

(vi)  $\ln \nu_N(2N) = o(\delta_N^{-2})$ .

Тоді  $\sqrt{N}(\tilde{\vartheta}_N - \vartheta)$  слабо збігається до нормального розподілу з нульовим середнім і дисперсією  $\sigma_\infty^2 / (G'(\vartheta, \alpha_\infty))$ .

Теорема фактично стверджує, що коефіцієнт розсіювання адаптивної оцінки  $\tilde{\vartheta}_N$ , визначеної (5.22), буде таким самим, як у моментної оцінки  $g(\cdot, \alpha_\infty)$ . Доведення теореми див. далі, наприкінці параграфу.

Розглянемо застосування цієї теореми для аналізу даних з прикладу 2 (зі зрізаним експоненційним розподілом  $H_1$ ). На роль пілотної оцінки  $\tilde{\vartheta}_N$  для  $\vartheta$  можна використати моментну оцінку з пробною функцією  $g(x) = x$ . Як оцінку для  $h_2$  використаємо зважену гістограму. Точніше, задамо  $K_N$  — кількість підінтервалів розбиття інтервалу  $[0, T)$ , покладемо  $t_k = kT/K_N$ ,  $k = 0, \dots, K_N$ ,  $A_k = [t_{k-1}, t_k)$  —  $k$ -тий підінтервал розбиття. Тоді зважена гістограма  $\hat{h}_{2,N}$  визначається як

$$\hat{h}_{2,N}(x) = \frac{K_N}{NT} \sum_{j=1}^N \sum_{k=1}^{K_N} a_{j:N}^2 \mathbb{I}\{\xi_{j:N} \in A_k\}, \quad (5.23)$$

де  $a_{j:N}^2$  — мінімаксні вагові коефіцієнти для оцінювання другої компоненти (домішки):

$$a_{j:N}^2 = (\langle (w)^2 \rangle_N - \langle w \rangle_N w_{j:N}) / \Delta_N. \quad (5.24)$$

**Наслідок 5.2.2** *Нехай виконуються наступні умови.*

(i) Основна компонента має зрізаний експоненційний розподіл (5.9).

(ii) Виконана умова (ii) теореми 5.2.3.

(iii) У розподілу домішки існує щільність  $h_2$ , яка є неперервно диференційовною функцією на  $[0, T]$ .

(iv)  $K_N = CN^\beta$  для деяких  $C > 0$ ,  $0 < \beta < 1/4$ .

(v) Пілотна оцінка  $\tilde{\vartheta}_N$  є моментною оцінкою з пробною функцією  $g(x) = x$ .

Тоді виконано твердження теореми 5.2.3.

У випадку необмеженого носія розподілу  $H_1$ , як от — експоненційного, у прикладі 1, можна використовувати гістограму на інтервалі, що розширюється зі збільшенням обсягу вибірки, скажімо, на  $[0, T_N)$ . Якщо  $T_N$  прямує до нескінченності досить повільно, щоб забезпечити виконання умов (v) та (vi) теореми 5.2.3, і, в той же час, досить швидко, щоб забезпечити  $P\{\sup_j \xi_{j:N} > T_N\} \rightarrow 0$  при  $N \rightarrow \infty$ , то твердження теореми буде виконуватись. Зокрема, у схемі прикладу 1, якщо хвіст розподілу  $H_2$  субекспоненційний, (тобто  $H_2([x, +\infty)) < Ce^{-\alpha x}$  для деяких  $C, \alpha > 0$ ) то можна покласти  $T_N = C(\ln N)^{\ln N}$ .

**Доведення теорем.**

Для доведення теореми 5.2.3 нам будуть потрібні допоміжні леми.

**Лема 5.2.1** *Нехай  $\mathcal{F}$  — деякий клас множин на  $\mathbb{R}$  з функцією зростання  $\nu^{\mathcal{F}}(l)$ , виконана умова (5.6), а  $i$  задано (5.7) або (5.24). Тоді для деяких  $\lambda_0, C$  і  $\alpha$  при  $\lambda > \lambda_0/N$ ,*

$$P\{\sup_{A \in \mathcal{F}} |\hat{H}_i^N(A) - H_i(A)| > \lambda\} \leq C\nu^{\mathcal{F}}(2N) \exp(-\alpha\lambda^2 N),$$

причому константи  $\lambda_0 > 0, C < \infty$  і  $\alpha > 0$  залежать лише від концентрацій  $w_{j:N}$  і не залежать від  $H_i, \mathcal{F}$  і  $\lambda$ .

Ця лема є тривіальним наслідком нерівності Валника-Червоненкіса — теореми 2.2.4.

Нехай  $\mathcal{G}$  — клас функцій  $g : \mathbb{R} \rightarrow \mathbb{R}$ . Позначимо  $\mathcal{F}_{\mathcal{G}} = \{\{x \in \mathbb{R} : g(x) < c\}, \forall c \in \mathbb{R}, \forall g \in \mathcal{G}\}$ .

**Лема 5.2.2** *Нехай  $\mathcal{G}$  — деякий клас вимірних обмежених функцій  $f : \mathbb{R} \rightarrow \mathbb{R}, K = \sup_{x \in \mathbb{R}, f \in \mathcal{G}} |f(x)|$ . Тоді*

$$\sup_{f \in \mathcal{G}} \left| \frac{1}{N} \sum_{j=1}^N f(\xi_{j:N}) a_{j:N}^i - \int f(x) H_i(dx) \right| \leq 2K \sup_{A \in \mathcal{F}_{\mathcal{G}}} |\hat{H}_i^N(A) - H_i(A)|.$$

**Доведення.** Зафіксуємо  $f \in \mathcal{G}$ . Для  $n \in \mathbb{N}$  і  $j = \pm 1, \pm 2, \dots, \pm n$  позначимо

$$A_j^n = \left\{ x \in \mathbb{R} : f(x) \leq \frac{Kj}{n} \right\}.$$

Враховуючи, що  $|f(x)| \leq K$ , за означенням інтегралу Лебега маємо

$$\int f(x) H_i(dx) = \lim_{n \rightarrow \infty} \sum_{j=-n}^n \frac{Kj}{n} H_i(A_j^n \setminus A_{j-1}^n)$$

і, аналогічно,

$$\frac{1}{N} \sum_{j=1}^N f(\xi_{j:N}) a_{j:N}^i = \lim_{n \rightarrow \infty} \sum_{j=-n}^n \frac{Kj}{n} \hat{H}_i^N(A_j^n \setminus A_{j-1}^n).$$

Тому

$$\begin{aligned} & \sup_{f \in \mathcal{G}} \left| \frac{1}{N} \sum_{j=1}^N f(\xi_{j:N}) a_{j:N}^i - \int f(x) H_i(dx) \right| \\ & \leq \lim_{n \rightarrow \infty} \sum_{j=-n}^n \frac{Kj}{n} |H_i(A_j^n \setminus A_{j-1}^n) - \hat{H}_i^N(A_j^n \setminus A_{j-1}^n)| \\ & \leq \lim_{n \rightarrow \infty} \sum_{j=-n}^n \frac{Kj}{n} (|H_i(A_j^n) - \hat{H}_i^N(A_j^n)| + |H_i(A_{j-1}^n) - \hat{H}_i^N(A_{j-1}^n)|) \\ & \leq 2K \sup_{A \in \mathcal{F}_G} |\hat{H}_i^N(A) - H_i(A)|. \end{aligned}$$

Внаслідок довільності  $f$  отримуємо твердження леми.

Лема доведена.

**Доведення теореми 5.2.3.** Помітимо, що

$$\begin{aligned} \sqrt{N}(\check{\vartheta}_N - \vartheta) &= \sqrt{N}(\tilde{\vartheta}_N - \vartheta) + \frac{\sqrt{N}(\hat{g}_N(\alpha_N) - G(\vartheta, \alpha_N))}{G'(\tilde{\vartheta}_N, \alpha_N)} \\ &\quad + \frac{\sqrt{N}(G(\vartheta, \alpha_N) - G(\tilde{\vartheta}_N, \alpha_N))}{G'(\tilde{\vartheta}_N, \alpha_N)}. \end{aligned}$$

Враховуючи, що  $G(\vartheta, \alpha_N) - G(\tilde{\vartheta}_N, \alpha_N) = G'(\zeta_N, \alpha_N)(\vartheta - \tilde{\vartheta}_N)$ , де  $\zeta_N$  — проміжна точка між  $\vartheta$  і  $\tilde{\vartheta}_N$ , отримуємо

$$\sqrt{N}(\check{\vartheta}_N - \vartheta) = J_1 + J_2/G'(\tilde{\vartheta}_N, \alpha_N) + J_3,$$

де

$$J_1 = \sqrt{N}(\tilde{\vartheta}_N - \vartheta) \left( 1 - \frac{G'(\zeta_N, \alpha_N)}{G'(\tilde{\vartheta}_N, \alpha_N)} \right),$$

$$J_2 = \sqrt{N}(\hat{g}_N(\alpha_N) - G(\vartheta, \alpha_N) - \hat{g}_N(\alpha_\infty) + G(\vartheta, \alpha_\infty)),$$

$$J_3 = (\hat{g}_N(\alpha_\infty) - G(\vartheta, \alpha_\infty))/G'(\vartheta_N, \alpha_N).$$

За умовами (iii) та (iv)  $J_1 \rightarrow 0$  за ймовірністю. Враховуючи (i) та (ii), за теоремою 3.1.2 отримуємо, що  $J_3$  збігається слабо до нормального розподілу з нульовим середнім та дисперсією  $\sigma_\infty^2/(G'(\vartheta, \alpha_\infty))$ .

Залишилось показати, що  $J_2 \rightarrow 0$  за ймовірністю. Зробимо це. Позначимо  $f(x; \alpha) = g(x, \alpha_\infty) - g(x, \alpha)$  і розглянемо клас функцій  $\mathcal{G}_N := \{f(\cdot; \alpha), \alpha \in \mathcal{A}_N\}$ .

Тоді

$$\hat{g}_N(\alpha_\infty) - \hat{g}_N(\alpha) = \frac{1}{N} \sum_{j=1}^N a_{j:N} (g(\xi_{j:N}, \alpha_\infty) - g(\xi_{j:N}, \alpha)) = \frac{1}{N} \sum_{j=1}^N a_{j:N} f(\xi_{j:N}),$$

$$G(\vartheta, \alpha_\infty) - G(\vartheta, \alpha) = \int f(x, \alpha) H_1(dx, \alpha).$$

Тому, за лемою 5.2.2, для будь-яких  $\lambda$ ,

$$\begin{aligned} p_N := \mathbb{P}\{|J_2| > \lambda\} &\leq \mathbb{P}\left\{\sup_x |g(x, \alpha_N) - g(x, \alpha_\infty)| > \delta_N\right\} \\ &+ \mathbb{P}\left\{\sup_{A \in \mathcal{F}_G} |\hat{H}_{1,N}(A) - H_1(A)| > \frac{\lambda \delta^{-1}}{2\sqrt{N}}\right\}. \end{aligned}$$

За умовою (v), перший доданок праворуч прямує до 0, а за лемою 5.2.1 та умовою (vi) другий не перевищує

$$C \nu_N(2N) \exp(-\alpha \lambda^2 \delta_N^{-2}/4) \leq C \exp(-\alpha \lambda^2 \delta_N^{-2}/4 + \ln \nu_N(2N)) \rightarrow 0.$$

Теорема доведена.

Для доведення наслідку 5.2.2, нам буде потрібна наступна оцінка швидкості збіжності зваженої гістограми  $\hat{h}_{2,N}$ , визначеної (5.23), до оцінюваної щільності.

**Лема 5.2.3** *Нехай розподіли компонент зосереджені на  $[0, T]$  і*

(i)  $h_2$  є неперервно диференційовною на  $[0, T]$ ;

(ii) Виконана умова (ii) теореми 5.2.3;

(iii) Для деяких  $0 < C_i, \beta, \gamma < \infty$ ,  $\delta_N = C_1 N^{-\beta}$ ,  $K_N = C_2 N^\gamma$ .

Тоді, якщо  $\beta < \gamma$ ,  $\beta + \gamma < 1/2$ , то

$$\frac{1}{\delta_N} \sup_{x \in [0, T]} |\hat{h}_{2,N}(x) - h_2(x)| \rightarrow 0$$

за ймовірністю.

**Доведення.** Позначимо  $\bar{h}_k = K_N H_2(A_k)/T$ ,

$$\bar{h}_{2,N}(x) = \mathbb{E} \hat{h}_{2,N} = \sum_{k=1}^{K_N} \bar{h}_k \mathbb{I}\{x \in A_k\},$$

$$h' = \sup_{x \in [0, T]} \frac{d}{dx} h_2(x).$$

Фіксуємо довільне  $\lambda > 0$ . Твердження леми еквівалентне тому, що

$$J_N := \mathbb{P}\left\{ \sup_{x \in [0, T]} |\hat{h}_{2,N}(x) - h_2(x)| > \lambda \delta_N \right\} \rightarrow 0$$

при  $N \rightarrow \infty$ . Легко бачити, що  $J_N \leq J_N^1 + J_N^2$ , де

$$J_N^1 = \mathbb{P}\left\{ \sup_{x \in [0, T]} |\hat{h}_{2,N}(x) - \bar{h}_{2,N}(x)| > \lambda \delta_N / 2 \right\},$$

$$J_N^2 = \mathbb{P}\left\{ \sup_{x \in [0, T]} |\bar{h}_{2,N}(x) - h_2(x)| > \lambda \delta_N / 2 \right\}.$$

Нехай  $x \in A_k$ . Оцінимо

$$\begin{aligned} |\bar{h}_{2,N}(x) - h_2(x)| &\leq \frac{K_N}{T} \int_{t_{k-1}}^{t_k} |h_2(x) - h_2(t)| dt \leq \frac{K_N}{T} h' \int_{t_{k-1}}^{t_k} |x - t| dt \\ &\leq \frac{h' T}{2 K_N} = \frac{T h'}{2 C_2} N^{-\gamma} < C_1 N^{-\beta} \end{aligned}$$

при достатньо великих  $N$ . Отже, при великих  $N$ ,  $J_2^N = 0$ .

За лемою 5.2.1,

$$\begin{aligned} J_1^N &\leq \sum_{k=1}^{K_N} \mathbb{P}\left\{ \frac{K_N}{T} |\hat{H}_{2,N}(A_k) - H_2(A_k)| \geq \lambda \delta_N / 2 \right\} \\ &\leq C K_N \exp\left(-\alpha \left(\frac{T \lambda \delta_N}{2 K_N}\right)^2 N\right) = C N^\gamma \exp\left(-\alpha \left(\frac{T \lambda}{2}\right)^2 N^{1-2\beta-2\gamma}\right) \rightarrow 0 \end{aligned}$$

при  $N \rightarrow \infty$ . Отже і  $J_1 \rightarrow 0$ .

Лема доведена.

**Доведення наслідку 5.2.2.** перевіримо виконання умов теореми 5.2.3. Позначимо  $t \in \mathcal{T} := [\vartheta - \varepsilon, \vartheta + \varepsilon]$ , де  $\vartheta$  — справжнє значення параметру,  $\varepsilon$  — будь-яке додатне число, таке, що  $\vartheta - \varepsilon > 0$ . При  $N \rightarrow 0$ ,  $\mathbb{P}\{\tilde{\vartheta}_N \in \mathcal{T}\} \rightarrow 1$ , тому можна вважати, що множина можливих значень параметру  $\Theta = \mathcal{T}$ .

Тоді в.в.  $g^*(\eta_i, h_2(\eta_i), t)$  є обмеженими для всіх  $t \in \mathcal{T}$ , тому (i) — виконано. Умова (iii) виконана внаслідок теореми 5.2.2. Умова (ii) — внаслідок неперервності  $G'$  і збіжності  $\tilde{\vartheta}_N \rightarrow \vartheta$ ,  $\hat{h}_{2,N} \rightarrow h_2$ .

Покладемо  $\delta_N = N^{-\gamma}$ , де  $\gamma = 3\beta/2$  при  $\beta < 1/6$  і  $\gamma = 1/4$  при  $1/6 \leq \beta \leq 1/4$ . Виконання умови (iv) випливає з леми 5.2.3, теореми 5.2.2 і гладкості  $g^*(x, h, t)$  по  $h$  і  $t$  при  $h > 0$ ,  $t \in \mathcal{T}$ .

Оцінимо  $\nu_N(2N)$ . Легко бачити, що коли  $t = \text{const}$ ,  $h = \text{const}$ ,  $g^*(x, h, t)$  як функція  $x \in (0, +\infty)$  може мати не більше ніж  $C$  інтервалів монотонності, де  $C$  — деяке фіксоване число, що не залежить від  $h$  та  $t$ . Тому  $g^*(x, \hat{h}_{2,N}, t)$  має не більше ніж  $CK_N$  інтервалів монотонності і  $\nu_N(2N) \leq (2N)^{CK_N}$ . Тому

$$\ln \nu_N(2N) \leq CN^\gamma \ln(2N) \leq \delta_N^{-2} = (C_1)^{-2} N^{2\beta},$$

оскільки  $\gamma < 2\beta$ .

Наслідок доведено.



## Розділ 6

# Задачі класифікації

Класифікація об'єктів на групи на основі їх спостережуваних ознак — класична область застосування методів математичної статистики. Задачі класифікації поділяють на дві групи: класифікація з вчителем і без вчителя. При класифікації з вчителем, як правило, вважають, що основою для побудови класифікатора є навчаюча вибірка, у якій спостережувані об'єкти однозначно розкласифіковані за даними класами. У задачах класифікації без вчителя вибірка не розкласифікована і задача дослідника — розбити об'єкти по класах, виходячи з властивостей їх спостережуваних характеристик.

У цьому розділі ми розглядаємо задачу класифікації з “невпевненим” вчителем, який не може вказати точно, якому класу належить об'єкт з навчаючої вибірки, але здатен визначити апріорні ймовірності його належності до кожного з можливих класів. Таким чином, навчаюча вибірка є вибіркою зі змінними концентраціями.

## 6.1 Баєсова класифікація

Загальна задача класифікації має наступний вигляд. Об'єкт  $O$  може належати одній з  $M$  різних популяцій  $\mathcal{P}_1, \dots, \mathcal{P}_M$ . Номер популяції, якій належить  $O$  (як і раніше, ми позначаємо його  $\text{ind}(O)$ ) невідомий. Спостерігається набір характеристик  $O$  —  $\xi(O)$ . Потрібно за спостережуваними характеристиками  $\xi(O)$  визначити (вгадати, оцінити) до якої популяції належить  $O$ .

Розв'язком задачі класифікації є функція  $g : \mathcal{X} \rightarrow \{1, \dots, M\}$ , яка кожному можливому значенню спостережуваних характеристик  $x \in \mathcal{X}$  ставить у відповідність номер популяції  $g(x)$ , до якої слід віднести об'єкт з такими характеристиками. Функцію  $g$  називають класифікатором.

Прикладом задачі класифікації є задача медичної діагностики. У цій задачі  $O$  — пацієнт, якому потрібно встановити діагноз,  $\mathcal{P}_i$ ,  $i = 1, \dots, M$  — хвороби, якими він в принципі може бути хворий (точніше,  $\mathcal{P}_i$  — популяція всіх людей, хворих  $i$ -тою хворобою),  $\text{ind}(O)$  — номер справжньої хвороби, якою хворіє  $O$ ,  $\xi(O)$  — сукупність даних обстеження, за якими можна встановлювати діагноз (температура, артеріальний тиск, формула крові, кардіограма, рентгенівський знімок —  $\xi(O)$  зовсім не обов'язково має бути набором виключно числових даних). Класифікатор  $g$  у даному випадку — це сукупність правил діагностики, які за даними  $\xi(O)$  дозволяють встановити діагноз, тобто вказати номер хвороби  $g(\xi(O))$ , якою на думку діагноста хворіє  $O$ .

На роль класифікатора можна, взагалі кажучи, використати будь-яку вимірну функцію. Для вибору найкращого класифікатора можна скористатись яким-небудь критерієм якості. Найбільш поширеним серед цих критеріїв є ймовірність помилкової класифікації. Щоб ввести це поняття потрібно припустити, що  $\xi(O)$  є випадковими елементами  $\mathcal{X}$ , а  $\text{ind}(O)$  — випадковою величиною зі значеннями на множині  $\{1, \dots, M\}$ .

Отже, нехай

$p_i = \mathbf{P}\{\text{ind}(O) = i\}$  — “априорна” ймовірність того, що об'єкт, який спостерігається, належить  $i$ -тій популяції,

$H_i(A) = \mathbf{P}\{\xi(O) \in A \mid \text{ind}(O) = i\}$  — розподіл спостережуваних характеристик об'єкта, що належить  $i$ -тій популяції (тут  $A$  — будь-яка вимірна підмножина  $\mathcal{X}$ ).

Тоді, для будь-якого класифікатора  $g$  визначена ймовірність помилки:

$$\begin{aligned} L(g) &= \mathbf{P}\{g(\xi(O)) \neq \text{ind}(O)\} = 1 - \mathbf{P}\{g(\xi(O)) = \text{ind}(O)\} \\ &= 1 - \sum_{i=1}^M p_i \mathbf{P}\{g(\xi(O)) = i \mid \text{ind}(O) = i\} = 1 - \sum_{i=1}^M p_i H_i(A_i), \end{aligned} \quad (6.1)$$

де  $A_i = \{x \in \mathcal{X} : g(x) = i\}$  — множина значень характеристик, при яких класифікатор  $g$  відносить об'єкт до  $i$ -тої популяції<sup>1</sup>.  $L(g)$  називають також баєсовим ризиком класифікатора  $g$ .

Зрозуміло, що чим меншою є ймовірність помилки, тим кращим є класифікатор. Нехай фіксована деяка множина допустимих класифікаторів  $\mathfrak{G}$ . Класифікатор  $g_0$  називають баєсовим в класі  $\mathfrak{G}$ , якщо  $g_0 \in \mathfrak{G}$  і для будь-якого  $g \in \mathfrak{G}$ ,  $L(g_0) \leq L(g)$ . Будемо позначати баєсів класифікатор  $g_{\mathfrak{G}}^B$ .

Класифікатор, баєсів у класі всіх можливих класифікаторів, називають просто “баєсовим” і позначають  $g^B$ .

Надалі всюди у цьому розділі нас буде цікавити можливість побудови класифікаторів, баєсових у деяких класах, або наближених до баєсових.

У випадку, коли апіорні ймовірності  $p_i$  та розподіли характеристик  $H_i$  відомі, побудова баєсового класифікатора не викликає утруднень. Дійсно, нехай  $\nu$  — міра, відносно якої всі розподіли  $H_i$  є абсолютно неперервними<sup>2</sup>. Позначимо  $h_i(x) = \frac{\partial H_i}{\partial \nu}(x)$  — щільність  $h_i$  відносно  $\nu$ , тобто

$$H_i(A) = \int_A h_i(x) \nu(dx).$$

Тоді

$$g^B(x) = \underset{i}{\operatorname{argmax}} p_i h_i(x). \quad (6.2)$$

Доведення цього факту дуже просте. За (6.1), для будь-якого класифікатора  $g$ ,

$$L(g) = 1 - \sum_{i=1}^M p_i \int_{A_i} h_i(x) \nu(dx) = \int_{\mathcal{X}} p_{g(x)} h_{g(x)} \nu(dx). \quad (6.3)$$

Зрозуміло, що

$$1 - \int_{\mathcal{X}} p_{g(x)} h_{g(x)} \nu(dx) \geq 1 - \int_{\mathcal{X}} \max_i (p_i h_i(x)) \nu(dx) = L(g^B).$$

<sup>1</sup>Класифікатор можна навіть формально ввести як розбиття простору  $\mathcal{X}$  на області  $A_i$ . Таке означення буде еквівалентним характеристизації класифікатора функцією  $g$

<sup>2</sup>Така міра завжди існує, наприклад, можна покласти  $\nu = \sum_{i=1}^M H_i$ .

Отже, класифікатор  $g^B$ , визначений (6.2), дійсно є баєсовим.

Відмітимо, що апостеріорні ймовірності  $p_i^*(x) = P\{\text{ind}(O) = i \mid \xi(O) = x\}$  можна обчислювати за формулою Баєса:

$$p_i^*(x) = \frac{p_i h_i(x)}{\sum_{m=1}^M p_m h_m(x)},$$

і, оскільки у цій формулі знаменник не залежить від  $i$ ,

$$g^B(x) = \underset{i}{\operatorname{argmax}} p_i^*(x),$$

тобто баєсів класифікатор — це класифікатор, що обирає популяцію, яка має найбільшу апостеріорну ймовірність. Інколи цю властивість вибирають як означення баєсового класифікатора, але ми будемо дотримуватись початкового означення.

На практиці, як правило, апріорні ймовірності та/або розподіли характеристик є невідомими. У такій ситуації для побудови класифікаторів використовують навчаючі вибірки, тобто вибірки, що складаються з об'єктів, подібних до тих, які потрібно буде класифікувати.

У розглянутому нами прикладі медичної діагностики такою вибіркою може бути набір даних про пацієнтів, які раніше звертались до даної медичної установи, та про діагнози, які їм були поставлені. В такій вибірці будуть міститись значення  $\xi_j = \xi(O_j)$  — спостережуваних характеристик  $j$ -того пацієнта,  $j = 1, \dots, N$  та, наприклад, встановлені пацієнтам діагнози  $i_j$ . Якщо вважати, що  $i_j = \text{ind}(O_j)$  — безпомилковий остаточний діагноз, то отримана навчаюча вибірка  $(\xi_j, i_j, j = 1, \dots, N)$  буде повністю розкласифікованою, тобто розпадеться на  $M$  вибірок, кожна з яких відповідатиме одній популяції (одному діагнозу). По цих вибірках звичайними методами можна оцінити  $h_i$ , а за частотою появи різних діагнозів у навчаючій вибірці оцінити апріорні ймовірності  $p_i$ . Отримані оцінки, скажімо,  $\hat{p}_i$  та  $\hat{h}_i$ , можна підставити замість справжніх значень у формулу для баєсового класифікатора  $g^B$  і отримати класифікатор  $\hat{g}$ , який зветься емпірично-баєсовим:

$$\hat{g}(x) = \underset{i}{\operatorname{argmax}} \hat{p}_i \hat{h}_i(x). \quad (6.4)$$

Такий підхід називають емпірично-баєсовою класифікацією.

Однак для реальних даних часто немає впевненості в тому, що класифікація навчаючої вибірки проведена абсолютно вірно. (Наприклад, як

правило, ми маємо дані не про остаточні, а про попередні діагнози пацієнтів<sup>3</sup>). Замість точних  $i_j$  у цьому випадку природно розглядати деякі ймовірності того, що об'єкт  $O_j$  належить  $m$ -тій популяції, тобто  $P\{\text{ind}(O_j) = m\} = w_{j:N}^m$ . Скажімо, якщо попередній діагноз розглядати як результат певної класифікації на основі обмеженої інформації, то на роль  $w_{j:N}^i$  можна взяти апостеріорні ймовірності цієї попередньої класифікації.

При такій інтерпретації навчаюча вибірка виявляється вибіркою з суміші зі змінними концентраціями. Для побудови емпірично-баєсового класифікатора можна тепер використовувати оцінки щільностей розподілів компонент, описані у розділі 4. Оцінку апіорних ймовірностей можна виконувати по-різному — або з використанням додаткової інформації про конкретного пацієнта (об'єкта), або на основі статистичних даних про поширеність тої чи іншої хвороби (появи об'єктів з різних популяцій). У цьому розділі ми, як правило, будемо вважати апіорні ймовірності відомими (тобто ігноруватимемо помилку, що виникає при їх оцінюванні). Підставляючи отримані оцінки у (6.4), отримуємо емпірично-баєсів класифікатор, побудований по вибірці з суміші зі змінними концентраціями.

Процес побудови класифікатора на основі навчаючої вибірки називають статистичним навчанням (statistical learning). У цьому розділі ми розглядаємо різні алгоритми статистичного навчання на основі неповної розкласифікованих навчаючих вибірок та вивчаємо асимптотичні властивості отриманих класифікаторів при зростанні обсягу вибірки.

Якість емпірично-баєсового класифікатора  $\hat{g}$  (або іншого класифікатора, побудованого за вибіркою) прийнято також оцінювати у термінах баєсового ризику  $L(\hat{g})$ , який тепер є випадковою величиною, залежною від вибірки:

$$L(\hat{g}) = \int p_{\hat{g}(x)} h_{\hat{g}(x)} \nu(dx) = P\{\hat{g}(\xi(O)) \neq \text{ind}(O) \mid \Xi_N\}, \quad (6.5)$$

тут у другій рівності ймовірність трактуємо як умовну при фіксованій навчаючій вибірці  $\Xi_N$ , а об'єкт, що класифікується —  $O$  вважаємо незалежним від  $\Xi_N$ .

Безумовна ймовірність помилки  $P\{\hat{g}(\xi(O)) \neq \text{ind}(O)\} = E L(\hat{g})$  характеризує не конкретний класифікатор, отриманий за даною вибіркою, а середню якість класифікаторів, побудованих певним методом за даними певного вигляду.

<sup>3</sup>Остаточні діагнози ставлять паталогоанатоми.

Класифікатор  $\hat{g}_N$ , побудований за вибіркою  $\Xi_N$ , будемо називати консистентним у класі  $\mathfrak{G}$ , якщо  $\hat{g}_N \in \mathfrak{G}$  м.н. і  $L(\hat{g}_N) \rightarrow L(g^B)$  (за ймовірністю) при  $N \rightarrow \infty$ . Якщо має місце збіжність майже напевне, класифікатор будемо називати сильно консистентним.

Зв'язок між консистентністю емпірично-баєсового класифікатора та консистентністю оцінок щільності, за якими він побудований, встановлює наступна теорема.

**Теорема 6.1.1 (Межа Джорфї)** Нехай  $g^B(x)$  — баєсів класифікатор, визначений (6.2), а  $\hat{g}(x)$  — емпірично баєсів класифікатор, визначений (6.4). Тоді

$$0 \leq L(\hat{g}) - L(g^B) \leq 2 \sum_{i=1}^M \int_{\mathcal{X}} |p_i h_i(x) - \hat{p}_i \hat{h}_i(x)| \nu(dx).$$

**Доведення.** Нерівність  $0 \leq L(\hat{g}) - L(g^B)$  випливає з означення баєсового класифікатора. Для доведення правої нерівності використаємо (6.3):

$$L(\hat{g}) - L(g^B) \leq \int_{\mathcal{X}} |p_{\hat{g}(x)} h_{\hat{g}(x)} - \max_i (p_i h_i(x))| \nu(dx) \leq J_1 + J_2,$$

де

$$J_1 = \int_{\mathcal{X}} |\max_i (p_i h_i(x)) - \max_i (\hat{p}_i \hat{h}_i(x))| \nu(dx),$$

$$J_2 = \int_{\mathcal{X}} |\max_i (\hat{p}_i \hat{h}_i(x)) - p_{\hat{g}(x)} h_{\hat{g}(x)}| \nu(dx).$$

Оцінимо

$$J_1 \leq \int_{\mathcal{X}} \max_i |p_i h_i(x) - \hat{p}_i \hat{h}_i(x)| \nu(dx) \leq \sum_{i=1}^M \int_{\mathcal{X}} |p_i h_i(x) - \hat{p}_i \hat{h}_i(x)| \nu(dx).$$

Враховуючи, що за означенням  $\hat{g}$ ,  $\max_i (\hat{p}_i \hat{h}_i(x)) = \hat{p}_{\hat{g}(x)} \hat{h}_{\hat{g}(x)}$ ,

$$J_2 \leq \int_{\mathcal{X}} |\hat{p}_{\hat{g}(x)} \hat{h}_{\hat{g}(x)} - p_{\hat{g}(x)} h_{\hat{g}(x)}| \nu(dx) \leq \sum_{i=1}^M \int_{\mathcal{X}} |p_i h_i(x) - \hat{p}_i \hat{h}_i(x)| \nu(dx).$$

Об'єднуючи ці дві оцінки, отримуємо твердження теореми.

Теорема доведена.

**Наслідок 6.1.1** В умовах теореми 6.1.1, якщо  $|\hat{p}_i| < 1$ , для всіх  $i = 1, \dots, M$ , то

$$|L(\hat{g}) - L(g^B)| \leq 2 \sum_{i=1}^M \int_{\mathcal{X}} |h_i(x) - \hat{h}_i(x)| \nu(dx) + 2 \sum_{i=1}^M |p_i - \hat{p}_i|$$

**Доведення.** Враховуючи, що  $\int p_i(x) \nu(dx) = 1$ , за теоремою 6.1.1 отримуємо

$$\begin{aligned} |L(\hat{g}) - L(g^B)| &\leq 2 \sum_{i=1}^M \left( \int_{\mathcal{X}} |p_i h_i(x) - \hat{p}_i h_i(x)| \nu(dx) \right. \\ &\quad \left. + \int_{\mathcal{X}} |\hat{p}_i h_i(x) - \hat{p}_i \hat{h}_i(x)| \nu(dx) \right) \\ &\leq 2 \sum_{i=1}^M |p_i - \hat{p}_i| + 2 \sum_{i=1}^M \int_{\mathcal{X}} |h_i(x) - \hat{h}_i(x)| \nu(dx). \end{aligned}$$

Наслідок доведено.

Як приклад застосування нерівності Джорфі розглянемо емпірично-баєсів класифікатор, побудований за вибіркою з суміші зі змінними концентраціями  $(\xi_{j:N}, j = 1, \dots, N)$  у випадку, коли  $\mathcal{X} = \mathbb{R}^d$  і розподіли компонент мають неперервні обмежені щільності відносно міри Лебега  $h_i$ . У цьому випадку для оцінювання  $h_i$  можна використати ядерні оцінки щільності  $\hat{h}_i^N(x)$ , визначені (4.2). Емпірично-баєсів класифікатор  $\hat{g}_N$  задамо як звичайно

$$\hat{g}_N(x) = \hat{p}_i^N \hat{h}_i^N(x),$$

де  $\hat{p}_i^N$  деякі оцінки апіорних ймовірностей  $p_i$ .

Використовуючи наслідок 6.1.1 та теорему 4.1.1, отримуємо наступну теорему.

**Теорема 6.1.2** Нехай

1. Для деякої константи  $C > 0$ ,  $\det \Gamma_N > C$ .
2. Параметр згладжування  $s_N$  ядерних оцінок  $\hat{h}_i^N(x)$  обрано так, що  $s_N \rightarrow 0$ ,  $\sqrt{\frac{\ln N}{N}} s_N^{-d} \rightarrow 0$  при  $N \rightarrow \infty$ .
3. Оцінки  $\hat{p}_i^N$  для апіорних ймовірностей є консистентними. Тоді емпірично баєсів класифікатор  $\hat{g}_N$  є консистентним.

## 6.2 Класифікація за методом найближчого сусіда

Для багатовимірних спостережень та спостережень з більш загальних метричних просторів одним з найбільш популярних методів класифікації є метод  $k$  найближчих сусідів [45, 46]. У випадку повністю розкласифікованої вибірки ця процедура виглядає зовсім просто: для об'єкта, який хочуть розкласифікувати, знаходять  $k$  найближчих (у метриці простору спостережень) сусідів серед елементів навчаючої вибірки. Класифікація проводиться голосуванням цих сусідів — об'єкт відносять до того класу, до якого належить найбільше його сусідів. Коротко кажучи: “скажи, хто твої сусіди, і я скажу хто ти”.

При такому описі процедури класифікації непомітно, що насправді класифікатор  $k$  найближчих сусідів — це емпірично-баєсів класифікатор. Але насправді це так: якщо на роль оцінок апіорних ймовірностей взяти відповідні частоти у навчаючій вибірці, а щільності розподілу оцінювати оцінками методу  $k$  найближчих сусідів, то емпірично-баєсів класифікатор, побудований за цими оцінками як раз і буде класифікатором  $k$  найближчих сусідів.

У випадку, коли спостереження для побудови класифікатора обираються з суміші зі змінними концентраціями, можна використати оцінки  $k$  найближчих сусідів для щільностей компонент суміші, описані у п. 4.4. Ми покажемо, що отриманий в результаті класифікатор є консистентним за дуже слабких умов.

Отже, нехай простір спостережень  $\mathcal{X}$  є сепарабельним метричним простором з метрикою  $\rho$ , а дані  $\Xi_N = (\xi_{j:N} \ j = 1, \dots, N)$  описуються моделлю суміші зі змінними концентраціями (2.1), тобто

$$P\{\xi_{j:N} \in A\} = \mu_{j:N}(A) = \sum_{m=1}^M w_{j:N}^m H_m(A). \quad (6.6)$$

Будемо вважати, що  $\det \Gamma_N \neq 0$  і оцінювати розподіли компонент суміші мінімксними оцінками

$$\hat{H}_m^N(A) = \frac{1}{N} \sum_{j=1}^N a_{j:N}^m \mathbb{I}\{\xi_{j:N} \in A\},$$

де  $a_{j:N}^m$  — мінімаксні вагові коефіцієнти, визначені (2.10). Використаємо



$\hat{H}_m^N$  для побудови оцінок  $k$  найближчих сусідів для щільності  $H_m$  відносно міри  $\nu(A) = H_0(A) = \sum_{i=1}^M \langle w^i \rangle H_i(A)$ .

Для довільного  $x \in \mathcal{X}$  позначимо  $r(x, k)$  відстань від  $x$  до його  $k$ -того найближчого сусіда у вибірці  $\Xi_N$  (більш детальне пояснення див. у п. 4.4). Нехай  $B(x, r)$  — замкнена куля радіуса  $r$  з центром  $x$  у просторі  $(\mathcal{X}, \rho)$ . Як оцінку для  $h_i(x) = \frac{\partial H_i}{\partial H_0}$  використаємо

$$\tilde{h}_i^N(x) = \frac{\hat{H}_i^N(B(x, r(x, k_N)))}{\hat{H}_0(B(x, r(x, k_N)))}, \quad (6.7)$$

де  $k_N$  — деяка не випадкова числова послідовність,  $\hat{H}_0$  — зважена емпірична міра з ваговими коефіцієнтами  $a_{j:N}^0 \equiv 1$ . (Відмітимо, що  $h_i$  існує, якщо  $\langle w_i \rangle \neq 0$ , тобто якщо частка спостережень з  $i$ -тої компоненти у вибірці не прямує до 0 при  $N \rightarrow \infty$ ).

Припустимо, що апріорні ймовірності  $p_i$ ,  $i = 1, \dots, M$  для класифікації об'єкта  $O$  відомі. Як класифікатор для  $O$  за спостереженням  $x = \xi(O)$  використаємо

$$\hat{g}_N(x) = \operatorname{argmax}_i p_i \tilde{h}_i^N(x)$$

**Теорема 6.2.1** *Нехай*

(i)  $(\mathcal{X}, \rho)$  є сепарабельним метричним простором.

(ii)  $|\langle w_i w_m \rangle_N - \langle w_i w_m \rangle| = O(1/\sqrt{N})$  для всіх  $i, m = 1, \dots, M$  і  $\det \Gamma \neq 0$ .

(iii)  $k_N/N \rightarrow 0$ ,  $k_N/\sqrt{N \log N} \rightarrow \infty$  при  $N \rightarrow \infty$ .

Тоді  $\tilde{h}_i^N$  є консистентною оцінкою  $h_i$  у просторі  $L_1(H_0)$ , а класифікатор  $\hat{g}_N$  є консистентним класифікатором.

**Зауваження.** 1. Теорема не накладає жодних умов на розподіли спостережуваних характеристик для різних популяцій. Вони можуть бути будь-якими — неперервними, дискретними, сингулярними.

2. З умови (ii) випливає існування  $\langle w_i \rangle \neq 0$ , оскільки  $\langle w_i \rangle = \langle w_i \times 1 \rangle = \sum_{m=1}^M \langle w_i w_m \rangle$ .

Для доведення теореми нам будуть потрібні три леми.

**Лема 6.2.1** (див. [43]). *Позначимо*

$$D_1 = \operatorname{supp} H_0 = \{z \in \mathcal{X} : H_0(B(z, r)) > 0 \ \forall r > 0\}.$$

Тоді  $H_0(D_1) = H_0(\mathcal{X})$ .

**Лема 6.2.2** (див. теорему 2.9.8 у [44]). *Позначимо*

$$D_2^i = \left\{ z \in \mathcal{X} : \lim_{r \rightarrow 0} \frac{H_i(B(z, r))}{H_0(B(z, r))} \rightarrow \frac{\partial H_i}{\partial H_0}(z) \right\}.$$

Тоді  $H_0(D_2^i) = H_0(\mathcal{X})$ .

Позначимо

$$\beta_N = \sqrt{\frac{\log N}{N}}$$

**Лема 6.2.3** *Якщо  $k_N/N \rightarrow 0$  то для всіх  $z \in \mathcal{X} \pmod{H_0}$*

$$r(z, k_N) \rightarrow 0 \text{ м.н.} \quad (6.8)$$

*i*

$$H_0(B(z, r(z, k_N))) \geq k_N/N - \Lambda \beta_N, \quad (6.9)$$

де  $\Lambda$  — деяка випадкова величина,  $\Lambda < \infty$  м.н.

**Доведення леми.** Нехай  $z \in D_1 = \text{supp } H_0$ . Тоді, за лемою 6.2.1,  $H_0(D_1) = H_0(\mathcal{X})$ . Враховуючи означення  $\hat{H}_0^N$  та  $r(z, k_N)$  отримуємо, що

$$\hat{H}_0^N(B(z, r(z, k_N))) \geq k_N/N. \quad (6.10)$$

Розглянемо набір  $\zeta_{j:N} = \rho(z, \xi_{j:N})$ ,  $j = 1, \dots, N$ . Зрозуміло, що цей набір утворює вибірку з суміші зі змінними концентраціями. Застосовуючи до цієї вибірки п. 2 наслідку 2.2.4, отримуємо, що

$$\sup_r |\hat{H}_0^N(B(z, r)) - H_0(B(z, r))| \geq \Lambda \beta_N.$$

Ця нерівність разом з (6.10) забезпечує (6.9). Щоб довести (6.8) помітимо, що

$$\forall \varepsilon > 0 \quad \{r(z, k_N) > \varepsilon\} \subseteq \{\hat{H}_0^N(B(z, \varepsilon/2)) \leq \frac{k_N}{N}\}$$

і, отже,

$$\begin{aligned} \mathbb{P}\{r(z, k_N) > \varepsilon\} &\leq \\ \mathbb{P}\{H_0(B(z, \varepsilon/2)) - \hat{H}_0^N(B(z, \varepsilon/2)) \geq H_0(B(z, \varepsilon/2)) - k_N/N\} \\ &\leq \exp(-C(H_0(B(z, \varepsilon/2)) - k_N/N)^2 N). \end{aligned}$$

Остання нерівність випливає з того, що  $|I\{\xi_j \in B(z, \varepsilon/2)\}| \leq 1$  внаслідок нерівності Хьюфдінга (теорема 7.3.2). Але  $H_0(B(z, \varepsilon/2)) > 0$  для всіх  $\varepsilon > 0$  оскільки  $z \in \text{supp } H_0$ . Отже,  $\sum \mathbb{P}\{r(z, k_N) > \varepsilon\} < \infty$  і  $r(z, k_N) \rightarrow 0$  м.н.

**Доведення теореми.** Доведемо, що для майже всіх  $z \in \mathcal{X} \pmod{H_0}$ ,  $\tilde{h}_N^i(z) \rightarrow h_i(z)$  м.н. Позначимо  $B(z, r(z, k_N)) = B_N$ ,  $D = \text{supp } H_0 \cap (\cap_{i=1}^M D_2^i)$ . Тоді за лемами 6.2.1 та 6.2.2  $H_0(D) = H_0(\mathcal{X})$ . Будемо вважати, що  $z \in D$ . Використовуючи наслідок 2.2.4, отримуємо

$$|\hat{H}_i^N(B_N) - H_i(B_N)| \leq \Lambda \beta_N \quad (6.11)$$

Тоді за (6.7)

$$\begin{aligned} |\tilde{h}_N^i(z) - h_i(z)| \leq & \left| \frac{\hat{H}_i^N(B_N)H_0(B_N) - H_i(B_N)\hat{H}_0^N(B_N)}{\hat{H}_0^N(B_N)H_0(B_N)} \right| + \\ & + \left| \frac{H_i(B_N)}{H_0(B_N)} - h_i(z) \right| \end{aligned}$$

Другий доданок прямує до 0 внаслідок (6.8) та  $z \in D_2^i$ . Згідно з (6.11), перший доданок менший, ніж

$$C\Lambda\beta_N(k_N/N + \beta_N)/(k_N/N \cdot (k_N/N - \beta_N)) \rightarrow 0$$

за умовою теореми.

Отже,  $\tilde{h}_N^i(z) \rightarrow h_i(z)$  м.н. для всіх  $z \pmod{H_0}$ . За теоремою Фубіні маємо  $\mathbb{P}\{\tilde{h}_N^i \rightarrow h_i, \pmod{H_0}\} = 1$ , а за теоремою про мажоровану збіжність,  $\int |\tilde{h}_N^i(z) - h_i(z)| H_0(dz) \rightarrow 0$  м.н., оскільки  $h_i \leq 1/\bar{w}^i \pmod{H_0}$  і  $\hat{h}_N^i \leq \sup_{j,N} |a_{j,N}^i|$ .

Отже,  $\tilde{h}_N^i$  є консистентними у  $L_1(H_0)$  оцінками щільностей  $h_i$ . Консистентність емпірично-баєсового класифікатора випливає з наслідку 6.1.1.

## 6.3 Асимптотика порогових класифікаторів

У цьому параграфі ми розглянемо задачу класифікації об'єкта  $O$  за спостереженням його числової характеристики  $\xi = \xi(O) \in \mathbb{R}$ . Будемо вважати, що об'єкт може належати лише одному з двох класів, і обмежимося розглядом порогових класифікаторів вигляду

$$g_t(\xi) = \begin{cases} 1, & \text{при } \xi \leq t, \\ 2, & \text{при } \xi > t, \end{cases} \quad (6.12)$$

тобто об'єкт відносять до першого класу, якщо його характеристика не перевищує поріг  $t$ , і до другого класу — в іншому випадку. Елементарний приклад такої класифікації - визначення людини (об'єкт) як хворої (другий клас), якщо її температура (характеристика  $\xi$ ) перевищує  $37^\circ$  (поріг  $t$ ).

Найкращим (баєсовим) будемо вважати такий поріг  $t = t^B$ , при якому  $g_t$  має найменшу ймовірність помилки. При цьому виникає проблема вибору (оцінки) порогу на основі навчаючої вибірки. Найбільш поширеними методами оцінювання  $t^B$  за повністю класифікованою вибіркою є емпірично-баєсова класифікація (ЕБК) описана у п. 6.1 та метод мінімізації емпіричного ризику (МЕР) [6, 71]. У цьому параграфі розглядається узагальнення МЕР на випадок, коли навчаюча вибірка отримана з суміші зі змінними концентраціями і порівнюється асимптотична поведінка класифікаторів, отриманих за допомогою цих двох методів.

МЕР - порівняно проста техніка, яка спирається на емпіричні функції розподілу як оцінки відповідних справжніх розподілів. Ці оцінки мають “гарну” швидкість збіжності порядку  $N^{-1/2}$  (див. розділ 2). ЕБК використовує оцінки щільностей розподілу, які мають значно гіршу швидкість збіжності: у гладенькому випадку, який ми розглядаємо, — порядку  $N^{-2/5}$  (пор. п. 4.3). Тому на перший погляд здається, що оцінки методу МЕР для  $t^B$  повинні бути асимптотично кращими, ніж ЕБК-оцінки. Але ми покажемо, що для оцінок методу МЕР характерна швидкість збіжності порядку  $N^{-1/3}$ , в той час як оцінки ЕБК забезпечують порядок збіжності  $N^{-2/5}$ .

Для випадку повністю розкласифікованої навчаючої вибірки збіжність порядку  $N^{-1/3}$  для МЕР-оцінок отримана у [59]. Асимптотика порядку кубічного кореня характерна для максимізації негладеньких функціоналів від емпіричних функцій розподілу [42, 53]. Метод ЕБК дозволяє враховувати гладкість функцій розподілу і відповідно згладжувати оцінки щільностей, тому він і забезпечує більшу точність класифікатора.

Отже, нехай у об'єкта  $O$  спостерігається деяка числова характеристика  $\xi = \xi(O)$ . Цей об'єкт може належати одному з двох класів. Невідомий нам номер класу, якому належить  $O$ , позначаємо  $\text{ind}(O)$ . Вважаються відомими апріорні ймовірності  $p_i = P(\text{ind}(O) = i)$ ,  $i = 1, 2$ . Характеристика  $\xi$  вважається випадковою, її розподіл залежить від  $\text{ind}(O)$ :  $P(\xi(O) < x \mid \text{ind}(O) = i) = H_i(x)$ . Розподіли  $H_i$  невідомі, але ми вважаємо, що вони мають неперервні щільності  $h_i$  відносно міри Лебега.

Як ми вже відмітили, в принципі, класифікатором  $g : \mathbb{R} \rightarrow \{1, 2\}$  може бути будь-яка вимірна функція, але у даному параграфі розглядаються

лише порогові класифікатори вигляду (6.12). Множину всіх таких класифікаторів позначимо  $\mathcal{G} = \{g_t : t \in \mathbb{R}\}$ . Ймовірність помилки такого класифікатора

$$\begin{aligned} L(g_t) &= L(t) = \mathbb{P}\{g_t(\xi(O)) \neq \text{ind}(O)\} \\ &= \sum_{i=1}^2 \mathbb{P}\{\text{ind}(O) = i\} \mathbb{P}\{g_t(\xi(O)) = 3 - i \mid \text{ind}(O) = i\} \\ &= p_1(1 - H_1(t)) + p_2 H_2(t). \end{aligned}$$

Баєсовим класифікатором у класі  $\mathcal{G}$  є класифікатор  $g^B \in \mathcal{G}$ , для якого досягається мінімум  $L(g)$ :

$$g^B = \underset{g \in \mathcal{G}}{\text{argmin}} L(g).$$

Поріг  $t^B$  баєсового класифікатора будемо називати баєсовим порогом:  $g^B = g_{t^B}$ ,

$$t^B = \underset{t \in \mathbb{R}}{\text{argmin}} L(t). \quad (6.13)$$

(Надалі  $\underset{x \in U}{\text{argmin}} f(x)$  позначає будь-яке значення  $x_* \in U$ , для якого  $f(x_*) = \inf_{x \in U} f(x)$ ). Якщо функція  $f$  є випадковою, то додатково вимагається, щоб  $x_*$  було випадковою величиною, тобто вимірною функцією на основному ймовірністному просторі). Оскільки  $\frac{d}{dt} H_i(t) = H'_i(t) = h_i(t)$ , отримуємо, що  $t^B$  є розв'язком рівняння

$$L'(t) = -p_1 h_1(t) + p_2 h_2(t) = 0. \quad (6.14)$$

Функції  $H_i$  (та, відповідно,  $h_i$ ) вважаються невідомими. Їх можна оцінити за даними, що являють собою вибірку з суміші зі змінними концентраціями:  $\Xi_N = \{\xi_{j:N}\}_{j=1}^N$ , де  $\xi_{j:N}$  незалежні між собою при фіксованому  $N$  і

$$\mathbb{P}\{\xi_{j:N} < x\} = w_{j:N} H_1(x) + (1 - w_{j:N}) H_2(x),$$

де  $w_{j:N}$  — відома концентрація об'єктів першого класу у суміші в момент  $j$ -того спостереження.

Для оцінки ф.р.  $H_i$  скористаємося зваженими емпіричними функціями розподілу

$$\hat{H}_i^N(x) = \frac{1}{N} \sum_{j=1}^N a_{j:N}^i \mathbb{I}\{\xi_j < x\},$$

з мінімаксними ваговими коефіцієнтами  $a_{j:N}^i$ .

Для оцінки щільностей розподілів  $h_i$  можна скористатись ядерними оцінками

$$\hat{h}_i^N(x) = \frac{1}{Ns_N} \sum_{j=1}^N a_{j:N}^i K\left(\frac{x - \xi_{j:N}}{s_N}\right)$$

де  $K$  — ядро (щільність деякого ймовірнісного розподілу),  $s_N > 0$  — параметр згладжування (див. п. 4.1).

Виходячи з формул (6.13) та (6.14), можна запропонувати два підходи до оцінювання  $t^B$ . Оцінка MER визначається як

$$\hat{t}^{MER} = \operatorname{argmin}_{t \in \mathbb{R}} L_N(t), \quad (6.15)$$

де  $L_N(t) = p_1(1 - \hat{H}_1^N(t)) + p_2\hat{H}_2^N(t)$  — емпіричний ризик класифікатора  $g_t$ .

ЕБК-оцінка будується наступним чином: знаходиться множина  $T_N$  усіх розв'язків рівняння

$$-p_1\hat{h}_1^N(t) + p_2\hat{h}_2^N(t) = 0$$

і як оцінка використовується

$$\hat{t}_N^{EBC} = \operatorname{argmin}_{t \in T_N} L_N(t). \quad (6.16)$$

**Зауваження.** Оцінки, визначені (6.15) та (6.16) можуть не існувати або бути не єдиними. В умовах, що накладаються далі, ймовірність того, що ці оцінки не існують, прямує до 0 при  $N \rightarrow \infty$ . Тому для вивчення асимптотичної поведінки оцінок у розумінні слабкої збіжності несуттєво, як вони довізнаються, коли мінімуми не існують.

Аналогічно, в цих умовах, ймовірність того, що оцінка  $\hat{t}_N^{EBC}$  визначена не однозначно, прямує до 0. Мінімум у (6.15) завжди досягається на нескінченній множині точок (як правило, на інтервалі). Але при виконанні умов теореми 6.3.3 всі точки мінімуму мають однакову асимптотичну поведінку, тобто на роль оцінки можна вибрати будь-яку з них.

#### Основні теореми.

Будемо вважати, що виконуються наступні умови:

(А)  $t^B$ , визначене (6.13), існує і є єдиною точкою глобального мінімуму  $L(t)$ ,  $L(t^B) < \min(p_1, p_2)$ .

(Остання нерівність виключає випадок, коли баєсовим є класифікатор, який відносить всі спостереження до одного класу, незалежно від значень  $\xi$ ).

(В<sub>k</sub>) Існують границі  $\langle w \rangle$ ,  $\langle (w)^2 \rangle, \dots, \langle (w)^k \rangle$  і  $\Delta = \langle (w)^2 \rangle - \langle w \rangle^2 > 0$ .

**Теорема 6.3.1** *Нехай виконані умови (A) та (B<sub>2</sub>), H<sub>i</sub> — неперервні функції на  $\mathbb{R}$ . Тоді  $\hat{t}_N^{MER} \rightarrow t^B$  за ймовірністю при  $N \rightarrow \infty$ .*

**Теорема 6.3.2** *Нехай виконано (A), (B<sub>2</sub>), існують і є неперервними щільності  $h_i$ ,  $s_N \rightarrow 0$ ,  $Ns_N \rightarrow \infty$ ,  $K$  — неперервна функція,*

$$d^2 := \int_{-\infty}^{\infty} K^2(t) dt < \infty.$$

*Тоді  $\hat{t}^{EBC} \rightarrow t^B$  за ймовірністю.*

Нехай щільності  $h_i$  існують і є  $s$  разів неперервно диференційовними у деякому околі  $t^B$ . Позначимо

$$f_s(t) = (-1)^s \left( p_1 \frac{d^s h_1(t)}{dt^s} - p_2 \frac{d^s h_2(t)}{dt^s} \right)$$

Зокрема,  $f_1(t) = f(x) = p_2 h'_2(x) - p_1 h'_1(x)$ ,

$$r_N = \left[ \frac{1}{N} \sum_{j=1}^N (p_2 a_{j:N}^2 + p_1 a_{j:N}^1)^2 (w_{j:N} h_1(t^B) + (1 - w_{j:N}) h_2(t^B)) \right]^{1/2},$$

$$r = \lim_{N \rightarrow \infty} r_N.$$

**Зауваження.** Для того, щоб ця границя існувала, достатньо виконання умови (B<sub>3</sub>).

Знаком  $\Rightarrow$  будемо позначати слабку збіжність.

**Теорема 6.3.3** *Якщо виконані умови (A) та (B<sub>3</sub>),  $h_i$  існують, є неперервно диференційовними у деякому околі  $t^B$  і  $f(t^B) \neq 0$ , то*

$$N^{1/3}(\hat{t}_N^{MER} - t^B) \Rightarrow \left( \frac{2r}{f(t^B)} \right)^{2/3} Z,$$

де  $Z = \operatorname{argmin}_{t \in \mathbb{R}} (W(t) + t^2)$ ,  $W(t)$  — двосторонній стандартний вінерів процес.

**Зауваження.** Те, що  $\min_{t \in \mathbb{R}} (W(t) + t^2)$  м.н. досягається в єдиній (випадковій) точці  $Z \in \mathbb{R}$ , впливає з леми 2.6 у [53]

**Теорема 6.3.4** Нехай виконані умови  $(A)$ ,  $(B_3)$  і

(i) у деякому okolí  $t^B$  існують і є обмеженими  $h_i''(t) = d^2 h_i(t)/dt^2$ ,  $f(t^B) \neq 0$ ,

(ii)  $\int_{-\infty}^{\infty} zK(z)dz = 0$ ,  $D^2 := \int_{-\infty}^{\infty} z^2 K(z)dz < \infty$ ,  $d^2 < \infty$ ,

(iii)  $s_n = c/N^{1/5}$  для деякого не випадкового  $c > 0$ .

Тоді

$$N^{2/5}(\hat{t}^{EBC} - t^B) \Rightarrow A + B\eta,$$

де  $A = D^2 c^{2/5} f_2(t^B)/(2f(t^B))$ ,  $B = dr/(c^{1/10} f(t^B))$ ,

$\eta$  — стандартна нормальна випадкова величина.

**Доведення теореми 6.3.1.** Помітимо, що умова  $(B_2)$  еквівалентна невідродженості матриці  $\Gamma$ . Тому за наслідком 2.2.2,

$$\sup_x |\hat{H}_i^N(x) - H_i(x)| \rightarrow 0$$

за ймовірністю при  $N \rightarrow \infty$ . Звідси випливає, що і  $\sup_x |L^N(x) - L(x)| \rightarrow 0$  за ймовірністю. Фіксуємо довільні  $\lambda > 0$ ,  $\varepsilon > 0$ . Нехай  $A_N = \{\sup_x |L^N(x) - L(x)| < \varepsilon/2\}$ . При достатньо великих  $N$ ,  $P(A_N) > 1 - \lambda$ .

Оскільки  $L$  — неперервна функція на  $\mathbb{R}$ ,  $L(-\infty) = p_1$ ,  $L(+\infty) = p_2$  і виконана умова  $(A)$ , то  $\forall \delta > 0 \exists \varepsilon$ , таке, що для всіх  $t$ , для яких  $|t - t^B| > \delta$ , має місце нерівність  $L(t) > L(t^B) + \varepsilon$ . Нехай подія  $A_N$  виконана. Тоді

$$L(\hat{t}_N^{MER}) - \frac{\varepsilon}{2} \leq L_N(\hat{t}_N^{MER}) \leq L_N(t^B) \leq L(t^B) + \frac{\varepsilon}{2},$$

отже  $L(\hat{t}_N^{MER}) \leq L(t^B) + \varepsilon$  і  $|\hat{t}_N^{MER} - t^B| \leq \delta$ . Внаслідок довільності  $\delta$  та  $\lambda$  теорема доведена.

**Доведення теореми 6.3.2.** Згідно з теоремою 4.2.1 в наших умовах  $\hat{h}_i^N(x) \rightarrow h_i(x)$  за ймовірністю в кожній точці  $x \in \mathbb{R}$ . Отже,

$$u_N(x) := p_2 \hat{h}_2^N(x) - p_1 \hat{h}_2^N(x) \rightarrow u(x) := p_2 h_2(x) - p_1 h_1(x)$$

за ймовірністю. Для будь-якого  $\delta > 0$  розглянемо подію  $A_N(\delta) = \{\exists t : |t - t^B| \leq \delta, u_N(t) = 0\}$ . Покажемо, що

$$P(A_N(\delta)) \rightarrow 1 \tag{6.17}$$

при  $N \rightarrow \infty$ .



Оскільки  $t^B$  — точка мінімуму  $L(t)$ , а  $L'(t) = u(t)$  — неперервна функція, то  $u$  повинна змінювати знак в околі точки  $t^B$ , тобто існують такі  $t^-, t^+$ , що  $t^B - \delta < t^- < t^B < t^+ < t^B + \delta$  і  $u(t^-)u(t^+) < 0$ . Отже,  $P\{u_N(t^-)u_N(t^+) < 0\} \rightarrow 1$ . Але оскільки  $u_N$  — неперервна функція, то  $\{u_N(t^-)u_N(t^+) < 0\} \subseteq A_N(\delta)$ . Отже, (6.17) доведено.

Фіксуємо  $\delta$ . Так само, як у доведенні теореми 6.3.1, маємо, що  $\exists \varepsilon > 0$ ,  $L(t) > L(t^B) + \varepsilon$  для всіх  $t \notin [t^B - \delta, t^B + \delta]$ . Виберемо  $0 < \delta' < \delta$  так, щоб для всіх  $t \in [t^B - \delta', t^B + \delta']$  виконувалось  $L(t) < L(t^B) + \varepsilon/4$ . Розглянемо випадкову подію

$$B_N = \left\{ \inf_{t \in [t^B - \delta, t^B + \delta]} L_N(t) > L(t^B) + \varepsilon/2 > \sup_{t \in [t^B - \delta', t^B + \delta']} L_N(t) \right\}.$$

Фіксуємо довільне  $\lambda > 0$ . Використовуючи рівномірну збіжність  $L_N$  до  $L$ , отримуємо, що, при достатньо великих  $N$ ,  $P(B_N) > 1 - \frac{\lambda}{2}$ . Згідно з (6.17), при великих  $N$ ,  $P(A_N(\delta')) > 1 - \frac{\lambda}{2}$ . Якщо виконано  $A_N(\delta')$ , то існує  $t^* \in T_N \cap [t^B - \delta', t^B + \delta']$  і, при виконанні  $B_N$ ,  $L_N(t^*) < L_N(t)$  для всіх  $t \notin [t^B - \delta, t^B + \delta]$ . Отже, в цьому випадку  $\hat{t}^{EBC} \in [t^B - \delta, t^B + \delta]$ . Тому

$$P\{|\hat{t}^{EBC} - t^B| < \delta\} \geq P(A_N(\delta')) \cap B_N \geq 1 - \lambda$$

при великих  $N$ .

Враховуючи довільність  $\lambda$ , отримуємо твердження теореми.

**Доведення теореми 6.3.3.** Позначимо

$$W_N(\tau) = N^{2/3}(L_N(t^B + N^{-1/3}\tau) - L_N(t^B) - L(t^B + N^{-1/3}\tau) + L(t^B)).$$

Для доведення теореми нам будуть потрібні твердження про асимптотичну поведінку процесу  $W_N$ , які ми сформулюємо у вигляді двох лем.

**Лема 6.3.1** *На будь-якому скінченному інтервалі  $U = [\tau_-, \tau_+]$  випадкові процеси  $W_N$  слабо збігаються при  $N \rightarrow \infty$  до процесу  $rW$  у просторі  $\text{Var}(U)$  функцій без розривів другого роду з рівномірною метрикою.*

**Доведення** Згідно з [1], враховуючи неперервність траєкторій  $W$ , досить довести асимптотичну нормальність скінченновимірних розподілів  $W_N$ , збіжність других моментів приростів і щільність розподілів  $W_N$  у  $D(U)$  (пор. п.7.4).

Спочатку підрахуємо  $E(W_N(\tau_2) - W_N(\tau_1))^2$ . Нехай  $\tau_1 < \tau_2$ . Позначимо  $b_{j:N} = p_2 a_{j:N}^2 - p_1 a_{j:N}^1$ , тоді

$$W_N(\tau_2) - W_N(\tau_1) = N^{-1/3} \sum_{j=1}^N b_{j:N} (\mathbb{I}\{\xi_{j:N} \in A_N\} - \mathbb{P}\{\xi_{j:N} \in A_N\}), \quad (6.18)$$

де  $A_N = A_N(\tau_1, \tau_2) = [N^{-1/3}\tau_1, N^{-1/3}\tau_2]$ . Отже,

$$\begin{aligned} \mathbb{E}(W_N(\tau_2) - W_N(\tau_1))^2 &= \\ &= N^{-2/3} \sum_{j=1}^N (b_{j:N})^2 [(w_{j:N} H_1(A_N) + (1 - w_{j:N}) H_2(A_N)) \\ &\quad - (w_{j:N} H_1(A_N) + (1 - w_{j:N}) H_2(A_N))^2] \end{aligned}$$

Враховуючи, що  $H_i(A_N) \sim h_i(t^B) N^{-1/3}(\tau_2 - \tau_1)$ , отримуємо при  $N \rightarrow \infty$ ,

$$\mathbb{E}(W_N(\tau_2) - W_N(\tau_1))^2 \rightarrow r^2(\tau_2 - \tau_1) = \mathbb{E}(rW(\tau_2) - rW(\tau_1))^2.$$

Асимптотичну нормальність скінченновимірних розподілів  $W_N$  можна довести, використовуючи центральну граничну теорему з умовою Ліндеберга з урахуванням рівномірної обмеженості всіх доданків у сумі (6.18).

Залишилось перевірити щільність сім'ї розподілів  $W_N$ . Щоб використати відповідний критерій (лема 7.4.1), покажемо, що для всіх  $\tau_1 < \tau < \tau_2$ ,

$$J := \mathbb{E}(W_N(\tau) - W_N(\tau_1))^2 (W_N(\tau_2) - W_N(\tau))^2 \leq C_1(\tau_2 - \tau_1)^2, \quad (6.19)$$

де  $C_1$  не залежить від  $N$ ,  $\tau_1$ ,  $\tau$ ,  $\tau_2$ .

Покладемо

$$\eta_j(\tau_1, \tau_2) = b_{j:N} (\mathbb{I}\{\xi_{j:N} \in A_N(\tau_1, \tau_2)\} - \mathbb{P}\{\xi_{j:N} \in A_N(\tau_1, \tau_2)\}).$$

Тоді

$$\begin{aligned} J &\leq \frac{C_2}{N^{4/3}} \left( \sum_{j \neq k} \{ \mathbb{E}(\eta_j(\tau, \tau_2))^2 (\eta_k(\tau_1, \tau))^2 \right. \\ &\quad \left. + | \mathbb{E} \eta_j(\tau, \tau_2) \eta_j(\tau_1, \tau) \eta_k(\tau, \tau_2) \eta_k(\tau_1, \tau) | \} \right. \\ &\quad \left. + \sum_{j=1}^N \mathbb{E}(\eta_j(\tau, \tau_2))^2 (\eta_j(\tau_1, \tau))^2 \right) \\ &\leq \frac{C_2}{N^{4/3}} (N^2 C_3 (h^*)^2 N^{-1/3} (\tau - \tau_1) N^{-1/3} (\tau_2 - \tau) + N C_4 (h^*)^4 N^{-1/3} (\tau - \tau_1) (\tau_2 - \tau)) \end{aligned}$$

$$\leq C_1(\tau_2 - \tau_1)^2.$$

(Тут  $C_2$  — абсолютна константа,  $C_3, C_4$  залежать лише від  $\sup_{j,N} |b_{j,N}| < \infty$ ,  $h^* = \sup_{t \in U} (h_1(t) + h_2(t))$ ).

Тепер, використовуючи теореми 15.4 та 15.6 з [1], отримуємо твердження леми.

**Лема 6.3.2** Для будь-якого  $\varepsilon > 0$  існує  $D > 0$  таке, що для всіх  $N$

$$\mathbb{P}\{\exists \tau \in \mathbb{R} : |W_N(\tau)| > D(1 + |\tau|)\} < \varepsilon.$$

**Доведення.** За нерівністю 15.30 [1], з (6.19) випливає, що

$$\begin{aligned} J_1 &:= \mathbb{P}\left\{\sup_{\tau \in [\tau_-, \tau_+]} \min(|W_N(\tau) - W_N(\tau_1)|, |W_N(\tau_2) - W_N(\tau)|) > \varepsilon/2\right\} \\ &\leq \frac{16C_1}{\varepsilon^4}(\tau_+ - \tau_-)^2. \end{aligned}$$

Так само, як при доведенні (6.19), отримуємо, що

$$\mathbb{E} W_N^4(\tau) \leq C_5 \tau^2,$$

тому, за нерівністю Чебишова,

$$J_2(\tau) := \mathbb{P}\{|W_N(\tau)| > \varepsilon/2\} \leq \frac{16C_5\tau^2}{\varepsilon^4}.$$

Отже

$$\mathbb{P}\left\{\sup_{\tau \in [\tau_-, \tau_+]} |W_N(\tau)| > \varepsilon\right\} \leq J_1 + J_2(\tau_-) + J_2(\tau_+) \leq \frac{C_6}{\varepsilon^4}((\tau_+ - \tau_-)^2 + \tau_-^2 + \tau_+^2).$$

Таким чином,

$$\begin{aligned} &\mathbb{P}\{\exists \tau \in \mathbb{R} : |W_N(\tau)| > D(1 + |\tau|)\} \\ &\leq \sum_{j=0}^{\infty} [\mathbb{P}\{\exists \tau \in [j, j+1] : |W_N(\tau)| > D(1+j)\} \\ &\quad + \mathbb{P}\{\exists \tau \in [-j-1, -j] : |W_N(\tau)| > D(1+j)\}] \\ &\leq C_7 \sum_{i=0}^{\infty} \left( \frac{(i+1-i)^2}{D^4(i+1)^4} + \frac{i^2}{D^4(1+i)^4} + \frac{(i+1)^2}{D^4(1+i)^4} \right) \leq \frac{C_8}{D^4}, \end{aligned}$$

звідки і випливає твердження леми.

Тепер, для того, щоб від збіжності випадкових процесів перейти до збіжності їх точок мінімуму, нам будуть потрібні наступні позначення.

Фіксуємо довільне  $S > 0$ . Нехай  $f$  — функція на  $U = [-S, S]$ . Позначимо

$$\text{Am}^-(f, \varepsilon) = \inf\{x \in U : f(x) \leq \inf_{y \in U} f(y) + \varepsilon\},$$

$$\text{Am}^+(f, \varepsilon) = \sup\{x \in U : f(x) \leq \inf_{y \in U} f(y) + \varepsilon\},$$

$$\text{Am}^-(f) = \lim_{\varepsilon \downarrow 0} \text{Am}^-(f, \varepsilon), \quad \text{Am}^+(f) = \lim_{\varepsilon \downarrow 0} \text{Am}^+(f, \varepsilon).$$

(Оскільки  $\text{Am}^-(f, \varepsilon)$  та  $\text{Am}^+(f, \varepsilon)$  є монотонно незростаючими по  $\varepsilon$ , ці границі завжди існують.)

Легко бачити, що коли існує  $x_* = \operatorname{argmin}_{x \in U} f(x)$ , то

$$\text{Am}^-(f) \leq x_* \leq \text{Am}^+(f).$$

Позначимо  $|f|_\infty = \sup_{x \in U} |f(x)|$  — рівномірна метрика. У наступній лемі всі інфімуми беруться по  $x \in U$ , тобто  $\inf g = \inf_{x \in U} g(x)$  і т.д.

**Лема 6.3.3** *Функціонали  $\text{Am}^-$  та  $\text{Am}^+$  є неперервними у просторі  $\mathcal{D}(U)$  з метрикою  $|\cdot|_\infty$ .*

**Доведення.** Покажемо, що для  $f_n, g \in \mathcal{D}(U)$  з  $|f_n - g|_\infty \rightarrow 0$  при  $n \rightarrow \infty$  впливає  $\text{Am}^-(f_n) \rightarrow \text{Am}^-(g)$ . (Для  $\text{Am}^+$  доведення аналогічне.)

За означенням  $\text{Am}^-(g, \varepsilon)$ , для будь-якого  $\lambda > 0$  знайдеться таке  $x_g$ , що  $g(x_g) \leq \inf g + \varepsilon$  і  $x_g \leq \text{Am}^-(g, \varepsilon) + \lambda$ . Нехай  $f, g \in \mathcal{D}(U)$ ,  $|f - g|_\infty \leq \delta$ . Тоді  $\inf g - \delta \leq \inf f \leq \inf g + \delta$ . Отже,

$$f(x_g) \leq g(x_g) + \delta \leq \inf g + \delta + \varepsilon \leq \inf f + \varepsilon + 2\delta.$$

Тому

$$\text{Am}^-(f, \varepsilon + 2\delta) \leq \text{Am}^-(g, \varepsilon) \tag{6.20}$$

і, аналогічно,

$$\text{Am}^-(g, \varepsilon + 2\delta) \leq \text{Am}^-(f, \varepsilon). \tag{6.21}$$

Нехай  $|f_n - g|_\infty \rightarrow 0$ . Для будь-якого  $\gamma > 0$ , знайдеться  $\varepsilon_0 > 0$ , таке, що для всіх  $0 \leq \varepsilon \leq \varepsilon_0$ ,

$$\text{Am}^-(g, \varepsilon) - \gamma \leq \text{Am}^-(g) \leq \text{Am}^-(g, \varepsilon) + \gamma.$$

Виберемо довільне  $0 < \varepsilon' < \varepsilon_0$ ,  $\delta = \varepsilon = \varepsilon'/3$ . При достатньо великих  $n$ ,  $|f_n - g|_\infty \leq \delta$  і за (6.20),

$$\text{Am}^-(f_n, \varepsilon') = \text{Am}^-(f_n, \varepsilon + 2\delta) \leq \text{Am}^-(g, \varepsilon) \leq \text{Am}^-(g) + \gamma.$$

Тому  $\text{Am}^-(f_n) \leq \text{Am}^-(g) + \gamma$  при достатньо великих  $n$ . Використовуючи (6.21), маємо (при великих  $n$ )

$$\text{Am}^-(g) - \gamma \leq \text{Am}^-(g, \varepsilon + 2\delta) \leq \text{Am}^-(f, \varepsilon)$$

і

$$\text{Am}^-(g) - \gamma \leq \text{Am}^-(f_n) + \gamma.$$

Внаслідок довільності  $\gamma$  отримуємо  $\text{Am}^-(f_n) \rightarrow \text{Am}^-(g)$ .

Лема доведена.

**Продовження доведення теореми 6.3.3.** Нехай  $U$  — окіл точки  $t^B$ , на якому  $h_i$  є неперервно диференційовними. За теоремою 6.3.1,  $\hat{t}^{MER} \rightarrow t^B$  за ймовірністю, отже, для будь-якого  $\varepsilon > 0$  і достатньо великих  $N$ , для  $A_N = \{\hat{t}^{MER} \in U\}$  виконано  $P(A_N) > 1 - \varepsilon/2$ . Позначимо  $\tau_N = N^{1/3}(\hat{t}_N^{MER} - t^B)$ . Покажемо, що знайдеться таке  $S = S_\varepsilon$ , що для достатньо великих  $N$ ,

$$P\{|\tau_N| < S\} > 1 - \varepsilon. \quad (6.22)$$

Оскільки

$$L_N(t^B + N^{-1/3}\tau) - L_N(t^B) = N^{-2/3}W_N(\tau) + L(t^B + N^{-1/3}\tau) - L(t^B),$$

то

$$\tau_N = \underset{\tau}{\operatorname{argmin}} v(\tau),$$

де

$$v(\tau) = W_N(\tau) + N^{2/3}(L(t^B + N^{-1/3}\tau) - L(t^B)).$$

Будемо вважати, що подія  $\{\hat{t}^{MER} \in U\}$  виконана. Оскільки  $H_i$  — двічі диференційовні на  $U$  і  $t^B$  — точка мінімуму  $L$ , то

$$L(t^B + N^{-1/3}\tau) - L(t^B) = \frac{1}{2}L''(\zeta)N^{-2/3}\tau^2,$$

де  $\zeta \in U$  — проміжна точка,  $L''(\zeta) = f(\zeta)$ . Оскільки  $f(t^B) > 0$ , можна обрати таке  $U$ , щоб  $f(\zeta) > c > 0$  на  $U$ . Тоді  $v(\tau) \geq W_N(\tau) + \frac{c}{2}\tau^2$  при  $N^{-1/3}\tau + t^B \in U$ .

Позначимо  $B_N = \{|W_N(\tau)| < D(1 + |\tau|), \forall \tau \in \mathbb{R}\}$ . За лемою 6.3.2, для достатньо великих  $D$ ,  $\mathbb{P}(B_N) > 1 - \varepsilon/2$ . Але, якщо виконано  $B_N$ , то при  $c\tau^2/2 > D(1 + |\tau|)$ ,  $v(\tau) > 0 = v(0)$ , тому у таких точках мінімум  $v(\tau)$  досягатись не може. Отже, при виконанні  $A_N \cap B_N$ ,  $c\tau_N^2 < D(1 + |\tau_N|)$ , тобто  $|\tau_N| \leq \max(1, 2D/c) = S$ . Оскільки  $\mathbb{P}(A_N \cap B_N) > 1 - \varepsilon$ , то (6.22) доведено.

Аналогічно або з використанням закону повторного логарифму для  $W$  доводиться, що, для достатньо великих  $S$ ,

$$\mathbb{P}\{|\operatorname{argmin}_{\tau \in \mathbb{R}} \tilde{W}(\tau)| < S\} > 1 - \varepsilon, \quad (6.23)$$

де  $\tilde{W}(\tau) = rW(\tau) + \frac{f(0)}{2}\tau^2$ . За лемою 6.3.2, на інтервалі  $[-S, S]$  процес  $W_N$  слабо збігається до  $rW$ , а внаслідок неперервної диференційовності  $h_i$ ,  $N^{2/3}(L(t^B + N^{-1/3}\tau) - L(t^B))$  рівномірно збігається до  $f(0)\tau^2/2$ . Тому  $v$  слабо збігається до  $\tilde{W}$  у  $\operatorname{Var}[-S, S]$  з рівномірною метрикою.

Оскільки функція  $v(\tau)$  є сталою на інтервалах між стрибками, то мінімум  $v(\tau)$  завжди досягається, тобто  $\operatorname{argmin}_{\tau} v(\tau)$  існує, хоча і не є єдиним. При цьому  $\operatorname{Am}^-(v) \leq \operatorname{argmin}_{\tau \in [-S, S]} v(\tau) \leq \operatorname{Am}^+(v)$ . Враховуючи лему 6.3.3, отримуємо  $\operatorname{Am}^-(v) \Rightarrow \operatorname{Am}^-(\tilde{W})$ ,  $\operatorname{Am}^+(v) - \operatorname{Am}^-(v) \Rightarrow \operatorname{Am}^+(\tilde{W}) - \operatorname{Am}^-(\tilde{W})$ . За лемою 2.6 з [53], мінімум  $\tilde{W}$  м.н. досягається рівно в одній точці, тому  $\operatorname{Am}^-(\tilde{W}) = \operatorname{argmin}_{\tau \in [-S, S]} \tilde{W}(\tau)$  і  $\operatorname{Am}^+(\tilde{W}) - \operatorname{Am}^-(\tilde{W}) = 0$ .

Звідси випливає, що, для всіх  $x \in \mathbb{R}$ ,

$$\mathbb{P}\{\operatorname{argmin}_{\tau \in [-S, S]} v(\tau) < x\} \rightarrow \mathbb{P}\{\operatorname{argmin}_{\tau \in [-S, S]} \tilde{W}(\tau) < x\}.$$

Це, разом з (6.22-6.23) забезпечує збіжність  $\tau_N$  до  $\operatorname{argmin}_{\tau \in \mathbb{R}} \tilde{W}(\tau) = \operatorname{argmin}_{\tau \in \mathbb{R}} (W(\tau) + \alpha\tau^2)$ , де  $\alpha = f(t^B)/(2r)$ . Враховуючи, що  $W(\alpha\tau) \doteq \sqrt{\alpha}W(\tau)$  ( $\doteq$  позначає рівність за розподілом) отримуємо

$$\operatorname{argmin}_{\tau \in \mathbb{R}} (W(\tau) + \alpha\tau^2) \doteq \operatorname{argmin}_{\tau \in \mathbb{R}} (\alpha^{-1/3}W(\alpha^{2/3}\tau) + \alpha^{-1/3}(\alpha^{2/3}\tau)^2) = \alpha^{-2/3}Z$$

Теорема доведена.

#### Доведення теореми 6.3.4.

Нехай, як і раніше,  $u_N(t) = p_2\hat{h}_2^N(t) - p_1\hat{h}_1^N(t)$ . За означенням  $t_N^{EBC}$ ,  $u_N(t_N^{EBC}) = 0$ . Покладемо  $\delta_N = t_N^{EBC} - t^B$ . За теоремою 6.3.2,  $\delta_N \rightarrow 0$  за ймовірністю. Тому

$$0 = u_N(t^B + \delta_N) \approx u_N(t^B) + \delta_N u'_N(t^B).$$

Отже

$$\begin{aligned}\delta_N &\approx -\frac{u_N(t^B)}{u'_N(t^B)} = \frac{p_2\hat{h}_2^N(t^B) - p_1\hat{h}_1^N(t^B)}{(p_2\hat{h}_2^N(t^B) - p_1\hat{h}_1^N(t^B))'} \\ &\approx \frac{p_2(\hat{h}_2^N(t^B) - h_2(t^B)) - p_1(\hat{h}_1^N(t^B) - h_1(t^B))}{f_1(t^B)}.\end{aligned}$$

В останній рівності використано, що  $p_1h_1(t^B) - p_2h_2(t^B) = 0$ .

З теореми 4.3.5 випливає, що при  $s_N = c/N^{1/5}$ ,

$$N^{2/5}(p_1(\hat{h}_1^N(t^B) - h_1(t^B)) - p_2(\hat{h}_2^N(t^B) - h_2(t^B))) \Rightarrow D^2c^{2/5}f_2(t^B)/2 + dr/c^{1/10}\eta,$$

де  $\eta$  — стандартна нормальна величина.

Звідси отримуємо твердження теореми.

## 6.4 Класифікація на основі єдиного індекса

У цьому параграфі ми розглянемо класифікацію об'єктів за багатовимірними спостереженнями  $\mathcal{X} = \mathbb{R}^d$ , розподіли яких для всіх популяцій є абсолютно неперервними відносно міри Лебега. До таких даних застосована техніка емпірично-байєсової класифікації з ядерними оцінками (консистентна за теоремою 6.1.2) або за методом найближчого сусіда (див. п. 6.2). Але для багатовимірних спостережень емпірично-байєсові класифікатори мають дуже складну, непрозору форму, їх важко інтерпретувати. Тому на практиці часто використовують наступний підхід. Багатовимірні характеристики аналізованих об'єктів проєктують на деякий напрям, тобто обирають певну лінійну комбінацію всіх спостережуваних змінних, що характеризують об'єкт. У економетриці та соціальній статистиці такі комбінації звать "сумарними індексами", у психометриці — "шкалами". Класифікація об'єктів проводиться на основі подібного загального індекса.

Ми розглянемо задачу побудови такого "лінійного" індекса, який забезпечував би асимптотично мінімальну помилку класифікації. Для цього вектори спостережуваних даних проєктуються на різні напрямки, за кожною проєкцією будується емпірично-байєсів класифікатор, оцінюється ймовірність його помилки, і на роль "найкращого" індекса беруть той, для якого ця оцінка є найменшою. Виявляється, що ймовірність помилки емпірично-байєсової класифікації на основі індексів, вибраних таким способом, прямує до найменшої можливої ймовірності помилки класифікації за лінійними індексами.

Отже, нехай спостережувана характеристика  $\xi = \xi(O)$  є  $d$ -вимірним вектором  $\xi = (\xi^1, \dots, \xi^d)^T$ , а навчаюча вибірка є вибіркою зі змінними концентраціями, причому  $\xi_{j:N} = (\xi_{j:N}^1, \dots, \xi_{j:N}^d)^T$ .

Ми будемо розглядати "лінійні індекси" вигляду

$$S(b) = \sum_{i=1}^d b_i \xi^i, \quad (6.24)$$

та, відповідно, для елементів навчаючої вибірки:

$$S_j^N(b) = \sum_{i=1}^d b_i \xi_{j:N}^i, \quad (6.25)$$

де  $b = (b_1, b_2, \dots, b_d)^T \in R^d$ ; довжина не випадкового вектора  $b$  дорівнює одиниці:

$$b_1^2 + b_2^2 + \dots + b_d^2 = 1. \quad (6.26)$$

Індекс  $S(b)$  є, по суті, проекцією  $\xi$  на напрямок  $b$ . Ми будемо мати справу лише з класифікаторами вигляду  $\tilde{g}_b(\xi) = g(S(b))$ , де  $g: R \rightarrow \{1, \dots, M\}$  — довільна вимірна функція. Ймовірність помилки класифікатора  $\tilde{g}_b$

$$L_{\tilde{g}_b} = P\{\tilde{g}_b(\xi) \neq \text{ind}(O)\}.$$

Мінімум  $L_{\tilde{g}_b}$  по всіх  $g$  при фіксованому  $b$  досягається на байєсовому класифікаторі

$$g^*(x) = \arg \max_i (p_i u_i(x)),$$

де  $p_i = P\{\text{ind}(O) = i\}$  — апіорні ймовірності (концентрації компонент суміші під час спостереження  $O$ );  $u_k(x) = u_k(x, b)$  — щільності в.в.  $\eta_k(b) = \sum_{i=1}^d \eta_k^i b_i$ , де  $\eta_k = (\eta_k^1, \eta_k^2, \dots, \eta_k^d)$  має щільність  $h_k$ . Ймовірність помилки цього класифікатора має вигляд

$$L_g^* = L^*(b) = 1 - \int_{-\infty}^{\infty} \max_{1 \leq k \leq M} (p_k u_k(x)) dx.$$

Найкращим для побудови класифікатора є індекс, отриманий проекцією  $\xi$  на напрямок

$$b^* = \underset{b}{\operatorname{argmin}} L^*(b) = \underset{b}{\operatorname{argmax}} \int_{-\infty}^{\infty} \max_{1 \leq k \leq M} (p_k u_k(x)) dx.$$



Оскільки справжні значення  $u_k$  невідомі, замінімо їх ядерними оцінками, побудованими за вибіркою  $(S_j(b))_{j=1}^N$ . Ці оцінки мають вигляд:

$$\hat{u}_i^N(b, x) = \frac{1}{N\sigma_N} \sum_{j=1}^N a_{j:N}^i K\left(\frac{x - S_j(b)}{s_N}\right), \quad (6.27)$$

де  $a_{j:N}^i$  — мінмаксні вагові коефіцієнти, визначені (2.10). На параметр згладжування  $s_N$  накладаються звичайні умови:  $s_N \rightarrow 0$ ,  $Ns_N \rightarrow \infty$ .

За оцінку ймовірності помилкової класифікації  $L^*(b)$  беремо

$$\hat{L}^N(b) = 1 - \int_{-\infty}^{\infty} \max_{1 \leq k \leq M} (p_k \hat{u}_k^N(b, x)) dx, \quad (5)$$

а за оцінку  $b^*$  —

$$\hat{b} = \arg \min \hat{L}^N(b)$$

Емпірично-баєсів класифікатор для спостереження  $\xi$  будуємо за формулою

$$\hat{g}(\hat{b}, x) = \arg \max_{1 \leq i \leq M} (p_i \hat{u}_i^N(\hat{b}, x))$$

Для оцінки якості класифікатора служить умовна ймовірність помилки при фіксованій навчаючій вибірці (пор. (6.5)):

$$L_{\hat{g}}(\hat{b}) = 1 - \int_{-\infty}^{\infty} p_{\hat{g}(\hat{b}, x)} u_{\hat{g}(\hat{b}, x)} dx.$$

Припустимо, що виконуються наступні умови:

$$(i) \quad K(x) \leq a, \quad \int_{-\infty}^{\infty} |x| K(x) dx < \infty;$$

де  $a < \infty$  — деяка стала.

Сукупність множин  $S = \{A\}$ , де

$$(ii) \quad A = \left\{ y : K\left(\frac{x - \sum_{i=1}^d b_i y_i}{s_N}\right) \geq c \right\}$$

( $c$ —деяка стала), утворює VC-клас (див. п. 2.2).

Для усіх  $i = 1 \div M$  існують і обмежені усі частинні похідні щільностей  $h_i(x_1, \dots, x_d)$ :

$$(iii) \quad \left| \frac{\partial h_i}{\partial x_s}(x_1, \dots, x_d) \right| \leq c_{is}; \quad s = 1 \div d.$$

Вагові коефіцієнти  $a_{j,N}^i$  задовольняють умові

$$(iv) \quad |a_{j,N}^i| \leq \tilde{a},$$

де  $\tilde{a} < \infty$  — деяка стала.

$$(v) \quad \frac{1}{s_N} = o(N^{\frac{1}{4}}), \quad N \rightarrow \infty.$$

**Теорема 6.4.1** При виконанні умов (i) - (v)

$$\mathbb{E} L_{\hat{g}}(\hat{b}) \rightarrow L_g^*(b^*) \quad \text{при } N \rightarrow \infty.$$

**Наслідок 6.4.1** В умовах теореми 6.4.1,

$$L_{\hat{g}}(\hat{b}) \rightarrow L_g^*(b^*) \quad \text{при } N \rightarrow \infty$$

за ймовірністю.

Спочатку доведемо наступну лему.

**Лема 6.4.1** За умов (i) — (v) для усіх  $k = 1 \div M$

$$\mathbb{E} \sup_{b,x} |\hat{u}_k^N(b, x) - u_k(b, x)| \rightarrow 0 \quad \text{при } N \rightarrow \infty.$$

**Доведення леми 6.4.1.** Зробимо наступні оцінки

$$\begin{aligned} & \sup_{b,x} |\hat{u}_k^N(b, x) - u_k(b, x)| \leq \\ & \leq \sup_{b,x} \left| \frac{1}{Ns_N} \sum_{j=1}^N a_{j,N}^k K \left( \frac{x - S_j^N(b)}{s_N} \right) - \frac{1}{Ns_N} \mathbb{E} \sum_{j=1}^N a_{j,N}^k K \left( \frac{x - S_j^N(b)}{s_N} \right) \right| + \\ & + \sup_{b,x} \left| \frac{1}{Ns_N} \mathbb{E} \sum_{j=1}^N a_{j,N}^k K \left( \frac{x - S_j^N(b)}{s_N} \right) - u_k(b, x) \right| = S_1^N + S_2^N. \end{aligned} \quad (6.28)$$

Розглянемо  $S_1^N$ . Використаємо зважену емпіричну міру співвідношенням

$$\hat{\mu}_N(A) = \frac{1}{N} \sum_{j=1}^N a_{j,N}^k \mathbb{1}\{\xi_j^N \in A\}; \quad \bar{\mu}_N(A) = \mathbb{E} \hat{\mu}_N(A). \quad (6.29)$$

Нехай  $P(dy)$  — ймовірнісна міра. Тоді, використовуючи лему 5.2.2, можемо написати

$$S_1^N = \sup_{b,x} \left| \frac{1}{s_N} \int_{R^d} K \left( \frac{x - \sum_{i=1}^d b_i y_i}{s_N} \right) P(dy) - \frac{1}{s_N} \int_{R^d} K \left( \frac{x - \sum_{i=1}^d b_i y_i}{s_N} \right) \hat{\mu}_N(dy) \right| \leq \frac{a}{s_N} \sup_{A \in S} |P(A) - \hat{\mu}_N(A)|, \quad (6.30)$$

де  $S$  — система множин з умови (ii). Тут ми скористалися умовою (i). Далі, відзначимо, що

$$\sup_{A \in S} |P(A) - \hat{\mu}_N(A)| \leq \sup_{1 \leq j \leq N} |a_{j,N}^i| + 1 \leq \tilde{a} + 1,$$

а також введемо позначення

$$\tilde{C}_N^k = \sup_{1 \leq j \leq N} |a_{j,N}^i| + \sup_{1 \leq j \leq N} a_{j,N}^i - \inf_{1 \leq j \leq N} a_{j,N}^i.$$

де  $g^G$  — функція росту класу  $G$ .

Зробимо наступну оцінку:

$$E \sup_{A \in S} \left| \frac{P(A) - \hat{\mu}_N(A)}{\tilde{C}_N^k} \right| \leq \int_0^{\frac{\tilde{a}+1}{\tilde{C}_N^k}} P \left\{ \sup_{A \in S} \left| \frac{P(A) - \hat{\mu}_N(A)}{\tilde{C}_N^k} \right| \geq \lambda \right\} d\lambda$$

Оскільки  $\tilde{C}_N^k$  — обмежена знизу величина, то, не обмежуючи загальності можемо вважати, що

$$\frac{2M}{\sqrt[4]{N}} < \frac{\tilde{a} + 1}{\tilde{C}_N^k}.$$

Розіб'ємо інтеграл на дві частини, і до оцінки другої застосуємо наслідок 2.2.4:

$$\begin{aligned} E \sup_{A \in S} \left| \frac{P(A) - \hat{\mu}_N(A)}{\tilde{C}_N^k} \right| &\leq \int_0^{\frac{2M}{\sqrt[4]{N}}} d\lambda + \int_{\frac{2M}{\sqrt[4]{N}}}^{\frac{\tilde{a}+1}{\tilde{C}_N^k}} \left( 6MN g^S(2N) \exp \left( -\frac{\lambda^2 N}{32M^2} \right) + \right. \\ &\quad \left. + 2 \exp \left( -\frac{\lambda^2 N}{8M^2} \right) \right) d\lambda \leq 2MN^{-\frac{1}{4}} + \left( 6MN g^S(2N) \exp \left( -\frac{1}{8} \sqrt{N} \right) + \right. \\ &\quad \left. + 2 \exp \left( -\frac{1}{2} \sqrt{N} \right) \right) \left( \frac{\tilde{a} + 1}{\tilde{C}_N^k} - \frac{2M}{\sqrt[4]{N}} \right). \end{aligned} \quad (6.31)$$

(тут  $g^S$  — функція росту класу  $S$ ). З оцінок (6.30), (6.31) випливає, що

$$\begin{aligned} \mathbb{E} S_1^N \leq O\left(s_N^{-1} N^{-\frac{1}{4}}\right) + O\left(s_N^{-1} N g^S(2N) \exp\left(-\frac{1}{8}\sqrt{N}\right)\right) + \\ + O\left(s_N^{-1} \exp\left(-\frac{1}{2}\sqrt{N}\right)\right). \end{aligned} \quad (6.32)$$

При  $N \rightarrow \infty$  права частина (6.32) прямує до нуля: перший доданок через умову (v); другий і третій завдяки тому, що  $g^S(N)$ , як функція росту VC класу  $S$ , зростає як степенева функція,  $s_N^{-1}$  теж, а експоненційні множники прямують до нуля швидше, ніж степеневі. Отже,

$$\mathbb{E} S_1^N = o(1) \quad \text{при } N \rightarrow \infty. \quad (6.33)$$

Тепер оцінимо  $S_2^N$ .

$$\begin{aligned} \frac{1}{N s_N} \mathbb{E} \sum_{j=1}^N a_{j,N}^k K\left(\frac{x - S_j^N(b)}{s_N}\right) &= \frac{1}{N s_N} \sum_{j=1}^N a_{j,N}^k \mathbb{E} K\left(\frac{x - S_j^N(b)}{s_N}\right) = \\ &= \frac{1}{N s_N} \sum_{j=1}^N a_{j,N}^k \sum_{k=1}^M w_{j,N}^k \int_{-\infty}^{\infty} K\left(\frac{x-y}{s_N}\right) u_k(b, y) dy = \\ &= \int_{-\infty}^{\infty} K(z) u_k(b, x - s_N z) dz. \end{aligned}$$

Тут ми скористалися незміщеністю оцінок з мініміксними коефіцієнтами. Отже,  $S_2^N$  можна представити наступним чином

$$S_2^N = |\mathbb{E} u_k(b, x - s_N \eta) - u_k(b, X)|,$$

де  $\eta$  — в.в. зі щільністю  $K(x)$ . Оскільки  $\|b\| = 0$ , у вектора  $b = (b_1, b_2, \dots, b_d)$  хоча б одна компонента ненульова, не обмежуючи загальності, вважаємо  $b_d \neq 0$ . Тоді щільність випадкової величини  $\mu_k$  можна представити так:

$$u_k(b, x) = \frac{1}{b_d} \int_{-\infty}^{\infty} dx_1 \dots \int_{-\infty}^{\infty} h_k\left(x_1, \dots, x_{d-1}, \frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i\right)\right) dx_{d-1}.$$

Для спрощення далі будемо писати  $h_k(x_d)$  замість  $h_k(x_1, \dots, x_{d-1}, x_d)$ . Виберемо послідовність  $t = t_N \rightarrow \infty$ ,  $N \rightarrow \infty$ . Зробимо наступні оцінки

$$S_2^N = \frac{1}{|b_d|} \mathbb{E} \int_{-\infty}^{\infty} dx_1 \dots \int_{-\infty}^{\infty} \left| h_k\left(\frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i - s_N \eta\right)\right) - \right.$$

$$\begin{aligned}
& -h_k \left( \frac{1}{b_d} \left( x - \sum_{i=1}^{d-1} b_i x_i \right) \right) \Big| dx_{d-1} \leq \\
& \leq \frac{1}{|b_d|} \mathbb{E} \int_{x_1^2 + \dots + x_{d-1}^2 \leq t^2} dx_1 \dots dx_{d-1} \left| h_k \left( \frac{1}{b_d} \left( x - \sum_{i=1}^{d-1} b_i x_i - s_N \eta \right) \right) - \right. \\
& \quad \left. - h_k \left( \frac{1}{b_d} \left( x - \sum_{i=1}^{d-1} b_i x_i \right) \right) \right| + \\
& \quad + \frac{1}{|b_d|} \mathbb{E} \int_{x_1^2 + \dots + x_{d-1}^2 > t^2} dx_1 \dots dx_{d-1} h_k \left( \frac{1}{b_d} \left( x - \sum_{i=1}^{d-1} b_i x_i - s_N \eta \right) \right) + \\
& \quad + \frac{1}{|b_d|} \int_{x_1^2 + \dots + x_{d-1}^2 > t^2} dx_1 \dots dx_{d-1} h_k \left( \frac{1}{b_d} \left( x - \sum_{i=1}^{d-1} b_i x_i \right) \right) = A + B + C. \quad (6.34)
\end{aligned}$$

Для оцінки  $A$  застосуємо теорему Лагранжа:

$$A \leq \mathbb{E} \frac{s_N |\eta|}{b_d^2} \int_{x_1^2 + \dots + x_{d-1}^2 \leq t^2} dx_1 \dots dx_{d-1} \left| \frac{\partial h_k}{\partial x_d} \left( x_1, \dots, x_{d-1}, \frac{s - \sum_{i=1}^{d-1} b_i x_i}{b_d} \right) \right| \quad (6.35)$$

Тут  $s$  — деяка в.в., що знаходиться між величинами

$$\frac{1}{b_d} \left( x - \sum_{i=1}^{d-1} b_i x_i - s_N \eta \right)$$

та

$$\frac{1}{b_d} \left( x - \sum_{i=1}^{d-1} b_i x_i \right).$$

У інтегралі правої частини (6.35) робимо заміну  $x_1 = v b_d^2$ , а потім використовуємо умову обмеженості частинних похідних (iii). Маємо

$$\begin{aligned}
& A \leq s_N \mathbb{E} |\eta| \times \\
& \times \int_{\frac{v^2}{b_d^2} + \dots + x_{d-1}^2 \leq t^2} dv \dots dx_{d-1} \left| \frac{\partial h_k}{\partial x_d} \left( v b_d^2, \dots, x_{d-1}, \frac{1}{b_d} \left( s - \sum_{i=1}^{d-1} b_i x_i \right) \right) \right| \leq \\
& \leq s_N c_{kd} \mathbb{E} |\eta| \lambda \left\{ (v, x_2, \dots, x_{d-1}) \in R^{d-1} : \frac{v^2}{b_d^2} + x_2^2 + \dots + x_{d-1}^2 \leq t^2 \right\} \leq
\end{aligned}$$

$$\begin{aligned} &\leq s_N c_{kd} \mathbf{E} |\eta| \lambda \left\{ (v, x_2, \dots, x_{d-1}) \in R^{d-1} : v^2 + x_2^2 + \dots + x_{d-1}^2 \leq t^2 \right\} \leq \\ &\leq s_N c_{kd} \mathbf{E} |\eta| (2t)^{d-1}. \end{aligned} \quad (6.36)$$

(Тут  $\lambda(A)$  — міра Лебега множини  $A$ ). Тепер оцінимо  $B$ .

$$B = \frac{1}{|b_d|} \mathbf{E} \int_{x_1^2 + \dots + x_{d-1}^2 > t^2; x_d = x - \sum_{i=1}^{d-1} b_i x_i - s_N \eta} dx_1 \dots dx_{d-1} h_k \left( x_1, \dots, x_{d-1}, \frac{x_d}{b_d} \right).$$

Оскільки підінтегральна функція невід'ємна, то, збільшуючи область інтегрування, ми не зменшимо величину інтеграла.

$$B \leq \frac{1}{|b_d|} \int_{x_1^2 + \dots + x_{d-1}^2 > t^2; -\infty < x_d < \infty} dx_1 \dots dx_d h_k \left( x_1, \dots, x_{d-1}, \frac{x_d}{b_d} \right).$$

Зробимо заміну змінної  $v_d = \frac{x_d}{b_d}$  в інтегралі.

$$B \leq \int_{x_1^2 + \dots + x_{d-1}^2 > t^2; -\infty < v_d < \infty} dx_1 \dots dv_d h_k(x_1, \dots, v_d). \quad (6.37)$$

Доданок  $C$  оцінюється аналогічно  $B$ .

$$C \leq \int_{x_1^2 + \dots + x_{d-1}^2 > t^2; -\infty < v_d < \infty} dx_1 \dots dv_d h_k(x_1, \dots, v_d). \quad (6.38)$$

Отже, враховуючи (6.34)–(6.38), можемо написати

$$\begin{aligned} S_2^N &\leq c_{kd} \mathbf{E} |\eta| s_N (2t)^{d-1} + \\ &+ 2 \int_{x_1^2 + \dots + x_{d-1}^2 > t^2; -\infty < v_d < \infty} dx_1 \dots dv_d h_k(x_1, \dots, v_d). \end{aligned} \quad (6.39)$$

Зрозуміло, що можна вибрати  $t = t_N$  таким чином, щоб  $s_N (2t)^{d-1} \rightarrow 0$  при  $N \rightarrow \infty$ . При такому виборі права частина (16) буде прямувати до 0 при  $N \rightarrow \infty$  і отже,

$$S_2^N \rightarrow 0 \quad \text{при } N \rightarrow \infty \quad (6.40)$$

рівномірно по  $x$  та  $b$ . Збираючи докупи твердження (6.28), (6.33) та (6.40), завершуємо доведення лєми.

**Зауваження.** Якщо  $K(x)$  — кусково-монотонна функція зі скінченною кількістю інтервалів монотонності, то система подій  $S$ , визначена в (ii), являє собою VC-клас.

Дійсно, в цьому випадку існує скінченна кількість  $G$  відрізків  $[d_l, d_{l+1}]$  таких, що

$$\left\{ y : K \left( \frac{x - \sum_{i=1}^d b_i y_i}{s_N} \right) \geq c \right\} = \cup_{l=1}^G \left\{ x - s_N d_{l+1} \leq \sum_{i=1}^d b_i y_i \leq x - s_N d_l \right\}.$$

Кожна така множина — прошарок між двома гіперплощинами; об'єднання скінченної кількості таких множин є VC-класом [7], с.232.

**Теорема 6.4.2** В умовах (i)–(iv)

$$\mathbb{E} \sup_b \left| \hat{L}_g^N(b) - L_g^*(b) \right| \rightarrow 0 \quad \text{при } N \rightarrow \infty.$$

**Доведення.** За нерівністю Джорфі (наслідок 6.1.1),

$$\hat{L}_g^N - L_g^* \leq 2 \sum_{k=1}^M \int_{-\infty}^{\infty} |u_k(x) - \hat{u}_k^N(x)| dx. \quad (6.41)$$

Виберемо послідовність  $D_N \rightarrow \infty, N \rightarrow \infty$  та зробимо наступні оцінки

$$\begin{aligned} \mathbb{E} \int_{-\infty}^{\infty} |\hat{u}_k^N(b, x) - u_k(b, x)| dx &\leq \mathbb{E} \int_{-D_N}^{D_N} |\hat{u}_k^N(b, x) - u_k(b, x)| dx + \\ &+ \mathbb{E} \left( \int_{D_N}^{\infty} \hat{u}_k^N(b, x) dx + \int_{-\infty}^{-D_N} \hat{u}_k^N(b, x) dx \right) + \\ &+ \left( \int_{D_N}^{\infty} u_k(b, x) dx + \int_{-\infty}^{-D_N} u_k(b, x) dx \right) = A + B + C. \end{aligned} \quad (6.42)$$

Оцінимо кожний доданок у правій частині (6.42).

$$A \leq 2D_N \mathbb{E} \sup_{b, x} |\hat{u}_k^N(b, x) - u_k(b, x)|. \quad (6.43)$$

За лемою 6.4.1,  $\mathbb{E} \sup_{b, x} |\hat{u}_k^N(b, x) - u_k(b, x)| \rightarrow 0, N \rightarrow \infty$  і отже, можна вибрати таку послідовність  $D_N \rightarrow \infty$ , щоб права частина (6.43) прямувала до нуля.

Тепер оцінимо доданок  $C$ . За нерівністю Коші-Буняковського<sup>4</sup>

$$-\sqrt{\sum_{i=1}^d (\eta_k^i)^2} \leq \sum_{i=1}^d b_i \eta_k^i \leq \sqrt{\sum_{i=1}^d (\eta_k^i)^2} \quad (6.44)$$

<sup>4</sup>Нагадаємо, що  $\eta_k = (\eta_k^1, \dots, \eta_k^d)^T$  це випадкові вектори з розподілом  $H_k$ , а  $u_k(b, x)$  — щільність розподілу випадкової величини  $\sum_{i=1}^d b_i \eta_k^i$ .

Позначимо  $F(x)$  функцію розподілу, що відповідає щільності  $u_k(b, \cdot)$  та при оцінюванні скористаємося нерівністю (6.44).

$$\begin{aligned} \int_{D_N}^{\infty} u_k(b, x) dx &= 1 - F(D_N) \leq \mathbf{P} \left\{ \sqrt{\sum_{i=1}^d (\eta_k^i)^2} \geq D_N \right\}; \\ \int_{-\infty}^{-D_N} u_k(b, x) dx &= F(-D_N) \leq \\ &\leq \mathbf{P} \left\{ -\sqrt{\sum_{i=1}^d (\eta_k^i)^2} \leq -D_N \right\} = \mathbf{P} \left\{ \sqrt{\sum_{i=1}^d (\eta_k^i)^2} > D_N \right\}. \end{aligned}$$

З цих оцінок випливає, що

$$C \leq 2 \mathbf{P} \left\{ \sum_{i=1}^d (\eta_k^i)^2 \geq D_N^2 \right\} \rightarrow 0 \quad \text{при } N \rightarrow \infty. \quad (6.45)$$

Аналогічно оцінимо доданок  $B$ . Не обмежуючи загальності, вважаємо, що  $s_N \leq 1$ . За нерівністю Коші-Буняковського

$$\left( \sum_{i=1}^d b_i \eta_k^i + s_N \eta \right)^2 \leq (1 + s_N^2) \left( \sum_{i=1}^d (\eta_k^i)^2 + \eta^2 \right)$$

Отже,

$$\begin{aligned} -\sqrt{\sum_{i=1}^d (\eta_k^i)^2 + \eta^2} &\leq \sum_{i=1}^d b_i \eta_k^i + s_N \eta \leq \sqrt{\sum_{i=1}^d (\eta_k^i)^2 + \eta^2}. \\ \mathbf{E} \int_{D_N}^{\infty} \hat{u}_k^N(b, x) dx &= \frac{1}{N s_N} \mathbf{E} \int_{D_N}^{\infty} \sum_{j=1}^N a_{j,N}^k K \left( \frac{x - S_j^N(b)}{s_N} \right) dx = \\ &= \int_{D_N}^{\infty} \mathbf{E} u_k(b, x - s_N \eta) dx = \mathbf{E} (1 - F(D_N - s_N \eta)) = \\ &= \mathbf{P} \left\{ \sum_{i=1}^d b_i \eta_k^i + s_N \eta \geq D_N \right\} \leq \mathbf{P} \left\{ 2 \sqrt{\sum_{i=1}^d (\eta_k^i)^2 + \eta^2} \geq D_N \right\}; \\ \mathbf{E} \int_{-\infty}^{-D_N} \hat{u}_k^N(b, x) dx &= \mathbf{E} F(-D_N - s_N \eta) \leq \end{aligned}$$



$$\leq \mathbf{P} \left\{ -2 \sqrt{\sum_{i=1}^d (\eta_k^i)^2} + \eta^2 < -D_N \right\} = \mathbf{P} \left\{ 2 \sqrt{\sum_{i=1}^d (\eta_k^i)^2} + \eta^2 > D_N \right\}. \quad (6.46)$$

З оцінок (6.46) випливає

$$B \leq 2 \mathbf{P} \left\{ \sum_{i=1}^d (\eta_k^i)^2 + \eta^2 \geq \frac{D_N^2}{4} \right\} \rightarrow 0 \quad \text{при } N \rightarrow \infty. \quad (6.47)$$

Збираючи до купи граничні твердження (6.43), (6.45), (6.47) та оцінку (6.42), за нерівністю Джорфі отримуємо твердження теореми 6.4.2.

**Доведення теореми 6.4.1.** Зробимо такі перетворення

$$L_{\hat{g}}(\hat{b}) - L_g^*(b^*) = (L_{\hat{g}}(\hat{b}) - \hat{L}_g^N(\hat{b})) + (\hat{L}_g^N(\hat{b}) - \hat{L}_g^N(b^*)) + \hat{L}_g^N(b^*) - L_g^*(b^*). \quad (6.48)$$

Оскільки

$$\hat{b} = \operatorname{argmin}_b \hat{L}_g^N(b),$$

а класифікатор побудований так, що

$$L_{\hat{g}}(\hat{b}) \leq \hat{L}_g^N(\hat{b}),$$

то перші два доданки у (6.48) є від'ємними.

З (6.48) випливає

$$\begin{aligned} 0 \leq L_{\hat{g}}(\hat{b}) - L_g^*(b^*) &\leq \left| \hat{L}_g^N(b^*) - L_g^*(b^*) \right| \leq \\ &\leq \sup_b \left| \hat{L}_g^N(b) - L_g^*(b) \right|. \end{aligned} \quad (6.49)$$

За теоремою 6.4.2 математичне сподівання правої частини (26) прямує до 0 при  $N \rightarrow \infty$ . Теорема 6.4.1 доведена.

Твердження наслідку 6.4.1 випливає з теореми 6.4.1 та нерівності Чебишова.

## 6.5 Швидкість збіжності класифікаторів єдиного індекса

В п. 6.3 ми розглядали асимптотику порогових емпірично-байєсових класифікаторів для випадку, коли спостерігається одновимірна характеристика об'єкту. В цьому параграфі ми дослідимо асимптотику порогового ЕБК-класифікатора, побудованого на основі "сумарного індексу", введеного в п. 6.4.

Для спрощення викладу ми обмежимося випадком двовимірною вектора спостережуваних характеристик  $\xi = (\xi^1, \xi^2)^T$ , хоча отримані асимптотичні результати без принципових труднощів можна перенести на випадок більших вимірностей.

Постановка задачі має наступний вигляд. Нехай в нашому розпорядженні є вибірка з  $N$  об'єктів, кожен з яких може належати одному з двох класів. Спостерігається векторна характеристика  $\xi_{j:N} = (\xi_{j:N}^1, \xi_{j:N}^2) \in R^2$   $j$ -го об'єкту;  $\xi_{j:N}$  незалежні між собою. Концентрації  $k$ -ї компоненти суміші під час  $j$ -го спостереження, тобто безумовні ймовірності того, що  $j$ -й об'єкт відноситься до класу  $k$ , вважаються відомими і рівними  $w_{j:N}^k$ ;  $k = 1, 2$ . Щільності розподілу елементів з одного класу позначаємо  $h_k(x_1, x_2)$ ;  $k = 1, 2$  і вважаємо їх невідомими.

Наша мета — маючи в розпорядженні навчаючу вибірку  $\{\xi_{j:N}, j = \overline{1, N}\}$ , побудувати класифікатор  $g : R^2 \rightarrow \{1, 2\}$ , який за спостережуваними характеристиками об'єкту оцінював би номер класу, якому цей об'єкт належить.

Отже, проектуємо вектор спостережуваних характеристик  $(\xi^1, \xi^2)$  на деякий не випадковий напрямок  $\bar{b} = (\cos \alpha, \sin \alpha)$ , отримуємо "сумарний індекс"

$$S^N(\alpha) = \xi^1 \cos \alpha + \xi^2 \sin \alpha,$$

і на його основі будуємо пороговий класифікатор типу (6.12):

$$g_{t,\alpha}^B(\xi^1, \xi^2) = \begin{cases} 1, & S^N(\alpha) \leq t; \\ 2, & S^N(\alpha) > t. \end{cases} \quad (6.50)$$

Як було відмічено в п. 6.3, при баєсовому підході найкращим вважається класифікатор, який мінімізує ймовірність помилки

$$L^*(t) = p_1(1 - H_1(t)) + p_2 H_2(t),$$

де  $p_1, p_2$  - апіорно відомі концентрації компонентів суміші,  $H_i(t)$ ;  $i = 1, 2$  - функція розподілу спостережуваних характеристик об'єкту, що належить  $i$ -му класу. Оскільки розподіли  $H_1(t), H_2(t)$  невідомі, то за навчаючою вибіркою будуємо  $\hat{H}_1(t), \hat{H}_2(t)$  - оцінки для них, і, підставивши в формулу для баєсівського класифікатора (6.50), отримуємо емпірично-баєсів класифікатор. За оцінку ймовірності помилкової класифікації беруть

$$\hat{L}^N(t) = p_1(1 - \hat{H}_1(t)) + p_2 \hat{H}_2(t).$$

Оптимальний напрямок  $\alpha^*$  і поріг баєсівського класифікатора  $t^B$  обираються з умови

$$(\alpha^*, t^B) = \underset{(\alpha, t)}{\operatorname{argmin}} L^*(\alpha, t).$$

Оцінка напрямку та порогу визначається як їх вибірковий аналог:

$$(\hat{\alpha}, \hat{t}) = \underset{(\alpha, t)}{\operatorname{argmin}} \hat{L}(\alpha, t).$$

За оцінку  $H_i(t)$  беремо інтеграл від ядерної оцінки щільності, що задається формулою (4.2), при цьому мінімаксні вагові коефіцієнти  $a_{j:N}^1, a_{j:N}^2$  задаються (2.15).

Чому ми використовуємо інтеграли від  $\hat{u}_i^N$  замість того, щоб одразу підставити у вираз для ймовірності похибки емпіричні міри? Як ми переконались у п. 6.3, таке згладжування за допомогою ядерного оцінювання щільності дозволяє поліпшити точність наближення справжнього баєсового класифікатора його емпіричним аналогом.

Таким чином,

$$\hat{u}_i^N(\alpha, t) = \frac{1}{Ns_N} \sum_{j=1}^N a_{j:N}^i K\left(\frac{t - S_j^N(\alpha)}{s_N}\right); \quad i = 1, 2; \quad (6.51)$$

$$\hat{H}_i^\alpha(t) = \frac{1}{N} \sum_{j=1}^N a_{j:N}^i F\left(\frac{t - S_j^N(\alpha)}{s_N}\right),$$

де  $F(t) = \int_{-\infty}^t K(x)dx$ . При цьому вагові коефіцієнти  $a_{j:N}^i$  задовольняють умову незміщеності 2.7

$$\langle w^k a^i \rangle = \frac{1}{N} \sum_{j=1}^N w_{j:N}^k a_{j:N}^i = \mathbb{I}\{k = i\}. \quad (6.52)$$

Щільності компонент суміші сумарного індексу  $S_j^N(\alpha)$  позначимо як  $u_i(\alpha, t)$  і помітимо, що

$$\begin{aligned} u_i(\alpha, t) &= \left( \int_{-\infty}^{\infty} dx_2 \int_{-\infty}^{\infty} \mathbb{I}\{x_1 \cos \alpha + x_2 \sin \alpha < t\} h_i(x_1, x_2) dx_1 \right)'_t = \\ &= \int_{-\infty}^{\infty} h_i(t \cos \alpha - u \sin \alpha, t \sin \alpha + u \cos \alpha) du; \quad i = 1, 2. \end{aligned}$$

Припустимо, що виконуються наступні умови:

(a1) для всіх  $\delta > 0$  виконується:

$$\inf_{\alpha, t: (\alpha - \alpha^*)^2 + (t - t^B)^2 > \delta^2} L^*(\alpha, t) > L^*(\alpha^*, t^B);$$

(a2) вагові коефіцієнти  $a_{j:N}^i$  обмежені:  $\tilde{a} = \sup_{i,j,N} |a_{j:N}^i| < \infty$ ;

(a3) щільності компонент сумарного індексу  $S_j^N(\alpha)$  є обмеженими по  $t$  і  $\alpha$ :  $\max_{i=1,2} \sup_{(t,\alpha)} |u_i(\alpha, t)| < \infty$ ;

(a4)  $\int_{-\infty}^{\infty} |x| K(x) dx < \infty$ .

**Лема 6.5.1** *За умов (a1)-(a4)  $\hat{t} \rightarrow t^B, \hat{\alpha} \rightarrow \alpha^*$  за ймовірністю при  $N \rightarrow \infty$ .*

**Доведення леми 6.5.1.** Спершу доведемо, що  $L^*(\hat{\alpha}, \hat{t}) \rightarrow L^*(\alpha^*, t^B)$  за ймовірністю. Оскільки

$$\begin{aligned} 0 \leq L^*(\hat{\alpha}, \hat{t}) - L^*(\alpha^*, t^B) &= L^*(\hat{\alpha}, \hat{t}) - \hat{L}(\hat{\alpha}, \hat{t}) + \hat{L}(\hat{\alpha}, \hat{t}) - \hat{L}(\alpha^*, t^B) + \\ &+ \hat{L}(\alpha^*, t^B) - L^*(\alpha^*, t^B) \leq 2 \sup_{(\alpha, t)} \left| \hat{L}(\alpha, t) - L^*(\alpha, t) \right|, \end{aligned}$$

то для цього достатньо довести, що  $\sup_{(\alpha, t)} \left| \hat{L}(\alpha, t) - L^*(\alpha, t) \right| \rightarrow 0$  при  $N \rightarrow \infty$  за ймовірністю. Зробимо наступні оцінки.

$$\hat{L}(\alpha, t) - L^*(\alpha, t) = p_1 \left( H_1^\alpha(t) - \hat{H}_1^\alpha(t) \right) + p_2 \left( H_2^\alpha(t) - \hat{H}_2^\alpha(t) \right).$$

$$\begin{aligned} &\sup_{(\alpha, t)} \left| \hat{H}_i^\alpha(\alpha, t) - H_i^\alpha(\alpha, t) \right| \leq \\ &\leq \sup_{(\alpha, t)} \left| \frac{1}{N} \sum_{j=1}^N a_{j:N}^i F \left( \frac{t - S_j^N(\alpha)}{s_N} \right) - \frac{1}{N} \sum_{j=1}^N a_{j:N}^i \mathbf{E} F \left( \frac{t - S_j^N(\alpha)}{s_N} \right) \right| \quad (6.53) \\ &+ \sup_{(\alpha, t)} \left| \frac{1}{N} \sum_{j=1}^N a_{j:N}^i \mathbf{E} F \left( \frac{t - S_j^N(\alpha)}{s_N} \right) - H_i^\alpha(t) \right| = V_1^N + V_2^N \end{aligned}$$

Розглянемо  $V_1^N$ . Визначимо зважену емпіричну міру співвідношенням, як ми це робили у (2.3):

$$\hat{\mu}_N(A) = \frac{1}{N} \sum_{j=1}^N a_{j:N}^i \mathbf{1} \{ \xi_{j:N} \in A \}; \quad \bar{\mu}_N(A) = \mathbf{E} \hat{\mu}_N(A) = \mathbf{P}(A) = H_i(A);$$

$\Pr(dy)$  - ймовірнісна міра. Тоді, згідно з лемою 5.2.2,

$$\begin{aligned} V_1^N = \sup_{(\alpha, t)} \left| \int_{R^2} F \left( \frac{t - \cos \alpha y_1 - \sin \alpha y_2}{s_N} \right) H_i(dy) \right. \\ \left. - \int_{R^2} F \left( \frac{t - \cos \alpha y_1 - \sin \alpha y_2}{s_N} \right) \hat{\mu}_N(dy) \right| \leq \sup_{A \in S} |H_i(A) - \hat{\mu}_N(A)|. \end{aligned} \quad (6.54)$$

Тут  $S$  — сукупність множин:

$$S = \left\{ (y_1, y_2) : F \left( \frac{t - \cos \alpha y_1 - \sin \alpha y_2}{s_N} \right) \geq C \right\}$$

для всіх  $\alpha, t, C$ . Позначимо

$$C_N^i = \sup_{1 \leq j \leq N} |a_{j:N}^i| + \sup_{1 \leq j \leq N} a_{j:N}^i - \inf_{1 \leq j \leq N} a_{j:N}^i$$

і відмітимо, що  $\sup C_N^i < 2\tilde{a}$ .

Тепер застосуємо теорему 2.2.3. Оскільки функція  $F$  є монотонною, то  $S$  являє собою VC-клас (див. с. 40), тобто функція росту класу задовольняє  $g^G(N) \leq 3 \frac{(N-1)^2}{2}$ . Згідно з теоремою 2.2.3 маємо для довільної сталої  $\lambda > 4/N$ :

$$\begin{aligned} \mathbb{P}\{\sup_{A \in S} |\hat{\mu}_N(A) - \mathbb{P}(A)| \geq \lambda\} &\leq \mathbb{P}\left\{\sup_{A \in S} \frac{|\hat{\mu}_N(A) - \mathbb{P}(A)|}{C_N^i} \geq \frac{\lambda}{\sup C_N^i}\right\} \\ &\leq 4 \left( 3N g^G(2N) \exp\left(-\frac{\lambda^2 N}{(2\tilde{a})^2 128}\right) + \exp\left(-\frac{\lambda^2 N}{(2\tilde{a})^2 32}\right) \right). \end{aligned} \quad (6.55)$$

Права частина (6.55) при  $N \rightarrow \infty$  прямує до 0, бо  $g^G(2N)$  зростає як степенева функція, а експоненціальний множник прямує до нуля швидше. Отже, з (6.54), (6.55) маємо: за ймовірністю

$$V_1^N \rightarrow 0, \quad N \rightarrow \infty. \quad (6.56)$$

Тепер оцінимо  $V_2^N$ , застосувавши співвідношення (6.52).

$$\begin{aligned} &\frac{1}{N} \sum_{j=1}^N a_{j:N}^i \mathbb{E} F \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) = \\ &= \frac{1}{N} \sum_{j=1}^N a_{j:N}^i \sum_{k=1}^2 w_{j:N}^k \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} F \left( \frac{t - x_1 \cos \alpha - x_2 \sin \alpha}{s_N} \right) h_k(x_1, x_2) dx_2 \end{aligned}$$

$$= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} F\left(\frac{t - x_1 \cos \alpha - x_2 \sin \alpha}{s_N}\right) h_i(x_1, x_2) dx_2.$$

Отже, позначивши  $(\eta_1^i, \eta_2^i)$  в.в. зі щільністю  $h_i(x_1, x_2)$ , маємо:

$$\begin{aligned} V_2^N &= \sup_{(\alpha, t)} \left| \mathbb{E} F\left(\frac{t - \eta_1^i \cos \alpha - \eta_2^i \sin \alpha}{s_N}\right) - \int_{-\infty}^t u_i(\alpha, x) dx \right| = \\ &= \sup_{(\alpha, t)} \left| s_N \int_{-\infty}^{\infty} F(v) u_i(\alpha, t - v s_N) dv - \int_{-\infty}^t u_i(\alpha, x) dx \right| = \\ &= \sup_{(\alpha, t)} \left| \int_{-\infty}^{\infty} \int_t^{t - v s_N} u_i(\alpha, x) dx K(v) dv \right| \leq s_N \sup_{(\alpha, t)} |u_i(\alpha, t)| \int_{-\infty}^{\infty} |v| K(v) dv. \end{aligned} \quad (6.57)$$

Зі співвідношень (6.53), (6.56), (6.57) робимо висновок, що

$$\sup_{(\alpha, t)} \left| \hat{H}_i^\alpha(t) - H_i^\alpha(t) \right| \rightarrow 0$$

за ймовірністю при  $N \rightarrow \infty$ , а, отже, і  $\sup_{(\alpha, t)} \left| \hat{L}(\alpha, t) - L^*(\alpha, t) \right| \rightarrow 0$ .

Звідси випливає, що  $L^*(\hat{\alpha}, \hat{t}) \rightarrow L^*(\alpha^*, t^B)$  за ймовірністю.

Згідно з умовою (a1) для довільного  $\delta > 0$  існує таке  $c > 0$ , що

$$L^*(\alpha^*, t^B) < -c + \inf_{(\alpha, t): (\alpha - \alpha^*)^2 + (t - t^B)^2 > \delta^2} L^*(\alpha, t).$$

Задамо довільне  $\delta > 0$  і припустимо, що виконується нерівність  $(\hat{\alpha} - \alpha^*)^2 + (\hat{t} - t^B)^2 > \delta^2$ . Тоді існує таке  $c > 0$ , що  $L^*(\alpha^*, t^B) < -c + L^*(\hat{\alpha}, \hat{t})$ . Ймовірність такої події за доведеним вище збігається до 0 при  $N \rightarrow \infty$ . Звідси випливає, що  $\mathbb{P}\{(\hat{\alpha} - \alpha^*)^2 + (\hat{t} - t^B)^2 > \delta^2\} \rightarrow_{N \rightarrow \infty} 0$ . Отже, для довільних сталих  $c_1, c_2 > 0$

$$\mathbb{P}\{|\hat{\alpha} - \alpha^*| > c_1; |\hat{t} - t^B| > c_2\} \leq \mathbb{P}\{(\hat{\alpha} - \alpha^*)^2 + (\hat{t} - t^B)^2 > c_1^2 + c_2^2\} \rightarrow_{N \rightarrow \infty} 0$$

Лема 6.5.1 доведена.

Тепер оцінимо швидкість цієї збіжності.

**Теорема 6.5.1** Нехай на додачу до умов (a1)-(a4) виконуються наступні умови ( $k=1,2$ ):

(a5) функції  $\hat{L}(\alpha, t); L^*(\alpha, t)$  двічі неперервно диференційовні;

$$u_k(t) < c < \infty;$$

$$\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 (x_1^4 + x_2^4) h_k(x_1, x_2) < \infty; \quad h_k(x_1, x_2) \rightarrow 0, \quad x_1, x_2 \rightarrow \pm\infty;$$

функції  $h_k(x_1, x_2)$  тричі неперервно-диференційовні і мають обмежені похідні до третього порядку включно;

$$\int_{-\infty}^{\infty} v^s \frac{\partial^2 h_i}{\partial x_i^2} (t \cos \alpha - v \sin \alpha, t \sin \alpha + v \cos \alpha) dv < \infty; \quad i = 1, 2 \quad \text{для всіх } \alpha, t;$$

$$\int_{-\infty}^{\infty} v^s \frac{\partial^2 h_i}{\partial x_1 \partial x_2} (t \cos \alpha - v \sin \alpha, t \sin \alpha + v \cos \alpha) dv < \infty; \quad s = 1, 2 \quad \text{для всіх } \alpha, t.$$

(a6)

$$\sigma_k^2(\alpha, t) = \lim_{N \rightarrow \infty} \sum_{r=1}^2 \langle (a^k)^2 w^r \rangle_N u_r(\alpha, t) < \infty;$$

$$\sigma^2(\alpha, t) = \lim_{N \rightarrow \infty} \sum_{r=1}^2 \langle (-p_1 a^1 + p_2 a^2)^2 w^r \rangle_N u_r(\alpha, t) < \infty;$$

$$\Delta^2(\alpha, t) = \lim_{N \rightarrow \infty} \sum_{r=1}^2 \langle (-p_1 a^1 + p_2 a^2)^2 w^r \rangle_N \times$$

$$\times \int_{-\infty}^{\infty} v^2 h_r(t \cos \alpha - v \sin \alpha, t \sin \alpha + v \cos \alpha) dv < \infty;$$

$$\omega^2(\alpha, t) = \lim_{N \rightarrow \infty} \sum_{r=1}^2 \langle (-p_1 a^1 + p_2 a^2)^2 w^r \rangle_N a_r(\alpha, t) < \infty;$$

$$\sigma^2(\alpha^*, t^B) \Delta^2(\alpha^*, t^B) - \omega^4(\alpha^*, t^B) > 0.$$

(a7)

$$\exists \frac{\partial u_k(t)}{\partial t} < \bar{c} < \infty;$$

існують і обмежені рівномірно по  $p, q$  інтеграли

$$\int_{-\infty}^{\infty} \frac{\partial h_k}{\partial x_i} (p - v \sin \alpha, q - v \cos \alpha) dv, \quad i = 1, 2;$$

$$\int_{-\infty}^{\infty} (1 + v^2) h_k(p - v \sin \alpha, q - v \cos \alpha) dv, \quad k = 1, 2;$$

$$\int_{-\infty}^{\infty} v^s h_k(t \cos \alpha - v \sin \alpha, t \sin \alpha + v \cos \alpha) dv < \infty; \quad k = 1, 2; \quad s = 1, 2.$$

(a8) умови, що накладаються на ядро:

$$d_1^2 = \int_{-\infty}^{\infty} (K'(z))^2 dz < \infty; \quad |K'(z)| < \infty; \quad K(z) \rightarrow 0, \quad z \rightarrow \pm\infty;$$

$$\int_{-\infty}^{\infty} zK(z)dz = 0; \quad D^2 = \int_{-\infty}^{\infty} z^2 K(z)dz < \infty;$$

$$d^2 = \int_{-\infty}^{\infty} K^2(z)dz < \infty; \quad \int_{-\infty}^{\infty} |z|^3 K(z)dz < \infty; \quad \int_{-\infty}^{\infty} z^2 K^2(z)dz < \infty.$$

(a9)

$$s_N = cN^{-1/5}; \quad \text{де } c - \text{деяка стала.}$$

Тоді при  $N \rightarrow \infty$ :

$$N^{2/5} \begin{pmatrix} \hat{\alpha} - \alpha^* \\ \hat{t} - t^B \end{pmatrix} \Rightarrow \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}$$

де  $(\eta_1, \eta_2)^T$  гауссів випадковий вектор з математичним сподіванням  $A$  та коваріаційною матрицею  $S$ , причому

$$A = D^2 B \begin{pmatrix} \frac{\partial S}{\partial t}(-p_1 T_1 + p_2 T_2) + \frac{1}{2} \frac{\partial r}{\partial t} \left( p_1 \frac{\partial^2 u_1}{\partial t^2} - p_2 \frac{\partial^2 u_2}{\partial t^2} \right) \\ \frac{1}{2} \frac{\partial r}{\partial \alpha} \left( -p_1 \frac{\partial^2 u_1}{\partial t^2} + p_2 \frac{\partial^2 u_2}{\partial t^2} \right) - \frac{\partial r}{\partial t} (p_1 T_1 - p_2 T_2) \end{pmatrix} \Big|_{(\alpha^*, t^B)}$$

$$B = \frac{1}{\left( \frac{\partial r}{\partial t} \right)^2 - \frac{\partial S}{\partial t} \frac{\partial r}{\partial \alpha}} \Big|_{(\alpha^*, t^B)};$$

$T_i(\alpha, t)$  задаються в формулюванні лема 6.5.2;

$$S = d^2 B^2 \begin{pmatrix} \beta^2 \sigma^2 - 2\beta\gamma\omega^2 + \gamma^2 \Delta^2 & \beta\mu\sigma^2 + \gamma^2 \Delta^2 - \gamma(\mu + \beta)\omega^2 \\ \beta\mu\sigma^2 + \gamma^2 \Delta^2 - \gamma(\mu + \beta)\omega^2 & \mu^2 \sigma^2 - 2\mu\gamma\omega^2 + \gamma^2 \Delta^2 \end{pmatrix} \Big|_{(\alpha^*, t^B)};$$

$$\beta = \frac{\partial S}{\partial t} \Big|_{(\alpha^*, t^B)}; \quad \gamma = \frac{\partial r}{\partial t} \Big|_{(\alpha^*, t^B)}; \quad \mu = \frac{\partial r}{\partial \alpha} \Big|_{(\alpha^*, t^B)};$$

$S(\alpha, t)$ ,  $r(\alpha, t)$  задаються формулами (6.60), (6.61).

**Доведення теореми 6.5.1.** Запишемо необхідні умови екстремуму для  $L^*$  та  $\hat{L}$  відповідно.

$$\begin{cases} \frac{\partial L^*}{\partial t}(\alpha^*, t^B) = 0; \\ \frac{\partial L^*}{\partial \alpha}(\alpha^*, t^B) = 0 \end{cases} \quad (6.58)$$



$$\begin{cases} \frac{\partial \hat{L}}{\partial t}(\hat{\alpha}, \hat{t}) = 0; \\ \frac{\partial \hat{L}}{\partial \alpha}(\hat{\alpha}, \hat{t}) = 0 \end{cases} \quad (6.59)$$

Беручи похідні, об'єднуючи перші і другі рівності формул (6.58), (6.59) відповідно і вводячи нові позначення, маємо

$$\begin{cases} S(\alpha^*, t^B) = -p_1 u_1(\alpha^*, t^B) + p_2 u_2(\alpha^*, t^B) = 0; \\ \hat{S}_N(\hat{\alpha}, \hat{t}) = -p_1 \hat{u}_1(\hat{\alpha}, \hat{t}) + p_2 \hat{u}_2(\hat{\alpha}, \hat{t}) = 0. \end{cases} \quad (6.60)$$

$$\begin{cases} r(\alpha^*, t^B) = -p_1 a_1(\alpha^*, t^B) + p_2 a_2(\alpha^*, t^B) = 0; \\ \hat{r}_N(\hat{\alpha}, \hat{t}) = -p_1 \hat{a}_1(\hat{\alpha}, \hat{t}) + p_2 \hat{a}_2(\hat{\alpha}, \hat{t}) = 0. \end{cases} \quad (6.61)$$

Тут  $\hat{u}_i(\alpha, t)$  задається формулою (6.51);

$$\begin{aligned} a_i(\alpha, t) &= \frac{\partial H_i^\alpha(t)}{\partial \alpha} = \int_{-\infty}^t dt \int_{-\infty}^{\infty} \frac{\partial h_i}{\partial x_2} (s \cos \alpha - v \sin \alpha, s \sin \alpha + v \cos \alpha) \times \\ &\times (s \cos \alpha - v \sin \alpha) dv - \int_{-\infty}^t dt \int_{-\infty}^{\infty} \frac{\partial h_i}{\partial x_1} (s \cos \alpha - v \sin \alpha, s \sin \alpha + v \cos \alpha) \times \\ &\times (s \sin \alpha + v \cos \alpha) dv = \\ &= - \int_{-\infty}^{\infty} v h_i(t \cos \alpha - v \sin \alpha, t \sin \alpha + v \cos \alpha) dv. \end{aligned} \quad (6.62)$$

(Тут ми врахували, що  $\int_{-\infty}^{\infty} \left( \cos \alpha \frac{\partial h_i}{\partial x_2} - \sin \alpha \frac{\partial h_i}{\partial x_1} \right) dv = 0$ ;

$$\int_{-\infty}^t \left( \cos \alpha \frac{\partial h_i}{\partial x_1} - \sin \alpha \frac{\partial h_i}{\partial x_2} \right) ds = h_i(t \cos \alpha - v \sin \alpha, t \sin \alpha + v \cos \alpha).$$

$$\begin{aligned} \hat{a}_i(\alpha, t) &= \frac{\partial \hat{H}_i^\alpha(t)}{\partial \alpha} = \frac{1}{N S_N} \sum_{j=1}^N a_{j:N}^i K \left( \frac{t - \xi_{j:N}^i \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) \times \\ &\times (\xi_{j:N}^1 \sin \alpha - \xi_{j:N}^2 \cos \alpha). \end{aligned} \quad (6.63)$$

Далі, позначимо  $\delta_N = \hat{\alpha} - \alpha^*$ ;  $\Delta_N = \hat{t} - t^B$  і розкладемо за формулою Тейлора ліву частину другої рівності (6.60):

$$\begin{aligned} 0 &= \hat{S}_N(\hat{\alpha}, \hat{t}) = \hat{S}_N(\alpha^* + \delta_N, t^B + \Delta_N) \approx \\ &\approx \hat{S}_N(\alpha^*, t^B) + \delta_N \frac{\partial \hat{S}_N}{\partial \alpha}(\alpha^*, t^B) + \Delta_N \frac{\partial \hat{S}_N}{\partial t}(\alpha^*, t^B). \end{aligned}$$

Ліву частину другої рівності (6.61) розкладаємо аналогічно, отримуємо систему наближених рівностей

$$\begin{cases} \delta_N \frac{\partial \hat{S}_N}{\partial \alpha}(\alpha^*, t^B) + \Delta_N \frac{\partial \hat{S}_N}{\partial t}(\alpha^*, t^B) \approx S(\alpha^*, t^B) - \hat{S}_N(\alpha^*, t^B); \\ \delta_N \frac{\partial \hat{r}_N}{\partial \alpha}(\alpha^*, t^B) + \Delta_N \frac{\partial \hat{r}_N}{\partial t}(\alpha^*, t^B) \approx r(\alpha^*, t^B) - \hat{r}_N(\alpha^*, t^B). \end{cases} \quad (6.64)$$

(Праворуч до першої рівності ми додали  $S(\alpha^*, t^B)$ , до другої  $r(\alpha^*, t^B)$ , бо вони рівні 0.) Розв'язуємо лінійну систему (6.64), спрощуємо позначення і помічаємо, що  $\frac{\partial \hat{S}_N}{\partial \alpha} = \frac{\partial \hat{r}_N}{\partial t}$ . Отже,

$$\begin{cases} \delta_N = \frac{(S - \hat{S}) \frac{\partial \hat{r}}{\partial t} - (r - \hat{r}) \frac{\partial \hat{S}}{\partial t}}{\left(\frac{\partial \hat{r}}{\partial t}\right)^2 - \frac{\partial \hat{S}}{\partial t} \frac{\partial \hat{r}}{\partial \alpha}}; & \Delta_N = \frac{(r - \hat{r}) \frac{\partial \hat{r}}{\partial t} - (S - \hat{S}) \frac{\partial \hat{S}}{\partial \alpha}}{\left(\frac{\partial \hat{r}}{\partial t}\right)^2 - \frac{\partial \hat{S}}{\partial t} \frac{\partial \hat{r}}{\partial \alpha}} \end{cases} \quad (6.65)$$

Тепер треба знайти порядок спадання правих частин (6.65). Завдяки виконанню умов (а5)-(а8) можна застосувати теорему 4.3.4. Отже, справедливе зображення

$$\frac{\partial \hat{u}_k}{\partial t}(t) = \bar{u}_k^{(1)}(t) + \frac{1}{\sqrt{N s_N^3}} \zeta_N^k, \quad k = 1, 2,$$

де  $\bar{u}_k^{(1)}(t) = \mathbb{E} \frac{\partial \hat{u}_k}{\partial t}(t) \rightarrow \frac{\partial u_k}{\partial t}(t)$ ,  $s_N \rightarrow \infty$ ; — є не випадковою функцією, а в.в.  $\zeta_N^k$  асимптотично нормальні з параметрами  $N(0, d_1^2 \sigma_k^2(t))$ . Згідно з цим зображенням

$$\begin{aligned} \frac{\partial \hat{S}_N}{\partial t} - \frac{\partial S}{\partial t} &= -p_1 \left( \bar{u}_1^{(1)}(t) - \frac{\partial u_1}{\partial t} \right) + p_2 \left( \bar{u}_2^{(1)}(t) - \frac{\partial u_2}{\partial t} \right) + \\ &+ \frac{1}{\sqrt{N s_N^3}} (p_2 \zeta_N^2 - p_1 \zeta_N^1) = o(1); \quad N \rightarrow \infty. \end{aligned}$$

Отже,

$$\frac{\partial \hat{S}_N}{\partial t}(\alpha^*, t^B) \rightarrow \frac{\partial S}{\partial t}(\alpha^*, t^B), \quad N \rightarrow \infty \quad (6.66)$$

за ймовірністю. Тепер впевнімося, що

$$\frac{\partial \hat{r}_N}{\partial t}(\alpha^*, t^B) \rightarrow \frac{\partial r}{\partial t}(\alpha^*, t^B), \quad N \rightarrow \infty \quad (6.67)$$

за ймовірністю. Для цього достатньо довести, що

$$\frac{\partial \hat{a}_i}{\partial t}(\alpha^*, t^B) \rightarrow \frac{\partial a_i}{\partial t}(\alpha^*, t^B), \quad N \rightarrow \infty; \quad i = 1, 2$$

за ймовірністю. Зробимо такі перетворення

$$\frac{\partial \hat{a}_i}{\partial t} - \frac{\partial a_i}{\partial t} = B_N^1 + B_N^2, \quad (6.68)$$

де

$$\begin{aligned} B_N^1 &= \frac{1}{N s_N^2} \sum_{j=1}^N a_{j:N}^i K' \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) (\xi_{j:N}^1 \sin \alpha - \xi_{j:N}^2 \cos \alpha) - \\ &\quad - \frac{1}{N s_N^2} \sum_{j=1}^N a_{j:N}^i \mathbb{E} K' \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) (\xi_{j:N}^1 \sin \alpha - \xi_{j:N}^2 \cos \alpha); \\ B_N^2 &= \frac{1}{N s_N^2} \sum_{j=1}^N a_{j:N}^i \mathbb{E} K' \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) \times \\ &\quad \times (\xi_{j:N}^1 \sin \alpha - \xi_{j:N}^2 \cos \alpha) - \frac{\partial a_i}{\partial t}. \end{aligned}$$

Скористаємось законом великих чисел (теорема 7.3.6) для аналізу  $B_N^1$ .

$$\begin{aligned} X_N &= \frac{a_{j:N}^i}{s_N^2} K' \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) (\xi_{j:N}^1 \sin \alpha - \xi_{j:N}^2 \cos \alpha); \\ \mathbb{E} X_N^2 &= \frac{(a_{j:N}^i)^2}{s_N^4} \sum_{l=1}^2 w_{j:N}^l \times \\ &\times \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 K' \left( \frac{t - x_1 \cos \alpha - x_2 \sin \alpha}{s_N} \right) (x_1 \sin \alpha - x_2 \cos \alpha) h_i(x_1, x_2). \end{aligned}$$

Цей інтеграл існує завдяки умовам (а5), (а8) і, як ми бачимо,  $\mathbb{E} X_N^2 = O\left(\frac{1}{s_N^4}\right)$ . Згідно з умовою (а9)  $\frac{\text{Var } X_N}{N} \rightarrow 0, N \rightarrow \infty$ . Отже, за ймовірністю

$$B_N \rightarrow 0; \quad N \rightarrow \infty. \quad (6.69)$$

Тепер розглянемо  $B_N^2 = \mathbb{E} \frac{\partial \hat{a}_i}{\partial t} - \frac{\partial a_i}{\partial t}$ . Після заміни змінної і інтегрування частинами (завдяки (а8)) маємо:

$$\begin{aligned} \mathbb{E} \frac{\partial \hat{a}_i}{\partial t} &= \frac{1}{s_N^2} \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 K' \left( \frac{t - x_1 \cos \alpha - x_2 \sin \alpha}{s_N} \right) (x_1 \sin \alpha - \\ &\quad - x_2 \cos \alpha) h_i(x_1, x_2) = - \int_{-\infty}^{\infty} K(z) \int_{-\infty}^{\infty} v \left( \cos \alpha \frac{\partial h_i}{\partial x_1} + \sin \alpha \frac{\partial h_i}{\partial x_2} \right) dv dz, \end{aligned} \quad (6.70)$$

де частинні похідні беруться в точці  $(t \cos \alpha - v \sin \alpha - z s_N \cos \alpha, t \sin \alpha + v \cos \alpha - z s_N \sin \alpha)$ . Відповідно

$$\begin{aligned} \frac{\partial a_i}{\partial t} &= \int_{-\infty}^{\infty} \left( (t \cos \alpha - v \sin \alpha) \frac{\partial h_i}{\partial x_2} - (t \sin \alpha + v \cos \alpha) \frac{\partial h_i}{\partial x_1} \right) dv = \\ &= - \int_{-\infty}^{\infty} v \left( \cos \alpha \frac{\partial h_i}{\partial x_1} + \sin \alpha \frac{\partial h_i}{\partial x_2} \right) dv, \end{aligned} \quad (6.71)$$

де похідні беруться в точці  $(t \cos \alpha - v \sin \alpha, t \sin \alpha + v \cos \alpha)$ . Згідно з умовою (a7) підінтегральна функція в інтегралі (6.70) має інтегровні мажоранти і, очевидно, прямує до підінтегральної функції (6.71) при  $s_N \rightarrow \infty$ . За теоремою Лебега про мажоровану збіжність  $B_N^2 \rightarrow 0$  при  $N \rightarrow \infty$ , і в сукупності з (6.68), (6.69) це дає твердження (6.67). Збіжність

$$\frac{\partial \hat{r}_N}{\partial \alpha}(\alpha^*, t^B) \rightarrow \frac{\partial r}{\partial \alpha}(\alpha^*, t^B), \quad N \rightarrow \infty \quad (6.72)$$

за ймовірністю можна довести за тією ж схемою.

З (6.66), (6.67) та (6.72) випливає, що знаменник формул (6.65) задовольняє:

$$\left( \frac{\partial \hat{r}_N}{\partial t} \right)^2 - \frac{\partial \hat{S}_N}{\partial t} \frac{\partial \hat{r}_N}{\partial \alpha} \Big|_{(\alpha^*, t^B)} \rightarrow \left( \frac{\partial r}{\partial t} \right)^2 - \frac{\partial S}{\partial t} \frac{\partial r}{\partial \alpha} \Big|_{(\alpha^*, t^B)}, \quad N \rightarrow \infty. \quad (6.73)$$

Тепер розглянемо чисельники формул (6.65).

**Лема 6.5.2** В умовах (a1)-(a8) справедливе граничне твердження

$$N^{2/5} \begin{pmatrix} \hat{S}_N - S \\ \hat{r}_N - r \end{pmatrix} \Big|_{(\alpha^*, t^B)} \Rightarrow \begin{pmatrix} \frac{D^2}{2} \left( -p_1 \frac{\partial^2 u_1}{\partial t^2} + p_2 \frac{\partial^2 u_2}{\partial t^2} \right) + \zeta_1 \\ \frac{D^2}{2} (-p_1 T_1 + p_2 T_2) + \zeta_2 \end{pmatrix} \Big|_{(\alpha^*, t^B)},$$

де

$$T_i(\alpha^*, t^B) = \int_{-\infty}^{\infty} (-v) \left( \frac{\partial^2 h_i}{\partial x_1^2} + 2 \frac{\partial^2 h_i}{\partial x_1 \partial x_2} + \frac{\partial^2 h_i}{\partial x_2^2} \right) \Big|_{x(\alpha^*, t^B)} dv;$$

$$x(\alpha, t) = (t \cos \alpha - v \sin \alpha, t \sin \alpha + v \cos \alpha); \quad i = 1, 2.$$

При цьому випадковий вектор  $\begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}$  розподілений нормально  $N(0, B^2)$ , де

$$B^2 = d^2 \begin{pmatrix} \sigma^2(\alpha^*, t^B) & \omega^2(\alpha^*, t^B) \\ \omega^2(\alpha^*, t^B) & \Delta^2(\alpha^*, t^B) \end{pmatrix}.$$

**Доведення леми 6.5.2.** Спершу доведемо, що

$$\hat{a}_i(\alpha, t) = \bar{a}_i^N(\alpha, t) + \frac{1}{\sqrt{N}s_N} \nu_N^i, \quad i = 1, 2,$$

де  $\bar{a}_i^N(\alpha, t) = \mathbb{E} \hat{a}_i(\alpha, t)$  - не випадкова функція;  $\bar{a}_i^N(\alpha, t) \rightarrow a_i(\alpha, t)$ ,  $N \rightarrow \infty$ ;  $\mathbb{E} \nu_N^i = 0$ .

Після застосування (6.52) і заміни змінної маємо

$$\begin{aligned} \bar{a}_i^N(\alpha, t) = \mathbb{E} \hat{a}_i(\alpha, t) &= \frac{1}{s_N} \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} K \left( \frac{t - x_1 \cos \alpha - x_2 \sin \alpha}{s_N} \right) (x_1 \sin \alpha - \\ &- x_2 \cos \alpha) h_1(x_1, x_2) dx_2 = - \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} K(z) v \times \\ &\times h_i(t \cos \alpha - v \sin \alpha - z s_N \cos \alpha, t \sin \alpha + v \cos \alpha - z s_N \sin \alpha) dv. \end{aligned}$$

Розкладемо за формулою Тейлора функцію  $h_i$  в точці  $x = x(v)$   $= (t \cos \alpha - v \sin \alpha, t \sin \alpha + v \cos \alpha)$ , також врахуємо (а8).

$$\begin{aligned} \bar{a}_i^N(\alpha, t) &= \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} K(z) \left( (-v) \left( h_i - z s_N \frac{\partial h_i}{\partial x_1} - z s_N \frac{\partial h_i}{\partial x_2} \right) \Big|_{x(v)} + \right. \\ &\left. + \frac{1}{2} z^2 s_N^2 (-v) \left( \frac{\partial^2 h_i}{\partial x_1^2} + 2 \frac{\partial^2 h_i}{\partial x_1 \partial x_2} + \frac{\partial^2 h_i}{\partial x_2^2} \right) \Big|_{x(v) - z s_N(\theta, \eta)} \right) dv, \end{aligned}$$

де  $\theta, \eta \in (0, 1)$ .

Отже,

$$\begin{aligned} \bar{a}_i^N(\alpha, t) - a_i(\alpha, t) &= \frac{s_N^2}{2} \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} K(z) z^2 (-v) \times \\ &\times \left( \frac{\partial^2 h_i}{\partial x_1^2} + 2 \frac{\partial^2 h_i}{\partial x_1 \partial x_2} + \frac{\partial^2 h_i}{\partial x_2^2} \right) \Big|_{x(v) - z s_N(\theta, \eta)} dv. \end{aligned} \quad (6.74)$$

Оскільки ми припускаємо існування цих інтегралів, то нескладно переконатися, що права частина (6.74) являє собою вираз

$$\frac{s_N^2 D^2}{2} \int_{-\infty}^{\infty} (-v) \left( \frac{\partial^2 h_i}{\partial x_1^2} + 2 \frac{\partial^2 h_i}{\partial x_1 \partial x_2} + \frac{\partial^2 h_i}{\partial x_2^2} \right) \Big|_{x(v)} dv + o(s_N^2).$$

Справді, позначимо  $R(x_1, x_2) = \frac{\partial^2 h_i}{\partial x_1^2} + 2 \frac{\partial^2 h_i}{\partial x_1 \partial x_2} + \frac{\partial^2 h_i}{\partial x_2^2}$  і оберемо таку послідовність  $C_N$ : при  $N \rightarrow \infty$   $C_N \rightarrow \infty$ , але  $C_N^2 s_N \rightarrow 0$ . Тоді

$$\begin{aligned} & \int_{-\infty}^{\infty} K(z) z^2 dz \int_{-\infty}^{\infty} (-v) (R|_{x(v)-zs_N(\theta, \eta)} - R|_{x(v)}) dv = \\ & = \int_{-\infty}^{\infty} K(z) z^2 dz \int_{|v| < C_N} (-v) (R|_{x(v)-zs_N(\theta, \eta)} - R|_{x(v)}) dv + \\ & + \int_{-\infty}^{\infty} K(z) z^2 dz \left( \int_{|v| > C_N} (-v) R|_{x(v)-zs_N(\theta, \eta)} dv - \int_{|v| > C_N} (-v) R|_{x(v)} dv \right). \end{aligned} \quad (6.75)$$

Другий і третій інтеграли як хвости збіжних інтегралів прямують до 0 при  $N \rightarrow \infty$ , перший інтеграл не перевищує  $2C_N^2 s_N L \int_{-\infty}^{\infty} |z|^3 K(z) dz \rightarrow 0$ . (Тут  $L = 3 \max \left\{ \sup_{x_1, x_2} \left| \frac{\partial R}{\partial x_1} \right|, \sup_{x_1, x_2} \left| \frac{\partial R}{\partial x_2} \right| \right\}$  - стала величина внаслідок (а5)).

Далі, з доведеного вище випливає, що

$$\hat{r} - r = -p_1(\hat{a}_1 - a_1) + p_2(\hat{a}_2 - a_2) = \frac{D^2 s_N^2}{2} (-p_1 T_1 + p_2 T_2) + \frac{\zeta_2^N}{\sqrt{N s_N}} + o(s_N^2),$$

де

$$\begin{aligned} \zeta_2^N &= -p_1 \nu_N^1 + p_2 \nu_N^2 = \frac{1}{\sqrt{N s_N}} \sum_{j=1}^N (-p_1 a_{j:N}^1 + p_2 a_{j:N}^2) \times \\ & \times \left( K \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) (\xi_{j:N}^1 \sin \alpha - \xi_{j:N}^2 \cos \alpha) - \right. \\ & \left. - \mathbb{E} K \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) (\xi_{j:N}^1 \sin \alpha - \xi_{j:N}^2 \cos \alpha) \right) = \frac{1}{\sqrt{N s_N}} \sum_{j=1}^N Y_j^N. \end{aligned}$$

Скористаємося лемою, доведеною в роботі [13]:

**Лема 6.5.3** [13], с.31. Нехай  $P_N(x) = \lambda_1(\hat{u}_1^N(t) - u_1(t)) + \lambda_2(\hat{u}_2^N(t) - u_2(t))$ , де  $\hat{u}_1^N(t)$  - оцінки виду (6.51),  $\lambda_1, \lambda_2$  - сталі. За припущеннями, що існують  $d^2 < \infty$  та  $\sigma^2(t) < \infty$  з умов (а8) та (а6) відповідно, а також умов (а3), (а9), при  $s_N = c/N^{1/5}$

$$N^{2/5} P_N \Rightarrow \left( \frac{D^2 c^{2/5}}{2} \phi_2(t) + \frac{d}{c^{1/10}} \sigma(t) \zeta \right),$$

де  $\zeta$  є нормальною  $N(0, 1)$ ,  $\phi_2(t) = \lambda_1 u_1''(t) + \lambda_2 u_2''(t)$ .

Отже, за цією лемою  $N^{2/5}(\hat{S}_N - S)|_{(\alpha, t)} = \phi^N(\alpha, t) + \zeta_1^N$ , де

$$\phi^N(\alpha, t) \rightarrow \frac{D^2}{2} \left( -p_1 \frac{\partial^2 u_1}{\partial t^2} + p_2 \frac{\partial^2 u_2}{\partial t^2} \right),$$

$$\begin{aligned} \zeta_1^N &= \frac{1}{\sqrt{N s_N}} \sum_{j=1}^N (-p_1 a_{j:N}^1 + p_2 a_{j:N}^2) \left( K \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) - \right. \\ &\quad \left. - \mathbb{E} K \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) \right) = \frac{1}{\sqrt{N s_N}} \sum_{j=1}^N X_j^N. \end{aligned}$$

Для завершення доведення леми 6.5.2 достатньо довести асимптотичну нормальність випадкового вектора  $\begin{pmatrix} \zeta_1^N \\ \zeta_2^N \end{pmatrix}$ . Для цього скористаємося ЦГТ в схемі серій (теорема 7.3.8).

Отже, з леми 6.5.3 випливає, що  $\lim_{N \rightarrow \infty} \text{Var } \zeta_1^N = d^2 \sigma^2(\alpha, t)$ . Побачимо, що

$$\begin{aligned} \text{Var } \zeta_2^N &= \frac{1}{N s_N} \sum_{j=1}^N (-p_1 a_{j:N}^1 + p_2 a_{j:N}^2)^2 \left( \mathbb{E} K^2 \left( \frac{t - S_j(\alpha)}{s_N} \right) (S'_j(\alpha)_\alpha)^2 - \right. \\ &\quad \left. - \left( \mathbb{E} K \left( \frac{t - S_j(\alpha)}{s_N} \right) (S'_j(\alpha)_\alpha) \right)^2 \right) = A_N^1 + A_N^2. \end{aligned}$$

Розглянемо спершу  $A_N^2$ .

$$\begin{aligned} &\frac{1}{s_N} \left( \mathbb{E} K \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) (\xi_{j:N}^1 \sin \alpha - \xi_{j:N}^2 \cos \alpha) \right)^2 = \\ &= s_N \left( \sum_{k=1}^2 w_{j:N}^k \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} K(z) v h_k(t \cos \alpha - v \sin \alpha - z s_N \cos \alpha, \right. \\ &\quad \left. t \sin \alpha + v \cos \alpha - z s_N \sin \alpha) dv \right)^2. \end{aligned}$$

Як ми це вже показали, інтеграл під квадратом прямує до  $a_k(\alpha, t) < \infty$ , і оскільки  $w_{j:N}^k, a_{j:N}^k$  обмежені, то  $A_N^2 \rightarrow 0$ ,  $N \rightarrow \infty$ . Тепер розглянемо  $A_N^1$ .

$$\frac{1}{s_N} \mathbb{E} K^2 \left( \frac{t - \xi_{j:N}^1 \cos \alpha - \xi_{j:N}^2 \sin \alpha}{s_N} \right) (\xi_{j:N}^1 \sin \alpha - \xi_{j:N}^2 \cos \alpha)^2 =$$

$$= \sum_{k=1}^2 w_{j:N}^k \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} K^2(z) v^2 h_k(t \cos \alpha - v \sin \alpha - z s_N \cos \alpha, \\ t \sin \alpha + v \cos \alpha - z s_N \sin \alpha) dv.$$

Завдяки умовам (а5)-(а8) можна скористатися методом доведення формули (6.74) і нескладно отримати, що

$$\lim_{N \rightarrow \infty} A_N^1 = d^2 \lim_{N \rightarrow \infty} \sum_{r=1}^2 \langle (-p_1 a^1 + p_2 a^2)^2 w^r \rangle \int_{-\infty}^{\infty} v^2 \times \\ \times h_r(t \cos \alpha - v \sin \alpha, t \sin \alpha + v \cos \alpha) dv = d^2 \Delta^2(\alpha, t).$$

Отже,  $\lim_{N \rightarrow \infty} \text{Var } \zeta_2^N = d^2 \Delta^2(\alpha, t)$ .

Знайдемо граничне значення  $\text{Cov}(\zeta_1^N, \zeta_2^N)$ .

$$\text{Cov}(\zeta_1^N, \zeta_2^N) = \text{Cov} \left( \frac{1}{\sqrt{N s_N}} \sum_{j=1}^N X_j^N, \frac{1}{\sqrt{N s_N}} \sum_{j=1}^N Y_j^N \right) = \\ \frac{1}{N s_N} \sum_{j=1}^N \text{Cov}(X_j^N, Y_j^N) = \frac{1}{N s_N} \sum_{j=1}^N (-p_1 a_{j:N}^1 + p_2 a_{j:N}^2)^2 \times \\ \times \left( \mathbb{E} K^2 \left( \frac{t - S_j(\alpha)}{s_N} \right) (S'_j(\alpha)_\alpha) - \right. \\ \left. - \mathbb{E} K \left( \frac{t - S_j(\alpha)}{s_N} \right) (S'_j(\alpha)_\alpha) \mathbb{E} K \left( \frac{t - S_j(\alpha)}{s_N} \right) \right) = C_1^N + C_2^N.$$

При цьому

$$C_2^N = \frac{s_N}{N} \sum_{j=1}^N (-p_1 a_{j:N}^1 + p_2 a_{j:N}^2)^2 \left( \sum_{k=1}^2 w_{j:N}^k \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} K(z) (-v) \right. \\ \left. h_k(t \cos \alpha - v \sin \alpha - z s_N \cos \alpha, t \sin \alpha + v \cos \alpha - z s_N \sin \alpha) dv \right) \\ \times \left( \sum_{k=1}^2 w_{j:N}^k \int_{-\infty}^{\infty} dz \times \int_{-\infty}^{\infty} K(z) (-v) \right. \\ \left. \times h_k(t \cos \alpha - v \sin \alpha - z s_N \cos \alpha, t \sin \alpha + v \cos \alpha - z s_N \sin \alpha) dv \right)$$



(Через те, що перший повторний інтеграл під знаком суми збіжний до  $a_k(\alpha, t)$ , другий до  $u_k(\alpha, t)$ , отже, множник при  $s_N$  — обмежений по  $N$  вираз.) Далі, аналогічно формулі (6.74) доводиться, що

$$C_1^N = \frac{1}{N} \sum_{j=1}^N (-p_1 a_{j:N}^1 + p_2 a_{j:N}^2)^2 \sum_{k=1}^2 w_{j:N}^k \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} K^2(z)(-v)$$

$$h_k(t \cos \alpha - v \sin \alpha - z s_N \cos \alpha, t \sin \alpha + v \cos \alpha - z s_N \sin \alpha) dv \rightarrow_{N \rightarrow \infty} d^2 \omega^2(\alpha, t).$$

Тепер залишається перевірити виконання умови Ліндеберга. Використовуючи нерівність  $(x - y)^2 \leq 2(x^2 + y^2)$  і позначаючи  $l_j^N = (-p_1 a_{j:N}^1 + p_2 a_{j:N}^2)^2$  бачимо, що

$$\begin{aligned} & \frac{1}{N s_N} \sum_{j=1}^N \mathbb{E} ((X_j^N)^2 + (Y_j^N)^2) \mathbb{I} \{ (X_j^N)^2 + (Y_j^N)^2 > \tau^2 \} = \\ & = \frac{1}{N s_N} \sum_{j=1}^N l_j^N \mathbb{E} \left( \left( K \left( \frac{t - S_j(\alpha)}{s_N} \right) - \mathbb{E} K \left( \frac{t - S_j(\alpha)}{s_N} \right) \right)^2 + \right. \\ & + \left. \left( K \left( \frac{t - S_j(\alpha)}{s_N} \right) (S_j(\alpha)'_{\alpha}) - \mathbb{E} K \left( \frac{t - S_j(\alpha)}{s_N} \right) (S_j(\alpha)'_{\alpha}) \right)^2 \right) \times \\ & \times \mathbb{I} \{ (X_j^N)^2 + (Y_j^N)^2 > \tau^2 \} \leq \frac{2}{N s_N} \sum_{j=1}^N l_j^N \mathbb{E} \left( K^2 \left( \frac{t - S_j(\alpha)}{s_N} \right) + \right. \\ & + K^2 \left( \frac{t - S_j(\alpha)}{s_N} \right) (S_j(\alpha)'_{\alpha})^2 + \left. \left( \mathbb{E} K \left( \frac{t - S_j(\alpha)}{s_N} \right) \right)^2 + \right. \\ & + \left. \left( \mathbb{E} K \left( \frac{t - S_j(\alpha)}{s_N} \right) (S_j(\alpha)'_{\alpha}) \right)^2 \right) \mathbb{I} \{ (X_j^N)^2 + (Y_j^N)^2 > \tau^2 \} \end{aligned} \quad (6.76)$$

Як було встановлено вище,

$$\left( \mathbb{E} K \left( \frac{t - S_j(\alpha)}{s_N} \right) \right)^2 = O(s_N^2); \quad \left( \mathbb{E} K \left( \frac{t - S_j(\alpha)}{s_N} \right) (S_j(\alpha)'_{\alpha}) \right)^2 = O(s_N^2).$$

Отже, двома відповідними доданками в правій частині (6.76) можна знехтувати - відповідні суми прямують до нуля при  $N \rightarrow \infty$ . Нам залишається розглянути

$$\mathbb{E} K^2 \left( \frac{t - S_j(\alpha)}{s_N} \right) \left( 1 + (S_j(\alpha)'_{\alpha})^2 \right) \mathbb{I} \left\{ \left( K \left( \frac{t - S_j(\alpha)}{s_N} \right) - \right. \right.$$

$$\begin{aligned}
& -\mathbb{E} K\left(\frac{t-S_j(\alpha)}{s_N}\right)\Big)^2 + \left(K\left(\frac{t-S_j(\alpha)}{s_N}\right)(S_j(\alpha)'_\alpha) - \right. \\
& \quad \left. -\mathbb{E} K\left(\frac{t-S_j(\alpha)}{s_N}\right)(S_j(\alpha)'_\alpha)\right)^2 > \tau^2 N s_N \Big\} \\
& \leq \mathbb{E} K^2\left(\frac{t-S_j(\alpha)}{s_N}\right) \left(1 + (S_j(\alpha)'_\alpha)^2\right) \times \\
& \quad \times \mathbb{I}\left\{K^2\left(\frac{t-S_j(\alpha)}{s_N}\right) \left(1 + (S_j(\alpha)'_\alpha)^2\right) > \frac{\tau^2 N s_N}{2} - \right. \\
& \quad \left. - \left(\mathbb{E} K\left(\frac{t-S_j(\alpha)}{s_N}\right)\right)^2 - \left(\mathbb{E} K\left(\frac{t-S_j(\alpha)}{s_N}\right)(S_j(\alpha)'_\alpha)\right)^2\right\} = \\
& = \sum_{k=1}^2 w_{j:N}^k \int_{B_N} K^2(z)(1+v^2) \times \\
& \quad \times h_k(t \cos \alpha - v \sin \alpha - z s_N \cos \alpha, t \sin \alpha + v \cos \alpha - z s_N \sin \alpha) dv,
\end{aligned}$$

де

$$\begin{aligned}
B_N = & \left\{ (z, v) : K^2(z)(1+v^2) > \frac{\tau^2 N s_N}{2} - \right. \\
& \left. - \left(\mathbb{E} K\left(\frac{t-S_j(\alpha)}{s_N}\right)\right)^2 - \left(\mathbb{E} K\left(\frac{t-S_j(\alpha)}{s_N}\right)(S_j(\alpha)'_\alpha)\right)^2 \right\}.
\end{aligned}$$

Завдяки виконанню умови (а7) цей вираз не перевищує

$$\sum_{k=1}^2 w_{j:N}^k \int_{-\infty}^{\infty} K^2(z) dz \sup_{p,q} \int_{\bar{B}_N} (1+v^2) h_k(p - v \sin \alpha, q + v \cos \alpha) dv, \quad (6.77)$$

де

$$\bar{B}_N = \left\{ (z, v) : (1+v^2) > \frac{\tau^2 N s_N}{2 \sup_z K^2(z)} + O(s_N^2) \right\}.$$

Другий інтеграл в правій частині (6.77) прямує до нуля при  $N \rightarrow \infty$  як залишок збіжного інтегралу. Отже, права частина (6.76) теж прямує до нуля, і умова Ліндеберга виконується.

Лема 6.5.2 доведена. Отже, підставляючи формулу (6.73) і результати леми 6.5.2 у праві частини (6.65) і підсумовуючи всі результати, завершуємо доведення теореми 6.5.1.

## Розділ 7

### Допоміжні відомості

У цьому розділі для зручності читача зібрані допоміжні відомості, на які є посилання у основному тексті книги. В основному, вони є добре відомими фактами з математичного аналізу, теорії ймовірностей та математичної статистики. Новим нетривіальним результатом можна вважати теорему 7.2.1, яка встановлює нерівність для визначників двох матриць.

У параграфі 7.4 описано класичний підхід до трактовки слабкої збіжності випадкових функцій. Інший підхід, що використовує техніку єдиного ймовірнісного простору див. у п. (iii) параграфу 2.2.

## 7.1 Формули інтегрування і пов'язані з ними нерівності

Нехай вектори  $u_i = (u_i^1, \dots, u_i^d) \in \mathbb{R}^d$   $i = 0, 1, 2$  задовольняють нерівність  $u_0 \leq u_1 \leq u_2$ . Позначимо  $K = [u_0, u_2] = \{x \in \mathbb{R}^d : u_0 \leq x \leq u_2\}$ . Для будь-якого  $\alpha \in \{0, 1\}^d$  і будь-якої гладенької функції  $f : K \rightarrow \mathbb{R}$  позначимо  $D^\alpha f(u)$  частинну похідну функції  $f$  по всіх координатах  $u^j$  для яких  $\alpha_j = 1$ , тобто

$$D^\alpha = \prod_{\alpha_j=1} \frac{\partial}{\partial u^j},$$

$K_\alpha = \bigotimes_{\alpha_j=1} [u_0^j, u_2^j]$  паралелепіпед у  $\mathbb{R}^d$  з “найменшою” вершиною  $u_0$  і “найбільшою” —  $u_2$ ,  $u_\alpha = (u_{\alpha_1}^1, \dots, u_{\alpha_d}^d)$ ,  $(du_1)^\alpha = \prod_{\alpha_j=1} du_1^j$  — диференціал по тих змінних, для яких  $\alpha_j = 1$ ,  $V(K_\alpha)$  — об'єм паралелепіпеда  $K_\alpha$ , тобто  $V(K_\alpha) = \prod_{\alpha_j=1} (u_2^j - u_0^j)$

**Лема 7.1.1** *Якщо  $f : K \rightarrow \mathbb{R}$  є  $d$  разів диференційовною функцією то*

$$\sup_{u \in K} |f(u)| \leq \|f\|_H \left( \sum_{\alpha \in \{0,1\}^d} V(K_\alpha) \right)^{1/2}$$

де

$$\|f\|_H = \left( f^2(u_0) + \sum_{\alpha \in \{0,1\}^d, \alpha \neq 0} \int_{K_\alpha} (D^\alpha f(u_\alpha))^2 (du_1)^\alpha \right)^{1/2}.$$

**Доведення.** Спочатку отримаємо наступну рівність, яка є узагальненням формули Ньютона-Лейбниці на багатовимірний випадок:

$$f(u_2) = \sum_{\alpha \in \{0,1\}^d} \int_{K_\alpha} D^\alpha f(u_1) (du_1)^\alpha. \quad (7.1)$$

Тут у випадку, коли  $K_\alpha = \emptyset$ , вважаємо інтеграл рівним підінтегральному виразу.

Наприклад, у випадку  $d = 2$  маємо наступну формулу (для спрощення тут прийняті позначення  $t_1 = u_1^1$ ,  $t_2 = u_1^2$ ):

$$\begin{aligned} f(u_2^1, u_2^2) = & \int_{u_0^2}^{u_2^2} \int_{u_0^1}^{u_2^1} \frac{\partial^2}{\partial t_1 \partial t_2} f(t_1, t_2) dt_1 dt_2 \\ & + \int_{u_0^2}^{u_2^2} \frac{\partial}{\partial t_2} f(u_N, t_2) dt_2 + \int_{u_0^1}^{u_2^1} \frac{\partial}{\partial t_1} f(t_1, u_N) dt_1 + f(u_0^1, u_0^2). \end{aligned} \quad (7.2)$$

Доведемо (7.1). Для  $\gamma \in \{0, 1, 2\}^d$  позначимо  $\bar{\gamma} = (\bar{\gamma}^1, \dots, \bar{\gamma}^d)$ , де  $\bar{\gamma}^j = 1$ , якщо  $\gamma = 1$  і  $\bar{\gamma}^j = 0$  в усіх інших випадках. Позначимо

$$p_j^k \gamma = (\gamma^1, \dots, \gamma^{k-1}, j, \gamma^{k+1}, \dots, \gamma^d),$$

$$J(\gamma) = \int_{S_{\bar{\gamma}}} D^{\bar{\gamma}} f(u_{\gamma})(du_1)^{\bar{\gamma}}.$$

Якщо  $\gamma^k = 2$ , то

$$J(\gamma) = J(p_0^k \gamma) + J(p_1^k \gamma). \quad (7.3)$$

Дійсно,

$$f(u_{\gamma}) = f(u_{p_0^k \gamma}) + \int_{u_0^j}^{u_1^j} f(u_{p_1^k \gamma}) du_1^k \frac{\partial}{\partial u_1^k}$$

(це звичайна формула Ньютона-Лейбниця, застосована до  $k$ -того аргумента функції  $f$ ). Диференціюючи її за іншими змінними, а потім інтегруючи, отримуємо (7.3).

Тепер, починаючи з  $f(u_2^1, \dots, u_2^d) = J(2, \dots, 2)$ , застосуємо (7.3) послідовно при  $k$  рівних  $d, d-1, \dots, 1$ . Отримуємо:

$$\begin{aligned} f(u_2) &= J(2, \dots, 2, 2, 2) \\ &= J(2, \dots, 2, 2, 0) + J(2, \dots, 2, 2, 1) \\ &= J(2, \dots, 2, 0, 0) + J(2, \dots, 2, 1, 0) + J(2, \dots, 2, 0, 1) + J(2, \dots, 2, 1, 1) = \dots \end{aligned}$$

На кроці  $d$  отримаємо (7.1).

Тепер, розширюючи у (7.1) там де потрібно, межі інтегрування і використовуючи нерівність Коші-Буняківського, оцінимо

$$\sup_{u \in K} |f(u)| \leq \sum_{\alpha \in \{0,1\}^d} \int_{K_{\alpha}} |D^{\alpha} f(u_{\alpha})|$$

$$\leq \sum_{\alpha \in \{0,1\}^d} \left( \int_{K_\alpha} (D^\alpha f(u_\alpha))^2 \right)^{1/2} \left( \sum_{\alpha \in \{0,1\}^d} V(K_\alpha) \right)^{1/2}.$$

Лема доведена.

## 7.2 Нерівність для визначників

У цьому параграфі ми доведемо теорему про нерівність для визначників двох матриць. Одна з цих матриць — матриця Грама скінченного набору функцій у просторі  $L_2(\nu)$  відносно деякої міри  $\nu$ . Другу також можна трактувати як матрицю Грама того ж набору функцій, однак міра, по якій підраховується скалярний добуток є матричнозначною. Тому і матриця будується з блоків, кожен з яких відповідає одному елементу першої матриці. Виявляється, що з невиродженості першої матриці випливає невиродженість другої. Інтуїтивно це легко зрозуміти, оскільки невиродженість матриці Грама рівносильна лінійній незалежності набору функцій, а ця властивість практично не залежить від вибору міри для простору  $L_2(\nu)$ , в якому ці функції розглядаються. Хіба що, поклавши  $\nu$  рівною 0 на множині, на якій розрізняються функції, можна зробити їх лінійно залежними у  $L_2(\nu)$ . Але такі вироджені випадки відсікаються умовою теореми.

Введемо позначення.

Нехай  $T \subseteq \mathbb{R}$ ,  $w_1, \dots, w_M$  — вимірні дійснозначні функції на  $T$ . Розглянемо набір зарядів (знакозмінних мір) на  $T$ , —  $\mu_{l_1, l_2}$ ,  $l_1, l_2 = 1, \dots, r$ , де  $r$  — фіксоване ціле число. Символом  $\mu(A)$ ,  $A \subseteq T$ , будемо позначати матрицю  $(\mu_{l_1, l_2}(A))_{l_1, l_2=1}^r$ . Інакше кажучи,  $\mu$  можна розглядати як матричнозначну міру, визначену на  $T$ .

Позначимо

$$\langle w_{m_1}, w_{m_2} \rangle_{l_1, l_2} = \int w_{m_1}(t) w_{m_2}(t) \mu_{l_1, l_2}(dt)$$

— “скалярний добуток” функцій  $w_{m_1}$  та  $w_{m_2}$  відносно заряду  $\mu_{l_1, l_2}$ . Покладемо

$$\langle w_{m_1}, w_{m_2} \rangle_\mu = (\langle w_{m_1}, w_{m_2} \rangle_{l_1, l_2})_{l_1, l_2=1}^r.$$

Матрицю  $\langle w_{m_1}, w_{m_2} \rangle_\mu$  можна інтерпретувати як “матричнозначний скалярний добуток” відносно матричнозначної міри  $\mu$ . (Надалі ми припускаємо, що всі розглядувані функції та міри такі, що введені нами скалярні добутки існують і є скінченними).

Нехай  $\nu$  — деяка (додатнозначна) міра на  $T$  і

$$\langle w_{m_1}, w_{m_2} \rangle_\nu = \int w_{m_1}(t) w_{m_2}(t) \nu(dt).$$

Позначимо через  $\Gamma_\nu = (\langle w_{m_1}, w_{m_2} \rangle_\nu)_{m_1, m_2=1}^M$  матрицю Грама для системи функцій  $w_1, \dots, w_M$  у просторі  $L_2(\nu)$ , а через

$$\Gamma_\mu = \begin{pmatrix} \langle w_1, w_1 \rangle_\mu & \langle w_1, w_2 \rangle_\mu & \cdots & \langle w_1, w_M \rangle_\mu \\ \langle w_2, w_1 \rangle_\mu & \langle w_2, w_2 \rangle_\mu & \cdots & \langle w_2, w_M \rangle_\mu \\ \vdots & \vdots & \ddots & \vdots \\ \langle w_M, w_1 \rangle_\mu & \langle w_M, w_2 \rangle_\mu & \cdots & \langle w_M, w_M \rangle_\mu \end{pmatrix}$$

— блочна матриця Грама для  $w_1, \dots, w_M$  у матричнозначному скалярному добутку  $\langle \cdot, \cdot \rangle_\mu$ . Матрицю  $\Gamma_\mu$  будемо розглядати як звичайну квадратну матрицю вимірності  $(Mr) \times (Mr)$  з дійснозначними елементами. Відповідно, визначник цієї матриці  $\det \Gamma_\mu$  є дійсним числом.

**Теорема 7.2.1** *Нехай  $\det \Gamma_\nu > 0$  і існує число  $Q > 0$  таке, що*

$$\lambda_{\min}(\mu(A)) > Q\nu(A)$$

*для всіх вимірних множин  $A$ .*

$$\text{Тоді } \det \Gamma_\mu \geq Q^{Mr} \det \Gamma_\nu.$$

(Нагадаємо, що  $\lambda_{\min}(Z)$  — найменше власне число матриці  $Z$ ).

**Доведення теореми.** Ми будемо називати елементами формальні суми  $U = \sum_{m=1}^M A_m w_m$ , де  $A_m$  — довільні дійснозначні матриці розміру  $r \times r$ ,  $w_m$  — елементи заданого в умовах теореми набору функцій. Для елементів очевидним способом визначені операції додавання та множення на дійснозначні  $r \times r$  матриці.

Нехай  $U$  та  $V = \sum_{m=1}^M B_m w_m$  — деякі елементи. Покладемо

$$\langle U, V \rangle = \sum_{m_1, m_2=1}^M A_{m_1} \langle w_{m_1}, w_{m_2} \rangle_\mu B_{m_2}$$

і для довільних елементів  $U_1, \dots, U_M$ ,

$$\Gamma_\mu(U_1, \dots, U_M) = (\langle U_{m_1}, U_{m_2} \rangle)_{m_1, m_2=1}^M.$$

Якщо  $U = Ew_m$ , де  $E$  — одинична матриця, позначатимемо  $U = w_m$ . У цих позначеннях  $\Gamma_\mu = \Gamma_\mu(w_1, \dots, w_m)$ .

Для доведення теореми ми скористаємось процедурою ортогоналізації елементів  $w_1, \dots, w_M$  відносно матричнозначного скалярного добутку  $\langle \cdot, \cdot \rangle$ . Ця процедура подібна до класичної процедури ортогоналізації Грама-Шмітта для звичайних скалярних добутків. При цьому нам будуть потрібні три допоміжні леми.

**Лема 7.2.1** *Нехай для деякої дійснозначної матриці  $A$ ,  $U'_{m_1} = AU_{m_1}$  і, при  $m \neq m_1$ ,  $U'_m = U_m$ . Тоді*

$$\Gamma_\mu(U'_1, \dots, U'_M) = S^T \Gamma_\mu(U_1, \dots, U_M) S, \quad (7.4)$$

де  $S$  є блочною матрицею складеною з блоків  $S_{m_1, m_2}$  розміру  $r \times r$ , причому  $S_{m_1, m_1} = A$ ,  $S_{m, m} = E$  для  $m \neq m_1$  і  $S_{m, m_2} = 0$  для всіх пар  $m, m_2$ , таких, що  $m \neq m_2$ .

**Лема 7.2.2** *Нехай для деякої дійснозначної матриці  $A$  і деяких  $m_1 \neq m_2$ ,  $U'_{m_1} = U_{m_1} + AU_{m_2}$ , при  $m \neq m_1$ ,  $U'_m = U_m$ . Тоді (7.4) виконано для  $S$  що є блочною матрицею, складеною з блоків  $S_{ij}$  розміру  $r \times r$ , причому  $S_{m_1, m_2} = A$ ,  $S_{i, i} = E$  для всіх  $i = 1, \dots, M$  і  $S_{ij} = 0$  для всіх пар  $i, j$ , таких, що  $i \neq j$ ,  $(i, j) \neq (m_1, m_2)$ .*

Ці дві леми доводяться безпосереднім підрахунком відповідних матриць.

**Лема 7.2.3** *Нехай в умовах теореми для деякого  $m_1$  і деякого набору індексів  $J$ , такого, що  $m_1 \notin J$ ,*

$$U = w_{m_1} + \sum_{m \in J} B_m w_m,$$

де  $B_m$  — дійснозначні  $r \times r$  матриці. Тоді

$$\langle U, U \rangle \geq Q \|z\|_\nu^2 E, \quad (7.5)$$

де  $z$  є ортогональним доповненням  $w_{m_1}$  до лінійного простору, натягнутого на вектори  $\{w_m, m \in J\}$  у  $L_2(\nu)$ ,  $\|z\|_\nu$  — норма  $z$  у  $L_2(\nu)$ .

(Інакше кажучи,  $\|z\|_\nu$  — довжина перпендикуляра, опущеного з кінця вектора  $w_{m_1}$  на простір лінійних комбінацій векторів  $\{w_m, m \in J\}$ , причому



всі вектори розглядаються у  $L_2(\nu)$ . Нерівність 7.5 слід трактувати у матричному розумінні — різниця між лівою і правою частинами є додатньо-визначеною матрицею.)

**Доведення.** Нехай  $c = (c_1, \dots, c_r)^T$  — довільний  $r$ -вимірний вектор стовпчик одиничної довжини (у звичайній евклідовій нормі  $\|c\| = 1$ ). Покладемо  $B_m c = b_m = (b_1^m, \dots, b_r^m)^T$ . Тоді

$$\begin{aligned}
 c^T \langle U, U \rangle c &= c^T \left\langle w_{m_1} + \sum_{m \in J} B_m w_m, w_{m_1} + \sum_{m \in J} B_m w_m \right\rangle c \\
 &= c^T \langle w_{m_1}, w_{m_1} \rangle_\mu c + \sum_{m \in J} b_m^T \langle w_m, w_{m_1} \rangle_\mu c \\
 &\quad + \sum_{m \in J} c^T \langle w_m, w_{m_1} \rangle_\mu b_m + \sum_{m_2, m_3 \in J} b_{m_2}^T \langle w_{m_2}, w_{m_1} \rangle_\mu b_{m_3} \\
 &= \int \left( c^T w_{m_1}(t) + \sum_{m \in J} b_m^T w_m(t) \right) \mu(dt) \left( c^T w_{m_1}(t) + \sum_{m \in J} b_m^T w_m(t) \right)^T \\
 &\geq Q \int \left\| c^T w_{m_1}(t) + \sum_{m \in J} b_m^T w_m(t) \right\|_{\mathbb{R}^r}^2 \nu(dt) \\
 &= Q \sum_{i=1}^r \int \left( c_i w_{m_1}(t) + \sum_{m \in J} b_i^m w_m(t) \right)^2 \nu(dt) \\
 &\geq Q \sum_{i: c_i \neq 0} (c_i)^2 \int \left( w_{m_1}(t) + \sum_{m \in J} \frac{b_i^m}{c_i} w_m(t) \right)^2 \nu(dt) \\
 &\geq Q \sum_{i=1}^r (c_i)^2 \|z\|_\nu^2 = Q \|z\|_\nu^2
 \end{aligned}$$

Щоб отримати останню нерівність, ми скористались тим фактом, що довжина перпендикуляра — найкоротша відстань від кінця вектора до лінійного простору.

Лема доведена.

**Закінчення доведення теореми.** Побудуємо набір елементів  $U_1, \dots, U_M$ , таких, що

1.  $U_m = A_m(w_m - \sum_{m_1=1}^{m-1} B_{m_1, m} w_{m_1})$ , де  $A_m$  та  $B_{m_1, m_2}$  деякі дійснозначні матриці.

2.  $\langle U_{m_1}, U_{m_2} \rangle = E$  якщо  $m_1 = m_2$  і  $\langle U_{m_1}, U_{m_2} \rangle = 0$  для  $m_1 \neq m_2$ .

Інакше кажучи,  $U_1, \dots, U_M$  це результат ортогоналізації набору  $w_1, \dots, w_M$ . Ортогоналізацію проведемо узагальненим алгоритмом Грама-Шміда, послідовно визначаючи  $U_1, U_2, \dots, U_M$ .

Спочатку покладемо  $A_1 = (\langle w_1, w_1 \rangle_\mu)^{-1/2}$ ,  $U_1 = A_1 w_1$ . За лемою 7.2.3

$$\langle w_1, w_1 \rangle_\mu \geq Q \|w_1\|_\nu^2 E,$$

тому

$$A_1 \leq Q^{-1/2} \|w_1\|_\nu^{-1} E.$$

Нехай  $U_1, \dots, U_m$ , що задовольняють умови 1-2, вже побудовані. Покладемо

$$\tilde{U}_{m+1} = w_{m+1} - \sum_{m_1=1}^m \langle w_{m+1}, U_{m_1} \rangle U_{m_1}.$$

Тоді для всіх  $m_1 < m$  отримуємо

$$\langle \tilde{U}_{m+1}, U_{m_1} \rangle = \langle w_{m+1}, U_{m_1} \rangle - \langle w_{m+1}, U_{m_1} \rangle = 0$$

внаслідок “ортонормованості”  $U_1, \dots, U_m$ . Покладемо  $U_{m+1} = A_{m+1} \tilde{U}_{m+1}$ , де  $A_{m+1} = (\langle \tilde{U}_{m+1}, \tilde{U}_{m+1} \rangle)^{-1/2}$ . Знову за лемою 7.2.3 маємо

$$A_{m+1} \leq Q^{-1/2} \|z_{m+1}\|_\nu^{-1} E,$$

де  $\|z_{m+1}\|$  — довжина перпендикуляра, опущеного з кінця вектора  $w_{m+1}$  на лінійний простір, натягнутий на вектори  $w_1, \dots, w_m$  у  $L_2(\nu)$ .

Таким чином, набір  $U_1, \dots, U_M$  побудовано. Згідно з лемами 7.2.1 та 7.2.2,

$$\Gamma_\mu(U_1, \dots, U_M) = S \Gamma_\mu(w_1, \dots, w_M) S^T$$

де

$$S = \begin{pmatrix} A_1 & B_{21} & B_{31} & \cdots & B_{M1} \\ 0 & A_2 & B_{32} & \cdots & B_{M2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_M \end{pmatrix}$$

Зрозуміло, що  $\Gamma_\mu(U_1, \dots, U_M)$  є одиничною матрицею, отже

$$1 = \det \Gamma_\mu(U_1, \dots, U_M) = \det \Gamma_\mu(\det S)^2,$$

звідки отримуємо

$$\det \Gamma_\mu = (\det S)^{-2} = \prod_{m=1}^M (\det A_m)^2 \geq Q^{Mr} \prod_{m=2}^M \|z_m\|_\nu^2 \langle w_1, w_1 \rangle_\nu = Q^{Mr} \det \Gamma_\nu.$$

Останню рівність легко отримати використовуючи звичайну процедуру ортогоналізації до набору функцій  $w_1, \dots, w_M$  у  $L_2(\nu)$ .

Теорема доведена.

## 7.3 Ймовірнісні нерівності і граничні теореми

У цьому параграфі зібрані теореми про поведінку сум незалежних випадкових величин і векторів. Вони розподіляються на три групи: нерівності для ймовірностей відхилення від математичного сподівання, теореми про збіжність середніх до математичних сподівань (закони великих чисел) та теореми про слабку збіжність нормованих відхилень від середнього (центральна гранична теорема, ЦГТ). Крім того, тут розглянуті теореми про слабку збіжність, які дозволяють застосовувати ЦГТ для асимптотичного аналізу статистичних оцінок — такі як теорема Слуцького та теореми неперервності.

**Нерівності для сум випадкових величин.** У цьому пункті  $\xi_1, \dots, \xi_n, \dots$ , позначають незалежні випадкові величини,  $S_n = \sum_{i=1}^n \xi_i$ .

**Теорема 7.3.1 Нерівність Прохорова.** Нехай для деякого  $c > 0$ ,  $|\xi_i| < c$ ,  $E \xi_i = 0$ . Покладемо  $B_n = \sum_{i=1}^n \text{Var } \xi_i$ . Тоді для всіх  $x \in \mathbb{R}$

$$P\{S_n \geq x\} \leq \exp \left\{ -\frac{x}{c} \operatorname{arcsch} \frac{cx}{2B_n} \right\}.$$

Тут  $\operatorname{arcsch}(x) = \log(x + \sqrt{x^2 + 1})$  — гіперболічний арксинус (ареасинус). Доведення див. у [33] (пор. також [28], с.93).

**Теорема 7.3.2 Нерівність Хьофдінга.** Нехай для деяких  $a_i, b_i \in \mathbb{R}$ ,  $a_i < \xi_i < b_i$ . Тоді для всіх  $x > 0$ ,

$$P\{S_n - E S_n \geq nx\} \leq \exp \left\{ -\frac{2n^2 x^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}.$$

Доведення див. [50] (пор. також [28], с.93).

Трохи інший варіант цієї нерівності можна отримати використовуючи теорію субгауссових випадкових величин (див. леми 2.3.2-2.3.4 у [26], пор. [5]).

**Теорема 7.3.3** *Нехай для деякого  $c > 0$ ,  $|\xi_i| < c$ ,  $E\xi_i = 0$ . Тоді для всіх  $x > 0$ ,*

$$P\left\{\frac{1}{n}\sum_{i=1}^n \xi_i \geq x\right\} \leq \exp\left\{-\frac{x^2 n}{2c^2}\right\}.$$

**Види збіжності випадкових величин.** Нехай  $\xi, \xi_n, n = 1, 2, \dots$  — випадкові вектори у  $\mathbb{R}^d$ ,  $|\cdot|$  — евклідова норма у  $\mathbb{R}^d$ .

Кажуть, що  $\xi_n \rightarrow \xi$  при  $n \rightarrow \infty$  майже напевно (скорочення м.н.), якщо

$$P\{\lim_{n \rightarrow \infty} \xi_n = \xi\} = 1.$$

Кажуть, що  $\xi_n \rightarrow \xi$  при  $n \rightarrow \infty$  за ймовірністю, якщо для всіх  $\varepsilon > 0$ ,

$$P\{|\xi_n - \xi| > \varepsilon\} \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Кажуть, що  $\xi_n$  збігається слабо до  $\xi$  (або розподіл  $\xi_n$  збігається слабо до  $\xi$ , запис:  $\xi_n \Rightarrow \xi$ ) якщо для всіх неперервних, обмежених функцій  $g: \mathbb{R}^d \rightarrow \mathbb{R}$

$$Eg(\xi_n) \rightarrow Eg(\xi) \text{ при } n \rightarrow \infty.$$

Зі збіжності м.н. випливає збіжність за ймовірністю, а зі збіжності за ймовірністю — слабка збіжність.

Для будь-якої множини  $A \subseteq \mathbb{R}^d$  позначимо  $\partial A$  — границю  $A$ . Кажуть, що вимірна множина  $A$  є множиною неперервності розподілу випадкового вектора  $\xi$ , якщо  $P\{\xi \in A\} = 0$ .

**Теорема 7.3.4** *Наступні твердження еквівалентні:*

1.  $\xi_n \Rightarrow \xi$  при  $n \rightarrow \infty$ ;
2.  $P\{\xi_n \in A\} \rightarrow P\{\xi \in A\}$  при  $n \rightarrow \infty$  для всіх  $A$ , що є множинами неперервності  $\xi$ ;
3. Функції розподілу  $F_n$  випадкових векторів  $\xi_n$  збігаються до функції розподілу  $F$  вектора  $\xi$  в усіх точках неперервності  $F$ .

Доведення див. у [8], теорема 4 п.4 розділ 6.

**Закони великих чисел.** Найпростішу форму має підсилений закон великих чисел для незалежних, однаково розподілених випадкових величин.

**Теорема 7.3.5** Нехай  $\xi_i$ ,  $i = 1, 2, \dots$  — послідовність незалежних однаково розподілених векторів у  $\mathbb{R}^d$  зі скінченним математичним сподіванням  $E\xi_i = a$ . Тоді

$$\frac{1}{n} \sum_{i=1}^n \xi_i \rightarrow a \text{ м.н. при } n \rightarrow \infty.$$

Доведення див. у [8], с. 148.

Для послідовностей неоднаково розподілених випадкових величин справедливий наступний закон великих чисел

**Теорема 7.3.6** Якщо послідовність незалежних в.в.  $\{X_N, N \geq 1\}$  така, що  $\text{Var } X_N$  існує і  $\frac{\text{Var } X_N}{N} \rightarrow 0, N \rightarrow \infty$ , то

$$\frac{1}{N} \sum_{k=1}^N X_k - \frac{1}{N} \sum_{k=1}^N E X_k \rightarrow 0; \quad N \rightarrow \infty$$

за ймовірністю.

(див. [15], с.47).

Розглянемо тепер схему серій випадкових величин  $\zeta_{i,n}$ ,  $i = 1, \dots, n$ ,  $n = 1, 2, \dots$ . Будемо вважати  $\zeta_{i,n}$  незалежними всередині кожної серії (тобто при фіксованому  $n$ ). Позначимо  $S_n = \sum_{i=1}^n \zeta_{i,n}$ .

Умови виконання закону великих чисел у схемі серій дає

**Теорема 7.3.7** Нехай  $E \zeta_{i,n} = 0$ ,

$$M_1 = \sum_{i=1}^N E |\zeta_{i,n}| \leq c < \infty,$$

і для всіх  $\tau > 0$ , при  $n \rightarrow \infty$ ,

$$M_2(\tau) = \sum_{i=1}^n E |\zeta_{i,n}| \mathbb{I}\{|\zeta_{i,n}| > \tau\} \rightarrow 0.$$

Тоді  $S_n \rightarrow 0$  за ймовірністю при  $n \rightarrow \infty$ .

(див. [2], п.3 розділу 8).

**Центральна гранична теорема.** Нехай  $\xi_{i,n}$   $i = 1, \dots, n$ ,  $n = 1, 2, \dots$  — схема серій випадкових векторів<sup>1</sup>, незалежних всередині кожної серії,  $\zeta_n = \sum_{k=1}^n \xi_{k,n}$ . Позначимо  $|\xi_{i,n}| = \sqrt{\xi_{i,n}^T \xi_{i,n}}$  евклідову норму  $\xi_{i,n}$ .

<sup>1</sup>Всі вектори, як правило, вважаємо векторами-стовпчиками.

Припустимо, що  $E \xi_{i,n} = 0$ ,  $E |\xi_{i,n}|^2 < \infty$ . Позначимо  $s_{k,n} = E \xi_{k,n} \xi_{k,n}^T$  коваріаційну матрицю  $\xi_{k,n}$ ,  $S_n = \sum_{k=1}^n s_{k,n}$  коваріаційну матрицю  $\zeta_n$ .

**Теорема 7.3.8** (див. [2], с.201) *Нехай крім вказаних вище обмежень виконуються умови*

1. При  $n \rightarrow \infty$ ,

$$S_n \rightarrow S, \quad (7.6)$$

де  $S$  — невироджена матриця;

2. Для всіх  $\tau > 0$  при  $n \rightarrow \infty$

$$B = \sum_{k=1}^n E |\xi_{k,n}|^2 \mathbb{I}\{|\xi_{k,n}| > \tau\} \rightarrow 0 \quad (7.7)$$

(умова Ліндеберга).

Тоді має місце слабка збіжність  $\zeta_n$  до розподілу  $N(0, S)$ .

(Зрозуміло, що у випадку одновимірних  $\xi_{k,n}$ ,  $S_n$  є дисперсією  $\zeta_n$ ).

**Теореми неперервності.** Наступні теореми разом з центральною граничною теоремою дозволяють аналізувати швидкість збіжності оцінок, які можна представити як функції від сум незалежних випадкових величин.

**Теорема 7.3.9** (Теорема 3B у [2], п.5, розділ 1). *Нехай  $\eta_n$  — послідовність випадкових векторів у  $\mathbb{R}^s$ ,  $a \in \mathbb{R}^s$  — фіксований вектор,  $b_n \in \mathbb{R}$  — числова послідовність,  $b_n \rightarrow 0$ ,  $H : \mathbb{R}^s \rightarrow \mathbb{R}^k$  — не випадкова функція.*

Припустимо, що

1.  $\eta_n \Rightarrow \eta$  при  $N \rightarrow \infty$ .

2. Для  $j = 1, \dots, k$  існують похідні

$$H'_j = \left( \frac{\partial}{\partial x_1} H_j(x_1, \dots, x_s), \dots, \frac{\partial}{\partial x_s} H_j(x_1, \dots, x_s) \right)^T,$$

неперервні в точці  $a$ .

Тоді

$$(H(a + b_n \eta_n) - H(a))/b_n \Rightarrow H'(a)\eta,$$

де  $H'(x) = (\frac{\partial}{\partial x_i} H_j(x_1, \dots, x_s), i = 1, \dots, s; j = 1, \dots, k)$  — матриця з  $s$  стовпчиків і  $k$  рядочків.

**Теорема 7.3.10 (Слуцького)** Нехай  $X, X_1, X_2, \dots, Y_1, Y_2, \dots$  — випадкові величини, причому  $X_n \Rightarrow X$  і  $Y_n \rightarrow c$ , при  $n \rightarrow \infty$ , де  $c$  — не випадкове число. Тоді

- (i)  $X_n + Y_n \Rightarrow X + c$ ;
- (ii)  $Y_n X_n \Rightarrow cX$ ;
- (iii)  $X_n/Y_n \Rightarrow X/c$  якщо  $c \neq 0$ .

Доведення див. теорему 1.11 у [64].

## 7.4 Слабка збіжність випадкових функцій

У цьому параграфі розглядається теорія слабкої збіжності випадкових функцій, як випадкових елементів певних функціональних просторів. Відмітимо, що можливі і інші трактування поняття слабкої збіжності. Наприклад, її можна визначати для функцій, які не задовольняють умовам вимірності, тобто не є випадковими елементами. Для таких функцій слабка збіжність визначається у термінах зовнішньої ймовірності та зовнішніх математичних сподівань. Цей підхід виявляється особливо плідним для аналізу емпіричних мір у випадку незалежних однаково розподілених спостережень (див., наприклад [70]). Інший можливий підхід — використання для визначення слабкої збіжності методу одного ймовірнісного простору на основі теореми Скорохода (див. с.43). Однак у цьому параграфі ми будемо дотримуватись класичного підходу до визначення слабкої збіжності, слідуючи в основному книжкам [1, 4] та роботі [39].

В основному ми будемо розглядати функції  $f : Z \rightarrow \mathbb{R}$  де  $Z \subseteq \mathbb{R}^d$  як елементи просторів  $C(Z)$  та  $D(Z)$ .

Простір  $C(Z)$  — це простір функцій, неперервних на множині  $Z$  з рівномірною нормою  $\|f\|_\infty = \sup_{x \in Z} |f(x)|$ . Якщо  $Z$  — компакт у  $\mathbb{R}^d$ , то  $C(Z)$  — сепарабельний банахів простір.

Нехай  $Z = [a, b] \subseteq \mathbb{R}$ . Простір  $\mathcal{D}(Z) = \mathcal{D}[a, b]$  складається з обмежених функцій  $f$ , неперервних зліва на  $[a, b]$ . На цьому просторі можна задати різні топології. Перша задається рівномірною нормою. У цій нормі простір  $(\mathcal{D}[a, b], \|\cdot\|)$  не є сепарабельним.

Інший варіант визначення топології у  $\mathcal{D}[a, b]$  — використання метрики Скорохода — див. [1]. Використовуючи певний варіант цієї метрики, простір  $\mathcal{D}[a, b]$  можна зробити сепарабельним. Цей метричний простір будемо позначати  $\mathcal{D}$ . (Означення простору  $D(Z)$  для  $Z \subseteq \mathbb{R}^d$  див. [39]).

Простір  $\mathcal{D}(\mathbb{R})$  утворюється з простору  $\mathcal{D}[0, 1]$  “розтягуванням” відрізка  $[0, 1]$  на  $\mathbb{R}$ , наприклад, за допомогою логістичної функції  $z(t) = 1/(1 + e^{-t})$ . Тобто  $\mathcal{D}(\mathbb{R}) = \{y(z^{-1}(\cdot)) : y \in \mathcal{D}[0, 1]\}$ .

Нехай  $S$  — будь-який метричний простір,  $\{\zeta_n\}$  — деяка послідовність його випадкових елементів. Кажуть, що  $\{\zeta_n\}$  слабо збігається до випадкового елемента  $\zeta$ , якщо для будь-якої неперервної, обмеженої функції  $g : S \rightarrow \mathbb{R}$  виконується

$$\lim_{n \rightarrow \infty} E g(\zeta_n) = E g(\zeta).$$

Слабку збіжність позначають  $\zeta_n \Rightarrow \zeta$ .

Множина  $\{\zeta_n\}$  випадкових елементів  $S$  зветься компактною, якщо будь-яка послідовність випадкових елементів з цієї множини має слабо збіжну підпослідовність.

Множина  $\{\zeta_n\}$  випадкових елементів  $S$  зветься щільною, якщо для будь-якого  $\varepsilon > 0$  знайдеться компактна множина  $K_\varepsilon \subset S$ , така, що

$$\sup_n P\{\zeta_n \in S \setminus K_\varepsilon\} \leq \varepsilon.$$

**Теорема 7.4.1** (Теорема Прохорова) У будь-якому метричному просторі  $S$  для будь-якої множини випадкових елементів з щільності випливає компактність.

Доведення див. [4] с. 33 або [32].

Нехай  $\zeta(t)$  — випадкова функція. Скінченновимірними розподілами  $\zeta$  називають розподіли випадкових векторів  $(\zeta(t_1), \dots, \zeta(t_k))$ , де  $t_1, \dots, t_k$  — довільні точки з області визначення  $\zeta$ .

Функцію  $\zeta$  називають вибірково неперервною на  $Z$ , якщо вона майже напевне є неперервною у всіх точках, тобто  $P\{\zeta \in C(Z)\} = 1$ .

**Теорема 7.4.2** Нехай  $\{\zeta_n\}$  є послідовністю випадкових елементів  $\mathcal{D}(Z)$  (відповідно  $C(Z)$ ) і  $\zeta$  є випадковим елементом  $\mathcal{D}(Z)$  (відповідно  $C(Z)$ ). Якщо всі скінченновимірні розподіли  $\zeta_n$  слабо збігаються до відповідних скінченновимірних розподілів  $\zeta$  і набір  $\{\zeta_n\}$  є компактним у  $\mathcal{D}(Z)$  (відповідно  $C(Z)$ ), то  $\zeta_n \Rightarrow \zeta$  у  $\mathcal{D}(Z)$  (відповідно  $C(Z)$ ).

**Теорема 7.4.3** (див. Bikel, Wichura, 1971) Якщо  $\{\zeta_n\}$  — набір елементів з  $\mathcal{D}(Z)$  і

- (i)  $\zeta_n \Rightarrow \zeta$  у  $\mathcal{D}(Z)$ ,
  - (ii)  $\zeta$  є вибірково неперервною на  $Z$ ,
- то  $\zeta_n \Rightarrow \zeta$  у  $\mathcal{D}(Z)$  у рівномірній нормі.



**Теорема 7.4.4** Нехай  $\zeta_n$  — послідовність випадкових елементів у  $C(Z)$ , де  $Z$  — компакт у  $\mathbb{R}^d$  і виконуються наступні умови.

1. Для деяких  $C > 0$ ,  $\alpha > 0$ ,  $\beta > 0$  і всіх  $n = 1, 2, \dots$ , всіх  $t, s \in Z$ ,

$$\mathbb{E} |\zeta_n(t) - \zeta_n(s)|^\alpha \leq C |t - s|^{d+\beta}.$$

2. Сім'я випадкових величин  $\zeta_n(t_0)$  є щільною при деякому  $t_0 \in Z$ .

Тоді набір  $\{\zeta_n\}$  є щільним у  $C(Z)$ .

Доведення див. наслідок 2.7.1 у [4].

**Лема 7.4.1** Нехай  $X_n$  — послідовність випадкових процесів з  $\mathcal{D}[a, b]$ , таких, що для деяких  $\gamma > 0$ ,  $\alpha > 1/2$  і всіх  $a \leq t_1 < t < t_2 \leq b$ ,

$$\mathbb{E} |X_n(t) - X_n(t_1)|^\gamma |X(t_2) - X(t)|^\gamma \leq (t_2 - t_1)^{2\alpha}.$$

Тоді сім'я розподілів  $X_n$  є щільною в  $\mathcal{D}[a, b]$ .

Доведення див. теорему 15.6 у [1].

**Лема 7.4.2** Нехай  $X$  — випадковий процес на інтервалі  $[a, b]$  і для деяких  $\gamma > 0$ ,  $\alpha > 1/2$  та неспадної неперервної функції  $F$  на  $[a, b]$ ,  $\forall a \leq t_1 < t < t_2 \leq b$

$$\mathbb{E} |X(t) - X(t_1)|^\gamma |X(t_2) - X(t)|^\gamma \leq (F(t_2) - F(t_1))^{2\alpha}.$$

Тоді існує така константа  $C$ , яка залежить лише від  $\alpha$  і  $\gamma$ , що для всіх  $\varepsilon > 0$

$$\mathbb{P}\left\{\sup_{t \in [a, b]} |X(t)| \geq \varepsilon\right\} \leq \mathbb{P}\{|X(a)| > \varepsilon/2\} + \mathbb{P}\{|X(b)| > \varepsilon/2\} + \frac{C}{\varepsilon^{2\gamma}} |F(b) - F(a)|^{2\alpha}.$$

**Доведення.** Згідно з нерівністю (15.30) з [1], в умовах леми існує така константа  $K < \infty$ , що

$$\begin{aligned} & \mathbb{P}\left\{\sup_{t \in [a, b]} \min\{|X(t) - X(t_1)|, |X(t_2) - X(t)|\} \geq \varepsilon\right\} \\ & \leq \frac{2K}{\varepsilon^{2\gamma}} |F(b) - F(a)| \sup_{t, s \in [a, b]} |F(t) - F(s)|^{2\alpha-1} \\ & = \frac{2K}{\varepsilon^{2\gamma}} |F(b) - F(a)|^{2\alpha}. \end{aligned} \quad (7.8)$$

Для того, щоб у деякій точці  $t \in [a, b]$  мало місце  $|X(t)| > \varepsilon$ , необхідно щоб або  $|X(a)| > \varepsilon/2$ , або  $|X(t) - X(a)| > \varepsilon/2$ , або  $|X(b) - X(t)| > \varepsilon/2$ , або  $|X(b)| > \varepsilon/2$ . Тому, враховуючи (7.8), отримуємо твердження леми.

## 7.5 Ефективність. Мінімаксність. Інформація.

**Інформаційна матриця Фішера.** Нехай спостережувані дані  $X$  є випадковим елементом деякого вимірного простору  $\mathcal{X}$  з розподілом  $P\{X \in A\} = P_{\vartheta}(A)$  визначеним з точністю до невідомого параметра

$$\vartheta = (\vartheta_1, \dots, \vartheta_d)^T \in \Theta \subseteq \mathbb{R}^d.$$

Будемо припускати, що для деякої міри  $\mu$  та для всіх  $\vartheta \in \Theta$  існує щільність  $f_{\vartheta}(\cdot)$  розподілу  $P_{\vartheta}$  відносно  $\mu$ , тобто

$$P_{\vartheta}(A) = \int_A f_{\vartheta}(x) \mu(dx).$$

Позначимо  $f'_{\vartheta}(x) = \left( \frac{\partial}{\partial \vartheta_1} f_{\vartheta}(x), \dots, \frac{\partial}{\partial \vartheta_d} f_{\vartheta}(x) \right)^T$  — градієнт  $f_{\vartheta}(x)$  як функції від  $\vartheta$  (припускаючи, звичайно, що цей градієнт існує). Інформаційною матрицею Фішера (або просто інформацією) що міститься у даних  $X$  про параметр  $\vartheta$  називають

$$I_{\vartheta}^X = \int f'_{\vartheta}(x) (f'_{\vartheta}(x))^T \frac{\mu(dx)}{f_{\vartheta}(x)} = \mathbb{E}(\log f_{\vartheta}(X))' ((\log f_{\vartheta}(X))')^T.$$

Інакше кажучи,  $I_{\vartheta}^X = (i_{ik})_{i,k=1}^d$ , де

$$i_{jk} = \int \frac{\partial f_{\vartheta}(x)}{\partial \vartheta_j} \frac{\partial f_{\vartheta}(x)}{\partial \vartheta_k} \frac{\mu(dx)}{f_{\vartheta}(x)} = \mathbb{E} \frac{\partial \log f_{\vartheta}(X)}{\partial \vartheta_j} \frac{\partial \log f_{\vartheta}(X)}{\partial \vartheta_k}.$$

Інформація не залежить від вибору мажоруючої міри  $\mu$ .

Якщо дані  $X$  складаються з незалежних спостережень:  $X = (\xi_1, \dots, \xi_N)$ , то

$$I_{\vartheta}^X = I_{\vartheta}^{\xi_1} + I_{\vartheta}^{\xi_2} + \dots + I_{\vartheta}^{\xi_N}$$

(властивість адитивності інформації, див. теорему 7.1 розділу 1 [10]).

Набір  $(X, P_{\vartheta}, \Theta)$  називають статистичним експериментом. Експеримент називають регулярним, якщо

1.  $f_{\vartheta}(x)$  є неперервною функцією  $\vartheta$  для майже всіх  $x$  відносно міри  $\mu$ .
  2. Існує скінченна інформація Фішера  $I_{\vartheta}^X$  для всіх  $\vartheta \in \Theta$ .
  3. Існує неперервна як елемент  $L_2(\mu)$  похідна функції  $f_{\vartheta}^{1/2}(\cdot)$  у середньому квадратичному (відносно міри  $\mu$ ).
- (Більш докладно див. у п. 7 розділу 1 [10]).

**Нерівність Крамера для незміщених оцінок.** Нехай  $(X, P_\vartheta, \Theta)$  — регулярний статистичний експеримент,  $\hat{\vartheta}$  — незміщена оцінка  $\vartheta$  (тобто  $E \hat{\vartheta} = \vartheta$ ). Тоді

$$\text{Cov}_\vartheta(\hat{\vartheta}) \geq (I_\vartheta^X)^{-1}, \quad (7.9)$$

де  $\text{Cov}_\vartheta(\hat{\vartheta})$  — коваріаційна матриця  $\hat{\vartheta}$  у випадку, коли справжнє значення невідомого параметра дорівнює  $\vartheta$ . (Див. теорему 7.3, п. 7 розділу 1 [10]).

**Локальна асимптотична нормальність і мінімаксність.** Сім'ю (послідовність) статистичних експериментів  $(X^N, P_\vartheta^N, \Theta)$ ,  $N = 1, 2, \dots$  називають асимптотично локально нормальною (ЛАН) у точці  $t \in \Theta$  при  $N \rightarrow \infty$  якщо для деякої невідродженої матриці  $\varphi_N = \varphi_N(t)$  і будь-якого  $u \in \mathbb{R}^d$  справедливе зображення

$$\frac{\partial P_{t+\varphi_N u}^N}{\partial P_t^N}(X^N) = \exp \left( u^T \Delta_{N,t} - \frac{1}{2} |u|^2 + \psi_N(u, t) \right),$$

де  $\Delta_{N,t} \Rightarrow \zeta$  при  $N \rightarrow \infty$ ,  $\zeta$  — гауссів вектор з нульовим середнім та одиничною коваріаційною матрицею,  $\psi_N(u, t) \rightarrow 0$  за ймовірністю при  $N \rightarrow \infty$ .

При цьому  $\varphi_N$  називають нормуючою матрицею послідовності експериментів  $(X^N, P_\vartheta^N, \Theta)$ .

Нехай  $X^N = (\xi_1, \dots, \xi_N)$ , де  $\xi_j$  — незалежні випадкові елементи з щільностями  $f_j(x) = f_j(x, \vartheta)$  відносно мір  $\mu_j$ . Позначимо

$$\Psi^2(\vartheta, N) = \sum_{j=1}^N I_\vartheta^{\xi_j}$$

**Теорема 7.5.1** Нехай  $\Theta \subseteq \mathbb{R}^d$ , матриця  $\Psi^2(\vartheta, N)$  строго додатнєвизначена, і виконані наступні умови:

1. Для всіх  $k > 0$ ,

$$\lim_{n \rightarrow \infty} \sup_{|u| < k} \sum_{j=1}^n \int \left( \left( \frac{\partial f_j^{1/2}(x, t + \Psi^{-1}(n, t)u)}{\partial t} - \frac{\partial f_j^{1/2}(x, t)}{\partial t} \right)^T \Psi^{-1}(n, t)u \right)^2 \mu_j(dx) = 0.$$

2. Умова Ліндеберга: для всіх  $\varepsilon > 0$ ,  $u \in \mathbb{R}^d$ ,

$$\lim_{n \rightarrow \infty} \sum_{j=1}^N E_t \left[ \left( u^T \Psi^{-1}(n, t) \frac{\partial f_j(\xi_j, t)}{\partial t} \right)^2 \right] = 0.$$

$$\times \mathbb{I} \left( \left| u^T \Psi^{-1}(n, t) \frac{\partial f_j(\xi_j, t)}{\partial t} \right| > \varepsilon \right) = 0,$$

Тоді сім'я  $X_N$  є ЛАН з нормуючою матрицею  $\varphi(n, t) = \Psi^{-1}(n, t)$ .

(Див. теорему 6.1 розділу 2 [10]).

**Теорема 7.5.2 (Гаєка)** Нехай сім'я  $X_N$  задовольняє ЛАН у точці  $\vartheta = t$  з нормуючою матрицею  $\varphi_N \rightarrow 0$  при  $N \rightarrow \infty$ . Тоді для будь-якої послідовності оцінок  $T_N = T_N(X_N)$ , будь-якого  $\delta > 0$  і будь-якого  $h \in \mathbb{R}^d$ ,

$$\liminf_{N \rightarrow \infty} \sup_{|\vartheta - t| < \delta} \mathbb{E}_{\vartheta}(h^T \varphi_N(T_N - \vartheta))^2 \geq \mathbb{E}(h^T \zeta)^2, \quad (7.10)$$

де  $\zeta$  — гауссів вектор з нульовим середнім та одиничною коваріаційною матрицею.

(Див. теорема 12.1 розділу 2 у [10]).

Послідовність оцінок  $T_N$  для яких у (7.10) має місце рівність називають локально асимптотично мінімаксними (асимптотично ефективними у розумінні Гаєка).

## 7.6 Оцінювання щільності за кратними вибірками

Використовуючи теорію, описану у попередньому параграфі, можна знаходити нижню межу для швидкості збіжності параметричних оцінок у регулярних випадках. Виявляється, що ця теорія дозволяє також отримувати подібні результати у задачах параметричного оцінювання. Ми зупинимось тут на оцінках щільності розподілу.

Нехай дані  $X = (\xi_1, \dots, \xi_N)$  являють собою вибірку з незалежних, однаково розподілених випадкових величин, причому існує щільність розподілу  $\xi_j$  (відносно міри Лебега). Позначимо цю щільність через  $f$ . Функція  $f$  невідома, але вважається, що вона належить класу Гьольдера  $\Sigma(\beta, L)$  для деяких  $0 < \beta < \infty$ ,  $0 < L < \infty$ .

Нагадаємо означення  $\Sigma(\beta, L)$ . Нехай  $\beta = k + \alpha$  де  $k$  — невід'ємне ціле число,  $\alpha \in (0, 1]$ . Функція  $f : \mathbb{R} \rightarrow \mathbb{R}$  належить  $\Sigma(\beta, L)$  тоді і тільки тоді, коли у неї існує неперервна  $k$ -та похідна  $f^{(k)}(x)$  і для всіх  $x_1, x_2 \in \mathbb{R}$  виконана нерівність

$$|f^{(k)}(x_1) - f^{(k)}(x_2)| \leq L|x_1 - x_2|^\alpha.$$

По суті,  $\beta$  є показником гладкості функції  $f$ . Наприклад, при  $\beta = 1$  клас Гьольдера  $\Sigma(1, L)$  являє собою клас функцій, для яких виконується умова Ліпшиця з константою  $L$ :

$$|f(x_1) - f(x_2)| \leq L|x_1 - x_k|.$$

При  $\beta = 2$ ,  $\Sigma(2, L)$  це клас функцій, таких, що їх перші похідні задовольняють умову Ліпшиця. (Тобто умова  $f \in \Sigma(2, L)$  є трохи слабшою, ніж умова обмеженості другої похідної  $f$ ).

Нехай  $\mathcal{F}_N$  — клас всіх можливих оцінок щільності, побудованих по спостереженнях  $X$ .

**Теорема 7.6.1** Для будь-яких  $L > 0$ ,  $\beta > 0$ ,  $x \in \mathbb{R}$ ,

$$\liminf_{N \rightarrow \infty} \inf_{\tilde{f}_N \in \mathcal{F}_N} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left[ (\tilde{f}_N(x_0) - f(x_0)) N^{\beta/(2\beta+1)} \right]^2 > 0.$$

(Див. теорему 5.1 розділу 4 у [10]).

Інакше кажучи, швидкість збіжності оцінок щільності не може бути кращою ніж  $\tilde{N}^{-\beta/(2\beta+1)}$  де  $\beta$  — показник гладкості оцінюваної щільності. Точніше, для будь-якої оцінки  $\tilde{f}_n$  знайдеться така  $f \in \Sigma(\beta, L)$ , для якої  $\left[ \mathbb{E}_f (\tilde{f}_N(x_0) - f(x_0))^2 \right]^{1/2} > c N^{-\beta/(2\beta+1)}$  для деякого  $c > 0$  і всіх  $N$ . Таку швидкість збіжності забезпечують ядерні оцінки щільності при правильному виборі параметра згладжування ( $h_N = N^{-1/(2\beta+1)}$ ).

У випадку багатовимірних однаково розподілених спостережень  $X = (\xi_1, \dots, \xi_N)$ ,  $\xi_j \in \mathbb{R}^d$ , для двічі диференційовних щільностей ядерні оцінки забезпечують оптимальну швидкість збіжності  $\tilde{N}^{-2/(4+d)}$  (див. [49, 72]).

# Література

- [1] Биллингсли П. Сходимость вероятностных мер.- М.: Наука, 1977.— 357с.
- [2] Боровков А.А. Математична статистика.- М.: Наука, 1984.— 472с.
- [3] Боровков А.А. Теория вероятностей.- М.: Наука, 1986.
- [4] Булдыгин В.В. Сходимость случайных элементов в топологических пространствах. — Киев.: Наукова думка, 1980. — 240с.
- [5] Булдыгин В.В., Козаченко Ю.В. Метрические характеристики случайных величин и процессов.- Киев: TViMS, 1998.— 289с.
- [6] В.Н.Вапник Индуктивные принципы поиска эмпирических закономерностей., в кн. -Распознавание -Классификация -Прогноз, Вып. 1. “Наука”, Москва, 1989, 17–81.
- [7] Вапник В.Н., Червоненкис А.Я. Теория распознавания образов.- М: Наука, 1974.— 416с.
- [8] Гихман, И.И., Скороход А.В., Ядренко М.И. Теория вероятностей и математическая статистика.- Киев: Вища школа, 1979.— 408с.
- [9] Л.Деврой, Л.Дьерфи Непараметрическое оценивание плотности. Мир, Москва,1988.—408 с.
- [10] Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. — М. : Наука, 1979. — 528с.
- [11] Ю.О.Іванько, Р.Є.Майборода Експоненціальні оцінки емпірично-баєсового ризику при класифікації суміші зі змінними концентраціями, Український математичний журнал.— т. 54 (2002), №10.— 1421–1428.

- [12] Іванько Ю.Н., Майборода Р.Є. Асимптотика порогових класифікаторів побудованих за вибіркою з суміші зі змінними концентраціями.— Теорія ймовірн. та математ. статист.— 2006, Вип. 74.— с.34-43.
- [13] Ю.О.Іванько Асимптотика ядерних оцінок щільностей та їх похідних, побудованих за спостереженнями із суміші зі змінними концентраціями, Вісник КНУ, сер. Математика. Механіка 2003.— вип. 9-10, 29-35.
- [14] Козаченко Ю.В. Лекції з теорії вейвлетів.— К.: ТВіМС, 2004.— 147с.
- [15] Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике. — М.:Наука, 1985. — 640 с.
- [16] Кубайчук О.О. Оцінювання моментів по сумішах з використанням виправленої зваженої емпіричної функції розподілу.— Вісник КНУ. Математика. Механіка.— 2003, Вип. 9-10. — с. 48-52.
- [17] Кравец Т.Н., Майборода Р.Е. Алгоритм нейронной классификации по наблюдениям из смеси с переменными концентрациями.— Кибернетика и системный анализ.— 1998, N6 с.133-140
- [18] Лодатко А., Майборода Р. Оцінювання щільності розподілу по спостереженнях з домішкою.— Теорія ймовірн. та мат. статист.— 2005, Вип. 73,с.88-96.
- [19] Лодатко А., Майборода Р. Адаптивна моментна оцінка параметру розподілу по спостереженнях з домішкою.— Теорія ймовірн. та мат. статист.—Вип. 75, 2006,с.61-70
- [20] Майборода Р.Є. Проекційні оцінки концентрацій сумішей що змінюються.— Теорія ймовірн. та математ. статист, 1992, N46.— с. 71-75.
- [21] Майборода Р.Є. Оцінка розподілів компонентів сумішей з концентраціями, що змінюються .— Український мат. журнал, 1996, v.48 N4.— с.с.562–566.
- [22] Майборода Р.Є. Статистика раптової смерті.— Теорія ймовірн. та математ. статист N 58, 1998.— с.113-122.

- [23] Майборода Р.Є. Асимптотично ефективна оцінка ймовірностей, побудована по спостереженням із суміші. - Теорія ймовірн. та математ. статист. 1998а, N 59.— с.121-128.
- [24] Майборода Р.Є. Оцінка границь розподілу по спостереженнях з суміші.— Теорія ймовірн. та математ. статист., 1999, N 60.— с.131-135.
- [25] Майборода Р.Є., Сугакова О.В. Мінімізаційна оцінка параметрів концентрацій двокомпонентних сумішей.— Теорія ймовірн. та математ. статист.— 2001, Вип. 65.— с. 121-132.
- [26] Майборода Р.Є. Статистичний аналіз сумішей.— К.: ВПЦ “Київський університет”, 2003.— 176с.
- [27] Э.А.Надарая Об интегральной среднеквадратической ошибке некоторых непараметрических оценок плотности вероятности.//Теория вероятностей и ее применения.— 1974 т. 19.— с.131-139.
- [28] Петров В.В. Предельные теоремы для сумм независимых случайных величин.— М., Наука, 1987.— 320с.
- [29] Похилько Д.І. Вейвлет-оцінки щільності по спостереженням з суміші// Теорія ймовірн. та математ. статист.— 2004.— Т. 70.— с. 121-130.
- [30] Похилько Д.І. Адаптивна оцінка щільності компоненти суміші.— Теорія ймовірн. та математ. статист.— 2006.— Т. 74.— с. 129-142.
- [31] Похилько Д.І. Експоненційні нерівності для швидкості збіжності в рівномірній нормі вейвлет-оцінки щільності компоненти суміші.— Вісник КНУ. Серія: фізико-математичні науки.— 2006.— N 1.— с/40–47.
- [32] Прохоров Ю.В. Сходимость случайных процессов и предельные теоремы теории вероятностей.— Теория вероятностей и ее применения, 1956, т.1.— с. 177-238.
- [33] Прохоров Ю.В. Одна экстремальная задача теории вероятностей.— Теория вероятностей и ее применения, 1959, т.4, N.2, с.211-214.
- [34] Рижов А.Ю. Оцінки розподілів компонент суміші по цензурованим даним.- Теорія ймовірностей та математична статистика.— 2003, Вип. 69.— с.154-161.



- [35] Рижов А.Ю. Перевірка гіпотези про однорідність компонент суміші зі змінними концентраціями за цензурованими даними.- Теорія ймовірностей та математична статистика.— 2005, Вип. 72.— с.129-139.
- [36] Рижов А.Ю. Оцінка параметрів концентрацій двокомпонентних сумішей за цензурованими даними.- Теорія ймовірностей та математична статистика.— 2007, Вип. 76.— с.150-159.
- [37] Скороход А.В. Исследования по теории случайных процессов.— Киев, Изд-во КГУ, 1961.— 216с.
- [38] Сугакова О.В. Асимптотика ядерної оцінки щільності розподілу по спостереженнях суміші зі змінними концентраціями.— Теорія ймовірностей та математична статистика.— 1998, Вип. 59, с.156–166.
- [39] Bickel P.J. The 1980 Wald memorial lectures on adaptive estimation.— Ann. Statist., 1982, v. 10, N.3.— p.647–671.
- [40] Bordes L., Mottelet S., Vandekerkhove P. Semiparametric estimation of a two-component mixture model.- Ann. Statist.- 2006.- v.34, No 3.- p.1204-1232.
- [41] Bordes L., Delmas C., Vandekerkhove P. Semiparametric Estimation of a two-component Mixture model where one component is known.- Scand. J. Statist.- 2006.- v. 33.- p. 733-752.
- [42] H.Chernoff Estimation of the mode.— Ann. Inst. Statist. Math.— 1964, **16**.— 31-41p.
- [43] Cover T.M., Hart P.E. Nearest neighbor pattern classification.— IEEE Trans. on Information Theory, 1967.— V.IT-13, p.21-27.
- [44] Federer H. Geometric Measure Teory.— Springer Verlag: 1969.
- [45] Györfi L. On the rate of convergence of nearest neighbor rules.— IEEE Trans. on Information Theory, 1978.— V.IT-24, p.509-512.
- [46] Györfi L. Recent results on nonparametric regression estimate and multiple classification.— Problems of Control and Information Theory, 1981.— V.10, N1, p.43-52.

- [47] Hall P., Zhou X.-H. Nonparametric estimation of component distributions in a multivariate mixture.— *Ann.Statist.* 2003, V. 31, No 1, 201-224.
- [48] Hall P., Titterington D. M. The use of uncategorized data to improve the performance of a nonparametric estimator of a mixture density.— *J. Roy. Statist. Soc. Ser. B.* 1985, v. 47.— p. 155–161.
- [49] Hardle W., Muller M., Sperlich S., Werwatz A. Nonparametric and Semiparametric Models.— Berlin, Springer, 2004.— 300p.
- [50] Hoeffding W. Probability inequalities for sums of bounded random variables.— *J. Amer. Math. Assoc.*— 1963, V. 58, N. 301.— p. 13-30.
- [51] Holzmann H., Munk A., Gneiting T. Identifiability of Finite Mixtures of Elliptical Distributions.— *Scand. J. Statist.*, 2006, v. 33.— p. 753–763.
- [52] Hunter D.R., Wang S., Hettmansperger T.R. Inference for mixtures of symmetric distributions.- Technical Report 04-01, Penn State University, Philadelphia, 2004.- 39p.
- [53] J.Kim, D.Pollard Cube root asymptotics.— *Annals of statistics* 1990 V.18, №1.— 191-219.
- [54] Kubaychuk O.O. Estimation of moments by observations from mixtures with varying concentrations.— *Theory of Stoch. Proc.*— 2002.— v.8(24), N3-4.— p.226–232.
- [55] Maiboroda R.E. Estimation and classification by mixtures with time-dependent concentrations .- VI international Vilnius conference on probability theory and math. statistics. Abstracts of communications. 1993. V. 2.— p.48.
- [56] Maiboroda R. Consistency of the weighted empirical distribution function based on the observations from mixtures with varying concentrations.— *Theory of Stochastic processes.*— 1999, v 5 (21), N3-4.— p.152-161.
- [57] McLachlan G. J., Basford K.E. Mixture models: Inference and Applications to Clustering.— New York: Dekker, 1988.— 312p.
- [58] McLachlan G. J., Peel D. Finite Mixture Models.— NY, Wiley.— 2000.

- [59] L. Mohammadi, S van de Geer, On threshold-based classification rules.— Institute of Mathematical Statistics, Lecture Notes Monograph Series, Mathematical Statistics and Applications: Festschrift for Constance van Eeden. **42** (2003).— p.261–280.
- [60] Newcomb S. A generalized theory of combination of observations so as to obtain the best result.— Amer. J. Math.— 1894.— V.8.— p. 343–366.
- [61] Pearson K. Contribution to the mathematical theory of evolution.— Phil. Trans. Roy. Soc. A.— 1894.— v. 185.— p. 71–110.
- [62] Pollard D. Convergence of Stochastic Processes.— New-York: Springer-Verlag, 1984.— 456p.
- [63] Teicher H. Identifiability of mixtures.— Ann. Statist.— 1961.— v. 32, N1.— p. 244–248.
- [64] Shao J. Mathematical statistics.- NY Berlin Heidelberg: Springer-Verlag, 1999.— 530p.
- [65] H. Shen On methods of sieves and penalization.— Ann. Statist. 1997 V.25, No 6.— p.2555–2591.
- [66] Stone C.J. Consistent nonparametric regression.- Ann. Statist., 1977, V.5.— p.595–645.
- [67] Stone C. Adaptive maximum likelihood estimation of a location parameter.— Ann. Statist., 1975, v.3.— p. 267–284.
- [68] Tarasenko F.P. On the evaluation of an unknown probability density function, the direct estimation of the entropy from independent observations of a continuous random variable and the distribution-free test of goodness-of-fit.— Proceedings IEEE, 1968, v. 56, N 1, p.2052–2053.
- [69] Titterington D.M., Smith A.F. and Makov U.E. Analysis of Finite Mixture Distributions.- New York: Wiley, 1985.— 364p.
- [70] Vaart A. W. van der, Welner J.A. Weak convergence and empirical processes.— Springer-Verlag, NY, 1996, 512p.
- [71] V.N.Vapnik The nature of Statistical Learning Theory. New York, Springer, 1996.

- [72] Wand, M.P., Jones, M.C. Kernel Smoothing, Vol 60 of Monographs on Statistics and Applied Probability.— London, Chapman and Hall, 1995.— 212 p.
- [73] Yakovitz S.A., Spragins J. On the identifiability of finite mixtures.— Ann. Math. Statist., 1968, V. 39, N1.— p.209-214.

## СПИСОК ПОЗНАЧЕНЬ

$\Rightarrow$ — слабка збіжність	$E(\xi \mid \eta)$ — умовне математичне
$\xrightarrow{\text{Sk}}$ — збіжність за Скороходом	сподівання випадкової вели-
$\mathbb{I}\{A\}$ — індикатор події $A$ ( $\mathbb{I}\{A\} = 1$ якщо $A$ виконано, $\mathbb{I}\{A\} = 0$ якщо $A$ не виконано)	чини $\xi$ при фіксованому $\eta$
$\langle a \rangle_N = \frac{1}{N} \sum_{j=1}^N a_{j:N}$ — оператор усереднення $N$ -того рядка трикутного масиву $a$	$P\{A\}$ — ймовірність події $A$
$\langle a \rangle = \lim_{N \rightarrow \infty} \langle a \rangle_N$	$P\{A \mid B\}$ — умовна ймовірність події $A$ за умови $B$
$a^k$ — масив мінімаксних вагових коефіцієнтів для оцінки розподілу $k$ -тої компоненти	$\mathbb{R}$ — множина дійсних чисел
$\mathcal{D}$ — простір функцій, неперервних зліва	$P\{A \mid \eta\}$ — умовна ймовірність події $A$ при фіксованому $\eta$
$\hat{F}_N(x, a)$ — зважена емпірична функція розподілу з ваговими коефіцієнтами $a$	$\mathcal{X}$ — простір можливих значень даних
$H_k(A)$ — розподіл спостережуваних характеристик $k$ -тої компоненти суміші ( $A \in \mathfrak{A}$ )	$\mathfrak{A}$ — $\sigma$ -алгебра вимірних множин з $\mathcal{X}$
$H_k(x)$ — функція розподілу спостережуваних характеристик $k$ -тої компоненти суміші ( $x \in \mathbb{R}^d$ )	$w_{j:N}^k$ — концентрація $k$ -тої компоненти суміші під час $j$ -того спостереження
$h_k(x)$ — щільність розподілу спостережуваних характеристик $k$ -тої компоненти суміші	$\Gamma_N = (\langle w^k w^m \rangle_N)_{k,m=1}^M$ — матриця Грама концентрацій для вибірки з $N$ спостережень
$M$ — кількість компонент суміші	$\Gamma = (\langle w^k w^m \rangle)_{k,m=1}^M$ — гранична матриця Грама при $N \rightarrow \infty$
$E\xi$ — математичне сподівання випадкової величини $\xi$	$\hat{\mu}(A, a)$ — зважена емпірична міра з ваговими коефіцієнтами $a$
	$\Xi_N = (\xi_{j:1}, \dots, \xi_{j:N})$ — вибірка з суміші зі змінними концентраціями
	$\Xi = (\Xi_N, N = 1, 2, \dots)$ — трикутний масив вибірок
	з.е.ф.р. — зважена емпірична функція розподілу
	м.н. — майже напевне