Теорія ймовір. та матем. статист. Teor. Ĭmovir. ta Matem. Statyst.
Вип. 84, 2011, стор. 142–154 No. 84, 2011, pp. 142–154

ОЦІНЮВАННЯ СЕРЕДНЬОГО У МОДЕЛІ СУМІШІ ЗІ ЗМІННИМИ КОНЦЕНТРАЦІЯМИ

УДК 519.21

А. ЩЕРБІНА

Анотація. Розглядається модель спостережень з двокомпонентної суміші. З кожним об'єктом пов'язана певна числова характеристика. Її розподіл вважається невідомим. Спостереження отримуються в результаті вибіркового обстеження об'єктів, розподілених за декількома групами. Це призводить до появи залежності між характеристиками об'єктів, що і є відмінністю від класичної задачі суміші. Загальні кількості об'єктів першого та другого класів у групах відомі, а класи спостережених об'єктів невідомі. Ставиться задача оцінювання середніх значень характеристик об'єктів у класах. Розглядаються дві постановки такої задачі, одержано вирази середньоквадратичних похибок відповідних оцінок. Показано консистентність та асимптотична нормальність оцінок.

Abstract. Estimation of the mean value in the model of two-component mixture with dependent observations is considered. Two kinds of models are analyzed. Linear and adaptive nonparametric estimators are constructed. Consistency and asymptotic normality of estimators is shown.

Аннотация. Рассматривается модель наблюдений из двухкомпонентной смеси. С каждым объектом связана определенная числовая характеристика. Её распределение считается неизвестным. Наблюдения получаются в результате выборочного обследования объектов, распределенных по нескольким группам. Это приводит к появлению зависимостей между характеристиками объектов, что и является отличием от классической модели смеси. Общие количества объектов первого и второго класса в группах известны, а классы объектов — нет. Ставится задача оценивания средних значений характеристик объектов в классах. Рассматриваются две постановки такой задачи, получены выражения для среднеквадратических ошибок соответствующих оценок. Показана консистентность и асимптотическая нормальность оценок.

1. Вступ і огляд літератури

Задачі аналізу сумішей природно виникають у задачах астрономії, екології, економіки, біології та соціології.

Часто спостережувані об'єкти належать до різних класів, що мають різні ймовірнісні характеристики. В результаті досліджувана вибірка є сумішшю таких об'єктів. Методи дослідження таких задач почали розвиватися з кінця XIX сторіччя завдяки роботам С. Ньюкомба [7] та К. Пірсона [8]. Останнім часом починають розглядатися моделі сумішей зі змінними концентраціями компонентів. Приклади таких задач з'являються у багатьох випадках, коли умови спостереження об'єктів змінюються в процесі отримання даних. Розгляд таких прикладів та подальше розвинення теорії можна знайти у [4, 5].

У даній статті розглядається певне узагальнення задачі аналізу суміші зі змінними концентраціями. На відміну від класичної моделі сумішей, спостереження стають залежними між собою. Залежність виникає через те, що вибіркова сукупність

2010 *Mathematics Subject Classification.* Primary 62G05; Secondary 62D05.

Ключові слова і фрази. Оцінювання у моделі суміші, вибірковий метод, непараметричне оцінювання.

об'єктів отримується в результаті випадкового відбору зі скінченної генеральної сукупності. І при невеликих обсягах генеральної сукупності ефекти цих залежностей стають достатньо значними.

Ми обмежимось розглядом суміші з двох компонентів. Тобто будемо вважати, що об'єкти можуть належати до двох різних класів та розподілені за декількома групами. Кількості об'єктів першого та другого класів у групах вважаємо відомими. Метою обстеження є оцінювання середніх значень характеристик об'єктів обох класів. Розглядаються два підходи до визначення моделі утворення даних.

Для застосування запропонованої моделі необхідно, щоб генеральна сукупність складалася з об'єктів двох різних класів. Причому безпосереднє визначення класу конкретного об'єкта є неможливим або небажаним. Визначення сукупного розподілу об'єктів за класами здійснюється лише за певними групами цих об'єктів. У випадку соціального обстеження питання належності респондентів до певного класу може бути "дражливим". Тобто при безпосередньому опитуванні можна одержати багато помилкових відповідей. У такому випадку доцільніше провести анонімне обстеження цього питання за групами респондентів і одержати сукупні кількості респондентів першого та другого класів за цими групами. Далі проводиться персональне вибіркове обстеження цільових характеристик. Застосування до отриманих даних оцінок, що розглядаються у цій статті, дозволить визначити середні значення цільових характеристик за класами респондентів.

Прикладом такої ситуації є соціологічне обстеження успішності школярів в залежності від того, палять вони чи ні. Маємо дві сукупності учнів, що розподілені за декількома групами — шкільними класами. Однією з можливостей реалізації такого обстеження є проведення анонімного обстеження щодо паління за шкільними класами. Це дозволить отримати пропорції курців за класами. Далі можна провести звичайне вибіркове обстеження для отримання необхідних характеристик.

Середні значення оцінюються за допомогою зважених сум характеристик об'єктів. Завдання оцінювання полягає у визначенні найкращих з точки зору якості оцінок.

У розділі 2 вводяться основні позначення, розглядаються два варіанти постановки задачі. У 3-му розділі наведено загальний вид оцінок та їх дисперсій. У 4-му розділі розглядаються оцінки з мінімаксними коефіцієнтами, визначеними у [5]. У 5-му розділі розглядаються також адаптивні оцінки, досліджується асимптотична поведінка оцінок, доводиться їх консистентність та асимптотична нормальність. Далі наведено певні висновки та доведення теорем.

2. Постановка задачі

Нехай ми маємо два класи об'єктів, що розподілені за K групами, причому в i-тій групі міститься N_i^1 об'єктів першого класу та N_i^2 об'єктів другого класу. Загальну кількість об'єктів у i-й групі позначимо N_i, а загальні кількості об'єктів першого та другого класів — $N_{(1)}$ та $N_{(2)}$ відповідно. Пропорції об'єктів першого та другого класів у групах будемо називати концентраціями та позначати наступним чином:

$$w_i^l = \frac{N_i^l}{N_i}, \qquad i = 1, 2, \dots, K, \ l = 1, 2.$$

Нехай у кожній групі спочатку йдуть N_i^1 об'єктів першого класу, а потім N_i^2 об'єктів другого класу. З кожним об'єктом пов'язана певна числова характеристика. Їх значення у групі i дорівнюють $X_{i1}, X_{i2}, \dots, X_{iN_i}$.

Далі з кожної групи за допомогою простого випадкового відбору без повторення відбирається $n_i \geq 1$ об'єктів. Позначимо класи цих об'єктів $C_{i1}, C_{i2}, \dots, C_{in_i}$, а значення їх характеристик — $Y_{i1}, Y_{i2}, \dots, Y_{in_i}$.

При цьому будемо розглядати два механізми породження характеристик: зі скінченної популяції та за допомогою випадкового механізму.

У першому випадку множина усіх об'єктів фіксована, можливий лише їх перерозподіл за групами. Всі такі розподіли вважаємо рівноймовірними. Введемо позначення для середніх значень та дисперсій цих характеристик за класами:

$$m_1 = \frac{1}{N_{(1)}} \sum_{i=1}^{K} \sum_{j=1}^{N_i^1} X_{ij}, \qquad m_2 = \frac{1}{N_{(2)}} \sum_{i=1}^{K} \sum_{j=N_i^1+1}^{N_i} X_{ij},$$

$$s_1^2 = \frac{1}{N_{(1)}-1} \sum_{i=1}^{K} \sum_{j=1}^{N_i^1} (X_{ij} - m_1)^2, \qquad s_2^2 = \frac{1}{N_{(2)}-1} \sum_{i=1}^{K} \sum_{j=N_i^1+1}^{N_i} (X_{ij} - m_2)^2.$$

Для позначення усереднення за всіма можливими перестановками об'єктів за групами та вибірками з них до знаків математичного сподівання, дисперсії та коваріації будемо додавати індекс **p**. Таке усереднення ми будемо називати умовним або при фіксованій популяції об'єктів, а значення середніх та дисперсій — популяційними.

Розглянемо задачу оцінювання середніх значень m_1 та m_2 за вибіркою

$$\{Y_{ij}, i = 1, 2, \ldots, K, j = 1, 2, \ldots, n_i\}$$

при відомих обсягах N_i^l, $i = 1, 2, \ldots, K$, $l = 1, 2$.

У другому випадку ми вважаємо, що характеристики об'єктів

$$\{X_{ij}, i = 1, 2, \ldots, K, j = 1, 2, \ldots, N_i\}$$

є незалежними випадковими величинами, розподіл яких однаковий для всіх об'єктів одного класу. У цих розподілів існують скінченні другі моменти. Середні значення та дисперсії цих розподілів ми будемо позначати наступним чином:

$$\mu_1 = \mathsf{E}\, X_{ij}, \qquad \sigma_1^2 = \mathsf{D}\, X_{ij}, \quad i = 1, 2, \ldots, K, \ j = 1, 2, \ldots, N_i^1,$$
$$\mu_2 = \mathsf{E}\, X_{ij}, \qquad \sigma_2^2 = \mathsf{D}\, X_{ij}, \quad i = 1, 2, \ldots, K, \ j = N_i^1 + 1, \ldots, N_i.$$

Надалі знаки математичного сподівання, дисперсії та коваріації означають усереднення за значеннями характеристик $\{X_{ij}\}$ та усіма можливими вибірками з груп. Таке усереднення ми будемо називати безумовним, а відповідні значення середніх та дисперсій — модельними.

Розглянемо задачу оцінювання модельних середніх значень μ_1 та μ_2 за вибіркою $\{Y_{ij}, i = 1, 2, \ldots, K, j = 1, 2, \ldots, n_i\}$ при відомих обсягах N_i^l, $i = 1, 2, \ldots, K$, $l = 1, 2$.

3. Побудова оцінок

Ідея оцінювання полягає в тому, щоб використати відмінності між концентраціями w_i^1 у різних групах. Для цього ми будемо використовувати оцінки наступного вигляду:

$$\hat{m}_l(a^l) = \frac{1}{K} \sum_{i=1}^{K} a_i^l T_i, \qquad l = 1, 2, \tag{1}$$

де T_i — це середнє значення характеристик у i–й групі:

$$T_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}, \qquad i = 1, 2, \ldots, K,$$

а $a^l = \left(a_1^l, a_2^l, \ldots, a_K^l\right)$, $l = 1, 2$ — певні набори коефіцієнтів.

Оскільки оцінювання для першого та другого класів виконується аналогічно, то будемо розглядати оцінки лише для першого класу, а індекс 1 у a^1 опускати.

Математичне сподівання при фіксованій популяції величини T_i дорівнює

$$\begin{aligned}\mathsf{E}_{\mathbf{p}} T_i &= \frac{1}{n_i}\sum_{j=1}^{n_i}\mathsf{E}_{\mathbf{p}} Y_{ij} = \mathsf{E}_{\mathbf{p}} Y_{i1}\\ &= \mathsf{P}_{\mathbf{p}}(C_{i1}=1)\,\mathsf{E}_{\mathbf{p}}(Y_{i1}\mid C_{i1}=1)+\mathsf{P}_{\mathbf{p}}(C_{i1}=2)\,\mathsf{E}_{\mathbf{p}}(Y_{i1}\mid C_{i1}=2)\\ &= w_i^1 m_1 + w_i^2 m_2, \qquad i=1,2,\ldots,K.\end{aligned}$$

Математичне сподівання при фіксованій популяції оцінки $\hat{m}_1(a)$ дорівнює

$$\begin{aligned}\mathsf{E}_{\mathbf{p}}\,\hat{m}_1(a) &= \frac{1}{K}\sum_{i=1}^{K} a_i\,\mathsf{E}_{\mathbf{p}} T_i = \frac{1}{K}\sum_{i=1}^{K} a_i\left(w_i^1 m_1 + w_i^2 m_2\right)\\ &= \frac{m_1}{K}\sum_{i=1}^{K} w_i^1 a_i + \frac{m_2}{K}\sum_{i=1}^{K} w_i^2 a_i.\end{aligned}$$

Для безумовного математичного сподівання можемо записати

$$\begin{aligned}\mathsf{E}\,\hat{m}_1(a) = \mathsf{E}\,\mathsf{E}_{\mathbf{p}}\,\hat{m}_1(a) &= \frac{\mathsf{E}\,m_1}{K}\sum_{i=1}^{K} w_i^1 a_i + \frac{\mathsf{E}\,m_2}{K}\sum_{i=1}^{K} w_i^2 a_i\\ &= \frac{\mu_1}{K}\sum_{i=1}^{K} w_i^1 a_i + \frac{\mu_2}{K}\sum_{i=1}^{K} w_i^2 a_i.\end{aligned}$$

В обох випадках для незміщенності оцінки $\hat{m}_1(a)$ необхідне виконання наступних умов:

$$\frac{1}{K}\sum_{i=1}^{K} w_i^1 a_i = 1, \qquad \frac{1}{K}\sum_{i=1}^{K} w_i^2 a_i = 0. \tag{2}$$

Надалі будемо вважати ці умови виконаними.

Серед усіх незміщених оцінок природно вибрати оцінку з найменшою можливою дисперсією. Для цього необхідно вміти рахувати дисперсію оцінки для довільного набору коефіцієнтів $a_1, a_2, \ldots, a_K$. У наступних двох лемах наводяться вирази для дисперсій у випадку фіксованої популяції та для випадкового породження характеристик.

Теорема 1. *У випадку фіксованої популяції дисперсія оцінки $\hat{m}_1(a)$ дорівнює:*

$$\begin{aligned}\mathsf{D}_{\mathbf{p}}\,\hat{m}_1(a) = \frac{1}{K^2}\sum_{i=1}^{K}\frac{a_i^2}{n_i}\Bigg(& w_i^1 s_1^2 + w_i^2 s_2^2 + \frac{N_i - n_i}{N_i - 1} w_i^2 w_i^1\\ &\times\left((m_1 - m_2)^2 - \frac{s_1^2}{N_{(1)}} - \frac{s_2^2}{N_{(2)}}\right)\Bigg) - \frac{s_1^2}{N_{(1)}}.\end{aligned} \tag{3}$$

Теорема 2. *У випадку випадкового породження характеристик дисперсія оцінки $\hat{m}_1(a)$ дорівнює:*

$$\mathsf{D}\,\hat{m}_1(a) = \frac{1}{K^2}\sum_{i=1}^{K}\frac{a_i^2}{n_i}\left(w_i^1\sigma_1^2 + w_i^2\sigma_2^2 + \frac{N_i - n_i}{N_i - 1} w_i^2 w_i^1 (\mu_1 - \mu_2)^2\right). \tag{4}$$

4. Мінімаксні коефіцієнти

Як бачимо, вирази для дисперсій (3) та (4) залежать від невідомих параметрів моделей, а тому оптимальні коефіцієнти $a_1, a_2, \ldots, a_K$ не можуть бути визначені за вибіркою.

У [5] для оцінювання функцій розподілу компонентів сумішей за незалежними спостереженнями використовуються мінімаксні коефіцієнти. Вони задовольняють умови (2) та мінімізують наступний вираз:

$$\sum_{i=1}^{K} a_i^2 \to \min .$$

Застосування методу множників Лагранжа показує, що в невиродженому випадку ($\bar{r}_2\bar{r}_0 - \bar{r}_1^2 \neq 0$) мінімаксні коефіцієнти мають вигляд:

$$\bar{a}_i = \frac{(1-\bar{r}_1)w_i^1 + \bar{r}_2 - \bar{r}_1}{\bar{r}_2 - \bar{r}_1^2}, \qquad i = 1, 2, \ldots, K, \tag{5}$$

де $\bar{r}_j$ — моменти концентрацій w_i:

$$\bar{r}_j = \frac{1}{K}\sum_{i=1}^{K} \left(w_i^1\right)^j, \qquad j = 1, 2.$$

Вироджений випадок відповідає рівності усіх w_i. Тоді модель стає неідентифікованою, а оцінювання неможливим.

Хоча мінімаксні коефіцієнти не забезпечують мінімум дисперсій у (3) або (4), вони не є надто поганими. У наступному розділі показано консистентність та асимптотичну нормальність оцінок з мінімаксними коефіцієнтами.

5. Асимптотична поведінка оцінок

Асимптотичну поведінку будемо досліджувати для безумовного випадку при $K \to \infty$. При цьому розміри груп N_i, концентрації об'єктів обох класів w_i^l, $l = 1, 2$ та обсяги вибірок з груп n_i вважаємо фіксованими послідовностями, визначеними для всіх $i \geq 1$.

Будемо використовувати оцінку вигляду (1). Таким чином, оцінка μ_1 з мінімаксними коефіцієнтами (5) має вигляд:

$$\bar{\mu}_{1,K} = \hat{m}_1(\bar{a}_K) = \frac{1}{K}\sum_{i=1}^{K} \bar{a}_{i,K} T_i.$$

Наступна теорема показує консистентність та асимптотичну нормальність цієї оцінки.

Теорема 3. *Нехай існують $C > 0$, $M \geq 1$, такі що $\bar{r}_{2,K} - \bar{r}_{1,K}^2 > C$ для всіх K та $n_i \leq M$ для всіх $i \geq 1$. Тоді оцінка $\bar{\mu}_{1,K}$ є консистентною, а розподіл величин*

$$\frac{1}{\sqrt{\mathsf{D}\,\bar{\mu}_{1,K}}}(\bar{\mu}_{1,K} - \mu_1)$$

слабко збігаються при $K \to \infty$ до стандартного нормального розподілу.

Для визначення найкращих коефіцієнтів a_i перепишемо вираз для $\mathsf{D}\,\hat{m}_1(a)$ наступним чином:

$$\mathsf{D}\,\hat{m}_1(a) = \frac{1}{K^2}\sum_{i=1}^{K} a_i^2 d_i,$$

$$d_i = \frac{1}{n_i}\left(w_i^1\sigma_1^2 + w_i^2\sigma_2^2 + \frac{N_i - n_i}{N_i - 1} w_i^2 w_i^1 (\mu_1 - \mu_2)^2\right), \qquad i = 1, 2, \ldots, K.$$

Враховуючи рівняння незміщенності (2), мінімізуємо вираз для дисперсії $\mathsf{D}\,\hat{m}_1(a)$ за коефіцієнтами $a_1, a_2, \ldots, a_K$. Застосування методу множників Лагранжа показує, що в невиродженому випадку ($\tilde{r}_{2,K}\tilde{r}_{0,K} - \tilde{r}_{1,K}^2 \neq 0$) найкращі ваги розраховуються за формулою

$$\tilde{a}_{i,K} = \frac{(\tilde{r}_{0,K} - \tilde{r}_{1,K})w_i^1 + \tilde{r}_{2,K} - \tilde{r}_{1,K}}{d_i\big(\tilde{r}_{2,K}\tilde{r}_{0,K} - \tilde{r}_{1,K}^2\big)}, \qquad i = 1, 2, \ldots, K, \tag{6}$$

де $\tilde{r}_{j,K}$ — зважені моменти концентрацій w_i:

$$\tilde{r}_{j,K} = \frac{1}{K}\sum_{i=1}^{K}\frac{\left(w_i^1\right)^j}{d_i}, \qquad j = 0, 1, 2.$$

Отже найкраща незміщена оцінка вигляду (1) має вигляд:

$$\tilde{\mu}_{1,K} = \hat{m}_1(\tilde{a}_K) = \frac{1}{K}\sum_{i=1}^{K}\tilde{a}_{i,K}T_i.$$

При цьому мінімальне значення дисперсії дорівнює

$$\mathsf{D}\,\tilde{\mu}_{1,K} = \frac{1}{K}\frac{\tilde{r}_{2,K} - 2\tilde{r}_{1,K} + \tilde{r}_{0,K}}{\tilde{r}_{2,K}\tilde{r}_{0,K} - \tilde{r}_{1,K}^2}.$$

Наступна теорема показує, що отримана оцінка $\tilde{\mu}_{1,K}$ є асимптотично нормальною.

Теорема 4. *В умовах теореми 3 оцінка $\tilde{\mu}_{1,K}$ є консистентною, а розподіл величин*

$$L_K := \frac{1}{\sqrt{\mathsf{D}\,\tilde{\mu}_{1,K}}}(\tilde{\mu}_{1,K} - \mu_1)$$

слабко збігаються при $K \to \infty$ до стандартного нормального розподілу.

Проте ваги $\tilde{a}$ неможливо застосувати, адже величини d_i залежать від вектору невідомих параметрів $\gamma = (\mu_1, \mu_2, \sigma_1, \sigma_2)$. Тому, за допомогою мінімаксних ваг $\bar{a}_i$ спочатку побудуємо пілотну оцінку $\bar{\gamma}_K = (\bar{\mu}_{1,K}, \bar{\mu}_{2,K}, \bar{\sigma}_{1,K}, \bar{\sigma}_{2,K})$:

$$\bar{\mu}_{l,K} = \frac{1}{K}\sum_{i=1}^{K}\bar{a}_{i,K}^l\frac{1}{n_i}\sum_{j=1}^{n_i}Y_{ij}, \qquad l = 1, 2, \tag{7}$$

$$\bar{\sigma}_{l,K}^2 = \frac{1}{K}\sum_{i=1}^{K}\bar{a}_{i,K}^l\frac{1}{n_i}\sum_{j=1}^{n_i}Y_{ij}^2 - \bar{\mu}_{l,K}^2, \qquad l = 1, 2. \tag{8}$$

Далі, можна побудувати оцінки дисперсій за групами

$$\bar{d}_{i,K} = \frac{1}{n_i}\left(w_i^1\bar{\sigma}_{1,K}^2 + w_i^2\bar{\sigma}_{2,K}^2 + \frac{N_i - n_i}{N_i - 1}w_i^1w_i^2(\bar{\mu}_{1,K} - \bar{\mu}_{2,K})^2\right), \qquad i = 1, 2, \ldots, K$$

та зважених моментів

$$\bar{r}_{i,K} = \frac{1}{K}\sum_{i=1}^{K}\frac{(w_i^1)^j}{\bar{d}_{i,K}}, \qquad i = 1, 2, \ldots, K.$$

Тепер можна розрахувати адаптивні ваги $\hat{a}_{i,K}$:

$$\hat{a}_{i,K} = \frac{(\bar{r}_{0,K} - \bar{r}_{1,K})w_i^1 + \bar{r}_{2,K} - \bar{r}_{1,K}}{\bar{r}_{0,K}\bar{r}_{2,K} - \bar{r}_{1,K}^2}, \qquad i = 1, 2, \ldots, K.$$

Отже, адаптивна оцінка середнього має вигляд:

$$\hat{\mu}_{1,K} = \hat{m}_1(\hat{a}_K) = \frac{1}{K}\sum_{i=1}^{K}\hat{a}_{i,K}T_i.$$

Покажемо, що при виконанні умов теореми 3 асимптотична поведінка адаптивної оцінки така ж сама, як асимптотична поведінка оцінки з найкращими вагами.

Розглянемо вагові коефіцієнти $\tilde{a}_{i,K}$, $i = 1, 2, \ldots, K$, як функції від невідомих параметрів γ. Сформулюємо частковий випадок леми 3.2.1 з [5] для $g(x) = x$ та $m = 2$:

Твердження 1. *Нехай для деякого $C > 0$, $\bar{r}_{2,K} - \bar{r}_{1,K}^2 > C$ для всіх K. Тоді*

$$\forall \delta > 0 : \quad \sup_{K \in \mathbb{N}} \mathsf{P}\left(\tau_\varepsilon(K) > \delta\right) \to 0, \qquad \varepsilon \to \infty,$$

$$\tau_\varepsilon(K) = \sup_{\alpha : |\alpha - \gamma| < \varepsilon} \frac{1}{\sqrt{\mathsf{D}\,\tilde{\mu}_{1,K}}} |\hat{m}_1(\tilde{a}_K(\alpha)) - \hat{m}_1(\tilde{a}_K(\gamma))|.$$

Тепер перейдемо до теореми про асимптотичну нормальність адаптивної оцінки.

Теорема 5. *В умовах теореми 3 оцінка $\hat{\mu}_{1,K}$ є консистентною, а розподіл величин*

$$Z_K := \frac{1}{\sqrt{\mathsf{D}\,\tilde{\mu}_{1,K}}} (\hat{\mu}_{1,K} - \mu_1)$$

слабко збігаються при $K \to \infty$ до стандартного нормального розподілу.

Доведення. Розглянемо випадкові величини L_K, визначені у теоремі 4. Для будь–якого $\delta > 0$

$$J_K := \mathsf{P}\left(|Z_K - L_K| > \delta\right) \le \mathsf{P}\left(\tau_\varepsilon(K) > \delta\right) + \mathsf{P}\left(|\bar{\gamma}_K - \gamma| > \varepsilon\right).$$

Зафіксуємо довільне $\lambda > 0$. За твердженням 1 можна обрати $\varepsilon > 0$ так, що для всіх K

$$\mathsf{P}\left(\tau_\varepsilon(K) > \delta\right) < \lambda.$$

З теореми 3 випливає, що оцінки середнього та дисперсії з мінімаксними коефіцієнтами (7)–(8) є консистентними, отже $\bar{\gamma}_K \to \gamma$ за ймовірністю при $K \to \infty$. Тому, при достатньо великому K, маємо

$$\mathsf{P}\left(|\bar{\gamma}_K - \gamma| > \varepsilon\right) < \lambda.$$

Тому для таких K виконується $J_K < 2\lambda$. Внаслідок довільності $\lambda > 0$ звідси випливає, що $J_K \to 0$, тобто $|Z_K - L_K| \to 0$ за ймовірністю.

За теоремою 4, розподіл L_K слабко збігається до стандартного нормального розподілу. Отже, розподіл Z_K також збігається до стандартного нормального. □

6. Висновки

Ця стаття узагальнює техніку дослідження сумішей зі змінними концентраціями на випадок залежних спостережень.

Застосування техніки адаптивного оцінювання дозволяє покращити асимптотичні властивості оцінок. Проте при невеликих K додаткова похибка при оцінюванні γ може призвести до погіршення якості адаптивних оцінок у порівнянні з мінімаксними. Отже застосування адаптивних оцінок у конкретній ситуації потребує окремого дослідження.

В якості подальшого розвитку можна зазначити узагальнення розвинутої техніки для сумішей з більше ніж двох компонентів.

7. Доведення теорем

Доведення теореми 1. Для обчислення дисперсії запишемо

$$\mathsf{D}_{\mathbf{p}}\,\hat{m}_1(a) = \mathsf{D}_{\mathbf{p}}\,\frac{1}{K}\sum_{i=1}^{K} a_i T_i = \frac{1}{K^2}\sum_{i=1}^{K} a_i^2\,\mathsf{D}_{\mathbf{p}}\,T_i + \frac{1}{K^2}\sum_{\substack{i,j=1\\ i\ne j}}^{K} a_i a_j \operatorname{Cov}(T_i, T_j). \tag{9}$$

Позначимо усереднення лише за вибірками за допомогою індексу $\mathbf{s}$. Тоді матимемо:

$$\mathsf{D}_{\mathbf{p}}\,T_i = \mathsf{E}_{\mathbf{p}}\,\mathsf{E}_{\mathbf{s}}(T_i - \mathsf{E}_{\mathbf{p}}\,T_i)^2 = \mathsf{E}_{\mathbf{p}}\,\mathsf{D}_{\mathbf{s}}\,T_i + \mathsf{E}_{\mathbf{p}}(\mathsf{E}_{\mathbf{s}}\,T_i - \mathsf{E}_{\mathbf{p}}\,T_i)^2.$$

Позначимо множину характеристик елементів l-го класу $Z_1^l, Z_2^l, \ldots, Z_{N_{(l)}}^l$, $l = 1, 2$, а також $\tilde{m}_i = \mathsf{E}_{\mathbf{s}} T_i$ та $\bar{m}_i = \mathsf{E}_{\mathbf{p}} T_i$, $i = 1, 2, \ldots, K$. З теорії вибіркових обстежень отримуємо:

$$\tilde{m}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij}, \qquad \mathsf{D}_{\mathbf{s}} T_i = \frac{N_i - n_i}{N_i(N_i-1)n_i}\sum_{j=1}^{N_i}(X_{ij} - \tilde{m}_i)^2, \qquad i = 1, 2, \ldots, K.$$

Доведемо наступну лему:

Лема 1. *Нехай $\{Z_1, Z_2\}$ — проста випадкова вибірка без повернення із сукупності характеристик об'єктів l-того класу $Z_1^l, Z_2^l, \ldots, Z_{N_{(l)}}^l$, $l = 1, 2$. Тоді виконуються наступні рівності:*

1. $\mathrm{Cov}\,(Z_1, Z_2) = -s_l^2/N_{(l)}$.
2. $\mathrm{Cov}\,(Y_{i1}, Y_{j1}) = -w_i^1 w_j^1 s_1^2/N_{(1)} - w_i^2 w_j^2 s_2^2/N_{(2)}$.
3. $\mathsf{E}_{\mathbf{p}}(\tilde{m}_i - \bar{m}_i)^2 = \frac{1}{N_i}\left(w_i^1 s_1^2 + w_i^2 s_2^2\right) - \left(w_i^1\right)^2 s_1^2/N_{(1)} - \left(w_i^2\right)^2 s_2^2/N_{(2)}$.

Доведення. Справедливість тверджень випливає з наступних рівностей:

$$\begin{aligned}
\mathrm{Cov}\,(Z_1, Z_2) &= Z_1 Z_2 - \mathsf{E}_{\mathbf{p}} Z_1\, \mathsf{E}_{\mathbf{p}} Z_2 \\
&= \frac{1}{N_{(l)}(N_{(l)}-1)} \sum_{i,j=1, i\neq j}^{N_{(l)}} Z_i^l Z_j^l - m_l^2 \\
&= N_{(l)}(N_{(l)}-1)\left(\left(N_{(l)}\right)^2 m_l^2 - \sum_{i=1}^{N_{(l)}} \left(Z_i^l\right)^2\right) - m_l^2 \\
&= N_{(l)}(N_{(l)}-1)\left(\left(N_{(l)}\right)^2 m_l^2 - (N_{(l)}-1)s_1^2 - N_{(l)} m_l^2\right) - m_l^2 \\
&= -\frac{s_l^2}{N_{(l)}},
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}\,(Y_{i1}, Y_{j1}) &= \mathrm{Cov}\left(Y_{i1}\,\mathbf{1}_{\{C_{i1}=1\}} + Y_{i1}\,\mathbf{1}_{\{C_{i1}=2\}}, Y_{j1}\,\mathbf{1}_{\{C_{j1}=1\}} + Y_{j1}\,\mathbf{1}_{\{C_{j1}=2\}}\right) \\
&= \mathrm{Cov}\left(Y_{i1}\,\mathbf{1}_{\{C_{i1}=1\}}, Y_{j1}\,\mathbf{1}_{\{C_{j1}=1\}}\right) + \mathrm{Cov}\left(Y_{i1}\,\mathbf{1}_{\{C_{i1}=2\}}, Y_{j1}\,\mathbf{1}_{\{C_{j1}=2\}}\right) \\
&= -w_i^1 w_j^1 \frac{s_1^2}{N_{(1)}} - w_i^2 w_j^2 \frac{s_2^2}{N_{(2)}},
\end{aligned}$$

$$\begin{aligned}
\mathsf{E}_{\mathbf{p}}(\tilde{m}_i - \bar{m}_i)^2 &= \mathsf{E}_{\mathbf{p}}\left(\frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij} - \bar{m}_i\right)^2 \\
&= \mathsf{E}_{\mathbf{p}}\left(w_i^1\left(\sum_{j=1}^{N_i^1} X_{ij} - m_1\right) + w_i^2\left(\sum_{j=N_i^1+1}^{N_i} X_{ij} - m_2\right)\right)^2 \\
&= \left(w_i^1\right)^2 \mathsf{E}_{\mathbf{p}}\left(\sum_{j=1}^{N_i^1} X_{ij} - m_1\right)^2 + \left(w_i^2\right)^2 \mathsf{E}_{\mathbf{p}}\left(\sum_{j=N_i^1+1}^{N_i} X_{ij} - m_2\right)^2 \\
&= \left(w_i^1\right)^2\left(1 - \frac{N_i^1}{N_{(1)}}\right)\frac{s_1^2}{N_i^1} + \left(w_i^2\right)^2\left(1 - \frac{N_i^2}{N_{(2)}}\right)\frac{s_2^2}{N_i^2} \\
&= w_i^1 s_1^2 + w_i^2 s_2^2 - \left(w_i^1\right)^2 \frac{s_1^2}{N_{(1)}} - \left(w_i^2\right)^2 \frac{s_2^2}{N_{(2)}}. \quad \square
\end{aligned}$$

Отже, для величини $\mathsf{D}_{\mathbf{p}}\,T_i$ маємо:

$$\begin{aligned}
\mathsf{D}_{\mathbf{p}}\,T_i &= \frac{N_i-n_i}{N_i(N_i-1)n_i}\sum_{j=1}^{N_i}\mathsf{E}_{\mathbf{p}}(X_{ij}-\tilde m_i)^2+\mathsf{E}_{\mathbf{p}}(\bar m_i-\tilde m_i)^2\\
&= \frac{N_i-n_i}{N_i(N_i-1)n_i}\sum_{j=1}^{N_i}\left(\mathsf{E}_{\mathbf{p}}(X_{ij}-\bar m_i)^2-\sum_{j=1}^{N_i}\mathsf{E}_{\mathbf{p}}(\bar m_i-\tilde m_i)^2\right)+\mathsf{E}_{\mathbf{p}}(\bar m_i-\tilde m_i)^2\\
&= \frac{N_i-n_i}{N_i(N_i-1)n_i}\sum_{j=1}^{N_i}\mathsf{E}_{\mathbf{p}}(X_{ij}-\bar m_i)^2+\frac{N_i(n_i-1)}{(N_i-1)n_i}\,\mathsf{E}_{\mathbf{p}}(\bar m_i-\tilde m_i)^2\\
&= \frac{N_i-n_i}{(N_i-1)n_i}\left(w_i^1\,\mathsf{E}_{\mathbf{p}}\left((Y_{i1}-\bar m_i)^2\mid C_{i1}=1\right)+w_i^2\,\mathsf{E}_{\mathbf{p}}\left((Y_{i1}-\bar m_i)^2\mid C_{i1}=2\right)\right)\\
&\quad+\frac{N_i(n_i-1)}{(N_i-1)n_i}\,\mathsf{E}_{\mathbf{p}}(\bar m_i-\tilde m_i)^2\\
&= \frac{N_i-n_i}{(N_i-1)n_i}\Big[w_i^1\left(\mathsf{E}_{\mathbf{p}}\left((Y_{i1}-m_1)^2\mid C_{i1}=1\right)+\mathsf{E}_{\mathbf{p}}(\bar m_i-m_1)^2\right)\\
&\qquad\qquad+w_i^2\left(\mathsf{E}_{\mathbf{p}}\left((Y_{i1}-m_2)^2\mid C_{i1}=2\right)+\mathsf{E}_{\mathbf{p}}(\bar m_i-m_2)\right)\Big]\\
&\quad+\frac{N_i(n_i-1)}{(N_i-1)n_i}\,\mathsf{E}_{\mathbf{p}}(\bar m_i-\tilde m_i)^2\\
&= \frac{N_i-n_i}{(N_i-1)n_i}\Bigg(w_i^1\frac{N_{(1)}-1}{N_{(1)}}s_1^2+w_i^2\frac{N_{(2)}-1}{N_{(2)}}s_2^2\\
&\qquad\qquad+w_i^1\left(w_i^2\right)^2(m_1-m_2)^2+w_i^2\left(w_i^1\right)^2(m_1-m_2)^2\Bigg)\\
&\quad+\frac{n_i-1}{(N_i-1)n_i}\left(w_i^1s_1^2+w_i^2s_2^2\right)-\frac{N_i(n_i-1)}{(N_i-1)n_i}\left(\left(w_i^1\right)^2\frac{s_1^2}{N_{(1)}}+\left(w_i^2\right)^2\frac{s_2^2}{N_{(2)}}\right)\\
&= \frac{1}{n_i}\left(w_i^1s_1^2+w_i^2s_2^2+\frac{N_i-n_i}{N_i-1}\left(w_i^2w_i^1(m_1-m_2)^2-w_i^1\frac{s_1^2}{N_{(1)}}-w_i^2\frac{s_2^2}{N_{(2)}}\right)\right)\\
&\quad-\frac{N_i(n_i-1)}{(N_i-1)n_i}\left(\left(w_i^1\right)^2\frac{s_1^2}{N_{(1)}}+\left(w_i^2\right)^2\frac{s_2^2}{N_{(2)}}\right).
\end{aligned}$$

Суму з коваріаціями у (9) можна перетворити наступним чином:

$$\begin{aligned}
\frac{1}{K^2}\sum_{\substack{i,j=1\\ i\neq j}}^{K}a_ia_j\,\mathrm{Cov}\,(T_i,T_j) &= \frac{1}{K^2}\sum_{\substack{i,j=1\\ i\neq j}}^{K}a_ia_j\,\mathrm{Cov}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}Y_{ik},\frac{1}{n_i}\sum_{j=1}^{n_i}Y_{jk}\right)\\
&= \frac{1}{K^2}\sum_{\substack{i,j=1\\ i\neq j}}^{K}a_ia_j\,\mathrm{Cov}\,(Y_{i1},Y_{j1})\\
&= -\frac{1}{K^2}\sum_{\substack{i,j=1\\ i\neq j}}^{K}a_ia_j\left(w_i^1w_j^1\frac{s_1^2}{N_{(1)}}-w_i^2w_j^2\frac{s_2^2}{N_{(2)}}\right)\\
&= -\frac{s_1^2}{N_{(1)}}+\frac{1}{K^2}\sum_{i=1}^{K}a_i^2\left(\left(w_i^1\right)^2\frac{s_1^2}{N_{(1)}}-\left(w_i^2\right)^2\frac{s_2^2}{N_{(2)}}\right).
\end{aligned}$$

Остаточно, вираз (9) можна переписати так:

$$\frac{1}{K^2}\sum_{i=1}^{K} a_i^2 \,\mathsf{D}_{\mathbf{p}}\, T_i + \frac{1}{K^2}\sum_{\substack{i,j=1\\ i\neq j}}^{K} a_i a_j \operatorname{Cov}(T_i,T_j)$$

$$= \frac{1}{K^2}\sum_{i=1}^{K} a_i^2 \Bigg[\frac{1}{n_i}\Bigg(w_i^1 s_1^2 + w_i^2 s_2^2 + \frac{N_i-n_i}{N_i-1}\left(w_i^2 w_i^1 (m_1-m_2)^2 - w_i^1\frac{s_1^2}{N_{(1)}} - w_i^2\frac{s_2^2}{N_{(2)}}\right)\Bigg) - \frac{N_i(n_i-1)}{(N_i-1)n_i}\left(\left(w_i^1\right)^2\frac{s_1^2}{N_{(1)}} + \left(w_i^2\right)^2\frac{s_2^2}{N_{(2)}}\right)\Bigg]$$

$$+ \frac{1}{K^2}\sum_{i=1}^{K} a_i^2 \left(\left(w_i^1\right)^2\frac{s_1^2}{N_{(1)}} - \left(w_i^2\right)^2\frac{s_2^2}{N_{(2)}}\right) - \frac{s_1^2}{N_{(1)}}$$

$$= \frac{1}{K^2}\sum_{i=1}^{K} a_i^2 \Bigg[\frac{1}{n_i}\Bigg(w_i^1 s_1^2 + w_i^2 s_2^2 + \frac{N_i-n_i}{N_i-1}\left(w_i^2 w_i^1 (m_1-m_2)^2 - w_i^1\frac{s_1^2}{N_{(1)}} - w_i^2\frac{s_2^2}{N_{(2)}}\right)\Bigg) + \frac{N_i-n_i}{(N_i-1)n_i}\left(\left(w_i^1\right)^2\frac{s_1^2}{N_{(1)}} + \left(w_i^2\right)^2\frac{s_2^2}{N_{(2)}}\right)\Bigg] - \frac{s_1^2}{N_{(1)}}$$

$$= \frac{1}{K^2}\sum_{i=1}^{K} \frac{a_i^2}{n_i}\left(w_i^1 s_1^2 + w_i^2 s_2^2 + \frac{N_i-n_i}{N_i-1} w_i^2 w_i^1\left((m_1-m_2)^2 - \frac{s_1^2}{N_{(1)}} - \frac{s_2^2}{N_{(2)}}\right)\right) - \frac{s_1^2}{N_{(1)}}. \quad \square$$

Доведення теореми 2. Зауважимо, що безумовне математичне сподівання можна рахувати через умовне математичне сподівання при фіксованій популяції, так що виконується

$$\mathsf{D}\,\hat{m}_1(a) = \mathsf{E}(\hat{m}_1(a) - m_1)^2 + \mathsf{E}(m_1-\mu_1)^2 = \mathsf{E}\,\mathsf{D}_{\mathbf{p}}\,\hat{m}_1(a) + \mathsf{D}\,m_1.$$

Далі, з означення популяційних середніх та дисперсій випливають рівності:

$$\mathsf{E}\,m_l = \mu_l, \qquad \mathsf{D}\,m_l = \frac{\sigma_1^2}{N_{(1)}}, \qquad \mathsf{E}\,s_l^2 = \sigma_l^2, \qquad l = 1,2.$$

Обчислимо також безумовне математичне сподівання від квадрату різниці популяційних середніх:

$$\mathsf{E}(m_1-m_2)^2 = \mathsf{E}\left(\frac{1}{N_{(1)}}\sum_{i=1}^{N_{(1)}} X_i^1 - \frac{1}{N_{(2)}}\sum_{i=1}^{N_{(2)}} X_i^2\right)^2$$

$$= (\mu_1-\mu_2)^2 + \mathsf{E}\left(\frac{1}{N_{(1)}}\sum_{i=1}^{N_{(1)}} X_i^1 - \mu_1\right)^2 + \mathsf{E}\left(\frac{1}{N_{(2)}}\sum_{i=1}^{N_{(2)}} X_i^2 - \mu_2\right)^2$$

$$= (\mu_1-\mu_2)^2 + \mathsf{D}\,\frac{1}{N_{(1)}}\sum_{i=1}^{N_{(1)}} X_i^1 + \mathsf{D}\,\frac{1}{N_{(2)}}\sum_{i=1}^{N_{(2)}} X_i^2$$

$$= (\mu_1-\mu_2)^2 + \frac{\sigma_1^2}{N_{(1)}} + \frac{\sigma_2^2}{N_{(2)}}.$$

Тепер ми можемо обчислити безумовну дисперсію оцінки $\hat{m}_1(a)$:

$$\begin{aligned}\mathsf{D}\,\hat{m}_1(a) = \mathsf{E}\Bigg[\frac{1}{K^2}\sum_{i=1}^{K}\frac{a_i^2}{n_i}\Bigg(&w_i^1 s_1^2 + w_i^2 s_2^2 + \frac{N_i - n_i}{N_i - 1}w_i^2 w_i^1\\ &\times\left((m_1 - m_2)^2 - \frac{s_1^2}{N_{(1)}} - \frac{s_2^2}{N_{(2)}}\right)\Bigg) - \frac{s_1^2}{N_{(1)}}\Bigg] + \frac{\sigma_1^2}{N_{(1)}}\\ = \frac{1}{K^2}\sum_{i=1}^{K}\frac{a_i^2}{n_i}&\left(w_i^1\sigma_1^2 + w_i^2\sigma_2^2 + \frac{N_i - n_i}{N_i - 1}w_i^2 w_i^1(\mu_1 - \mu_2)^2\right). \quad \square\end{aligned}$$

Доведення теореми 3. Зауважимо, що з умови теореми випливає обмеженість коефіцієнтів $a_{i,K}$:

$$\begin{aligned}|\bar{a}_{i,K}| &= \left|\frac{(1-\bar{r}_{1,K})w_i^1 + \bar{r}_{2,K} - \bar{r}_{1,K}}{\bar{r}_{2,K} - \bar{r}_{1,K}^2}\right| \le \frac{1 - \bar{r}_{1,K} + |\bar{r}_{2,K} - \bar{r}_{1,K}|}{\bar{r}_{2,K} - \bar{r}_{1,K}^2}\\ &= \frac{1 - \bar{r}_{2,K}}{\bar{r}_{2,K} - \bar{r}_{1,K}^2} \le \frac{1}{C}, \qquad i = 1, 2, \ldots, K.\end{aligned}$$

Для доведення асимптотичної нормальності оцінок з мінімаксними коефіцієнтами використаємо наступну граничну теорему [2].

Твердження 2 (Центральна гранична теорема Ліндеберга для стандартних серій)**.** *Нехай для послідовності ζ_{ni} виконуються наступні умови:*

(1) *Незалежність. Випадкові величини в кожній серії $\{z_{n1}, z_{n2}, \ldots, z_{nn}\}$ незалежні в сукупності.*
(2) *Центрованість.* $\mathsf{E}\,\zeta_{ni} = 0$, $n \ge 1$, $i = 1, 2, \ldots, n$.
(3) *Нормованість.* $\exists\,\mathsf{D}\,\zeta_{ni} = \sigma_{ni}^2$ *і* $\sum_{i=1}^{n}\mathsf{D}\,\zeta_{ni} = 1$.
(4) *Умова Ліндеберга* $L_n(\varepsilon) := \sum_{i=1}^{n}\mathsf{E}\,\zeta_{ni}^2\,\mathbf{1}_{\{|\zeta_{ni}|>\varepsilon\}} \to 0$, $n \to \infty$, $\forall\varepsilon > 0$.

Тоді суми $S_n = \sum_{i=1}^{n}\zeta_{ni}$ слабко збігаються до стандартного нормального розподілу при $n \to \infty$.

Позначимо

$$\xi_{i,K} = \frac{\bar{a}_{i,K}}{K\sqrt{\mathsf{D}\,\bar{\mu}_{1,K}}}(T_i - \mathsf{E}\,T_i).$$

Перевіримо виконання умов центральної граничної теореми Ліндеберга для стандартних серій для

$$S_K = \frac{\bar{\mu}_{1,K} - \mu_1}{\sqrt{\mathsf{D}\,\bar{\mu}_{1,K}}} = \sum_{i=1}^{K}\xi_{i,K}.$$

Умова (1) твердження 2 виконується згідно постановки задачі, оскільки формування та відбір з груп відбувається незалежним чином. Умови (2-3) твердження 2 випливають з визначення $\xi_{i,K}$. Щоб оцінити $L_n(\varepsilon)$, оцінимо дисперсію оцінки $\hat{m}_1(\bar{a})$:

$$\begin{aligned}\mathsf{D}\,\bar{\mu}_{1,K} &= \frac{1}{K^2}\sum_{i=1}^{K}\frac{\bar{a}_{i,K}^2}{n_i}\left(w_i^1\sigma_1^2 + w_i^2\sigma_2^2 + \frac{N_i - n_i}{N_i - 1}w_i^2 w_i^1(\mu_1 - \mu_2)^2\right)\\ &\ge \frac{\min(\sigma_1^2, \sigma_2^2)}{MK^2}\sum_{i=1}^{K}\bar{a}_{i,K}^2.\end{aligned}$$

Далі, використовуючи вираз для коефіцієнтів $\bar{a}_{i,K}$ та рівняння (2), запишемо:

$$\frac{1}{K}\sum_{i=1}^{K}\bar{a}_{i,K}^2 = \frac{1}{K}\sum_{i=1}^{K}\frac{(1-\bar{r}_{1,K})w_i^1 a_{i,K} + (\bar{r}_{2,K}-\bar{r}_{1,K})a_{i,K}}{\bar{r}_{2,K}-\bar{r}_{1,K}^2} = \frac{\bar{r}_{2,K}-2\bar{r}_{1,K}+1}{\bar{r}_{2,K}-\bar{r}_{1,K}^2}$$
$$= 1 + \frac{(\bar{r}_{1,K}-1)^2}{\bar{r}_{2,K}-\bar{r}_{1,K}^2} \geq 1.$$

Отже, для деякої константи $C_1 > 0$ виконується

$$\mathsf{D}\,\hat{\mu}_{1,K} \geq \frac{C_1}{K}.$$

Тому

$$\mathsf{E}\,\xi_{i,K}^2\,\mathbf{1}_{\{|\xi_{i,K}|>\tau\}} = \mathsf{E}\,\frac{\bar{a}_{i,K}^2(T_i-\mathsf{E}\,T_i)^2}{K^2\,\mathsf{D}\,\bar{\mu}_{1,K}}\,\mathbf{1}_{\left\{|T_i-\mathsf{E}\,T_i|>\frac{\tau K\sqrt{\mathsf{D}\,\bar{\mu}_{1,K}}}{|\bar{a}_{i,K}|}\right\}}$$
$$\leq \frac{1}{C^2C_1K}\,\mathsf{E}(T_i-\mathsf{E}\,T_i)^2\,\mathbf{1}_{\left\{|T_i-\mathsf{E}\,T_i|>C\sqrt{C_1K}\varepsilon\right\}}.$$

Нехай $Z_1^1, Z_2^1, \ldots, Z_M^1$ та $Z_1^2, Z_2^2, \ldots, Z_M^2$ — незалежні випадкові величини, розподілені як характеристики об'єктів першого та другого класів відповідно. Розглянемо центровану суму їх модулів:

$$\eta = \sum_{j=1}^{M}|Z_j^1-\mu_1| + \sum_{j=1}^{M}|Z_j^2-\mu_2|.$$

Оскільки T_i — це середнє значення не більше ніж M характеристик об'єктів першого та другого класів, то

$$\mathsf{E}(T_i-\mathsf{E}\,T_i)^2\,\mathbf{1}_{\left\{|T_i-\mathsf{E}\,T_i|>C\sqrt{C_1K}\varepsilon\right\}} \leq \mathsf{E}\,\eta^2\,\mathbf{1}_{\left\{\eta>C\sqrt{C_1K}\varepsilon\right\}}.$$

Отже, маємо виконання умови Ліндеберга:

$$L_n(\varepsilon) \leq \frac{1}{C^2C_1}\,\mathsf{E}\,\eta^2\,\mathbf{1}_{\left\{\eta>C\sqrt{C_1K}\varepsilon\right\}} \to 0, \qquad K\to\infty,$$

оскільки величина η має скінченний другий момент.

Дисперсія оцінки $\bar{\mu}_{1,K}$ прямує до нуля:

$$\mathsf{D}\,\bar{\mu}_{1,K} \leq \frac{\sigma_1^2+\sigma_2^2+(\mu_1-\mu_2)^2}{K^2}\sum_{i=1}^{K}\bar{a}_{i,K}^2 \leq \frac{\sigma_1^2+\sigma_2^2+(\mu_1-\mu_2)^2}{C^2K} \to 0, \qquad K\to\infty.$$

Тому з асимптотичної нормальності оцінки $\bar{\mu}_{1,K}$ випливає її консистентність. □

Доведення теореми 4. Відмітимо, що з вигляду d_i випливає існування констант $0 < C_1, C_2 < \infty$, що $C_1 < d_i < C_2$ для всіх i та K. Розпишемо умову на коефіцієнти $\bar{r}_{1,K}$ та $\bar{r}_{2,K}$ наступним чином:

$$C < \bar{r}_{2,K}-\bar{r}_{1,K}^2 = \frac{1}{K}\sum_{i=1}^{K}(w_i-\bar{r}_{1,K})^2 = \frac{1}{K}\sum_{|w_i|\geq\varepsilon}(w_i-\bar{r}_{1,K})^2 + \frac{1}{K}\sum_{|w_i|<\varepsilon}(w_i-\bar{r}_{1,K})^2$$
$$\leq \frac{1}{K}\sum_{|w_i|\geq\varepsilon}1 + \frac{1}{K}\sum_{|w_i|<\varepsilon}\varepsilon^2 \leq \frac{h_1}{K}+\varepsilon^2,$$

де h_1 — це кількість груп з концентраціями w_i, що знаходяться на відстані більше ніж ε від середнього значення. Отже, маємо наступну умову на h_1:

$$h_1 > K\left(C-\varepsilon^2\right).$$

Зрозуміло, що принаймні половина таких концентрацій є меншою або більшою за середнє значення. Припустимо, для визначеності, що для кількості концентрацій h_2, більших за $\bar r_{1,K}+\varepsilon$, виконується нерівність $h_2 > h_1/2 = K(C-\varepsilon^2)/2$. Оцінимо також кількість груп h_3 з концентраціями меншими за $\bar r_{1,K}$:

$$0=\sum_{i=1}^{K}(w_i-\bar r_{1,K})=\sum_{w_i\le \bar r_{1,K}}(w_i-\bar r_{1,K})+\sum_{w_i>\bar r_{1,K}}(w_i-\bar r_{1,K})\ge -h_3+h_2\varepsilon.$$

Отже, $h_3 \ge h_2\varepsilon = K\varepsilon(C-\varepsilon^2)/2$, та маємо наступну оцінку:

$$\tilde r_{2,K}\tilde r_{0,K}-\tilde r_{1,K}^2=\tilde r_{0,K}\frac{1}{K}\sum_{i=1}^{K}\frac{(w_i-\tilde r_{1,K}/\tilde r_{0,K})^2}{d_i}\ge\frac{1}{KC_2^2}\sum_{i=1}^{K}(w_i-\tilde r_{1,K}/\tilde r_{0,K})^2$$
$$\ge\frac{h_3}{4KC_2^2}\varepsilon^2=\frac{\varepsilon^3(C-\varepsilon^2)}{8C_2^2}.$$

Виберемо $\varepsilon = C/2$, тоді

$$\tilde r_{2,K}\tilde r_{0,K}-\tilde r_{1,K}^2\ge\frac{C^5}{256C_2^2}.$$

Звідси випливає обмеженість найкращих вагових коефіцієнтів:

$$|\tilde a_{i,K}|=\frac{|(\tilde r_{0,K}-\tilde r_{1,K})w_i^1+\tilde r_{2,K}-\tilde r_{1,K}|}{d_i\big(\tilde r_{2,K}\tilde r_{0,K}-\tilde r_{1,K}^2\big)}\le\frac{\tilde r_{0,K}}{d_i\big(\tilde r_{2,K}\tilde r_{0,K}-\tilde r_{1,K}^2\big)}\le\frac{256C_2^2}{C^5C_1^2}.$$

Позначимо

$$\xi_{i,K}=\frac{\tilde a_{i,K}}{K\sqrt{\mathsf{D}\,\tilde\mu_{1,K}}}(T_i-\mathsf{E}\,T_i).$$

Тоді $L_K=\sum_{i=1}^{K}\xi_i$. Твердження теореми випливає з центральної граничної теореми для ξ_i з умовою Ліндеберга. Як і в теоремі 3 можна перевірити виконання центральної граничної теореми з умовою Ліндеберга. Також, оскільки дисперсії оцінок $\mathsf{D}\,\tilde\mu_{1,K}$ прямують до нуля при $K\to\infty$, то отримаємо консистентність. □

Література

1. А. А. Боровков, *Математическая статистика*, "Наука", Москва, 1977.
2. М. В. Карташов, *Імовірність, процеси, статистика*, Посібник, ВПЦ "Київський університет", Київ, 2008.
3. Р. Є. Майборода, *Оцінка розподілів компонентів сумішей що змінюються*, Укр. мат. журнал **48** (1996), №4, 562–566.
4. Р. Є. Майборода, *Статистичний аналіз сумішей*, ВПЦ "Київський університет", Київ, 2003.
5. Р. Є. Майборода, О. В. Сугакова, *Оцінювання та класифікація за спостереженнями із суміші*, ВПЦ "Київський університет", Київ, 2008.
6. O. O. Kubaychuk, *Estimation of moments by observations from mixtures with varying concentrations*, Theory of Stochastic Processes **8(24)** (2002), no. 3–4, 226–232.
7. S. Newcomb, *A generalized theory of the combination of observations so as to obtain the best result*, Amer. J. Math. **8** (1894), 343–366.
8. K. Pearson, *Contribution to the mathematical theory of evolution*, Trans. Roy. Sos. A **185** (1894), 71–110.
9. D. M. Titterington, A. F. Smith, and O. E. Makov, *Analysis of Finite Mixture Distributions*, Wiley, New York, 1985.

Кафедра теорії ймовірностей, математичної статистики та актуарної математики механіко–математичний факультет, Національний університет ім. Тараса Шевченка, проспект Глушкова, 2, Київ 03127, Україна

Адреса електронної пошти: artshcherbina@gmail.com

Надійшла 08/06/2010