Київський національний університет імені Тараса Шевченка
Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка
Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

Мірошниченко Віталій Олегович

УДК 519.22

ДИСЕРТАЦІЯ

Регресійний аналіз сумішей зі змінними концентраціями

112 - Статистика

Подається на здобуття наукового ступеня *доктора філософії*

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

______________________ *В. О. Мірошниченко*

Науковий керівник: *Майборода Ростислав Євгенович*
доктор фізико-математичних наук, професор

Електронний підпис
Мірошниченко
Віталій
Олегович
Ідентифікаційний
код
3439009251
Підписано у Вчасно

Київ — 2022

Анотація

Мірошниченко В. О. Регресійний аналіз сумішей зі змінними концентраціями. — Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 112 “Статистика” — Київський національний університет імені Тараса Шевченка, Київ, 2022.

Дисертаційна робота присвячена регресійному аналізу сумішей зі змінними концентраціями, який включає в себе: оцінювання параметрів розподілу, дослідження асимптотичних властивостей оцінок параметрів, та побудова довірчих інтервалів.

У дисертаційному дослідженні розглядаються багатовимірні дані, що описуються моделлю суміші зі змінними концентраціями, причому розподіл кожного компонента суміші відповідає певній моделі регресії (або лінійній або нелінійній). Метою дослідження є розробка техніки статистичного аналізу, яка дозволяла б робити висновки про параметри моделей для кожного компонента суміші окремо, та застосувати розроблені методи до аналізу соціологічних даних.

Дисертація складається з анотацій українською та англійською мовами, переліку умовних позначень, вступу, п’яти розділів основної частини, висновків, списку використаних джерел та додатку.

У вступі обґрунтовано актуальність теми дослідження, сформульовано мету, об’єкт, предмет, завдання і методи дослідження, зазначено наукову новизну отриманих результатів, їх практичне значення, зв’язок роботи з науковими темами й особистий внесок здобувача, Наведено зміст роботи, та вказано також де було апробовано та опубліковано результати дисертації.

В Розділі 1 зроблено огляд літератури, наводяться основні означення та допоміжні твердження, які будуть використані в наступних розділах.

Розділ 2 присвячений продовженню вивчення параметричній моделі лінійної регресійної суміші, яка була описана в роботах R. DeVeaux (1986), W. DeSarbo, R. Cron (1988), та продовженню вивчення непараметричної моделі регресійної суміші за змінними концентраціями, яка описана у роботі Р. Майбороди, та Д. Любашенко. Для асимптотичної коваріаційної матриці оцінок параметрів регресії побудована оцінка методом складаного ножа, та доведено її консистентність.

Розділ 3 присвячено вивченню моделі нелінійної регресійної суміші зі змінними концентраціями. Оцінки параметрів регресії для окремих компонент (УОР-оцінки) визначаються як розв'язок оціночного рівняння. У цьому розділі вивчалися асимптотичні властивості оцінок параметрів регресії, та оцінок коваріаційної матриці оцінок, доведено загальні теореми про консистентність і асимптотичну нормальність оцінок параметрів регресії, які також застосовано для випадку оціночних рівнянь у МНК, та додатково для логістичної функції регресії. Як і для моделі лінійної регресії, для асимптотичної коваріаційної матриці оцінок параметрів нелінійної регресії, побудовано оцінку методом складаного ножа, та доведено її консистентність.

Розділ 4 даного дослідження присвячено аналізу залишків лінійної та нелінійних моделей: доведено консистентність оцінки дисперсії похибок для різних моделей (лінійної або нелінійної) сумішей зі змінними концентраціями, описано побудову діаграми квантиль-проти-квантилю для залишків регресії, та доведено консистентність оцінки функції розподілу, та оцінок квантилів розподілу похибок регресії у моделі регресійної суміші зі змінними концентраціями.

Розділ 5 присвячено застосуванням моделі регресійної суміші зі змінними концентраціями до аналізу соціологічних даних. Ці дані складено із результатів ЗНО-2016 року, об'єднаними із результатами виборів до Парламенту України 2014 року.

У дисертаційній роботі отримано наступні нові результати:

1 Доведено консистентність і асимптотичну нормальність оцінок параметрів регресії у нелінійних моделях регресійної суміші.
2 Доведено консистентність оцінки дисперсії залишків у моделі нелінійної регресійної суміші.
3 Для оцінок параметрів нелінійної регресії отримано оцінку коваріаційної матриці методом складаного ножа і доведено її консистентність.
4 Побудовано оцінки квантилів розподілу похибок регресії і доведена їх консистентність.
5 Розроблено обчислювальні процедури знаходження оцінок коваріаційних матриць, їхня ефективність досліджена за допомогою імітаційного моделювання.

6 На основі отриманих оцінок побудовано довірчі еліпсоїди для параметрів регресії у лінійних і нелінійних моделях регресійних сумішей.
7 На соціологічних даних продемонстровано можливості прикладного застосування розробленого методу.

Дисертаційне дослідження носить теоретичний. Отримані результати можуть бути використані в статистиці та застосовані для опису даних у медичних, економічних, соціологічних дослідженнях, та актуарній математиці.

Ключові слова: суміш, мінімаксні навантаження, регресійна суміш, лінійна регресія, нелінійна регресія, оцінювання параметрів, консистентність оцінок параметрів, асимптотична нормальність оцінок, консистентність оцінок коваріацій для оцінок регресії, складаний ніж, сендвіч-формула, довірчі еліпсоїди, суміш нормальних розподілів, непараметричні оцінки розподілів, УОР-оцінки, узагальнений метод найменших квадратів

Abstract

Miroshnychenko V. O. Regression analysis of mixtures with varying concentrations. — Qualifying scientific work on the rights of the manuscript. The thesis presented for the academic degree Doctor of Philosophy in specialty 112 “Statistics”. — Taras Shevchenko National University of Kyiv, Kyiv, 2022.

The thesis is devoted to the study of regression analysis of a mixture with varying concentrations including parameter estimation, research of asymptotics for parameter estimators, and confidence ellipsoids construction. Multidimensional data which can be described by a model of mixtures with varying concentrations are considered in this study. Each mixture’s component has its own distribution and is described by a regression model (linear or nonlinear). The goal of a study is to develop statistical methods to make conclusions about the parameters of the model for different components of the mixture separately and apply these methods to analyze real life data.

The thesis consists of an abstract in Ukrainian and English, a list of symbols, an introduction, five chapters of the main part, conclusions, a reference list, and an appendix.

In the introduction the research topic is motivated, the purpose, object, subject, tasks, and methods of research are formulated, and the scientific novelty of the obtained results and their practical significance is indicated.

Chapter 1 provides a review of the literature, main definitions, and additional statements. All of these are used in the following chapters. Chapter 2 continues the study of parametric models for linear regression mixtures. This model is described in the works R. DeVeaux (1986), W. DeSarbo, R. Cron (1988). A study for nonparametric models that is based on R. Maiboroda and D. Luibashenko (2017) is also added to this chapter. For the regression parameter estimator’s asymptotic covariance matrix estimator is built. The consistency of this estimator is proven in this chapter.

The Chapter 3 is devoted to the study of nonlinear regression mixture model with varying concentrations. In this model, the nonlinear regression parameter estimates are defined separately for different components as solutions to estimating equations. In this chapter asymptotical properties of GEE estimators for regression mixture and their asymptotic covariance matrix estimator are researched:

- General theorems about the estimator’s consistency and asymptotical normality of GEE estimators for regression parameters are proven.

- Analogs of these theorems are proven for least square estimators and logistic regression functions.
- Jackknife estimator for GEE estimator's asymptotic covariance matrices is built.
- The consistency of this Jackknife estimator is demonstrated.

The Chapter 4 presents techniques of residual analysis for linear and nonlinear regression models and quantile-vs-quantile diagrams for regression residuals. Consistency of the error's term variance estimator, CDF estimator, and quantiles is demonstrated.

The Chapter 5 consists of an application of the regression mixture model with varying concentrations to sociological data. This data is a joined dataset of EIT-2016 and the Ukrainian parliamentary election in 2014.

The following new results were developed in the thesis in the framework of models of mixture with varying concentrations:

1 Consistency and asymptotic normality are proven for linear and nonlinear regression parameter estimators.
2 Consistency of the error's term variance estimator is demonstrated.
3 For the GEE estimator's asymptotic covariance jackknife estimator is built. The consistency of this estimator is demonstrated.
4 Consistent estimators for the error's term distribution quantile are built.
5 Computational procedures for estimates of covariance matrix calculation were developed. Their effectiveness was investigated in a simulation study.
6 Confidence sets for regression parameters are constructed.
7 Developed methods were applied to sociological data.

The research is of theoretical nature. The obtained results can be used in mathematical and applied statistics, in medical, economical, sociological, and actuarial statistical data analysis.

Keywords: mixture, minimax weights, regression mixture, linear regression, nonlinear regression, parameter estimation, consistent estimator, asymptotic normality, asymptotic covariation matrix consistency, jackknife, sandwich-formula, confidence ellipsoid, gaussian mixture model, nonparametric distribution estimation, generalized estimating equation, generalized least squares.

Перелік умовних позначень

У роботі використовуються наступні позначення та скорочення.

в.в.	випадкова величина
$\mathbb{P}$	ймовірність
$\mathbb{E}\,\xi$	математичне сподівання в.в. ξ
$\mathbb{D}\,\xi$	дисперсія в.в. ξ
$\mathbb{C}\mathrm{ov}(\xi)$	коваріаційна матриця в.в. ξ
$\xrightarrow{\mathbb{P}}$	збіжність за ймовірністю
$\xrightarrow{\mathrm{w}}$	слабка збіжність
$O_P(1)$	послідовність обмежених за ймовірністю в.в
$o_P(1)$	послідовність в.в, що прямує до нуля за ймовірністю
$\xi = \eta$ $a.s.$,	випадкова величина ξ майже напевно дорівнює в.в. η
м.н.	майже напевно
ОМВ	оцінка максимальної вірогідності
МНК	метод найменших квадратів
$\langle a, b\rangle$	скалярний добуток векторів $a, b \in \mathbb{R}^d$, $\langle a, b\rangle = \sum_{i=1}^d a_i b_i$
$\lvert x\rvert$	модуль величини $x \in \mathbb{R}$, та $\sqrt{\sum_{j=1}^d x_j^2}$ для $x \in \mathbb{R}^d$
$\lvert A\rvert$	операторна норма матриці A, $\sqrt{\sum_{i=1}^{d_1} \sum_{j=1}^{d_2} a_{ij}^2}$

Зміст

Вступ

Актуальність теми. Параметричні моделі суміші застосовуються у статистиці з початку XX століття. Такі моделі природно застосовувати до опису даних біологічної, медичної, економічної, актуарної статистики, коли аналізовані популяції складаються з кількох компонент з різним розподілом досліджуваних характеристик.

В останнє двадцятиліття набула розвитку непараметрична та семі-параметрична техніка аналізу моделей сумішей. Зокрема, у роботі Холла і Жоу, 2003 [33] розглянуто суміш двох компонент з багатовимірними спостереженнями, координати яких є незалежними для кожної компоненти окремо. У роботах Борде, Делма, Вандекерхофе, 2006 [14], Хантера, Ванга, Хетманшпергера, 2007 [36], Майбороди, 2008 [4] та Бутуче, Вандекерхофе, 2014 [17] розглядається оцінювання параметрів моделей сумішей симетричних розподілів. Непараметричне та семі-параметричне оцінювання у моделях сумішей зі змінними концентраціями розглянуто у роботах Майбороди, Сугакової, 2009 [5], Майбороди і Сугакової, 2011 [56], Майбороди, Сугакової, Дороніна, 2013 [57]. Перевірці гіпотез у рамках цієї моделі присвячені роботи Аутін, Пует, 2011 [8], Майбороди, Сугакової, 2012 [51].

У роботі Гран, та Ляйх, 2006 [28], Фарії, та Сормехено, 2010 [25] та ін. розглядаються параметричні моделі регресійного аналізу для сумішей зі змінними концентраціями. Непараметрична лінійна модель розглянута у роботі Любашенко і Майбороди, 2015 [47]. У роботі Майбороди і Мірошниченко, 2018 [49] порівнюються можливості параметричних та непараметричних моделей при побудові довірчих множин для коефіцієнтів регресії.

У дисертаційній роботі передбачається подальший розвиток цих досліджень з метою створення більш ефективних методів оцінювання, перевірки гіпотез та діагностики регресійних моделей за спостереженнями з суміші.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційну роботу виконано в рамках державних бюджетних дослідницьких наукових тем No 19БФ038-01 «Точні формули, оцінки, асимптотичні властивості і статистичний аналіз складних еволюційних систем з багатьма ступенями свободи» (номер державної реєстрації 0119U100317) та No 22БФ038-01 «Стохастичні динамічні системи, неоднорідні у часі або з випадковим часом: асимптотика та статистичний аналіз» (номер державної реєстрації 0122U000821) кафедри теорії ймовірностей, статистики та актуарної математики механіко-математичного факультету Київського національного університету імені Тараса Шевченка, що входять до комплексного тематичного плану науково-дослідних робіт «Сучасні математичні проблеми природознавства, економіки та фінансів».

Задачі дослідження. Метою роботи є вивчення асимптотичних властивостей оцінок: параметрів регресії, оцінок коваріаційної матриці цих параметрів, параметрів розподілу залишків регресії у моделі регресійної суміші зі змінними концентраціями для сімейств рівностепенево неперервних функцій регресії. Завдання дослідника є доведення теорем про асимптотичні властивості оцінок параметрів та перевірка отриманих результатів чисельними моделюваннями.

Наукова новизна одержаних результатів.

Усі результати, що отримані є новими і стосуються моделі регресійної суміші зі змінними концентраціями. Перелік основних результатів:

1 Доведено консистентність і асимптотичну нормальність оцінок параметрів регресії у нелінійних моделях регресійної суміші.
2 Доведено консистентність оцінки дисперсії залишків у моделі нелінійної регресійної суміші.
3 Для оцінок параметрів нелінійної регресії отримано оцінку матриці коваріації методом складаного ножа і доведено її консистентність.
4 Побудовано оцінки квантилів розподілу похибок регресії і доведена їх консистентність.
5 Розроблено обчислювальні процедури знаходження оцінок коваріаційних матриць, їхня ефективність досліджена за допомогою імітаційного моделювання.
6 На основі отриманих оцінок побудовано довірчі еліпсоїди для параметрів регресії у лінійних і нелінійних моделях регресійних сумішей.
7 На соціологічних даних продемонстровано можливості прикладного застосування розробленого методу.

Практичне значення одержаних результатів. Дисертаційна робота має як теоретичне, так і практичне значення. Результати та розроблені методи можуть бути застосовані у соціологічних, медико-біологічних та економічних дослідженнях, коли має місце розподіл об'єктів на популяції. Доведена консистентність і нормальність оцінок у регресійній моделі дозволяє проводити регресійний аналіз, а розвинуті непараметричні методи оцінки матриці коваріації оцінок дозволяють будувати довірчі множини для оцінок параметрів регресії у моделі суміші.

Особистий внесок здобувача. За результатами дисертації здобувач опублікував шість робіт у фахових виданнях. Три із них опубліковані автором самостійно, а три - у співавторстві з науковим керівником професором Р.Є. Майбородою, у яких Р.Є. Майбороді належить постановка задачі, керівництво роботою, участь у розробці остаточного тексту статті. У статті

Maiboroda R., Miroshnichenko V., Sugakova O. Jackknife for nonlinear estimating equations.- Modern Stochastics: Theory and Applications, 2022 https://doi.org/10.15559/22-VMSTA208

Р.Є. Майбороді належить постановка задачі і загальне керівництво роботою, О.В. Сугаковій належить ідея доведення рівності (57). Всі інші результати належать здобувачу.

Апробація результатів дисертації.

1 Miroshnychenko V.O. EM алгоритм для аналізу регресійної суміші. XV International Scientific-Practical Conference "Shevchenkivska Vesna - 2018". Kyiv, Ukraine, 2017.
2 Maiboroda R.E., Miroshnychenko V.O. Confidence Sets For Regression Coefficients By Observations From a Mixture. Modern Stochastics: Theory And Applications IV. Kyiv, Ukraine, 2018
3 Miroshnychenko V.O. Asymptotics of generalized least squares estimates for mixture of nonlinear regressions. International Conference of Young Mathematicians. Kyiv, Ukraine, 2019
4 Miroshnychenko V.O. Оцінки параметрів нелінійної регресії за спостереженнями з суміші. XVII International Scientific-Practical Conference "Shevchenkivska Vesna - 2019". Kyiv, Ukraine, 2019
5 Miroshnychenko V.O. Асимптотична нормальність коефіцієнтів у суміші регресій. XVIII International Scientific-Practical Conference "Shevchenkivska Vesna - 2020". Kyiv, Ukraine, 2020

6 Miroshnychenko V.O., Maiboroda R.E. Asymptotic normality of modified LS estimator for mixture of non-linear regressions. Scientific Conference "Actual Problems Of Stochastic Analysis" dedicated to the 80-th anniversary of the birth of academician Sh.K.Formanov. Tashkent, Uzbekistan, 2021

7 Maiboroda R.E., Miroshnychenko V.O. Consistency of jackknife covariance estimator for mixture of nonlinear regressions. International conference "Modern stochastics: theory and applications V". Kyiv, Ukraine, June 1-4, 2021

8 Семінар з математичної статистики механіко-математичного факультету Київського Національного університету імені Тараса Шевченка під керівництвом проф. О.Г. Кукуша та проф. Р.Є. Майбороди. Київ, Україна, 2021

9 Maiboroda R.E., Miroshnychenko V.O. Метод складаного ножа для суміші нелінійних регресій. XX International Scientific-Practical Conference "Shevchenkivska Vesna - 2022". Kyiv, Ukraine 2022

10 Miroshnychenko V.O. Jackknife for nonlinear estimating equations. The 4-th International Conference on Statistics: Theory and Applications, Virtual conference. Prague, Czech Republic, 2022
https://avestia.com/ICSTA2022_ Proceedings/files/paper/ICSTA_ 149.pdf

11 Семінар з математичної статистики механіко-математичного факультету Київського Національного університету імені Тараса Шевченка під керівництвом проф. О. Г. Кукуша та проф. Р. Є. Майбороди (м. Київ, 2021 р.)

12 Семінар "Статистичні проблеми для випадкових процесів і полів" кафедри математичного аналізу та теорії ймовірностей Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського" під керівництвом проф. О. І. Клесова, проф. О. В. Іванова (м. Київ, 2022 р.)

13 Семінар кафедри теорії ймовірностей і математичного аналізу Ужгородського Національного університету під керівництвом проф. Г. І. Сливки-Тилищак (м. Ужгород, 2022 р.)

Публікації.

1 Maiboroda R. E., Miroshnichenko V. O. Confidence ellipsoids for regression coefficients by observations from a mixture. Modern Stochastics: Theory and Applications. 2018. Vol.5, Iss.2, pp. 225-245

2 Miroshnichenko V. O. Generalized least squares estimates for mixture of nonlinear regressions. Bulletin of Taras Shevchenko National University of Kyiv. 2018. Vol. 3, pp. 25-29

3 Miroshnichenko V. O. Residual analysis in regression mixture model. Bulletin of Taras Shevchenko National University of Kyiv. 2019. Vol. 3, pp. 8-16

4 Miroshnychenko V. O., Maiboroda R. E. Asymptotic normality of modified LS estimator for mixture of nonlinear regressions. Modern Stochastics: Theory and Applications. 2020. Vol.7 Iss.4, pp. 435–448

5 Miroshnychenko V. O. Fast calculations of Jackknife covariance matrix estimator. Bulletin of Taras Shevchenko National University of Kyiv. 2021. Vol. 1, pp. 27-36

6 Maiboroda R., Miroshnichenko V., Sugakova O. Jackknife for nonlinear estimating equations. Modern Stochastics: Theory and Applications. 2022. https://doi.org/10.15559/22-VMSTA208

Структура та обсяг дисертації. Дисертація складається із вступу, п'яти розділів, розбитих на підрозділи, висновку, списку використаних джерел (92 найменування), та додатку, який містить список публікацій здобувача за темою дисертації та відомості про апробацію результатів. Повний обсяг дисертації становить 139 сторінок, з них основний текст займає 110 сторінок.

Зміст роботи. **Перший розділ** містить короткий історичний огляд літератури по тематиці дисертації і оглядає сучасні роботи, що вивчають теми близькі до теми даної дисертаційної роботи, а саме: моделі суміші, оцінки розподілів, моделі регресії, асимптотика оцінок методом найменших квадратів і оцінок максимальної вірогідності, оцінки коваріацій методом складаного ножа.

Дамо означення суміші зі змінними концентраціями. Позначимо об'єкт, що досліджується, літерою O. У об'єкта O присутня неспостережувана характеристика $\kappa(O) \in \{1, ..., M\}$ - номер компонента (популяції), до якого об'єкт належить. Також, у об'єкта O є спостережувані дані. Наприклад, для нього відомі ймовірності належати до фіксованого компонента $\mathbb{P}\{\kappa(O) = k\}$. Окрім цих ймовірностей, для об'єкта O, також спостерігається вектор характеристик $\xi(O) = \left(\xi^1(O), ..., \xi^D(O)\right)^T$. Розподіл цього вектора залежить від номера компонента $\kappa(O)$. Символом F_m, $1 \leqslant m \leqslant M$ позначимо розподіл характеристик об'єкта з m - го компонента $F_m(A) = P\Big(\xi(O) \in A | \kappa(O) = m\Big)$, для довільної вимірної $A \subset \mathbb{R}^D$.

Розглянемо об'єкти $O_{j;n}$, $j = 1, ..., n$, які належать різним компонентам з різними відомими ймовірностями $p^k_{j;n} = P\{\kappa(O_{j;n}) = k\}$, $\kappa_j = \kappa(O_{j;n})$. Символом $p^k_{j;n}$ позначатимемо ймовірність j-го об'єкта належати компоненту із номером k.

В цій моделі спостереження $\xi_{j;n} = (X_j, Y_j) \in \mathbb{R}^D$, $j = 1, ..., n$ є незалежними при різних j і складаються із регресорів X_j, та відгуків Y_j. Припускається, що відгуки залежать від регресорів, і ця залежність різна для різних компонент. Наприклад,

$$Y_j = g(X_j, b^{(\kappa_j)}) + \varepsilon_j. \tag{1}$$

Тут ε_j - похибка вимірювання, $g : \mathbb{R}^{D-1} \times \Theta \to \mathbb{R}$ - функція регресії, а $b^{(m)} \in \Theta \subseteq \mathbb{R}^d$ - невідомі параметри регресії. Припускається, що X_j і ε_j незалежні. У моделі лінійної регресійної суміші виконується рівність $d = D - 1$, і функція g має наступний вигляд $g(x, \gamma) = \sum_{j=1}^{D-1} x_j \gamma_j$, $x \in \mathbb{R}^{D-1}$, $\gamma \in \Theta$.

Функцією розподілу спостереження $\xi_{j;n} = (X_j, Y_j)$ $\mathbb{P}\{\xi_{j;n} \in A\} = \sum_{m=1}^{M} p^m_{j;n} F_m(A)$, де F_m - розподіл на $\mathbb{R}^D$ (F_m - розподіл m-го компонента суміші), $p^m_{j;n}$ - ймовірність спостерігати m-тий компонент при j-му спостереженні. Ймовірності $p^m_{j;n}$ є відомими.

У даному розділі дається означення мінімаксних навантажень $A_n = (A^{(1)}_n, ..., A^{(k)}_n) = \left(a^k_{j;n}\right)^{n,M}_{j=1,k=1} = P^T_n(\Gamma_n)^{-1}$, де $P_n = (p^k_{j;n})^{k=1,...,M}_{j=1,...,n}$ - матриця ймовірностей змішування, а $\Gamma_n = P_n P^T_n$. Позначимо $\Gamma_\infty = \lim_{n\to\infty} \frac{1}{n}\Gamma_n$

У **другому розділі** розглядаються два підходи до вивчення лінійної регресійної суміші: непараметричний і параметричний.

У підрозділі 2.1 оглядається оцінка параметрів

$$\hat{b}^{LS,(k)}_n = \left(X^T A^{(k)}_n X\right)^{-1} X^T A^{(k)}_n Y \tag{2}$$

лінійної регресії з мінімаксними навантаженнями. Тут X - матриця з регресорів, Y - вектор відгуків.

У підрозділі 2.2 розглядається параметрична модель суміші, у якій параметри розподілів та регресій у підігнані ЕМ алгоритмом (Expected Mean), $\hat{b}^{EM}(k, n) = (X^T W^{(k)}_n X)^{-1} X^T W^{(k)}_n Y$. W_n - діагональна матриця з апостеріорних ймовірностей для $k-$го компонента для моделі нормальної суміші. У цьому підрозділі доведено теореми про консистентність оцінок коваріацій для оцінок параметрів регресії параметричним, та непараметричним способами.

Теореми про асимптотичний рівень довірчих множин для параметрів регресії у непараметричній (Теорема 2.3), та параметричній моделі регресії (Теорема 2.5) доведено у підрозділі 2.3.

У підрозділі 2.4 розглядається непараметрична оцінка $\hat{V}_n^{(k)}$ методом складаного ножа (СН) для асимптотичної матриці коваріації $V^{(k)}$ оцінок регресійних параметрів $b^{(k)}$. Оцінка (СН) що будується за оцінками $(\hat{b}_{i-;n}^{(k)})_{i=1}^n$ (аналогічно оцінці $\hat{b}_n^{LS,(k)}$) параметрів регресії на вибірках, із яких яких вилучили відповідні спостереження (1-ше, 2-ге, і тд). У Теоремі 2.6 доведено наступне відношення

$$n\sum_{i=1}^{n}\left(\hat{b}_{i-;n}^{(k)}-\hat{b}_n^{LS,(k)}\right)\left(\hat{b}_{i-;n}^{(k)}-\hat{b}_n^{LS,(k)}\right)^T-V^{(k)}\xrightarrow{\mathbb{P}}0, n\to\infty.$$

У підрозділі 2.5 описані результати імітаційних експериментів, проведених для перевірки можливості застосування отриманих результатів для вибірок скінченного обсягу, та порівняти асимптотичну поведінку довірчих множин у різних умовах для оцінок параметрів лінійної регресійної суміші.

Третій розділ присвячено вивченню нелінійній регресійній суміші зі змінними концентраціями. На відміну від лінійної моделі, у нелінійній моделі не вдається виписати явний вигляд оцінок як у (2). Тому оцінки параметрів регресії k-го компонента $\theta^{(k)}\in\Theta$ (УОР-оцінки) шукаються як розв'язок оціночного рівняння по параметру γ :

$$S_n^{(k)}(\gamma)=\sum_{j=1}^{n}a_{j;n}^{k}s(\xi_{j;n},\gamma)=0.$$

Для випадку МНК оціночна функція s має вигляд $s(\xi_j,\gamma)=\left(Y_j-g(X_j,\gamma)\right)\dot{g}(X_j,\gamma)$, а $g(x,\gamma)$ - деяка відома диференційовна функція регресії, $\xi_{j;n}=(X_j,Y_j)\in\mathbb{R}^D$ - пара з регресорів та відгуку. У цьому розділі вивчалися асимптотичні властивості оцінок параметрів регресії, та оцінок матриці коваріації оцінок. Тут доведено як загальні теореми про консистентність і асимптотичну нормальність, які також застосовано для випадку оціночних рівнянь у МНК, та додатково для логістичної функції регресії. Далі символами $\hat{\theta}_n^{(k)}$ і $\hat{\theta}_{i-;n}^{(k)}$ позначатимемо УОР-оцінки параметрів нелінійної регресії $\theta^{(k)}$ для k-го компонента на всій вибірці, та із вилученим $i-$тим спостереженням.

У підрозділі 3.1 дається означення нелінійної регресійної суміші, та визначаються оцінки методом узагальнених оціночних рівнянь.

У підрозділі 3.2 доведено додаткові леми, що будуть використані у теоремах про консистентність оцінок матриць коваріацій.

У підрозділі 3.3 доведено Теорему 3.1 про консистентність УОР-оцінок, $\hat{\theta}_n^{(k)}\xrightarrow{\mathbb{P}}\theta^{(k)}$. Теорема 3.3 з цього підрозділу наводить умови, за яких ймовірність існування розв'язку оціночного рівняння прямує до 1.

У підрозділі 3.4 доведено Теорему 3.4 про асимптотичну нормальність УОР оцінок, яку можна звузити до оцінок k-го компонента $\sqrt{n}(\hat{\theta}_n^{(k)} - \theta^{(k)}) \xrightarrow{w} N(0, V^{(k)})$.

У підрозділі 3.5 розглядається оцінка коваріацій для оцінок параметрів регресії методом складаного ножа, та доводиться її консистентність. Для $k, m = 1, ...M$ позначимо $\hat{V}_n^{(k,m)} = n\sum_{i=1}^n \left(\hat{\theta}_{i-;n}^{(k)} - \hat{\theta}_n^{(k)}\right)\left(\hat{\theta}_{i-;n}^{(m)} - \hat{\theta}_n^{(m)}\right)^T$.

Теорема 3.6 *Нехай виконуються наступні умови:*

1 Θ - *компакт,* θ - *внутрішня точка* Θ.

2 *Існує така функція* $h : \mathbb{R}^D \to \mathbb{R}$, *що*

$$\sup_{\gamma\in\Theta} |s(x,\gamma)| \leqslant h(x); \ \sup_{\gamma\in\Theta} |\dot{s}(x,\gamma)| \leqslant h(x); \ \max_{i,j,l=1,\ldots,d} \sup_{\gamma\in\Theta} \left|\frac{\partial s^l(x,\gamma)}{\partial\gamma^i\partial\gamma^j}\right| \leqslant h(x) \text{ і},$$

$$\text{для деякого } \alpha \geqslant 4, \ \mathbb{E}\left(h(\xi_{(l)})\right)^\alpha < \infty \text{ для всіх } l = 1, ..., M.$$

3 *Матриця* Γ_∞ *існує і* $\det\Gamma_\infty \neq 0$.

4 *Існують границі* $\langle a^k a^m p^l p^i \rangle$ *для всіх* $i, l = 1, ..., M$.

5 *Матриці* $M^{(k)} = \mathbb{E}\,\dot{s}(\xi_{(k)}, \theta^{(k)})$, $M^{(m)} = \mathbb{E}\,\dot{s}(\xi_{(m)}, \theta^{(m)})$ *є невиродженими.*

6 $\hat{\theta}_n^{(k)}, \hat{\theta}_n^{(m)}$ *є* $\sqrt{n}$ - *консистентними оцінками* $\theta^{(k)}, \theta^{(m)}$.

7 $\sup_{i=1,\ldots,m} |\hat{\theta}_{i-;n}^{(l)} - \theta^{(l)}| \to 0$, *для* $l = k, m$ *за ймовірністю при* $n \to \infty$.

Тоді $\hat{V}_n^{(k,m)} \to V^{(k,m)}$ *за ймовірністю.*

Тут $V^{(k,m)} = (M^{(k)})^{-1}Z^{(k,m)}(M^{(m)})^{-T}$ - асимптотична матриця коваріації оцінок $\hat{\theta}_n^{(k)}$ для $\theta^{(k)}$, $M^{(k)} = \mathbb{E}\,\dot{g}(X_{(k)}, \theta^{(k)})\dot{g}(X_{(k)}, \theta^{(k)})^T$, та $Z^{(k,m)} = \lim_{n\to\infty} n\,\mathbb{C}\mathrm{ov}\left(S_n^{(k)}(\theta^{(k)}), S_n^{(m)}(\theta^{(

параметрів регресії, та наведено приклад застосування теореми для логістичної функції регресії.

Підрозділ 4.3 присвячено побудові діаграми квантиль проти квантилю для регресійної моделі суміші зі змінними концентраціями. У Теоремі 4.3 доведено рівномірну консистентність непараметричної оцінки розподілу похибок, а у Теоремі 4.4 наведено умови консистентності оцінок квантилів розподілу для похибок регресійної суміші. У підрозділі 4.4 проведено імітаційне моделювання, метою якого є перевірка отриманих теоретичних результатів у підрозділі 4.2

У **п'ятому розділі** йде мова про застосування моделі регресійної суміші зі змінними концентраціями до аналізу соціологічних даних.

У підрозділі 5.1 описуються соціологічні дані, які є поєднанням результатів ЗНО за 2016 рік, та виборів до Парламенту України 2014 року.

У підрозділах 5.2 і 5.3 застосовано лінійну та нелінійну моделі регресійної суміші зі змінними концентраціями до соціологічних даних. В якості нелінійної функції обрано логістичну функцію. Також тут наведено значення отриманих параметрів, та зображено графіки функціоналів МНК. У підрозділі 5.4 порівнюються криві цих залежностей, та значення навантажених функціоналів МНК.

У підрозділі 5.5 побудовано діаграми квантиль проти квантилю для залишків лінійної та нелінійної моделі регресійної суміші, а також зображено функції розподілу залишків різних моделей для різних компонент.

У підрозділі 5.6 побудовано довірчі множини для обох моделей. Для лінійної регресійної моделі зображено довірчі множини для параметричних та непараметричних оцінок параметрів. Для нелінійної функції регресії зображено довірчі множини на основі матриць коваріацій, які оцінено параметричним методом, та методом складаного ножа.

У **висновках** сформульовано основні результати дисертаційної роботи.

Автор висловлює вдячність науковому керівнику — професору Майбороді Ростиславу Євгеновичу — за постійну підтримку, та допомогу в роботі; завідувачці кафедри теорії ймовірностей, статистики та актуарної математики — професорці Мішурі Юлії Степанівні — за мотивацію до навчання; усьому колективу кафедри теорії ймовірностей, статистики та актуарної математики за всебічну допомогу під час навчання.

Розділ 1

Огляд літератури і допоміжні відомості

Цей розділ присвячено історичному огляду розвитку методів статистичного аналізу даних на основі моделі суміші різних ймовірнісних розподілів, та опису основних моделей даних, які будуть використані у роботі та методів їх аналізу: модель суміші зі змінними концентраціями, проста модель регресії, та методи: оцінки максимальної вірогідності, EM (Expected Mean) алгоритм, метод найменших квадратів (МНК), однокрокова процедура Ньютона-Рафсона, метод складаного ножа.

Розглянуті методи і моделі будуть використані для аналізу моделі регресійної суміші зі змінними концентраціями, та знаходження асимптотичних властивостей оцінок параметрів суміші. Окрім цього, в даному розділі, наведено допоміжні результати, що отримані в дослідженнях інших авторів.

1.1 Історичний огляд

У моделі скінченної суміші розглядається набір даних, у якому кожен об'єкт належить одному з компонент (популяції). Цих компонентів скінченна кількість. Розподіл спостережуваних характеристик об'єктів залежить від номеру компонента, до якого належить сам об'єкт. Об'єкт належить різним компонентам з відомими ймовірностями (концентраціями). Такий опис даних є зручним у задачах і дослідженнях природничих наук: медицині, економіці, соціології. У книзі [84] P. Schlattmann навів приклади застосування моделей суміші для аналізу захворювань, у психофізіології та дослідження експресії генів.

Класичним варіантом є модель суміші зі скінченною кількістю компонент, та сталими концентраціями. Наприклад, такою є модель суміші нормальних розподілів.

Методи аналізу моделей суміші поділяються на параметричні й непараметричні, при чому параметричні більш досліджені.

Зокрема параметрична модель суміші досліджувалася у роботах J. Bilmes [12], D. Titterington [88], G. Mclachlan [60], [62], [63]. У [88] для моделі сумішей наведено аналоги класичних статистичних методів оцінки невідомих параметрів розподілів: метод моментів, оцінки максимальної вірогідності, баєсівський підхід до оцінювання параметрів. Авторами R. A. Redner, та H. F Walker у [72] описано модель суміші для експоненційного сімейства розподілів регресорів, і описано EM (Expected Mean) алгоритм для оцінювання параметрів цих розподілу, в тому числі і нормального. Робота G. Mclachlan [63] розвиває EM алгоритм для застосування у задачі оцінки параметрів нормальної суміші, досліджує властивості отриманих оцінок та асимптотику оцінок, що отримані EM алгоритмом. Окрім нормальної суміші, G. Mclachlan [62] досліджував суміші розподілів t-Стьюдента і побудував (EM)-оцінки для цього випадку. У роботі J. Bilmes [12] параметрична модель суміші використовується для підгонки невідомих параметрів суміші різних нормальних розподілів, і для побудови мовних моделей у задачах обробки природних мов. У роботах M. Stephens [85], M. Bilancia [9], автори застосовують баєсівські методи до оцінювання параметрів моделі нормальної суміші, а у роботі T. Tatarinova і A. Schumitzky [87] застосовано баєсівські методи для оцінювання параметрів у нелінійних моделях сумішей.

Робота P. Hall, D. Titterington [70] про непараметричні моделі, а робота D. Titterington., A. Smith, U. Makov [89] для параметричних моделей розглядають суміші різних концентрацій, у яких концентрації можуть бути різними для різних спостережень. Авторами показано що можливо отримати консистентні оцінки параметрів без припущення про однакову розподіленість спостережень. Даний результат досягається з припущенням, що різні вибірки матимуть різні концентрації компонентів. Дані моделі можна узагальнити припущенням про змінні концентрації компонент у суміші.

Дослідження сумішей стосуються не лише побудови оцінок параметрів розподілу, а й регресійного аналізу. Інколи модель регресійної суміші називають “switching regression model”. Одні з перших кроків у цьому напряму зробили R. DeVeaux [21] (1986), і W. DeSarbo, R. Cron [19]. R. DeVeaux в роботах [21], [22], в якій він розглядає двокомпонентну регресійну суміш нормальних розподілів, із двовимірними регресорами. В наведених статтях змішуючі ймовірності однакові для різних об'єктів спостереження. В статті [21] автором доведено консистентність оцінок параметрів розподілу та параметрів регресії, що були побудовані (EM) алгоритмом. (EM) алгоритм потребує деяких пілотних оцінок. За допомогою методу моментів, R. DeVeaux [22] будує такі $\sqrt{n}$-консистентні пілотні оцінки параметрів. W. DeSarbo та R. Cron у [19] розглянули випадок багатовимірних регресорів, а для підгонки параметрів регресії також використали (EM) алгоритм. Пізніше, у роботі W. Desarbo , M. Wedel, [20] зроблено загальний огляд лінійної регресійної моделі з регресорами із експоненційного сімейства розподілів. У роботі P. Lenk і W. DeSarbo у [46] розглядається узагальнення моделі регресії, в якій для різних об'єктів робиться різна кількість спостережень, а розподіли регресорів належать до експоненційного сімейства. У [45] Kwon J. та Caramanis C. навели умови консистентності (EM)-оцінок параметрів регресійної суміші, та вказано порядок збіжності необхідної кількості кроків (EM)-алгоритму для досягнення заданої якості. Авторами Yao W., Wei Y., Yu C [92] продовжено дослідження [62] для випадку регресійної суміші. В цій статті авторами було зроблено припущення, що регресори мають t-розподіл Стьюдента, та додатково наведено метод оцінки параметру ступенів свободи t-розподілу.

Окрім моделей з незмінними концентраціями, також створено моделі регресійних сумішей зі змінними концентраціями. Такі моделі розглянуто у статтях B. Bettina, D. Leisch [28] і [29], S. Faria, G. Soromenho [25]. Автори дослідження [28] вивчали модель лінійної регресійної суміші зі змінними концентраціями компонент для нормальної суміші, а у [29] вони розглянули суміш для регресорів з розподілом із експоненційного сімейства розподілів. Додатково у дослідженні [28] робиться параметричне припущення про вигляд концентрацій кожного спостережуваного об'єкта в залежності від його характеристик та невідомих параметрів. У [28] та [29] запропоновано EM алгоритм як метод оцінювання невідомих параметрів регресії, параметрів розподілу похибок і характеристик тощо. Автори зазначили, як і у [12], що функціонал умовного математичного сподівання вірогідності можна максимізувати по параметрам регресії та по параметрам концентрацій окремо. У роботі [25] розглянуто EM-оцінки параметрів регресії і додатково побудовано їх дві модифікації: Класифікаційні та Стохастичні EM оцінки.

У випадках, коли відсутні припущення про розподіл характеристик у суміші також можливо будувати оцінки параметрів розподілів, як це показано у роботі R. Maiboroda, O. Sugakova, A. Doronin [57]. У статті розглядається семі-параметрична модель, і методом узагальнених оціночних рівнянь будуються оцінки моментів розподілів для різних компонентів суміші. Загальна теорія оціночних рівнянь викладена у розділі 5 книги J. Shao [76].

Побудова регресійних моделей можлива і з використанням семі-параметричної техніки. У роботі R. Maiboroda та D. Liubashenko [47] розглядається структурна модель лінійної регресійної суміші, у якій розподіли регресорів оцінюються непараметричним чином. У цій роботі також досліджено консистентність і нормальність отриманих оцінок. У роботі R. Maiboroda, O. Sugakova [51] розглядається непараметричний підхід до оцінювання у моделях ортогональної регресії, а у статті R. Maiboroda, H. Navara, O. Sugakova [54] доведено асимптотичні властивості оцінок методом ортогональної регресії.

В усіх раніше перелічених роботах робиться припущення про те, що кількість компонент суміші відома. Для випадку, коли це не так, G. Mclachlan [61] використовує тест відношення вірогідності для підбору числа компонент. Інший спосіб підбирати параметр числа компонент суміші наведено у статтях S. Richardson [73] і M. Stephens [86]. Обидві статті присвячені баєсівським методам оцінювання параметрів для сумішей з невідомим числом компонент, зокрема з використаннями метода Markov Chain Monte Carlo (MCMC). У роботі [58] використано метод максимізації функції вірогідності зі штрафним функціоналом для оцінки числа компонент. В цій же роботі наведено умови консистентності цієї оцінки, та наведено приклади для моделі нормальних сумішей, та моделі сумішей пуассонівських розподілів.

У дослідженні асимптотичної нормальності оцінок регресії важливо вміти оцінювати матрицю коваріації цих оцінок, оскільки не завжди вдається описати її граничний вигляд. У розділі 5.5 книги J. Shao [76] наведено декілька способів оцінки цієї матриці: метод підстановки, бутстреп, та складаний ніж (CH). Метод CH описано у роботі M. Quenouille [71], в якій його використано для корекції зміщення оцінок параметрів моделей. Далі CH розглянуто у роботах J. Tukey [90] і книзі Efron B. “The Jackknife, the Bootstrap, and Other Resampling Plans” [24]. У статті J. Shao [77] розглянуто метод складного ножа для зменшення дисперсії

М-оцінок, а у публікації [78] J. Shao довів консистентність оцінок найменших квадратів та оцінки коваріації СН для нелінійної регресії. Загальна теорія СН для оцінки дисперсії викладена у статті J. Shao [81], і також у книгах "Jackknife and bootstrap" J. Shao, D. Tu [79], "Mathematical Statistics" J. Shao [76]. Останні дослідження A. McIntosh [65] вивчають незміщенність і асимптотичні розклади оцінок, що отримані методом складаного ножа.

Метод СН окрім застосування у статистиці знайшов застосування і у області машинного навчання. Однією із задач машинного навчання є задача підбору регресійної моделі з деякої множини моделей, і одночасно перевіряти її на спостереженнях, які були вилучені з навчальної вибірки. J. Shao [82] (1997) описав і довів консистентність оцінок параметрів лінійних моделей регресії (МНК), які підбиралися так, щоб мінімізувати функціонал CV-1, який нагадує функціонал (МНК). Відмінність наведеного методу від класичного (МНК) в тому, що для кожного спостереження його залишок рахувався за лінійною моделлю, параметри якої були підігнані на вибірці, з якої цей об'єкт вилучили. У роботі [32] B. Hansena та J. Racine продовжили дослідження [82]. У [32] досліджуються асимптотичні властивості ансамблю лінійних регресійних моделей, для оцінки параметрів якого використовувався метод СН. Подібну модель ще називають "стековим узагальненням". Ця ідея була висловлена у роботі Wolpert D. [91] (1992), а Breiman L. [16] серіями моделювань показав, що стекові регресійні моделі дають значно менші значення функцій похибок на раніше невідомих даних, ніж аналогічні регресійні моделі без використання ансамблів. В одній з останніх робіт на цю тему, Kessy S., Sherris M. [43] (2021) застосували стекові регресійні ансамблі на практичній задачі, - передбачення смертності у Англії та Уельсі під час епідемії SARS-CoV-2.

1.2 Моделі суміші зі змінними концентраціями

Із [5] використаємо означення суміші. Позначимо об'єкт, що досліджується, літерою O. У об'єкта O присутня неспостережувана характеристика $\kappa(O) \in \{1, ..., M\}$ - номер компонента (популяції), до якого об'єкт належить. Також, у об'єкта O є спостережувані дані. Наприклад, для нього відомі ймовірності належати до фіксованого компонента $\mathbb{P}\{\kappa(O) = k\}$. Окрім цих ймовірностей, для об'єкта O, також спостерігається вектор характеристик $\xi(O) = \left(\xi^1(O), ..., \xi^D(O)\right)^T$. Розподіл цього вектора залежить від номера компонента $\kappa(O)$. Символом F_m, $1 \leqslant m \leqslant M$ позначимо розподіл характеристик об'єкта з m - го компонента:

$$F_m(A) = P\Big(\xi(O) \in A|\kappa(O) = m\Big),$$

для довільної вимірної $A \subset \mathbb{R}^D$.

Розглянемо об'єкти $O_{j;n}$, $j = 1, ..., n$, які належать різним компонентам з різними відомими ймовірностями $p^k_{j;n} = P\{\kappa(O_{j;n}) = k\}$, $\kappa_j = \kappa(O_{j;n})$. Символом $p^k_{j;n}$ позначатимемо ймовірність j-го об'єкта належати компоненту із номером k.

Символом Ξ_n позначатимемо вибірку, розміром n:

$$\Xi_n = \Big(\xi(O_{j;n})\Big)_{j=1}^n = \Big(\xi_{j;n}\Big)_{j=1}^n.$$

Ми припускаємо, що випадкові величини $(\xi_{j;n}, \kappa_j)$ незалежні при різних j. Символом P_n будемо позначати матрицю концентрацій компонент

$$P_n = \Big(p^k_{j;n}\Big)_{j=1,k=1}^{n,M}.$$

При різних $j = 1, ..., n$, $\xi_{j;n}$ мають наступний розподіл

$$\mathbb{P}\{\xi_{j;n} \in A\} = \sum_{m=1}^M p^m_{j;n} F_m(A),$$

для довільної вимірної $A \subset \mathbb{R}^D$.

1.2.1 Навантажені функції розподілу і їх виправлення

Для F_m ми розглядаємо непараметричну модель. Навантаженою емпіричною функцією розподілу (н.е.ф.р) з ваговими коефіцієнтами $a_{j;n}$ називають

$$\hat{F}_n(x, a) := \sum_{j=1}^n a_{j;n} I[\xi_{j;n} < x].$$

Прикладом такої (н.е.ф.р) є оцінки функцій розподілу F_m, із $a^m = (a^m_{1;n}, ..., a^m_{n;n})^T$

$$\hat{F}^{(m)}_n(x, a^m) := \sum_{j=1}^n a^m_{j;n} I[\xi_{j;n} < x], \tag{1.1}$$

з навантаженнями $a^m_{j;n}$, які визначені наступним чином

$$A_n = \Big(a^k_{j;n}\Big)_{j=1,k=1}^{n,M} = P^T_n (\Gamma_n)^{-1}. \tag{1.2}$$

Γ_n - матриця Грама для системи векторів, що утворюють P_n

$$\Gamma_n = P_n P_n^T = (\langle p^k, p^l \rangle)_{k,l=1}^M.$$

Помітимо, що матриця A_n є правою оберненою до матриці P_n, тобто наступна матриця є діагональною:

$$P_n A_n = \Big(\sum_{i=1}^n a_{i:n}^k p_i^t \Big)_{t,k=1}^M = \mathbb{E}_M \,.$$

У статті [51] показано, що $\hat{F}_n^{(m)}$ - мінімаксна оцінка для $F^{(m)}$. Насправді, функція $\hat{F}_n^{(m)}$ з (1.1) не є функцією розподілу. Серед навантажень $a_{j;n}^m$ обов'язково зустрічаються від'ємні значення, і тому $\hat{F}_n^{(m)}(x, a)$ не є монотонно неспадною. Оцінку $\hat{F}_n^{(m)}(x, a)$ можна використати для оцінки моментів розподілу компонентів суміші. Ці оцінки функцій розподілів можна використати для оцінювання математичних сподівань і дисперсії випадкових величин у моделі суміш, проте деякі навантаження $a_{j;n}^k$ є від'ємні і тому є можливість отримати від'ємні значення оцінок дисперсії, що є нонсенс. Саме тому треба ввести виправлені навантаження, які дадуть оцінки дисперсії із додатними значеннями.

Виправлення навантажень. Різні методи виправлення (н.е.ф.р) описано у роботі [5]. В цьому випадку, (н.е.ф.р) $\hat{F}_n^{(m)}(x, a)$ можна виправити і побудувати навантажену емпіричну функцію розподілу, поклавши

$$\hat{F}_n^{(m),+}(x, a) = \min\left(1, \sup_{y<x} \hat{F}_n^{(m)}(y, a)\right) = \sum_{j=1}^n b_{j;n}^{m,+} I[\xi_{j;n} < x].$$

Обмеження зверху одиницею є обов'язковим, оскільки $\hat{F}_n^{(m)}(x, a)$ може приймати значення більші 1.

Вираз $\hat{F}_n^{(m),+}(x, a)$ є “верхньою огинаючою” для функції $\hat{F}_n^{(m)}(x, a)$. Аналогічно, можна побудувати і “нижню огинаючу”, яка також буде функцією розподілу

$$\hat{F}_n^{(m),-}(x, a) = \max\left(0, \lim_{y\to x} \inf_{z>y} \hat{F}_n^{(m)}(z, a)\right) = \sum_{j=1}^n b_{j;n}^{m,-} I[\xi_{j;n} < x].$$

Навантаження $b_{j;n}^{m,+}$ і $b_{j;n}^{m,-}$ - виправлені навантаження $a_{j;n}^m$. Далі ці навантаження буде використано для оцінок параметрів суміші зі змінними концентраціями, зокрема для оцінювання дисперсій.

1.3 Моделі регресії і оцінки МНК

У даному підрозділі, ми розглянемо модель регресійної суміші зі змінними концентраціями, та наведемо теореми про консистентність і асимптотичну нормальність оцінок параметрів регресії зі статті R. Maiboroda та D. Liubashenko [47]). Найбільш поширеним способом побудови оцінок параметрів регресії є "метод найменших квадратів"(МНК), але є і інші (метод найменших модулів). Один із таких є метод ортогональної регресії (регресія Демінга), який описав W. Deming (див. розділ 4 книги [18]). Саме такі ортогональні моделі із класичною, та Берксонівською похибкою у регресорах розглянуто в роботах I. Fazekas, A. Kukush [26], та S. Masiuk, A. Kukush [59]. Авторами статті [26] доведено консистентність оцінок параметрів у моделі із Берксонівською похибкою для лінійної моделі, регресори в якій утворені нелінійними перетвореннями над спостережуваними характеристиками об'єктів. У роботі [59] також розглянуто і нелінійні моделі регресії, окрім лінійних. Щодо лінійних моделей, A. Kukush та S. Van Huffel у роботі [44] отримали умови строгої консистентності оцінок параметрів багатовимірної регресії в моделях із похибкою у регресорах. Для цієї моделі С. Шкляр у роботі [83] показав умови існування та єдиності оцінок регресії.

Ще один вид регресійної моделі - модель із автокорельованими похибками регресії, яку ще називають загальною моделлю (Generalized Least Squares). Вперше така модель розглянута A. Aitken (1935), а у роботах T. Kariya та Y. Toyooka [40], і T. Kariya, H. Kurata [41] доведено аналог теореми Гаусса-Маркова для (GLS) моделей. У роботі T. Kariya та H. Takeaki [39] оглянуто загальні (GLS) лінійні та нелінійні регресійні моделі.

Крім регресійного аналізу вибірок розглядають також задачі регресійного аналізу у моделях з неперервним часом, коли спостереження є випадковим процесом. Приклад такої моделі описано А. В. Івановим, та М. М. Леоненко у роботі "Statistical Analysis of Random Fields" [30]. Іванов, та Леоненко у [30] описали регресійне випадкове поле з нелінійною функцією регресії, а для оцінок методом найменших квадратів навели умови їх консистентності та асимптотичної нормальності. А. В. Іванов та І. В. Орловський у [1] продовжили вивчення даної моделі, та виписали умови строгої консистентності М-оцінок параметрів регресії, у [2] ними доведено асимптотичну нормальність, а у [3] наведено умови асимптотичної єдиності М-оцінки.

А. В. Іванов та І. В. Орловський у [31] продовжили дослідження регресійних полів для випадку великих значень дисперсій оцінок МНК регресійних параметрів, та виписали обмеження на ймовірності великих відхилень оцінок МНК від справжніх значень параметрів.

1.3.1 Однорідна вибірка

Модель лінійної регресії запропоновано у роботах Лежандра (1805) і Гаусса (1821/1823). Далі ці ідеї розвинуто у роботах Френсіса Галтона (1886) і роботі J. Kenney і E. Keeping [42]. Опишемо далі модель простої лінійної регресії.

У моделі простої лінійної регресії розглядається однорідна вибірка $\Xi_n = \left(\xi_j\right)_{j=1}^{n} = \left(X_j, Y_j\right)_{j=1}^{n}$, в якій спостереження $\xi_j \in \mathbb{R}^D$ незалежні між собою. Залежність відгуків $Y_j \in \mathbb{R}$ від регресорів $X_j \in \mathbb{R}^{D-1}$ має наступний вигляд:

$$Y_j = \langle X_j, b\rangle + \varepsilon_j.$$

$\langle X_j, b\rangle$ - скалярний добуток векторів, а ε_j - похибка вимірювання. $b \in \mathbb{R}^{D-1}$ - невідомі параметри регресії, які оцінюються за вибіркою Ξ_n. В найпростішому випадку ξ_j - н.о.р. вектори, а X_j і ε_j - незалежні. Розподіл ε_j однаковий для різних j (наприклад, нормальний $N(0, \sigma^2)$). Позначимо розподіл похибки наступним чином:

$$F_\varepsilon(t) = P\{\varepsilon < t\},\ t \in \mathbb{R}.$$

Існує багато способів оцінювати значення параметрів b: метод найменших квадратів (МНК), метод найменших модулів, метод ортогональної регресії. Вибір методу залежить від багатьох чинників. Найбільш популярним методом оцінювання є МНК, у якому вводиться функціонал найменших квадратів:

$$S(\Xi_n, \gamma) = \sum_{j=1}^{n} (Y_j - \langle X_j, \gamma\rangle)^2. \tag{1.3}$$

Оцінкою $\hat{b}_n$ параметрів регресії b методу МНК параметрів називається така вимірна функція від вибірки Ξ_n, при підстановці якої у (1.3) замість γ, функція $S(\Xi_n, \gamma)$ набуває свого найменшого значення, тобто

$$\hat{b}_n = \mathrm{argmin}_{\gamma\in\Theta} S(\Xi, \gamma).$$

Ця оцінка є єдиною точкою мінімуму функції (1.3). Позначимо матрицю $X = (X_1, ..., X_n)$, і вектор $Y = (Y_1, ..., Y_n)$. Якщо $\det X^T X \neq 0$, то оцінка МНК має наступний вигляд

$$\hat{b}_n = (X^T X)^{-1} X^T Y. \tag{1.4}$$

Вже давно відомі теоретичні результати про консистентність і асимптотичну нормальність оцінок методу МНК. Такі теореми наведено у J. Shao розд. 3.3 [76]. У наступних розділах ми розглянемо модель регресійної суміші, у якій оцінки параметрів мають вигляд, аналогічний (1.4).

1.3.2 Модель регресійної суміші

Раніше ми вже згадували модель суміші зі змінними концентраціями. У даному підрозділі буде розглянуто її частковий випадок, а саме модель регресійної суміші зі змінними концентраціями, яка описана у статті R. Maiboroda та D. Liubashenko [47].

В цій моделі спостереження $\xi_{j;n} = (X_j, Y_j) \in \mathbb{R}^D,\ j = 1, ..., n$ є незалежними при різних j і складаються із регресорів X_j, та відгуків Y_j. Припускається, що відгуки залежать від регресорів, і ця залежність різна для різних компонент. Наприклад,

$$Y_j = g(X_j, b^{(\kappa_j)}) + \varepsilon_j. \tag{1.5}$$

Тут ε_j - похибка вимірювання, $g : \mathbb{R}^{D-1} \times \Theta \to \mathbb{R}$ - функція регресії, а $b^{(m)} \in \Theta \subseteq \mathbb{R}^d$ - невідомі параметри регресії. Припускається, що X_j і ε_j незалежні.

Функцією розподілу спостереження $\xi_{j;n} = (X_j, Y_j)$ є

$$\mathbb{P}\{\xi_{j;n} \in A\} = \sum_{m=1}^{M} p_{j;n}^m F_m(A), \tag{1.6}$$

де F_m - розподіл на $\mathbb{R}^D$ (F_m - розподіл m-го компонента суміші), $p_{j;n}^m$ - ймовірність спостерігати m-тий компонент при j-му спостереженні. Ймовірності $p_{j;n}^m$ є відомими.

Для F_m ми розглядаємо семіпараметричну модель $F_m(A) = F(A; b^{(m)}, \nu^{(m)})$, де $b^{(m)} \in \Theta \subseteq \mathbb{R}^{D-1}$ - вектор невідомих коефіцієнтів регресії, а $\nu^{(m)} = (F_X^{(m)}, F_\varepsilon^{(m)})$ - це розподіли регресорів X_j і похибок регресії ε_j відповідно.

Для оцінювання невідомого евклідового параметра $b^{(m)}$ можна використати узагальнений метод найменших квадратів (МНК). Для цього записується узагальнений функціонал МНК

$$G_n^{(m)}(\gamma) = \sum_{j=1}^{n} a_{j;n}^m \left(Y_j - g(X_j, \gamma)\right)^2, \tag{1.7}$$

і шукається його точка мінімуму.

Означення. Така статистика $\hat{b}_n^{LS,(m)}$, що є вимірною функцією від Ξ_n, а також є точкою екстремуму (1.7), називається оцінкою методу МНК.

Оскільки серед навантажень $a_{j;n}^m$ є від'ємні значення, то точок екстремума може бути декілька. В такому випадку на роль оцінки МНК обирається будь-яка з них.

Деякі теореми, що описують граничну поведінку параметрів регресії $\hat{b}_n^{LS,(m)}$ потребують додаткових припущень про граничну поведінку навантажень A_n:

(Γ) $\exists \lim_{n\to\infty} n^{-1}\Gamma_n = \Gamma_\infty$ і $\det \Gamma_\infty > 0$

(a^2) $\exists \lim_{n\to\infty} n\langle (a_\cdot^k)^2 p_\cdot^t\rangle_n = \lim_{n\to\infty} n\sum_{i=1}^n a_{i:n}^{k2} p_i^t = \langle (a_\cdot^k)^2 p_\cdot^t\rangle_\infty > 0$ і
$\exists \lim_{n\to\infty} n\langle (a_\cdot^k)^2 p_\cdot^t p_\cdot^f\rangle_n = \lim_{n\to\infty} n\sum_{i=1}^n a_{i:n}^{k2} p_i^t p_i^f = \langle (a_\cdot^k)^2 p_\cdot^t p_\cdot^f\rangle_\infty > 0$, для довільних $1 \leqslant t, f \leqslant M$.

1.3.3 Асимптотична нормальність оцінки МНК у лінійній регресії

У статті R. Maiboroda та D. Liubashenko [47] доведено асимптотичну консистентність і нормальність оцінки $\hat{b}_n^{LS,(m)}$. Введемо позначення

$$D^{(k)} = \mathbb{E}\left[X(O)X^T(O)|\kappa(O) = k\right] = \mathbb{E}\,X_{(k)}(X_{(k)})^T,$$

та

$$\sigma_{(k)}^2 = \mathbb{E}\left[\varepsilon^2(O)|\kappa(O) = k\right] = \mathbb{E}\,{\varepsilon_{(k)}}^2$$

Твердження 1.1. *Припустимо, що*

1 $D^{(k)}$ та $\sigma_{(k)}^2$ скінченні для $1 \leqslant k \leqslant M$.

2 $D^{(k)}$ є невиродженою матрицею.

3 Виконується умова (Γ).

Тоді $\hat{b}_n^{LS,(k)} \xrightarrow{\mathbb{P}} b^{(k)}$, при $n \to \infty$

Для твердження про асимптотичну нормальність оцінок введемо наступні позначення:

$$D^{ik(s)} = \mathbb{E}\left[X^i(O)X^k(O)|\kappa(O) = s\right],$$

$$L^{ik(s)} = \Big(\mathbb{E}[X^i(O)X^k(O)X^q(O)X^l(O)|\kappa(O) = s]\Big)_{l,q=1}^{D-1},$$

$$M^{ik(s,p)} = \Big(D^{il(s)}D^{kq(p)}\Big)_{l,q=1}^{D-1}.$$

Твердження 1.2. *Припустимо, що*

1 $\mathbb{E}\left[\left(X^i(O)\right)^4|\kappa(O) = k\right] < \infty$ *та* $\mathbb{E}\left[(\varepsilon(O))^4|\kappa(O) = k\right] < \infty$ *для всіх* $1 \leqslant k \leqslant M$.

2 $D = D^{(k)}$ *є невиродженою матрицею.*

3 *Виконується умова* (Γ).

4 *Виконується умова* (a^2).

Тоді

$$\sqrt{n}(\hat{b}_n^{LS,(k)} - b^{(k)}) \xrightarrow{\mathrm{w}} N(0, V^{(k)}),\ n \to \infty.$$

Тут

$$V^{(k)} = (D^{(k)})^{-1}\Sigma(D^{(k)})^{-1}, \tag{1.8}$$

та $\Sigma = (\Sigma^{ij})_{i,j=1}^{D}$, *і*

$$\Sigma^{ij} = \sum_{s=1}^{M}\langle(a^k)^2p^s\rangle\Big(D^{ij(s)}(\sigma^{(s)})^2 + (b^{(s)} - b^{(k)})^T L^{ij(s)}(b^{(s)} - b^{(k)})\Big) - \tag{1.9}$$

$$-\sum_{s=1}^{M}\sum_{l=1}^{M}\langle(a^k)^2p^sp^l\rangle\Big(b^{(s)} - b^{(k)}\Big)^T M^{ij(s,l)}\Big(b^{(l)} - b^{(k)}\Big)$$

1.4 Оцінки максимальної вірогідності

У цьому підрозділі ми поставимо задачу оцінювання невідомих параметрів розподілу за однорідною вибіркою, щоб далі застосувати ці знання для побудови нових оцінок параметрів у моделі суміші за змінними концентраціями.

Розглядається деяка вибірка Ξ із щільністю розподілу f_θ відносно деякої міри ν, і $\theta \in \Theta \subset \mathbb{R}^k$. Для кожної вибірки Ξ, $f_\theta(\Xi)$ розглядається як функція від θ і називається функцією вірогідності. Позначимо її як $l(\theta) = f_\theta(\Xi)$

Означення. Вимірна функція від вибірки $\hat{\theta} = \hat{\theta}(\Xi) \in \Theta$, для якої

$$l(\hat{\theta}) = \sup_{\theta \in \Theta} l(\theta), \tag{1.10}$$

називається оцінкою максимальної вірогідності (ОМВ).

Якщо функція $l(\theta)$ диференційовна по θ і θ - внутрішня точка Θ, то можливі кандидати на оцінку ОМВ повинні задовольняти умову

$$\frac{\partial l(\theta)}{\partial \theta} = 0.$$

Для багатьох класичних розподілів, наприклад, нормального, оцінки параметрів можна записати в явному вигляді розв'язавши попереднє рівняння, проте для випадку моделі нормальної суміші таких оцінок не існує. Окрім попередньої проблеми, для більш складних розподілів не вдається виписати оцінки параметрів. Саме тому далі наведено дві процедури, які можна використати для оцінювання невідомих параметрів.

1.4.1 Оптимізаційна процедура Ньютона-Рафсона

Перший спосіб знайти таку оцінку ОМВ, для якої виконується (1.10) - використати метод Ньютона-Рафсона. Опис цього методу взято із книги [76], ст. 278.

Нехай далі, $\frac{\partial l(\theta)}{\partial \theta} = \left(\frac{\partial l(\theta)}{\partial \theta_1}, ..., \frac{\partial l(\theta)}{\partial \theta_D}\right)$ - градієнт функції $l(\theta)$ по параметрам θ, а $\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T}$ - гессіан функції $l(\theta)$ (невироджений для усіх $\theta \in \Theta$),

$$\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} = \begin{pmatrix} \frac{\partial^2 l(\theta)}{\partial \theta_1 \partial \theta_1} & \cdots & \frac{\partial^2 l(\theta)}{\partial \theta_1 \partial \theta_D} \\ \vdots & \cdots & \vdots \\ \frac{\partial^2 l(\theta)}{\partial \theta_D \partial \theta_1} & \cdots & \frac{\partial^2 l(\theta)}{\partial \theta_D \partial \theta_D} \end{pmatrix},$$

а $\theta^{(0)}$ - деяка пілотна оцінка θ.

Тоді оцінка вектора параметра θ методом Ньютона-Рафсона шукається наступною ітеративною процедурою, $t = 1, 2, ...$

$$\hat{\theta}^{(t+1)} = \hat{\theta}^{(t)} - \left[\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T}\Big|_{\theta=\hat{\theta}^{(t)}}\right]^{-1} \frac{\partial l(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}^{(t)}}.$$

Дана процедура припиняється за умови, коли $|\hat{\theta}^{(t+1)} - \hat{\theta}^{(t)}| < \delta$, для деякого наперед заданого $\delta > 0$.

1.4.2 EM алгоритм

Інший спосіб знайти оцінку максимальної вірогідності (1.10) - використати EM алгоритм. У статі J. Bilmes [12], розглядається параметрична модель суміші, проте концентрації вважаються однаковими для різних об'єктів спостереження. У наступних розділах ми розглянемо узагальнення цієї моделі для сумішей зі змінними концентраціями, а цей пункт присвячено опису EM алгоритму для оцінювання параметрів сумішей. Цей алгоритм зручно використовувати на практиці у задачах кластеризації, оскільки у певному сенсі є узагальненням методу k-means, та задачах сегментації зображень у задачах комп'ютерного зору. J. Bilmes [12] показує застосування на прикладі марковських ланцюгів для оцінки параметрів контекстновільних граматик (алгоритм Баума-Велша). Окрім цих задач, EM алгоритм широко застосовний у задачах факторизації матриці на невід'ємні компоненти (алгоритм латентного розміщення Діріхлє). Саме цю модель розглядають Hoffman M., Blei D., Bach F у [34], а Hoffman M., Blei D. та ін у [35] модифікують EM алгоритм для підгонки параметрів моделі. Модель (LDA) застосовується у так званому "тематичному моделюванні" на корпусах текстів, що зроблено Blei D.M., Ng A.Y., Jordan M.I. у [13]. Окрім роботи із текстами, EM алгоритм використовують у моделях суміші експертів. У статті [37] Jordan M., Jacobs. R використовують EM алгоритм для підгонки параметрів у моделі ієрархічної суміші експертів (ICE). Регресійна функція у моделі (ICE) є деревом рішень, де "експертні"нейронні мережі знаходяться у листах дерева, а шлюзові мережі - у вузлах дерева. Авторами статті [37] використано EM алгоритм для підгонки параметрів усіх мереж у моделі (ICE). У [38] Jordan M. і Xu L. довели консистентність оцінок параметрів для моделі суміші з експертів та моделі (ICE), у яких параметри моделей оцінювались EM алгоритмом.

Дамо означення EM алгоритму. Розглянемо випадкову величину $\xi = (x, y)$ та M - компонентну суміш. У парі ξ, x - спостережувана змінна, а y - номер компонента, до якого належить об'єкт. Від номера компонента залежить розподіл характеристики

x. У даній моделі припускається, що сукупна щільність ξ має такий вигляд

$$p(\xi|\theta) = p(x, y|\theta) = p(y|x, \theta)p(x|\theta).$$

Тут θ - невідомий параметр моделі. Набір $\Xi_n = (X, Y) = (X_j, Y_j)_{j=1}^n$ називатимемо повними даними, $X = (X_1, ..., X_n)$ - неповні, а $Y = (Y_1, ..., Y_n)$ - пропущені дані. Введемо функцію вірогідності на повних даних $L(\theta|X, Y) = p(X, Y|\theta)$, і функцію вірогідності X, $L(X|\theta)$. Задача полягає у оцінці невідомих параметрів θ, і для цього буде використано EM алгоритм, який складається із двох кроків:

Е-крок. Шукається математичне сподівання випадкової величини $L(\theta|X, Y)$, за умови, що Y - невідомі, а спостерігаються лише X, і якимось чином обрано пілотну оцінку $\theta^{(0)}$. Е-крок на i-й ітерації полягає у пошуку

$$Q(\theta, \theta^{(i-1)}) = \mathbb{E}[\log p(X, Y|\theta)|X, \theta^{(i-1)}].$$

Тут θ - параметр, по якому потрібно максимізувати функцію $Q(\theta, \theta^{(i-1)})$.

М-крок. На цьому кроці i-ї ітерації шукається нова оцінка параметрів θ^i наступним чином

$$\theta^{(i)} = \mathrm{argmax}_\theta Q(\theta, \theta^{(i-1)})$$

Ці два кроки потрібно повторювати, доки не буде досягнена умова збіжності. Прикладом такої умови може бути умова $|\theta^{(i)} - \theta^{(i+1)}| < \varepsilon$, для наперед заданого малого $\varepsilon > 0$.

У статті [23] наведено умови, за яких на кожній ітерації EM алгоритму значення функції $Q(\theta, \theta^{(i)})$ зростатиме, а оцінки $\theta^{(i)}$ збігатимуться до точок локального максимума функції вірогідності повних даних.

Наведений EM алгоритм буде нами використано для оцінки параметрів регресійної суміші зі змінними концентраціями.

1.4.3 Умови регулярності (RR) для моделі суміші зі змінними концентраціями

Нехай далі розглядається вибірка $\Xi = (\xi_{j;n})_{j=1}^n = (X_j, Y_j)_{j=1}^n$, яка описується моделлю регресійної суміші. $P_n = \left(p_{j;n}^m\right)_{j=1,k=1}^{n,M}$ - матриця концентрацій компонент, і $f_\theta(\xi_{j;n}, p_{j;n})$ - щільність розподілу j-го спостереження. Введемо логарифмічну функцію вірогідності для невідомих параметрів θ за вибіркою $\Xi = (\xi_{j;n})_{j=1}^{n} = (X_j, Y_j)_{j=1}^{n}$:

$$L(\theta) = \sum_{j=1}^{n} L(\xi_{j;n}, p_{j;n}, \tau),$$

тут $p_{j;n} = (p_{j;n}^1, ..., p_{j;n}^M)^T$, і

$$L(\xi_j, p_{j;n}, \tau) = \ln(f_\tau(\xi_j, p_{j;n}))$$

Якщо вибірка Ξ описується моделлю нормальної регресійної суміші, то

$$L(\xi_{j;n}, p_{j;n}, \tau) = \ln(\sum_{m=1}^{M} p_{j;n}^m f_{X,m}(X_j, \nu_m) f_{\varepsilon,m}(Y_j - X_j^T b^m)),$$

$$f_{\varepsilon,m} = N(0, \sigma_m^2),\ f_{X,m} = N(\mu_m, \Sigma_m),$$

Введемо

$$I(p_{j;n}, \tau) = \mathbb{E}\, \frac{L(\xi_{j;n}, p_{j;n}, \tau)}{\partial \tau} \left(\frac{L(\xi_{j;n}, p_{j;n}, \tau)}{\partial \tau} \right)^T.$$

Умови регулярності (RR)

(A_0) $f_t(z, \mathrm{p}) \neq f_\theta(z, \mathrm{p})$, якщо $t \neq \theta$.

(A_C) Θ - компактна множина.

(A_u) Існує $g > 0$ та $c_1 < \infty$ такі, що для усіх $\tau, \tau + u \in \Theta$ маємо

$$\frac{r_i(u)}{u^2} \geqslant g I(p_{j;n}, \tau);\ \ I(p_{j;n}, \tau + u) < c_1 I(p_{j;n}, \tau),$$

$$r_j(u) = \int_{\mathbb{R}^D} \left(\sqrt{f_{\tau+u}(z, p_{j;n})} - \sqrt{f_\tau(z, p_{j;n})} \right)^2 dz$$

(R_1) Функції $\sqrt{f_\tau(z, p_{j;n})}$, $j = 1, ..., n$ є неперервно диференційовними по τ для усіх значень x. Матриці Фішера $I(p_{j;n}, \tau)$ неперервні у τ, та визначник матриці $I(\tau, n) = \sum_{j=1}^{n} I(p_{j;n}, \tau)$ не нуль.

(R_2) Функції $L(\xi_j, p_{j;n}, \tau)$ двічі неперервно диференційовні по τ, та

$$\sup_{\tau \in \Theta} |\ddot{L}(\xi_{j;n}, p_{j;n}, \tau)| \leqslant l_j(x),\ \ l_j = \int_{\mathbb{R}^D} l_j(x) f_\theta(x, p_{j;n}) dx < \infty.$$

Для моделі нормальної суміші умова (R_1) виконується. Для виконання умови (A_0) потрібно припустити, що для усіх $j = 1, ..., n$ $p_{j;n}^k \neq p_{j;n}^m$ при $k \neq m$.

1.5 Оцінка складаного ножа

Для практичного застосування підігнаних регресійних моделей недостатньо лише вказати оцінки параметрів регресії. Важливо також вміти перевіряти гіпотези про ці параметри та про відповідність даних використаної моделі. Існує ряд загальних методів оцінювання коваріаційних матриць оцінок, з яких ми будемо використовувати метод "складаного ножа", який запропоновано у роботах M. Quenouille [71] і J. Tukey [90].

Нехай, далі, $\hat{b}_n$ - оцінка параметрів регресії на вибірці Ξ_n, а $\hat{b}_{i-;n}$ - оцінка, що порахована на вибірці Ξ_n, з якої вилучили i-те спостереження.

Позначимо

$$\hat{V}_n = n\sum_{i=1}^{n}\left(\hat{b}_{i-;n}-\hat{b}_n\right)\left(\hat{b}_{i-;n}-\hat{b}_n\right)^T.$$

Матрицю $\hat{V}_n$ будемо називати оцінкою методом складаного ножа для асимптотичної матриці коваріації $V_{(k)}$ (1.8) розподілу оцінок $\hat{b}_n$.

Більше уваги цим оцінкам, а зокрема умовам консистентності, буде приділено на 4-му розділі.

1.6 Допоміжні твердження

У роботі Р. Майбороди, та О. Сугакової [5] описана теорема про консистентність оцінок моментів у моделі сумішей. Нехай $g : X \to \mathbb{R}$ - невипадкова функція. Тоді функціональний момент для k-го компонента

$$\bar{g}_k = \int g(x)F_k(dx) = \mathbb{E}\, g(\xi_{\cdot}^{(k)})$$

може бути оцінений як

$$\hat{g}_{k:n} = \int g(x)\hat{F}_k(dx) = \sum_{j=1}^{n} a_{j:n}^{k} g(\xi_j).$$

Тут

$$\hat{F}_k(A) = \sum_{j=1}^{n} a_{j:n}^{k} I[\xi_i \in A]$$

є оцінкою (н.е.ф.р) $F_k(A)$.

Твердження 1.3. *Припустимо, що для усіх* $k = 1, ..., M$

1 $\bar{g}_k$ *існують.*

2 $\sup_{j,n} |a^k_{j:n}| < \infty$.

Тоді $\hat{g}_{k:n} \xrightarrow{\mathbb{P}} \bar{g}_k,\ n \to \infty$

Твердження 1.4. *(Теорема Ляпунова, [76] с.67). Нехай* ξ_i *- незалежні випадкові величини. Позначимо* $s_n^2 = \sum_{i=1}^n \sigma_i^2$ *і* $\mu_i = \mathbb{E}\,\xi_i$. *Якщо для деякого* $\delta > 0$ *виконується умова*

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}\,|\xi_i - \mu_i|^{2+\delta} = 0,$$

то має місце слабка збіжність

$$\tfrac{1}{s_n} \textstyle\sum_{i=1}^n (\xi_i - \mu_i) \xrightarrow{\text{W}} N(0,1)$$

Наступне твердження із R. Maiboroda, O. Sugakova [51] описує асимптотичні властивості мінімаксних навантажень.

Твердження 1.5. *За виконання умови* (Γ)*, мають місце наступні відношення:*

1 $\sup_{\substack{j=\overline{1,n},\\ k=\overline{1,M}}} |a^k_{j:n}| = O(n^{-1})$

2 $\sup_{\substack{i,j=\overline{1,n}, i\neq j,\\ k=\overline{1,M},}} |a^k_{j:n} - a^k_{j:i-}| = O(n^{-2})$

Наступне твердження взято із [55] і буде корисно для доведення консистентності оцінок матриці коваріації у моделі лінійної суміші.

Нехай далі вибірка $\Xi = (\xi_1, ..., \xi_n)$ описується моделлю суміші зі змінними концентраціями, та $\mathbb{E}\,\xi_{(k)} = \mu^{(k)}$. Розглянемо невідомий параметр компонента розподілу $F_\xi^{(k)}$, який можна записати у вигляді $\nu^{(k)} = H(\mu^{(k)})$. Тут $H : \mathbb{R}^d \to \mathbb{R}^q$ - деяка відома функція. Оцінку параметра $\nu^{(k)}$ позначимо $\hat{\nu}^{(k)}_{;n} = H(\sum_{i=1}^n a^{(k)}_{j;n} \xi^{(k)}_{j;n})$. Також у статті [55] вводяться матриці

$$V_\infty = V_\infty^{(k)} = H'(\mu^{(k)})\Sigma_\infty^{(k)}(H'(\mu^{(k)}))^T,$$

$$H' = \begin{pmatrix} \frac{\partial H^1}{\partial \mu^1} & \dots & \frac{\partial H^1}{\partial \mu^d} \\ \dots & \dots & \dots \\ \frac{\partial H^q}{\partial \mu^1} & \dots & \frac{\partial H^q}{\partial \mu^d} \end{pmatrix}.$$

$$\Sigma_{\infty}^{(k)} = \sum_{m=1}^{M} \langle (a^{(k)})^2 p^{(m)} \rangle_{\infty} \Sigma^{(m)} - \sum_{m.l=1}^{M} \langle (a^{(k)})^2 p^{(m)} p^{(l)} \rangle_{\infty} \mu^{(m)} (\mu^{(l)})^T,$$

$$\Sigma^{(m)} = \mathbb{E}\, \xi_{(m)} (\xi_{(m)})^T.$$

Матрицею $\hat{V}_{;n}^{(k)}$ позначимо оцінку асимптотичної матриці коваріації методом "складаного ножа"для оцінки $\hat{v}_{;n}^{(k)}$.

Твердження 1.6. *Припустимо, що виконуються наступні умови*

1 H є двічі неперервно диференційовною у деякому околі $\mu^{(k)}$.

2 Існує деяка $\alpha > 0$, така що $\mathbb{E}\,|\xi_{(m)}|^{\alpha} < \infty$ для усіх $m = 1, ..., M$.

3 Виконується умова (Γ).

4 Виконується умова (a^2).

Тоді

$$\hat{V}_{;n}^{(k)} \to V_{\infty}^{(k)}.$$

Наступне твердження - Теорема 2.2.4 із [5]. Розглянемо модель суміші зі змінними концентраціями і вибірку $\Xi_n = (\xi_{1;n}, ..., \xi_{n;n})$. Позначимо

$$\bar{F}_n^{(m)}(x, a) = \sum_{m=1}^{M} \langle a p^m \rangle_n F^{(m)}(x),$$

де $F^{(m)}$ - розподіл спостережуваних характеристик $\xi_{(m)}$ для m-го компонента, p^m - вектор концентрацій $m-$го компонента, $\hat{F}_n^{(m)}(x, a)$ - (з.е.ф.р) для розподілу $F^{(m)}(x)$,

$$\hat{F}_n^{(m)}(x, a) = \sum_{j=1}^{n} a_j I[\xi_{j;n} < x]$$

Нехай $(\mathcal{X}, \mathcal{U})$ - вимірний простір спостережень, а $\mathcal{S} \subseteq \mathcal{X}$ - деякий клас вимірних множин.

Твердження 1.7. *Якщо $\mathcal{S}$ є класом Вапника-Червоненкіса, то для довільного масиву навантажень $a = (a_1, ..., a_n)$ існує така випадкова величина $\Lambda < \infty$ м.н., що для довільної $A \in \mathcal{S}$*

$$\sup_x |\hat{F}_n^{(m)}(A, a) - \bar{F}_n^{(m)}(A, a)| \leqslant \Lambda \sqrt{\frac{\ln n}{n}}$$

Зокрема це твердження виконується для навантажень $a^k = (a_{1;n}^k, ... a_{n;n}^k)$

Розділ 2

Лінійна регресійна модель

У цьому розділі розглядається модель суміші лінійних регресій зі змінними концентраціями компонентів. Побудовані семі-параметричні оцінки невідомих параметрів регресії для випадку лінійної регресії. Для лінійної функції залежності розглянуто параметричні та непараметричні моделі суміші. Для оцінок параметрів цих моделей доведено консистентність і асимптотичну нормальність, після чого оцінюються матриці коваріації оцінок в обох моделях для побудови довірчих еліпсоїдів оцінок параметрів.

Результати цього розділу надруковані в статтях [49], [66], [67], [68].

2.1 Непараметричний підхід до лінійної регресії у сумішах

Далі ми розглядаємо модель регресійної суміші зі змінними концентраціями, де $\Xi_n = (X_j, Y_j)_{j=1}^n$ - вибірка, а (X_j, Y_j) - пара з регресора і відгуку, $Y_j \in \mathbb{R}$, та $X_j = (X_j^1, ..., X_j^{D-1}) \in \mathbb{R}^{D-1}$, $j = 1, ..., n$. Розподіл характеристик (X_j, Y_j) залежить від номера компонента, до якого належить j-й об'єкт. Номер компонента j-го об'єкта невідомий, проте відомі ймовірності належати k-му компоненту суміші. Позначимо цю матрицю ймовірностей $P_n = \left(p_{j;n}^k\right)_{j=1,...,n}^{k=1,...,M}$

Побудову методу регресійного аналізу почнемо із лінійної функції залежності між регресорами $X_{(m)}$ та відгуками $Y^{(m)}$, які належать m-му компоненту. Запишемо лінійну функцію регресії

$$Y_j = \sum_{l=1}^{D-1} X_j^l \, b_j^{(\kappa_j)} + \varepsilon_j$$

Позначимо матрицю значень регресорів $X = \left(X_i^d\right)_{i=1,d=1}^{n,D-1}$, вектор відгуків $Y = (Y_i)_{i=1}^n$. Введемо діагональну матрицю навантажень k-го компонента $A_n^{(k)}$, в якій на головній діагоналі розташовані мінімаксні навантаження $(a_{1:n}^k, ..., a_{n:n}^k)$, визначені (1.2).

У цьому випадку

$$\hat{b}_n^{LS,(k)} = (X^T A_n^{(k)} X)^{-1} X^T A_n^{(k)} Y, \tag{2.1}$$

є оцінкою вектору параметрів регресії $b^{(k)}$.

2.2 Оцінки у параметричній моделі, узагальнений метод найбільшої вірогідності

2.2.1 Параметричний підхід до лінійної регресії у сумішах

У цьому підрозділі ми розглянемо параметричний підхід до оцінювання $b^{(k)}$. Тут зробимо додаткові припущення про розподіл спостережуваних характеристик X та похибки ε. Припустимо, що існують щільності $f_{X,m}(x) = f(x, \nu_m)$ та $f_{\varepsilon,m}(t) = f_\varepsilon(t, \nu_m)$ і $\nu_m \in \Theta \subset \mathbb{R}^L$, такі, що для довільної вимірної A:

$$F_{X,m}(A) = \int_A f_{X,m}(x)dx; \;\; F_{\varepsilon,m}(A) = \int_A f_{\varepsilon,m}(x)dx$$

У найбільш популярній параметричній моделі суміші ці щільності нормальні, тобто

$$f_{\varepsilon,m} = N(0, \sigma_m^2), \;\; f_{X,m} = N(\mu_m, \Sigma_m), \tag{2.2}$$

тут $\mu_m \in \mathbb{R}^D$ - вектор середніх m-го компонента та $\Sigma_m \in \mathbb{R}^{D\times D}$ - матриця коваріації. Зазвичай всі ці параметри невідомі. В даному випадку невідомі параметри це:

$$\nu_m = (\mu_m, \Sigma_m, \sigma_m^2), \;\; 1 \leqslant m \leqslant M.$$

Для зручності позначимо $\beta = (b^{(m)}, m = 1, ..., M, m \neq k)$, та $\nu = (\nu_1, ..., \nu_M)$ і об'єднаємо їх у вектор параметрів $\tau = (b^{(k)}, \beta, \nu)$. Тут β, ν - параметри, що заважають, а $b^{(k)}$ - ті, що нас цікавлять. Введемо логарифмічну функцію вірогідності для невідомих параметрів τ за вибіркою Ξ_n:

$$L(\tau) = \sum_{j=1}^n L(\xi_{j;n}, p_{j;n}, \tau),$$

тут $p_{j;n} = (p^1_{j;n}, ..., p^M_{j;n})^T$, і

$$L(\xi_{j;n}, p_{j;n}, \tau) = \ln\left(f_\tau(\xi_{j;n}, p_{j;n})\right) = \ln\left(\sum_{m=1}^{M} p^m_{j;n} f_{X,m}(X_j, \nu_m) f_{\varepsilon,m}(Y_j - X_j^T b^m)\right).$$

Означення: Оцінкою максимальної вірогідності (MLE) називається

$$\tau_n^{MLE} = \mathrm{argmax}_\tau L(\tau),$$

де максимум береться по всім можливим значенням τ.

На жаль, ці оцінки не можуть бути використані у багатьох розповсюджених параметричних моделях сумішей, оскільки $L(\tau)$ не обмежена зверху. Наприклад, це так для нормальної моделі суміші. У цій моделі $L(\tau) \to \infty$, при $\sigma_1^2 \to 0$ та $Y_1 - X_1^T b^{(1)} = 0$ з фіксованими іншим параметрами. Зазвичай цю проблему можна вирішити за допомогою однокрокової ОМВ, яка може бути представлена як одна ітерація алгоритму Ньютона-Рафсона для приблизного розрахунку ОМВ із якимось вибором початкових оцінок. Точніше, нехай $\hat{\tau}_n^{(0)}$ - початкова оцінка для τ. Розглянемо τ як вектор розмірності $P = (D-1) \times M + M \times L$ і перепозначимо елементи вектора через τ:

$$\tau = (\tau_1, ..., \tau_P).$$

Градієнт функції $L(\tau)$ позначимо через

$$s_n(\tau) = \frac{\partial L(\tau)}{\partial \tau} = \left(\frac{\partial L(\tau)}{\partial \tau_1}, ..., \frac{\partial L(\tau)}{\partial \tau_P}\right)^T$$

та Гессіан $L(\tau)$ через

$$\gamma_n(\tau) = \left(\frac{\partial^2 L(\tau)}{\partial \tau_i \partial \tau_l}\right)_{i,l=1}^{P,P} = \begin{pmatrix} \frac{\partial^2 L(\tau)}{\partial \tau_1 \partial \tau_1} & \dots & \frac{\partial^2 L(\tau)}{\partial \tau_1 \partial \tau_P} \\ \vdots & \dots & \vdots \\ \frac{\partial^2 L(\tau)}{\partial \tau_P \partial \tau_1} & \dots & \frac{\partial^2 L(\tau)}{\partial \tau_P \partial \tau_P} \end{pmatrix}, \tag{2.3}$$

Тоді однокрокова оцінка параметру τ, із початковим приближенням $\hat{\tau}^{(0)}$ визначається наступним чином:

$$\tau_n^{OS} = \hat{\tau}^{(0)} - \left(\gamma_n(\hat{\tau}^{(0)})\right)^{-1} s_n(\hat{\tau}^{(0)})$$

Теорема 4.19 у [76] описує загальні умови, в яких оцінка τ_n^{OS}, що побудована на незалежних та однаково розподілених спостереженнях, є консистентною, асимптотично нормальною, та асимптотично ефективною. Граничний розподіл однокрокових оцінок такий самий, як і у консистентної версії ОМВ. Отже, якщо припущення теореми 4.19 виконуються, то немає необхідності використовувати ітеративну процедуру для отримання оцінок з асимптотично оптимальними властивостями. Проте, на вибірках помірного розміру, оцінки τ_n^{OS} може бути недостатньо. Інший поширений спосіб отримати стійкі оцінки τ - це використати деякі версії ЕМ алгоритму, який описан у статті Білмса Дж. (1998) [12]. Узагальнений ЕМ-алгоритм - це ітеративна процедура для наближеного розрахунку ОМВ, коли спостерігається не вся інформація про досліджувані об'єкти. Опишемо тут лише алгоритм розрахунку ЕМ-оцінок $\hat{\tau}_n^{EM}$ у припущенні, що дані описуються моделлю суміші нормальних розподілів (2.2). Робота алгоритму починається із пілотних (початкових) оцінок

$$\hat{\tau}^{(0)} = \left(\hat{b}_m^{(0)}, \hat{\sigma^2}_m^{(0)}, \hat{\mu}_m^{(0)}, \hat{\Sigma}_m^{(0)}, m = 1, ..., M\right)$$

для повного набору параметрів моделі.

Далі для $i = 1, 2, ...$ оцінки перераховуються наступним чином. Припустимо, що отримано оцінки $\hat{b}_m^{(i)}, \hat{\sigma}_m^{2,(i)}, \hat{\mu}_m^{(i)}, \hat{\Sigma}_m^{(i)}, m = 1, ..., M$ на i-й ітерації. На цій же ітерації рахуються ваги

$$w_j^{m(i)} = w_j^m(\xi_j, \hat{\tau}^{(i)}) = \frac{p_{j;n}^m f_{X,m}\left(X_j; \hat{\mu}_m^{(i)}, \hat{\Sigma}_m^{(i)}\right) f_{\varepsilon,m}\left(Y_j - X_j^T \hat{b}_m^{(i)}; \hat{\sigma_m^2}^{(i)}\right)}{\sum_{l=1}^M p_{j;n}^l f_{X,l}\left(X_j; \hat{\mu}_l^{(i)}, \hat{\Sigma}_l^{(i)}\right) f_{\varepsilon,l}\left(Y_j - X_j^T \hat{b}_l^{(i)}; \hat{\sigma_l^2}^{(i)}\right)}.$$

Зауважимо те, що $w_j^{m(i)}$ - апостеріорні ймовірності $\mathbb{P}(\kappa_j = m|\xi_{j;n})$, які розраховані для $\tau = \hat{\tau}^{(i)}$.

Позначимо $\overline{w}^m = \sum_{j=1}^n w_j^{m(i)}$. Тоді оцінки для $i + 1$-ї ітерації визначаються наступним чином:

$$\hat{\mu}_m^{(i+1)} = \frac{1}{\overline{w}^m} \sum_{j=1}^n w_j^{m(i)} X_j,$$

$$\hat{\Sigma}_m^{(i+1)} = \frac{1}{\overline{w}^m} \sum_{j=1}^n w_j^{m(i)} (X_j - \hat{\mu}_m^{(i)})(X_j - \hat{\mu}_m^{(i)})^T,$$

$$\hat{b}_m^{(i+1)} = (\sum_{j=1}^n w_j^{m(i)} X_j X_j^T)^{-1} \sum_{j=1}^n w_j^{m(i)} Y_j X_j$$

$$\hat{\sigma_m^2}^{(i+1)} = \frac{1}{\overline{w}^m} \sum_{j=1}^{n} w_j^{m(i)} (Y_j - X_j^T \hat{b}^{m(i)})^2.$$

Повторення ітерацій закінчується при виконанні деякої кінцевої умови, якою може бути умова

$$|\hat{\tau}^{(i+1)} - \hat{\tau}^{(i)}| < \delta.$$

Тут $\delta > 0$ - наперед задана точність.

Відомо, що ця процедура дає стійкі оцінки, яка (для достатньо великих вибірок) збігається до точки локального мінімуму $L(\tau)$, що є найближчою для пілотної оцінки $\hat{\tau}^{(0)}$.

Асимптотична поведінка оцінок $\hat{\tau}_n^{OS}$ та $\hat{\tau}_n^{EM}$ може бути описана у термінах матриці інформації Фішера $\mathrm{I}^*(n,\tau) = \left(I_{i,l}^*(n,\tau)\right)_{i,l=1}^{P}$, тут

$$I_{i,l}^*(n,\tau) = \sum_{j=1}^{n} I_{i,l}(p_{j;n},\tau); \;\; I_{i,l}(\mathrm{p},\tau) = \mathbb{E}\, \frac{\partial L(\xi_{\mathrm{p}},\mathrm{p},\tau)}{\partial \tau_i} \frac{\partial L(\xi_{\mathrm{p}},\mathrm{p},\tau)}{\partial \tau_l},$$

тут $\mathrm{p} = (p^1, ..., p^M)$, $p_{j;n} = (p_{j;n}^1, ..., p_{j;n}^M)$, а ξ_{p} - випадковий вектор з щільністю

$$f_{\mathrm{p}}(y,x;\tau) = \sum_{m=1}^{M} p^m f(x;\nu_m) f_{\varepsilon}\left(y - x^T b^m; \nu_m\right). \qquad (2.4)$$

За умов регулярності теореми 70.5 [15],

$$\mathrm{I}^*(n,\tau)^{1/2}(\tau - \hat{\tau}_n^{MLE}) \xrightarrow{\mathrm{W}} N(0,\mathbb{E}),$$

тут $\mathbb{E}$ - одинична матриця розміру $P \times P$, а $\mathrm{I}^*(n,\tau)^{1/2}$ - квадратний корінь з додатно визначеної симетричної матриці $\mathrm{I}^*(n,\tau)$.

Умови регулярності включають припущення про обмеження вірогідності, тобто вони не виконуються для моделі нормальної суміші. Проте, якщо пілотні оцінки $\hat{\tau}_n^{(0)}$ є консистентними із порядком $\sqrt{n}$, то вимагається лише локальна версія умов регулярності, щоб довести асимптотичну нормальність оцінок $\hat{\tau}_n^{OS}$ та $\hat{\tau}_n^{EM}$. Іншими словами, умови регулярності повинні виконуватися у деякому околі справжнього значення параметру τ. Ці локальні умови регулярності виконуються для суміші нормальних розподілів, якщо $\sigma_m^2 > 0$, та Σ_m - невироджена для всіх $m = 1, ..., M$.

Щоб використати ці результати у побудові оцінок $b^{(k)}$, треба побудувати $\sqrt{n}$ консистентні пілотні оцінки для заважаючих параметрів і тих, що нас цікавлять.

Вони можуть бути отримані непараметричною технікою, що розглянута далі у цьому підрозділі. Для побудови довірчих еліпсоїдів нам також потрібні оцінки асимптотичної матриці коваріації $V^{(k)}$ із (1.8) для непараметричного випадку, та оцінки матриці інформації $\mathrm{I}^*(n,\tau)$ у параметричному випадку.

2.2.2 Оцінювання заважаючих параметрів

У цьому розділі йде мова про оцінювання матриць коваріацій оцінок регресії EM алгоритмом та непараметричною моделлю із навантаженнями $A_n^{(k)}$. Почнемо розділ із оцінювання асимпототичної матриці коваріації $V^{(k)}$ із Твердження 1.2. Фактично для цього знадобиться оцінити матриці D та Σ. Зауважимо, що матриця D відповідає k-му компоненту, тобто $D = D^{(k)} = \left(D^{is(k)}\right)_{i,s=1}^{D-1}$, де

$$D^{is(k)} = \mathbb{E}[X^i(O)X^s(O)|\kappa O = k].$$

За твердженням 4.2 із [51],

$$\hat{D}_n^{is(k)} = \sum_{j=1}^{n} a_{j:n}^k X_j^i X_j^s$$

є консистентною оцінкою для $D^{is(k)}$, за умови якщо $\mathbb{E}[|X(O)|^2|\kappa(O) = m] < \infty$ для всіх $m = 1, ..., M$ і виконується умова (Γ) : $\det \Gamma_\infty \neq 0$. Тому можна використати $\hat{D}_n^{(k)} = \left(\hat{D}_n^{is(k)}\right)_{i,s=1}^{D-1}$ як консистентну оцінку для матриці D при виконанні умови 1 Твердження 1.1.

Аналогічно, $L^{is(m)}$ може бути оцінена консистентною матрицею

$$\hat{L}^{is(m)} = \sum_{j=1}^{n} a_{j:n}^m X_j^i X_j^s X_j X_j^T$$

за виконання умов теореми 1.2.

Ту ж саму ідею можна використати для оцінювання σ_m^2 через

$$\hat{\sigma}_{m;n}^2 = \sum_{j=1}^{n} a_{j:n}^m \left(Y_j - X_j^T \hat{b}_n^{LS,(m)}\right)^2$$

Коефіцієнти $\langle (a_\cdot^k)^2 p_\cdot^s p_\cdot^q \rangle_\infty$ можуть бути наближені через

$$\hat{\alpha}_{s,q}^{(k)} = n \sum_{j=1}^{n} (a_{j:n}^{(k)})^2 p_j^s p_j^q.$$

Далі $V^{(k)}$ - асимптотична матриця коваріації оцінок $b^{(k)}$ з твердження 1.2.

Теорема 2.1. *Нехай виконуються наступні умови:*

1 Умова (Γ) *виконується.*

2 $\exists L^{(k)} < \infty$, $\exists M^{(k)} < \infty$, $\exists D^{(k)} < \infty$, $\exists \sigma^2_{(k)} < \infty$.

3 $\hat{b}_n^{LS,(k)} \xrightarrow{\mathbb{P}} b^{(k)}$, *при* $n \to \infty$.

Тоді

$$\hat{V}_n = \hat{D}_n^{-1} \hat{\Sigma}_n \hat{D}_n^{-1}$$

буде консистентною оцінкою $V^{(k)}$.

Доведення. З умов 1 і 2, і теореми 1.3, наступні оцінки будуть консистентними $\hat{D}_n^{(k)}$, $\hat{L}_n^{(k)}$, $\hat{M}_n^{(m)}$.

Покажемо, що оцінка $\hat{\sigma}^2_{k;n}$ буде консистентною. За означенням,

$$\hat{\sigma}^2_{k;n} - \sigma^2_{(k)} = \sum_{j=1}^{n} a^k_{j:n} \left[\left(Y_j - X_j^T \hat{b}_n^{LS,(k)} \right)^2 - \left(Y_j - X_j^T b^{(k)} \right)^2 \right] =$$

$$\hat{b}_n^{LS,(k)} \sum_{j=1}^{n} a^k_{j:n} X_j^T X_j (\hat{b}_n^{LS,(k)})^T - b^{(k)} \sum_{j=1}^{n} a^k_{j:n} X_j^T X_j (b^{(k)})^T \tag{2.5}$$

$$+2(b^{(k)} - \hat{b}_n^{LS,(k)}) \sum_{j=1}^{n} a^{(k)}_{j;n} Y_j X_j^T. \tag{2.6}$$

З умов 1, 2, 3, вирази 2.5 і 2.6 прямують до 0 за ймовірністю. Доведення теореми випливає із означення Σ (1.9). □

Щоб отримати пілотні оцінки параметрів нормальної суміші можна використати той же підхід. Власне, визначимо

$$\mu^{(0)}_{m,n} = \sum_{j=1}^{n} a^m_{j:n} X_j; \quad \hat{\Sigma}^{(0)}_{m,n} = \sum_{j=1}^{n} a^m_{j:n} \left(X_j - \mu^{(0)}_{m,n} \right) \left(X_j - \mu^{(0)}_{m,n} \right)^T$$

як оцінки для μ_m, та Σ_m.

З твердження 4.3 із [51], $\mu^{(0)}_{m,n}$; $\Sigma^{(0)}_{m,n}$ та $\hat{\sigma}^{2,(0)}_{m,n}$ є $\sqrt{n}$-консистентними оцінками для відповідних параметрів моделі нормальної суміші. Це дозволяє використати у ролі пілотних оцінок для одного кроку методом Ньютона-Рафсона та ЕМ алгоритму.

Тепер розглянемо оцінку матриці інформації Фішера у випадку моделі нормальної суміші. Визначимо

$$\hat{I}_{il}(n,\tau) = \sum_{j=1}^{n} \frac{\partial L(\xi_{j;n}, p_{j;n}, \tau)}{\partial \tau_i} \frac{\partial L(\xi_{j;n}, p_{j;n}, \tau)}{\partial \tau_l}, \tag{2.7}$$

$$\hat{\mathrm{I}}(n,\tau) = \left(\hat{I}_{il}(n,\tau)\right)_{i,l=1}^{P}, \quad \hat{\mathrm{I}}(n) = \hat{\mathrm{I}}(n,\hat{\tau}), \tag{2.8}$$

тут $\hat{\tau}$ - довільна консистентна оцінка для τ ($\hat{\tau}^{OS}$ або $\hat{\tau}^{EM}$).

У моделі нормальної суміші запишемо $\tau_{(l)} = (b^{(l)}, \mu_l, \Sigma_l, \sigma_l^2)$ як множину всіх невідомих параметрів, які описують l-тий компонент суміші.

Теорема 2.2. *Розглянемо модель нормальної суміші. Якщо виконуються такі умови:*

1 $\sigma_m^2 > 0$, *та* Σ_m *- невироджена для довільних* $m = 1, ..., M$.

2 $\exists c > 0 : \forall 1 \leqslant j \leqslant m, \forall m = 1, ..., M, \forall n \geqslant 1$

$$p_j^m > c.$$

3 $\tau_{(l)} \neq \tau_{(m)}$ *для усіх* $l \neq m; l, m = 1, ..., M$.

Тоді:

1 Існують $0 < c_0 < C_1 < \infty$ *такі, що*

$$c_0 n \leqslant |\mathrm{I}^*(n,\tau)| \leqslant C_1 n$$

для усіх $n \geqslant 1$.

2 $\frac{1}{n}\left|\mathrm{I}^*(n,\tau) - \hat{I}(n)\right| \xrightarrow{\mathbb{P}} 0, \ n \to \infty$.

Зауважмо, що тут і далі для довільних квадратних матриць I вираз $|\mathrm{I}|$, означає оператор норми

$$|\mathrm{I}| = \sup_{|u|=1} |u^T \mathrm{I} u|.$$

Доведення. Покажемо спочатку що матриця I додатно визначена, тобто

$$u^T \mathrm{I}(\mathrm{p}, \tau) u > 0$$

для довільних τ та довільних $u \in \mathbb{R}^P$ із $|u| = 1$ та для довільних $\mathrm{p} = (p^1, ..., p^M)$ із $p^m > c$ для довільних $m = 1, ...M$. Нагадаємо позначення $\tau = \left(\tau_{(1)}, ..., \tau_{(M)}\right)$, де $\tau_{(m)}$ відповідає параметрам, що описують m-й компонент. Розділимо вектор u на аналогічні блоки $u = \left(u^T_{(1)}, ..., u^T_{(m)}\right)^T$.

Тоді

$$u^T \mathrm{I}(\mathrm{p}, \tau) u = \mathbb{E}\, u^T \left(\frac{\partial L(\xi_\mathrm{p}, \mathrm{p}, \tau)}{\partial \tau}\right) \left(\frac{\partial L(\xi_\mathrm{p}, \mathrm{p}, \tau)}{\partial \tau}\right)^T u$$

$$= \mathbb{E}\left(u^T \frac{\partial L(\xi_\mathrm{p}, \mathrm{p}, \tau)}{\partial \tau}\right)^2$$

$$= \mathbb{E}\left(\sum_{m=1}^{M} u^T_{(m)} \frac{\partial L(\xi_\mathrm{p}, \mathrm{p}, \tau)}{\partial \tau}\right)^2.$$

За припущень нормальної моделі суміші $\frac{\delta L(\xi_\mathrm{p}, \mathrm{p}, \tau)}{\delta \tau}$ може бути представлена у вигляді

$$\frac{\partial L(\xi_\mathrm{p}, \mathrm{p}, \tau)}{\partial \tau} = \mathrm{A}\left(\tau_{(m)}, \xi_\mathrm{p}\right) \frac{p^m \phi_{\tau_{(m)}}(\xi_\mathrm{p})}{f_\mathrm{p}(\xi_\mathrm{p}, \tau)}, \tag{2.9}$$

тут $\phi_{\tau_{(m)}}$ - нормальна щільність спостереження ξ із m-го компонента, f_p - це щільність суміші (2.4), $\mathrm{A}\left(\tau_{(m)}, \xi_\mathrm{p}\right)$ - це вектор з елементів, що є поліноміальними функціями від ξ_p.

Далі

$$u^T \mathrm{I}(\mathrm{p}, \tau) u = \mathbb{E}\left(\sum_{m=1}^{M} p^m u^T_{(m)} \mathrm{A}(\tau_{(m)}, \xi_\mathrm{p}) \phi_{\tau_{(m)}}(\xi_\mathrm{p})\right)^2 \frac{1}{f_\mathrm{p}(\xi_\mathrm{p}, \tau)^2}. \tag{2.10}$$

Зауважимо, що за припущення 1 та 2 теореми $f_\mathrm{p}(\xi, \tau) > 0, \forall \xi \in \mathbb{R}^D$. Тоді $u^T_{(m)} \mathrm{A}\left(\tau_{(m)}, \xi_\mathrm{p}\right)$ - поліноми від ξ_p, та $\phi_{\tau_{(m)}}(\xi_\mathrm{p})$ є експонентами від різних та невироджених квадратичних форм від ξ_p за умовами 1) і 3) теореми.

Припустимо, що $u^T_{(m)} \mathrm{I}(\mathrm{p}, \tau) u_{(m)} = 0$ для деякого u з одиничною нормою, $|u| = 1$. З (2.10) маємо

$$u^T_{(m)} \mathrm{A}(\tau_{(m)}, \xi_\mathrm{p}) = 0 \text{ м.н.}. \tag{2.11}$$

для $m = 1, ..., M$.

З іншої сторони, з (2.11) випливає

$$\mathbb{E}\left(u_{(m)}^T \mathrm{A}(\tau_{(m)}, \xi_{\mathrm{p}})\phi_{\tau_{(m)}}(\xi_{\mathrm{p}})\right)^2 = u_{(m)}^T \mathrm{I}_{\tau_{(m)}} u_{(m)} = 0,$$

тут $\mathrm{I}_{\tau_{(m)}}$ - це матриця інформації Фішера для невідомого $\tau_{(m)}$ за одним спостереженням з m-го компонента. За припущенням 1, $\mathrm{I}_{\tau_{(m)}}$ невироджена, тож $u_{(m)} = 0$ для усіх $m = 1, ..., M$. Це суперечить припущенню, що $|u| = 1$.

З суперечності маємо, що $u^T \mathrm{I}(\mathrm{p}, \tau) u > 0$. Оскільки $u^T \mathrm{I}(\mathrm{p}, \tau) u$ неперервна функція на компактній множині від $u : \ |u| = 1$, та p задовільняє умову 2, то маємо що $u^T \mathrm{I}(\mathrm{p}, \tau) u > c_0$ для деякого $c_0 > 0$. З іншої сторони, з (2.9), випливає, що $|\mathrm{I}(\mathrm{p}, \tau)| < C_1$.

Отже, з виразу $\mathrm{I}^*(n, \tau) = \sum_{j=1}^n I(\mathrm{p}_j, \tau)$ маємо перше твердження теореми.

2. Для доведення другого твердження запишемо закон великих чисел для матриці інформації

$$\Delta_n(\tau) = \frac{1}{n}\left(\hat{\mathrm{I}}(n, \tau) - \mathrm{I}^*(n, \tau)\right)$$
$$= \frac{1}{n}\sum_{j=1}^{n}\Bigg[\frac{\partial L(\xi_j, p_{j;n}, \tau)}{\partial \tau}\left(\frac{\partial L(\xi_j, p_{j;n}, \tau)}{\partial \tau}\right)^T$$
$$- \mathbb{E}\,\frac{\partial L(\xi_j, p_{j;n}, \tau)}{\partial \tau}\left(\frac{\partial L(\xi_j, p_{j;n}, \tau)}{\partial \tau}\right)^T\Bigg] \xrightarrow{\mathbb{P}} 0, \ n \to \infty,$$

оскільки

$$\mathbb{E}\left|\frac{\partial L(\xi_j, p_{j;n}, \tau)}{\partial \tau}\right|^4 \leqslant C < \infty$$

для усіх j.

Нехай далі $B \subset \mathbb{R}^P$ - достатньо малий відкритий окіл точки τ. Зауважимо, що

$$\mathbb{E}\sup_{\tau \in B}\left|\frac{\partial}{\partial \tau}\frac{\partial L(\xi_j, p_{j;n}, \tau)}{\partial \tau}\left(\frac{\partial L(\xi_j, p_{j;n}, \tau)}{\partial \tau}\right)^T\right| < C_2 < \infty$$

для усіх j.

Розглянемо

$$\frac{1}{n}\left|\mathrm{I}^*(n, \tau) - \hat{I}(\hat{\tau})\right| \leqslant \Delta_n(\tau) + \frac{1}{n}\left|\hat{I}(n, \tau) - \hat{I}(\hat{\tau})\right|.$$

Перший доданок прямує до 0 за ймовірністю. Для того, щоб показати це і для другого доданку, внаслідок консистентності $\hat{\tau}$, достатньо розглянути випадок $\hat{\tau} \in B$. За теоремою про середнє значення,

$$\frac{1}{n}\left|\hat{I}(n,\tau)-\hat{I}(\hat{\tau})\right| \leqslant \frac{1}{n}\sum_{j=1}^{n}\sup_{\tau\in B}\left|\frac{\partial}{\partial\tau}\frac{\partial L(\xi_j,p_{j;n},\tau)}{\partial\tau}\left(\frac{\partial L(\xi_j,p_{j;n},\tau)}{\partial\tau}\right)^T\right|\left|\hat{\tau}-\tau\right|.$$

Залишилося показати, що вираз

$$B_n = \frac{1}{n}\sum_{j=1}^{n}\sup_{\tau\in B}\left|\frac{\partial}{\partial\tau}\frac{\partial L(\xi_j,p_{j;n},\tau)}{\partial\tau}\left(\frac{\partial L(\xi_j,p_{j;n},\tau)}{\partial\tau}\right)^T\right| = O_n(1). \tag{2.12}$$

Дійсно, для довільного $\varepsilon > 0$, існує $C_\varepsilon = \frac{C_2}{\varepsilon} > 0$, що

$$\sup_n P\{B_n \geqslant C_\varepsilon\} \leqslant \frac{\varepsilon}{C_2}\sup_n \mathbb{E}\, B_n \leqslant \frac{C_2}{C_2}\varepsilon = \varepsilon,$$

що за означенням і є (2.12). Теорему доведено. □

2.3 Довірчі еліпсоїди для $b^{(k)}$

Нехай Ξ_n - довільна вибірка розміру n з розподілом, що залежить від невідомого параметра $\mathrm{b} \in \mathbb{R}^d$. Множину $B_\alpha = B_\alpha(\Xi_n) \subset \mathbb{R}^d$ будемо називати асимптотично довірчою множиною з рівнем довіри α якщо

$$\lim_{n\to\infty} P\{\mathrm{b} \notin B_\alpha(\Xi_n)\} = \alpha$$

Далі ми побудуємо довірчу множину для вектора коефіцієнтів регресії $\mathrm{b} = b^{(k)}$ за вибіркою з суміші Ξ_n. У непараметричному випадку для побудови $B_\alpha(\Xi_n)$ ми скористаємось статистикою

$$S^{LS}(\beta) = n\left(\beta - \hat{b}_n^{LS,(k)}\right)^T \hat{V}_n^{-1}\left(\beta - \hat{b}_n^{LS,(k)}\right).$$

У параметричному випадку ми візьмемо матрицю $\hat{\mathrm{I}}(n)$, що задається (2.8) і розглянемо її обернену матрицю $\hat{\mathrm{I}}(n)^{-1} = \left(\hat{I}_{(i,m)}\right)_{i,m=1}^{R}$.

Зауважимо, що з (2.7) та (2.8), елементи $\hat{I}_{im}$ матриці $\hat{\mathrm{I}}(n)$ відповідають координатам τ_i та τ_m вектора невідомих параметрів τ. Ми будемо розглядати множину індексів l_m, $m = 1, ..., D-1$ таких, що $\tau_{l_m} = b_m^{(k)}$ та розглянемо матрицю

$$[\hat{\mathrm{I}}^{-1}]_{(k)} = \left(\hat{I}_{(l_i,l_m)}\right)_{i,m=1}^{D-1}.$$

Тобто матриця $[\hat{\mathrm{I}}^{-1}]_{(k)}$ складена з елементів $\hat{\mathrm{I}}^{-1}$, що відповідають елементам $b^{(k)}$.

Далі ми ще раз обертаємо матрицю:

$$\hat{\mathrm{I}}_k(n)^+ = \left([\hat{\mathrm{I}}^{-1}]_{(k)}\right)^{-1}.$$

Ця матриця використовується для побудови статистик, які визначають довірчі множини:

$$S^{OS}(\beta) = n\left(\beta - \hat{b}^{OS}(k,n)\right)^T \hat{I}_k(n)^+ \left(\beta - \hat{b}^{LS}(k,n)\right),$$
$$S^{EM}(\beta) = n\left(\beta - \hat{b}^{EM}(k,n)\right)^T \hat{I}_k(n)^+ \left(\beta - \hat{b}^{EM}(k,n)\right).$$

Тут $\hat{b}^{OS}(k,n)$ та $\hat{b}^{EM}(k,n)$ - частини оцінок τ_n^{OS} та τ_n^{EM}, що оцінюють $b^{(k)}$.

Далі символ $*$ означатиме довільний символ із множини $\{LS, OS, EM\}$. Довірча множина $B_\alpha^*(\Xi_n)$ визначається наступним чином

$$B_\alpha^*(\Xi_n) = \{\beta \in \mathbb{R}^{D-1} : S^*(\beta) \leqslant Q^{\chi^2_{D-1}}(1-\alpha)\}, \tag{2.13}$$

тут $Q^{\chi^2_{D-1}}(1-\alpha)$ - це квантиль χ^2 розподілу з $D-1$ ступенями свободи і рівня $1-\alpha$.

У параметричному випадку $\hat{I}_k(n)^+$ є додатно визначеною матрицею, отже $B_\alpha^*(\Xi_n)$, що визначена (2.13), задає внутрішність еліпса із центром в $\hat{b}^*(k,n)$.

У непараметричному випадку, матриця $\hat{V}_n$ може не бути додатно визначеною для малих n, а отже і множина $B_\alpha^{LS}(\Xi_n)$ може бути необмеженою.

Теорема 2.3. *За виконання умов твердження* 1.2*,*

$$\lim_{n\to\infty} P\{b^k \notin B_\alpha^{LS}(\Xi_n)\} = \alpha.$$

Доведення. З твердження 1.2 та консистентності $\hat{V}_n$ випливає що $S^{LS}(b^k) \xrightarrow{\mathrm{W}} \chi^2_{D-1}$, тому

$$P\{b^k \notin B_\alpha^{LS}(\Xi_n)\} = P\{S^{LS}(b^k) > Q^{\chi^2_{D-1}}(1-\alpha)\} \to \alpha$$

при $n \to \infty$.

□

Нагадаємо, що

$$I(p_{j;n},\tau)=\mathbb{E}\,\frac{L(\xi_j,p_{j;n},\tau)}{\partial\tau}\left(\frac{L(\xi_j,p_{j;n},\tau)}{\partial\tau}\right)^T.$$

Наступне твердження є теоремою 70.5 із [15], яка доводить асимптотичну нормальність локальної ОМВ. Тут і далі використовуються умови регулярності для моделі суміші зі змінними концентраціями. Ці умови описано у 2-му розділі.

Теорема 2.4. *При виконанні умов регулярності (RR) має місце слабка збіжність*

$$\mathrm{I}^*(n,\tau)^{1/2}\left(\hat{\tau}^{lMLE}-\tau\right)\xrightarrow{\mathrm{w}}N(0,\mathbb{E}),\ n\to\infty.$$

Теорема 2.5. *Нехай виконуються умови* (A_C), (A_u), *та* (R_2) *у умов регулярності* (RR), *і для усіх* $i=1,...,n$ $p^k_{j;n}\neq p^m_{j;n}$ *при* $k\neq m$. *За виконання умов твердження* 2.2,

$$\lim_{n\to\infty}P\{b^k\notin B^{OS}_\alpha(\Xi_n)\}=\alpha.$$

Доведення. За теоремою 70.5 із [15] і умов (RR) можна отримати асимптотичну нормальність локальної ОМВ

$$\hat{\tau}^{lMLE}=\mathrm{argmax}_{\tau\in D}L(\tau),$$

тут D - достатньо малий окіл справжнього значення параметра τ. Покажемо, що наступна збіжність виконується

$$\mathrm{I}^*(n,\tau)^{1/2}(\hat{\tau}^{(OS)}_n-\tau)\xrightarrow{\mathrm{w}}N(0,\mathbb{E}_k). \tag{2.14}$$

Додамо та віднімемо $\hat{\tau}^{lMLE}$ до наступного виразу:

$$\mathrm{I}^*(n,\tau)^{1/2}\left(\hat{\tau}^{(OS)}_n-\tau\right)=\mathrm{I}^*(n,\tau)^{1/2}\left(\hat{\tau}^{(OS)}_n-\hat{\tau}^{lMLE}\right)+\mathrm{I}^*(n,\tau)^{1/2}\left(\hat{\tau}^{lMLE}-\tau\right).$$

Другий доданок $\mathrm{I}^*(n,\tau)^{1/2}\left(\hat{\tau}^{lMLE}-\tau\right)\xrightarrow{\mathrm{w}}N(0,\mathbb{E}_k)$. Покажемо, що перший доданок збігається до нуля за ймовірністю. За теоремою про середнє значення

$$\hat{\tau}^{(OS)}_n-\hat{\tau}^{lMLE}=-\gamma_n\left(\hat{\tau}^{(0)}_n\right)S_n\left(\tau^{lMLE}\right)+\left[\mathbb{E}_k-G_n(\hat{\tau}^{(0)}_n)\right]\left(\hat{\tau}^{(0)}_n-\tau^{lMLE}\right).$$

Тут $S_n(\tau^{lMLE})=0$ за означенням τ^{lMLE} та $G_n(\hat{\tau}^{(0)}_n)=\gamma_n\left(\hat{\tau}^{(0)}_n\right)^{-1}\gamma_n(\zeta_n)$, де γ_n визначено (2.3). Тут ζ_n - проміжна точка між $\hat{\tau}^{(0)}_n$ і $\hat{\tau}^{lMLE}$.

Щоб довести (2.14) залишилося дослідити збіжність множників виразу

$$\mathrm{I}^*(n,\tau)^{1/2}\left(\mathbb{E}_k - G_n(\hat{\tau}_n^{(0)})\right)\left(\hat{\tau}_n^{(0)} - \tau^{lMLE}\right) \tag{2.15}$$

Розглянемо другий множник (2.15) $\mathrm{I}^*(n,\tau)^{1/2}\left(\hat{\tau}_n^{(0)} - \tau^{lMLE}\right)$, та додамо і віднімемо τ.

Тут маємо наступні збіжності:

$$\mathrm{I}^*(n,\tau)^{1/2}\left(\hat{\tau}_n^{(0)} - \tau\right) = O_p(1), \text{ та}$$

$$\mathrm{I}^*(n,\tau)^{1/2}\left(\tau^{lMLE} - \tau\right) \xrightarrow{\mathrm{w}} N(0, \mathbb{E}_k).$$

Для доведення залишилося показати, що

$$\mathbb{E}_k - G_n(\hat{\tau}_n^{(0)}) \xrightarrow{\mathbb{P}} 0, \; n \to \infty.$$

$$\mathbb{E}_k - G_n(\hat{\tau}_n^{(0)}) = \gamma_n^{-1}(\hat{\tau}_n^{(0)})\left(\gamma_n(\hat{\tau}_n^{(0)}) - \gamma_n(\zeta_n)\right)$$

З теореми 2.2 маємо збіжність $\frac{1}{n}\left|\mathrm{I}^*(n,\tau) - \gamma_n(\hat{\tau}_n^{(0)})\right| \xrightarrow{\mathbb{P}} 0, \; n \to \infty$. Доведення факту (2.14) випливає із наступного і теореми про середнє значення

$$\frac{1}{n}\left|\gamma(\zeta_n) - \gamma_n(\hat{\tau}_n^{(0)})\right| \leqslant \frac{1}{n}\sup_{\gamma\in D}\left\|\frac{\partial\gamma_n(\gamma)}{\partial\gamma}\right\|\left|\zeta_n - \hat{\tau}_n^{(0)}\right| = O_p(1)\left|\hat{\tau}^{lMLE} - \hat{\tau}_n^{(0)}\right| \xrightarrow{\mathbb{P}} 0, \text{ при } n \to \infty. \tag{2.16}$$

з консистентності оцінок $\hat{\tau}^{lMLE}$, та $\hat{\tau}_n^{(0)}$. Отже, (2.14) доведено.

Позначимо

$$S_0^{OS}(\beta) = \left(\beta - \hat{b}^{OS}(k,n)\right)^T \mathrm{I}_k(n)^+\left(\beta - \hat{b}^{OS}(k,n)\right),$$

тут $\mathrm{I}_k(n)^+$ - теоретичний аналог матриці $\hat{\mathrm{I}}_k(n)^+$:

$$\mathrm{I}_k(n)^+ = \left([\mathrm{I}^*(n,\tau)^{-1}]_{(k)}\right)^{-1}.$$

Далі, із (2.14), $S_0^{OS}(b^k) \xrightarrow{\mathrm{w}} \chi^2_{D-1}$. Зауважимо, що з (2.14) та першого результату теореми 2.2 виводиться

$$\zeta_n = \hat{b}^{os}(k,n) - b^{(k)} = O_p(n^{-0.5}).$$

З другого результату теореми 2.2 виводиться

$$\frac{1}{n}\left|\hat{\mathrm{I}}_k(n)^+ - \mathrm{I}_k(n)^+\right| \xrightarrow{\mathbb{P}} 0.$$

Отже

$$S_0^{OS}(b^{(k)}) - S^{OS}(b^{(k)}) = \zeta_n^T \frac{1}{n}\left(\hat{\mathrm{I}}_k(n)^+ - \mathrm{I}_k(n)^+\right)\zeta_n \xrightarrow{\mathbb{P}} 0$$

та $S^{OS}(b^{(k)}) \xrightarrow{\mathrm{W}} \chi^2_{D-1}$.

Це і доводить теорему. □

2.4 Оцінки коваріацій методом складаного ножа

У минулому підрозділі досліджено властивості оцінок $\hat{V}_n$ асимптотичної матриці коваріації $V^{(k)}$ для оцінок $\hat{b}_n^{(k)}$ параметрів $b^{(k)}$. У цьому розділі мова йтиме про непараметричні оцінки методом "складаного ножа" для тієї ж асимптотичної матриці коваріацій. На початку цього розділу вже введено оцінку $\hat{b}_n^{(k)}$ (2.1),

$$\hat{b}_n^{(k)} = \left(X_n A_n^{(k)} X_n^T\right)^{-1} X_n A_n^{(k)} Y_n.$$

Введемо ще оцінки $\hat{b}_{i-;n}^{(k)}$ параметрів регресії $b^{(k)}$, $j = 1, ..., n$, де $\hat{b}_{i-;n}^{(k)}$ - розв'язок рівняння

$$S_{ni-}^k(\gamma) = \sum_{j\neq i, j=1}^{n} a_{ji-;n}^k s(\xi_j, \gamma) = 0.$$

Тобто, $\hat{b}_{i-;n}^{(k)}$ - оцінка параметрів регресії, у випадку, коли із вибірки вилучили $i-$те спостереження. Можна показати, що розв'язком рівняння $S_{ni-}^k(\gamma) = 0$, буде оцінка

$$\hat{b}_{i-;n}^{(k)} = \left(X_{i-;n} A_{i-;n}^{(k)} X_{i-;n}^T\right)^{-1} X_{i-;n} A_{i-;n}^{(k)} Y_{i-;n}. \tag{2.17}$$

Тут $X_{i-;n} = \left(X_j\right)_{j=1, j\neq i}^n$, $Y_{i-;n} = \left(Y_j\right)_{j=1, j\neq i}^n$, а символом $A_{i-;n}^{(k)}$ позначимо діагональну матрицю навантажень k-го компонента, в якій на головній діагоналі розташовані навантаження (1.2), $(a_{1i-:n}^k, ..., a_{ni-:n}^k)$. Ці навантаження отримано за набором концентрацій $\{p_{j;n}^m, m = 1, ..., M, j = 1, ..., i-1, i+1, ..., n\}$ (без $i-$го спостереження).

Оцінкою методу складаного ножа для $V^{(k)}$ називають матрицю

$$\hat{V}_n^{(k)} = n \sum_{i=1}^{n} \left(\hat{b}_{i-;n}^{(k)} - \hat{b}_n^{(k)}\right)\left(\hat{b}_{i-;n}^{(k)} - \hat{b}_n^{(k)}\right)^T. \tag{2.18}$$

2.4.1 Додаткові твердження

Для доведення наступної леми буде корисна матрична тотожність Шермана-Моріса-Вудбурі з розділу 2.1.3 книги G. H. Golub, та C. F. van Loan “Matrix Computations” [27].

Нехай A - матриця розміру $n \times n$, U - $n \times k$, C - $k \times k$, V - $k \times n$. Тоді має місце тотожність:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1} VA^{-1},$$

і її частковий випадок, при $n = k$, і $U = V = I$:

$$(A + C)^{-1} = A^{-1} - A^{-1}C\left(C^{-1} + CA^{-1}C\right)^{-1} CA^{-1} \tag{2.19}$$

Нехай $\psi_{j;n} : \mathbb{R}^D \to \mathbb{R}^{D-1}$ - деякі функції, визначені для $j = 1, ..., n;\ n = 1, 2, ...$.

Лема 2.1. *Нехай для деякої функції $H : \mathbb{R}^{D-1} \to \mathbb{R}$ виконано*

1 $\sup_{\gamma\in\Theta} |\psi_{j;n}(x, \gamma)| \leqslant H(x)$, *для всіх $j = 1, ..., n$, $n = 1, 2, ...$*

2 Для деякого $\alpha > 1$ і всіх $m = 1, ..., M$: $\mathbb{E}\left(H(\xi_{(m)})\right)^\alpha < \infty$.

Тоді $\max\limits_{j=1,...,n} \sup\limits_{\gamma\in\Theta} |\psi_{j;n}(x, \gamma)| = o_P\left(n^{\frac{1}{\alpha}}\right)$.

Доведення. Позначимо $\tilde{\psi}_{j;n} = \sup_{\gamma\in\Theta} |\psi_{j;n}(\xi_{j;n}, \gamma)|$, $A = \max_{m=1,...,M} \mathbb{E}\left(H(\xi_{(m)})\right)^\alpha$.

Розглянемо для довільного $c > 0$

$$P_n(c) = \mathbb{P}\{\max_{j=1,...,n} \tilde{\psi}_{j;n} > c\} = 1 - \mathbb{P}\{\text{Для всіх } j = 1, ..., n;\ \tilde{\psi}_{j;n} \leqslant c\} =$$

$$= 1 - \prod_{j=1}^{n} \mathbb{P}\{\tilde{\psi}_{j;n} \leqslant c\} = 1 - \prod_{j=1}^{n}\Big(1 - \mathbb{P}\{\tilde{\psi}_{j;n} > c\}\Big). \tag{2.20}$$

Покладемо $C_n = C_0 n^{\frac{1}{\alpha}}$, де C_0 - довільне фіксоване число. Покажемо, що

$$\tilde{p}_n = \max_{j=1,...,n} \mathbb{P}\{\tilde{\psi}_{j;n} > C_n\} = o\Big(\frac{1}{n}\Big). \tag{2.21}$$

Дійсно,

$$\mathbb{P}\{\tilde{\psi}_{j;n} > C_n\} \leqslant \mathbb{P}\{H(\xi_{j;n}) > C_n\} \leqslant$$

$$\leqslant \sum_{m=1}^{M} p_{j;n}^{m} \mathbb{P}\{H(\xi_{(m)}) > C_n\} \leqslant \sum_{m=1}^{M} \mathbb{P}\{H(\xi_{(m)}) > C_n\}.$$

Отже, для доведення (2.21) досить переконатись, що

$$n\mathbb{P}\{H(\xi_{(m)}) > C_n\} \to 0 \text{ при } n \to \infty \text{ для всіх } m = 1, ..., M. \tag{2.22}$$

Оцінимо

$$n\mathbb{P}\{H(\xi_{(m)}) > C_0 n^{\frac{1}{\alpha}}\} = n\,\mathbb{E}\,I_{H(\xi_{(m)}) > C_0 n^{\frac{1}{\alpha}}} \leqslant$$

$$n\,\mathbb{E}\,I\{H(\xi_{(m)}) > C_0 n^{\frac{1}{\alpha}}\} \frac{H(\xi_{(m)})^{\alpha}}{C_0^{\alpha} n} = \frac{1}{C_0^{\alpha}}\,\mathbb{E}\,I\{H(\xi_{(m)}) > C_0 n^{\frac{1}{\alpha}}\} \left(H(\xi_{(m)})\right)^{\alpha} \to 0$$

при $n \to \infty$ за теоремою Лебега про мажоровану збіжність внаслідок другої умови леми.

Звідси випливає (2.21), а отже і (2.22).

Продовжуючи (2.20) з урахуванням (2.21), отримуємо

$$P_n(c) \leqslant 1 - \exp\left(n \ln(1 - \tilde{p}_n)\right) = 1 - \exp\left(1 - o\left(\frac{1}{n}\right)\right) = o\left(1\right)$$

для будь-якого $C_0 > 0$. Лема доведена. □

Наступна теорема показує, що за досить простих умов можна отримати асимптотику оцінок $\hat{b}_n^{(k)}$ і $\hat{b}_{i-;n}^{(k)}$, $1 \leqslant i \leqslant n$.

Лема 2.2. *Нехай далі* $\hat{b}_n^{(k)}$, $\hat{b}_{i-;n}^{(k)}$ *- МНК-оцінки вектору параметрів* $b^{(k)}$, *для* $1 \leqslant k \leqslant M$. *Якщо виконуються умови твердження* 1.5, $\mathbb{E}\,|X_{(k)}|^2 \leqslant \infty$, *і* $\mathbb{E}\,|X_{(k)} Y_{(k)}| < \infty$ *для* $1 \leqslant k \leqslant M$, *тоді*

$$\sqrt{n} \max_{\substack{\leqslant i \leqslant n \\ 1 \leqslant k \leqslant M}} \left|\hat{b}_{i-;n}^{(k)} - \hat{b}_n^{(k)}\right| = o_P(1),\ n \to \infty.$$

Доведення. Розглянемо праву частину рівності (2.17). Позначимо

$$B_n = \sum_{j=1}^{n} a_{j;n}^{k} X_j X_j^T,$$

і

$$C_n = \sum_{j \neq i}^{n} (a_{ji-;n}^{k} - a_{j;n}^{k}) X_j X_j^T - a_{i;n}^{k} X_i X_i^T.$$

Тоді для довільного i

$$\left(X_{i-;n}^{T}A_{i-;n}^{(k)}X_{i-;n}\right)^{-1} = (B_n + C_n)^{-1} = B_n^{-1} - B_n^{-1}C_n\left(C_n^{-1} + C_nB_n^{-1}C_n\right)^{-1}C_nB_n^{-1}.$$

Помітимо, що за виконання умов леми і часткового випадку леми 2.1, а також твердження 1.5 $C_n = o_P\left(\frac{1}{\sqrt{n}}\right)$, при $n \to \infty$. Це означатиме, що

$$B_n^{-1}C_n\left(C_n^{-1} + C_nB_n^{-1}C_n\right)^{-1}C_nB_n^{-1} = o_P\left(\frac{1}{\sqrt{n}}\right),$$

і як наслідок,

$$\left(X_{i-;n}^{T}A_{i-;n}^{(k)}X_{i-;n}\right)^{-1} = \left(X_nA_n^{(k)}X_n^{T}\right)^{-1} + o_P\left(\frac{1}{\sqrt{n}}\right),\ n \to \infty. \qquad (2.23)$$

Аналогічно до попереднього, введемо

$$B_n^{Y} = \sum_{j=1}^{n} a_{j;n}^{k}X_jY_j,$$

і

$$C_n^{Y} = \sum_{j\neq i}^{n}(a_{ji-;n}^{k} - a_{j;n}^{k})X_jY_j - a_{i;n}^{k}X_iY_i.$$

Очевидно, що за виконання умов леми і часткового випадку леми 2.1, а також леми 3.1, для довільного $1 \leqslant i \leqslant n$

$$X_{i-;n}A_{i-;n}^{(k)}Y_{i-;n} = B_n^{Y} + C_n^{Y} = X_n^{T}A_n^{(k)}Y_n + o_P\left(\frac{1}{\sqrt{n}}\right) \qquad (2.24)$$

З (2.23) та (2.24) випливатиме, що

$$\max_{\substack{\leqslant i\leqslant n\\ 1\leqslant k\leqslant M}}\left|\hat{b}_{i-;n}^{(k)} - \hat{b}_n^{(k)}\right| =$$

$$\max_{\substack{\leqslant i\leqslant n\\ 1\leqslant k\leqslant M}}\left|\left(X_nA_n^{(k)}X_n^{T}\right)^{-1}X_nA_n^{(k)}Y_n - \left(\left(X_nA_n^{(k)}X_n^{T}\right)^{-1} + o_P\left(\frac{1}{\sqrt{n}}\right)\right)\left(X_n^{T}A_n^{(k)}Y_n + o_P\left(\frac{1}{\sqrt{n}}\right)\right)\right| =$$

$$o_P\left(\frac{1}{\sqrt{n}}\right),\ n \to \infty.$$

$\square$

Наступна теорема використовує твердження 1.6 і доводить консистентність оцінок методом складаного ножа у моделі регресійної суміші.

Теорема 2.6. *Якщо виконуються наступні умови*

1 *Умови* (Г) *і* (a^2),

2 $\det \mathbb{E} X_{(k)} X_{(k)}^T \neq 0$,

3 *Для деякого* $\alpha > 4$, $\mathbb{E}|X_{(k)}|^{2\alpha} < \infty$ *і* $\mathbb{E}|\varepsilon_{(k)}|^{\alpha} < \infty$,

тоді

$$\hat{V}_n^{(k)} \xrightarrow{\mathbb{P}} V^{(k)}.$$

Доведення. Для того, щоб використати твердження 1.6 далі розглядатимемо $\tilde{\xi}_j = \left(X_j^1, ..., X_j^{D-1}, Y_j, X_j^1 Y_j, ..., X_j^{D-1} Y_j, X_j^1 X_j^1, ..., X_j^{D-1} X_j^{D-1}\right)$. Позначимо $\theta_{xx}^{(k)} = \mathbb{E} X_{(k)} X_{(k)}^T$ і $\theta_{xy}^{(k)} = \mathbb{E} X_{(k)} Y_{(k)}$, та відображення $H(\gamma_1, \gamma_2) = \gamma_1^{-1}\gamma_2$, тут $\gamma_1 \in \mathbb{R}^{D-1 \times D-1}$, $\gamma_2 \in \mathbb{R}^{D-1}$.

З твердження 1.3 випливатиме, що

$$\hat{b}_n^{(k)} \xrightarrow{\mathbb{P}} H\left(\theta_{xx}^{(k)}, \theta_{xy}^{(k)}\right) = \left(\mathbb{E} X_{(k)} X_{(k)}^T\right)^{-1} \mathbb{E} X_{(k)} Y_{(k)}.$$

Для того, щоб застосувати Твердження 1.6, залишилося показати, що виконана умова 2 Твердження 1.6. Вона випливає із умови 3 даного твердження. Дійсно,

$$\mathbb{E}|\tilde{\xi}_k|^{\alpha} = \mathbb{E}\left(\sum_{l=1}^{D-1}(X_{(k)}^l)^2 + \sum_{l=1}^{D-1}(X_{(k)}^l)^2 Y_{(k)}^2 + (\sum_{l=1}^{D-1}(X_{(k)}^l)^2)^2\right)^{\alpha/2}.$$

Оскільки усі доданки під знаком математичного сподівання додатні, то

$$\mathbb{E}|\tilde{\xi}_k|^{\alpha} \leqslant \mathbb{E}\left(\sum_{l=1}^{D-1}(X_{(k)}^l)^2\right)^{\alpha/2} + \mathbb{E}\left(Y_{(k)}^2 \sum_{l=1}^{D-1}(X_{(k)}^l)^2\right)^{\alpha/2} + \mathbb{E}\left(\sum_{l=1}^{D-1}(X_{(k)}^l)^2\right)^{\alpha}.$$

З того, що $Y_{(k)} = \langle X_{(k)}, b^{(k)}\rangle + \varepsilon_{(k)}$, останньої нерівності, умови 3 теореми, та нерівності трикутника для норми, випливатиме доведення цієї теореми. □

2.5 Алгоритмічні питання і результати моделювання

2.5.1 Результати моделювання

Для перевірки якості отриманих довірчих еліпсоїдів описаних у п. 3.3, була проведена серія експериментів на модельованих даних. Дані для експериментів генерувалися на основі моделі двокомпонентної суміші ($M = 2$) із простою регресією. Регресійна модель має наступну форму

$$Y_{(k)} = b_0^k + b_1^{(k)} X_{(k)} + \varepsilon^{(k)}, \tag{2.25}$$

тут $X_{(k)} \sim N(\mu_k, \sigma_k^2)$ - регресори, а $Y_{(k)}$ - відгуки, k - номер компонента, який не спостерігається, і ε_k - похибка регресії. Ця похибка є незміщеною із дисперсією σ_k^2.

Ймовірності $p_{j;n}^m$ були згенеровані наступною стохастичною процедурою

$$p_{j;n}^m = \frac{u_j^m}{\sum_{k=1}^M u_j^k}, \tag{2.26}$$

де u_j^m – незалежні псевдовипадкові числа, розподілені рівномірно на [0, 1]. Надалі ця процедура генерування ймовірностей буде використана і в інших імітаційних експериментах.

Для кожної вибірки розміру n було проведено 1000 експериментів. Для кожної вибірки були побудовані параметричні (EM), та непараметричні (LS) довірчі еліпсоїди. Еліпсоїди у параметричному випадку базувалися на EM-оцінках, які використовували LS-оцінки у якості початкового наближення (як пілотні оцінки), та матриця $\hat{\mathrm{I}}_k(n)^+$ використана як матриця для квадратичної форми у S^{EM}.

Непараметричні довірчі еліпсоїди будувалися на основі LS-оцінок. Як було раніше сказано, матриця $\hat{V}_n^{(k)}$ не завжди буде додатно визначеною. Тому відповідна довірча множина може бути необмеженою. У випадку простої регресії (2.25) цей недолік може бути виправлений використанням покращених навантажень b_j^+, які визначені у п. 2.2.1. Також непараметричні еліпсоїди будувалися на основі методу “складаного ножа” (JK). В цьому випадку оцінки матриці $V^{(k)}$ будуть додатно визначеними, проте їх рахувати значно довше ніж усі інші.

Усі еліпсоїди були побудовані із номінальним довірчим рівнем $\alpha = 0.05$. У кожному експерименті пораховано частоти покриття справжніх значень параметрів b^k побудованими довірчими еліпсоїдами, та середній об’єм цих еліпсоїдів.

Експеримент 1. Значення параметрів для цього експерименту зазначено у таблиці 2.1. Похибки ε_k мали гаусів розподіл. Цей експеримент описується моделлю “повної роздільності”, коли візуально спостереження можуть бути розділені на дві групи відповідно до різних компонент суміші (див.Рисунок 2.1 ліворуч).

Частоти покриття та середні об’єми еліпсоїдів для вибірок різних розмірів n наведено у таблиці 2.2. Вони демонструють достатньо відповідність із номінальним

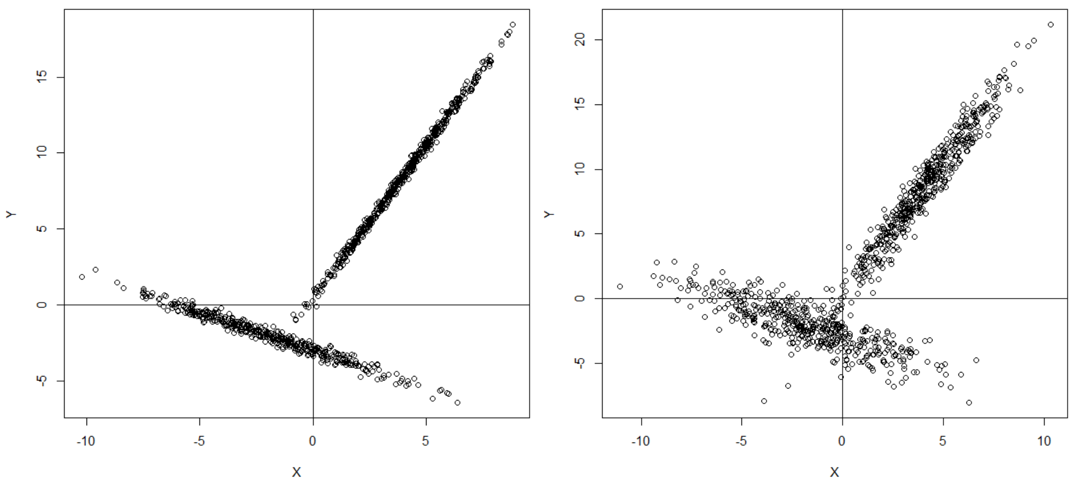


Рис. 2.1. Діаграма розсіювання пар у Експерименті 3 (ліворуч), Експерименті 4 (праворуч)

k	1	2
μ_k	-2	4
Σ_k	2	2
σ_k	0.05	0.05
b_0^k	0.5	-0.5
b_1^k	2	-1/3

Табл. 2.1. Параметри розподілів і регресії для моделювання експерименту 1

рівнем значущості для вибірок, розмір яких вищий за 1000. Екстремально великі середні об'єми для LS- і JK-еліпсоїдів при малих обсягах вибірки виникли внаслідок окремих викидів, пов'язаних з тим, що оцінки $\hat{V}_n$ виявились для деяких вибірок близькими до вироджених. З таблиці 2.3 видно, що об'єми EM-еліпсоїдів значно менші, ніж інших, а JK-еліпсоїди трошки менші, за LS.

Параметричні довірчі множини є значно меншими ніж непараметричні.

Експеримент 2. Для того, щоб подивитися вплив розкиду похибок регресії на якість нашого методу, ми збільшили цей параметр $\sigma_k = 0.25$ у другому експерименті, а всі інші параметри було залишено як у першому експерименті. Діаграму розсіювання пар для побудованих даних наведено на рисунку 2.1 праворуч.

Результати експерименту наведено у таблицях 2.4 (частоти покриття) і 2.5 (об'єми оцінених довірчих множин). Частоти покриття у цьому експерименті змінилися неістотно. У порівнянні із Експериментом 1, середні об'єми зменшилися істотно для EM-еліпсоїдів, але не для JK- та LS-еліпсів.

n	LS		EM		JK	
	k=1	k=2	k=1	k=2	k=1	k=2
100	0.954	0.821	0.946	0.951	0.917	0.803
500	0.925	0.925	0.955	0.953	0.924	0.879
1 000	0.949	0.941	0.947	0.952	0.923	0.915
5 000	0.962	0.935	0.953	0.945	0.927	0.932
7 500	0.955	0.947	0.947	0.949	0.924	0.913
10 000	0.936	0.96	0.95	0.95	0.944	0.934

Табл. 2.2. Ймовірності включення справжніх параметрів регресії у довірчі множини у експерименті 1

n	LS		EM		JK	
	k=1	k=2	k=1	k=2	k=1	k=2
100	806288	46369047	0.0649	0.0714	388128	216661
500	92296	722272210	0.0112	0.0110	9439	6.1×10^{7}
1 000	1242	8563	4.94×10^{-3}	5.38×10^{-3}	2.19	55.84
5 000	0.357	0.208	0.97×10^{-3}	1.01×10^{-3}	0.243	0.140
7 500	0.195	0.107	0.63×10^{-3}	0.66×10^{-3}	0.151	0.083
10 000	0.140	0.072	4.71×10^{-4}	4.74×10^{-4}	0.124	0.071

Табл. 2.3. Середній об'єм оцінених довірчих множин у експерименті 1

n	LS		EM		JK	
	k=1	k=2	k=1	k=2	k=1	k=2
100	0.886	0.922	0.95	0.943	0.846	0.842
500	0.918	0.931	0.945	0.965	0.890	0.915
1 000	0.95	0.94	0.945	0.948	0.887	0.904
5 000	0.948	0.941	0.943	0.958	0.960	0.944
7 500	0.941	0.937	0.944	0.943	0.938	0.945
10 000	0.954	0.946	0.951	0.955	0.947	0.950

Табл. 2.4. Ймовірності включення справжніх параметрів регресії у довірчі множини у експерименті 2

n	LS		EM		JK	
	k=1	k=2	k=1	k=2	k=1	k=2
100	23795290	14294378	0.01437	0.01289	358757	290535
500	655946.7	34102444.1	0.00249	0.00245	3517692	176667
1 000	132463.34	19568.26	0.00119	0.00125	46618	6554
5 000	0.329	0.201	2.392×10^{-4}	2.435×10^{-4}	0.271	0.144
7 500	0.185	0.111	1.595×10^{-4}	1.622×10^{-4}	0.168	0.087
10 000	0.143	0.079	1.189×10^{-4}	1.221×10^{-4}	0.121	0.062

Табл. 2.5. Середній об’єм оцінених довірчих множин у експерименті 2

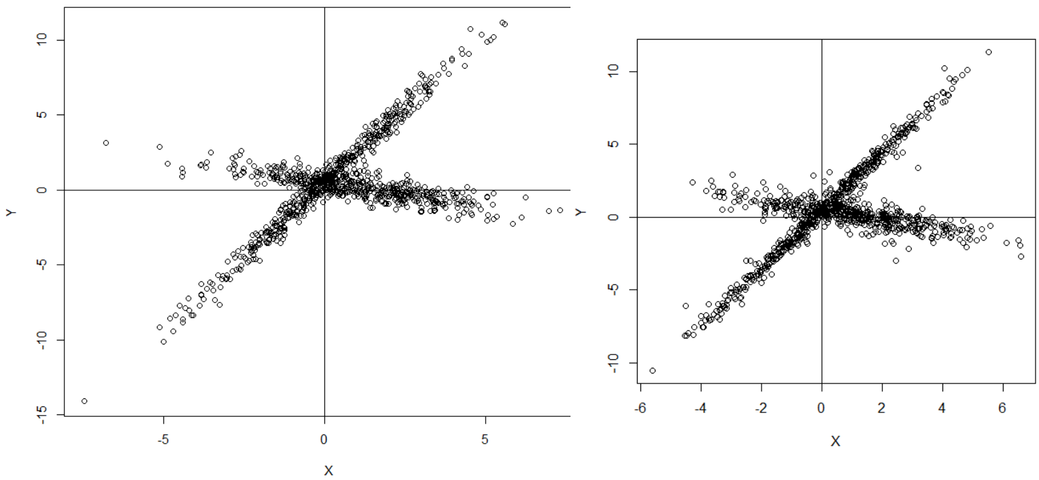


Рис. 2.2. Діаграма розсіювання пар у Експерименті 3 (ліворуч), Експерименті 4 (праворуч)

Експеримент 3. Тут ми розглянемо інший набір параметрів (див. таблицю 2.6). Похибки регресії розподілені нормально. у цій моделі спостереження не можуть бути однозначно візуально розподілені на два різних компонента (див. рисунок 2.2 ліворуч).

Результати наведено у таблицях 2.7 і 2.8. Знову, EM-еліпсоїди менші за LS-еліпсоїди. Слід зазначити, що у цьому експерименті об’єми LS, та JK-еліпсоїдів менші, ніж у попередніх експериментах, а аномально великі значення об’ємів спостерігаються при $n = 100$, а не 1000, як це було раніше. Окремо треба звернути увагу, що частоти покриття для JK-еліпсоїдів відрізняються від номінальних, навіть при збільшенні розміру вибірки.

Експеримент 4. У цьому експерименті використано ті ж самі параметри, що і у Експерименті 3, але похибки регресії не гаусові. Покладемо $\varepsilon^k = \sqrt{3/5}\sigma_k \nu$, де ν - Т-розподіл Стьюдента із 5-ма ступенями вільності. Такі похибки мають ту саму

k	1	2
μ_k	0	1
Σ_k	2	2
σ_k	0.5	0.5
b_0^k	0.5	-0.5
b_1^k	2	-1/3

Табл. 2.6. Параметри для моделювання експериментів 3 і 4

Таблиця 5. Результати експерименту 3 (k-номер компонента)

n	LS		EM		JK	
	k=1	k=2	k=1	k=2	k=1	k=2
100	0.857	0.882	0.949	0.935	0.969	0.985
500	0.898	0.924	0.950	0.936	0.952	0.974
1 000	0.929	0.914	0.948	0.946	0.936	0.968
5 000	0.955	0.933	0.944	0.947	0.929	0.977
7 500	0.953	0.938	0.939	0.953	0.963	0.981
10 000	0.951	0.957	0.933	0.921	0.964	0.942

Табл. 2.7. Ймовірності включення справжніх параметрів регресії у довірчі множини у експерименті 3

n	LS		EM		JK	
	k=1	k=2	k=1	k=2	k=1	k=2
100	521.61	876.77	0.053	0.066	90.85	178829.3
500	1.238	1.314	0.0103	0.0102	1.215	1.090
1 000	0.623	0.678	5.00×10^{-3}	5.18×10^{-3}	0.611	0.518
5 000	0.120	0.119	0.99×10^{-3}	1.01×10^{-3}	0.112	0.111
7 500	0.078	0.077	6.72×10^{-4}	6.66×10^{-4}	0.076	0.073
10 000	0.059	0.058	4.96×10^{-4}	5.06×10^{-4}	0.056	0.055

Табл. 2.8. Середній об’єм оцінених довірчих множин у експерименті 3

n	LS		EM		JK	
	k=1	k=2	k=1	k=2	k=1	k=2
100	0.912	0.915	0.943	0.937	0.965	0.989
500	0.949	0.959	0.931	0.941	0.990	0.982
1 000	0.948	0.945	0.943	0.948	0.948	0.971
5 000	0.948	0.969	0.952	0.959	0.958	0.952
7 500	0.952	0.960	0.950	0.943	0.961	0.956
10 000	0.937	0.0945	0.936	0.933	0.951	0.956

Табл. 2.9. Ймовірності включення справжніх параметрів регресії у довірчі множини у експерименті 4

n	LS		EM		JK	
	k=1	k=2	k=1	k=2	k=1	k=2
100	354	173687	0.013	0.014	89.11	11.82
500	1.481	1.300	2.45×10^{-3}	2.48×10^{-3}	1.338	1.221
1 000	0.563	0.607	1.21×10^{-3}	1.22×10^{-3}	0.673	0.623
5 000	0.115	0.119	2.38×10^{-4}	2.44×10^{-4}	0.112	0.116
7 500	0.079	0.078	1.60×10^{-4}	1.61×10^{-4}	0.079	0.073
10 000	0.062	0.058	1.20×10^{-4}	1.20×10^{-4}	0.055	0.055

Табл. 2.10. Середній об’єм оцінених довірчих множин у експерименті 4

дисперсію як і у Експерименті 3, але їх розподіл має важкі хвости. Зауважимо, що 5 - найменший ступінь вільності, для якого виконується припущення $\mathbb{E}(\varepsilon_k)^4 < \infty$ Теореми 1.2. При цьому $\mathbb{E}(\varepsilon_k)^5$ не існує.

Типовий приклад діаграми розсіювання пар для цієї моделі продемонстровано на рисунку 2.2 праворуч. Вона візуально відрізняється від типового розкиду для гаусових похибок із Експерименту 3, що зображено на рисунку 2.2 ліворуч.

Результати експерименту записано у таблицях 2.9 і 2.10. Зауважимо, що у цьому випадку частоти покритя ЕМ еліпсоїдів не збігаються до номінального значення $1 - \alpha = 0.95$ для великих n. Ймовірності покриття LS-еліпсоїдів значно ближчі до 0.95. Отже, для розподілів із важкими хвостами JK-, EM-еліпсоїди демонструють гіршу якість, ніж LS-еліпсоїди.

2.6 Висновки

У цьому розділі було розглянуто дві моделі регресійної суміші: параметрична та семі-параметрична. Для обох моделей доведено консистентність оцінки матриці коваріації і на основі цих матриць побудовано довірчі множини, та доведено теореми про їх асимптотичні властивості для лінійної регресійної моделі суміші зі змінними концентраціями. Окрім цього, для асимптотичних матриць коваріацій була побудована оцінка методом складаного ножа і доведено її консистентність. Якості побудованих оцінок також перевірено у імітаційних експериментах.

Показано, що у випадку, коли розподіл даних відповідає параметричній моделі, параметричні оцінки забезпечують кращу якість довірчих еліпсоїдів. Якщо параметрична модель не відповідає розподілу даних, семі-параметричні оцінки можуть давати точніші результати, ніж параметричні. Також моделюваннями показано, що об’єм довірчих множин , які оцінені методом складаного ножа, менший, ніж у множин, які побудовані для семі-параметричної моделі.

Розділ 3

Асимптотика оцінок методу оціночних рівнянь для суміші зі змінними концентраціями

У цьому розділі розглядається модель суміші нелінійних регресій зі змінними концентраціями компонентів. Побудовані семі-параметричні оцінки невідомих параметрів регресії для випадку нелінійної регресії методом оціночних рівнянь. Для оцінок параметрів цієї моделі доведено консистентність і асимптотичну нормальність, після чого оцінюються матриці коваріації оцінок для побудови довірчих еліпсоїдів оцінок параметрів.

Результати цього розділу надруковані в статтях [68], [69], [50].

3.1 Модель нелінійної регресії, МНК оцінки для суміші, УОР-оцінки

Ми розглядаємо модель суміші зі змінними концентраціями, у якій спостереження $\xi_j = \xi_{j;n} \in \mathbb{R}^D,\ j = 1, ..., n$ є незалежними при різних j і мають розподіл

$$\mathbb{P}\{\xi_{j;n} \in A\} = \sum_{m=1}^{M} p_{j;n}^{m} F_m(A), \tag{3.1}$$

де F_m - розподіл на $\mathbb{R}^D$ (F_m - розподіл m-го компонента суміші), $p_{j;n}^{m}$ - ймовірність спостерігати m-ту компоненту при j-му спостереженні. $p_{j;n}^{m}$ вважаються відомими. Для F_m ми розглядаємо семі-параметричну модель $F_m(A) = F(A; \theta^{(m)}, \nu^{(m)})$, де $\theta^{(m)} \in \Theta \subseteq \mathbb{R}^d$ невідомі параметри розподілу F_m, $\nu^{(m)}$ - непараметрична складова моделі (також відома). Наприклад, у регресійній моделі

$$Y_j = g(X_j, \theta^{\kappa_j}) + \varepsilon_j, \tag{3.2}$$

Y_j - відгук, $X_j = (X_j^1, ..., X_j^{D-1})$ - вектор регресорів для j-го спостереження, $\xi_j = (Y_j, X_j)$, $\kappa_j = \mathrm{ind}(\xi_j)$ - прихований номер компонента спостереження j, $\theta^{(m)}$ - вектор невідомих коефіцієнтів регресії і є евклідовим параметром. Непараметрична складова - це розподіл регресорів $F_X^{(m)}$ і розподіл похибок $F_\varepsilon^{(m)}$, тобто $\nu^{(m)} = (F_X^{(m)}, F_\varepsilon^{(m)})$.

Для оцінювання невідомого евклідового параметра $\theta^{(k)}$ ми використаємо метод узагальнених оціночних рівнянь (УОР). Для цього вибирається елементарна оціночна функція $s(\xi, \gamma)$, $\xi \in \mathbb{R}^D, \gamma \in \Theta$, така, що

$$\mathbb{E}\, s(\xi_{(k)}, \gamma) = 0 \text{ тоді і лише тоді, коли } \gamma = \theta^{(k)}. \tag{3.3}$$

Тут $\xi_{(k)}$ - випадковий вектор з розподілом $F_k = F(\cdot; \theta^{(k)}, \nu^{(k)})$.

Оціночна функція для оцінки $\theta^{(k)}$ має вигляд

$$S_n^{(k)}(\gamma) = \sum_{j=1}^{n} a_{j;n}^k s(\xi_{j;n}, \gamma), \tag{3.4}$$

де $a_{j;n}^k$ - мінімаксні вагові коефіцієнти.

Означення. Статистику $\hat{\theta}_n^{(k)}$ будемо називати оцінкою УОР з елементарною оціночною функцією s, якщо рівність

$$S_n^{(k)}(\hat{\theta}_n^{(k)}) = 0 \tag{3.5}$$

виконується майже напевно.

Наприклад, якщо дані описуються регресійною моделлю (3.2), то для оцінки параметрів регресії $\theta^{(k)}$ можна використати елементарну оціночну функцію

$$s(\xi_j, \gamma) = \left(Y_j - g(X_j, \gamma)\right) \dot{g}(X_j, \gamma), \tag{3.6}$$

де $\xi_j = \left(Y_j, X_j^1, ..., X_j^{D-1}\right)^T$, $\gamma = (\gamma^1, ..., \gamma^d)^T$, $\dot{g}(X_j, \gamma) = \begin{pmatrix} \frac{\partial g(X_j,\gamma)}{\partial \gamma^1} \\ \vdots \\ \frac{\partial g(X_j,\gamma)}{\partial \gamma^d} \end{pmatrix}$.

Зазвичай елементарна оціночна функція приймає значення у $\mathbb{R}^d$, тобто

$$s(\xi, \gamma) = \begin{pmatrix} s^1(\xi, \gamma) \\ \vdots \\ s^d(\xi, \gamma) \end{pmatrix} \tag{3.7}$$

Для того, щоб використовувати оцінки УОР, потрібно спочатку визначити, за яких умов вони існують (тобто існує розв'язок (3.5), що є вимірною функцією від даних) і є консистентними. У наступних підрозділах ми розглянемо ці питання. Далі ми покажемо, що за досить широких умов УОР оцінки є асимптотично нормальними і знайдемо у явному вигляді коваріаційну матрицю граничного нормального розподілу $V^{(k)}$. Оскільки ця матриця залежить від невідомих евклідових параметрів та непараметричних складових розподілів F_m, $m = 1, ..., M$, безпосередньо використовувати її для статистичних висновків неможливо. Тому ми розглядаємо оцінки для $V^{(k)}$, побудовані за методом складаного ножа. А саме, для кожного $i = 1, ..., n$, розглянемо вагові коефіцієнти $a^k_{ji-;n}$ які утворюються як мінімаксні коефіцієнти за набором концентрацій компонент $\{p^m_{j;n},\ m = 1, ..., M,\ j = 1, ..., i-1, i+1, ..., n\}$ (тобто i-те спостереження пропущене).

Позначимо

$$S^{(k)}_{i;n}(\gamma) = \sum_{i \neq j} a^k_{ji-;n} s(\xi_{j;n}, \gamma) \tag{3.8}$$

- оціночна функція, що побудована за даними з вилученим i-м спостереженням. $\hat{\theta}^{(k)}_{i-;n}$ - УОР оцінка, що відповідає оціночній функції $S^{(k)}_{i;n}$, тобто

$$S^{(k)}_{i;n}(\hat{\theta}^{(k)}_{i-;n}) = 0 \text{ м.н.} \tag{3.9}$$

Оцінкою методу складаного ножа для $V^{(k)}$ називають

$$\hat{V}^{(k)}_n = n \sum_{i=1}^{n} \left(\hat{\theta}^{(k)}_{i-;n} - \hat{\theta}^{(k)}_n\right)\left(\hat{\theta}^{(k)}_{i-;n} - \hat{\theta}^{(k)}_n\right)^T. \tag{3.10}$$

3.2 Допоміжні леми

Нехай $\psi_{j;n} : \mathbb{R}^D \to \mathbb{R}^d$ - деякі функції, визначені для $j = 1, ..., n;\ n = 1, 2, ...$ Позначимо

$$\Psi^k_n(\gamma) = \sum_{j=1}^{n} a^k_{j;n} \psi_{j;n}(\xi_{j;n}, \gamma),$$

$$\Psi^k_{ni-}(\gamma) = \sum_{j \neq i} a^k_{ji-;n} \psi_{j;n}(\xi_{j;n}, \gamma).$$

Нагадаємо позначення

$$\Gamma_n = p_{;n} p^T_{;n} = \left(\sum_{j=1}^n p^i_{j;n} p^k_{j;n} \right)^M_{i,k=1},$$

$$\Gamma_\infty = \lim_{n \to \infty} \frac{1}{n} \Gamma_n$$

(в припущенні, що ця границя існує).

Наступна лема є переформульованою лемою 1.5.

Лема 3.1. *Нехай* $\exists \Gamma_\infty$ *і* $\det \Gamma_\infty > 0$. *Тоді* $\exists C_a < \infty$, *що*

$$|a^m_{j;n}| < \frac{C_a}{n}, \text{ і } |a^m_{j;n} - a^m_{ji-;n}| < \frac{C_a}{n^2}$$

для всіх $m = 1, ..., M$, $1 \leqslant i \neq j \leqslant n$, $n = 1, 2, ...$

Лема 2.1 доведена без припущень про вигляд оціночної функції $\psi_{j;n}(z, \gamma)$. У випадку узагальненого методу найменших квадратів, коли оціночні функції приймають вигляд

$$\psi_{j;n}(\xi_{j;n}, \gamma) = \Big(Y_j - g(X_j, \gamma)\Big) \dot{g}(X_j, \gamma), \tag{3.2}$$

зручно використати наступну лему замість леми 2.1.

Лема 3.2. *Нехай* $\psi_{j;n}$ *мають вигляд* (3.2) *і виконуються такі умови для усіх* $m = 1, ..., M$ *і деякого* $\alpha > 1$

1 $\exists H_0(x) : |g(x, \gamma)| \leqslant H_0(x)$, $\forall \gamma \in \Theta$.

2 $\exists H_1(x) : |\dot{g}(x, \gamma)| \leqslant H_1(x)$, $\forall \gamma \in \Theta$.

3 $\exists\, \mathbb{E} \left(\left(|Y_{(m)}| + H_0(\xi_{(m)}) \right) H_1(\xi_{(m)}) \right)^\alpha < \infty$.

Тоді виконується лема 2.1

Доведення. Дійсно, перевіримо виконання умов леми 2.1. Перевіримо першу умову леми 2.1.

$$\sup_{\gamma \in \Theta} \left| \Big(Y_j - g(X_j, \gamma)\Big) \dot{g}(X_j, \gamma) \right| \leqslant \Big(|Y_j| + H_0(X_j)\Big) H_1(X_j) =: H(\xi_j)$$

за властивостями норми і модуля.

Другу умова леми 2.1 виконується внаслідок умов 1) і 3) леми. □

Для випадку, коли $\psi_{j;n}$ приймають вигляд (3.2), лему 3.2 можна застосувати, наприклад, для наступного вигляду функції g:

$$g(x,\gamma) = \frac{1}{1+e^{-\langle x,\gamma\rangle}}, \tag{3.3}$$

Приклад 3.1. Нехай $\psi_{j;n}, g(x,\gamma)$ мають вигляд (3.2), (3.3) відповідно, та існують для усіх $m = 1,...,M$ і деякого $\alpha > 1$ наступні математичні сподівання $\exists\,\mathbb{E}\,|X_{(m)}|^{\alpha} < \infty$. $\exists\,\mathbb{E}\,|\varepsilon_{(m)}|^{\alpha} < \infty$.

Тоді виконуються умови леми 2.1

Доведення. Доведення прикладу полягає у перевірці умов леми 3.2. З означення функції $g(x,\gamma)$ випливають умови 1) і 2) леми 3.2. Дійсно, для всіх γ із Θ

$$|g(x,\gamma)| \leqslant 1 =: H_0, \text{ та } |\dot{g}(x,\gamma)| = g(x,\gamma)\,(1-g(x,\gamma))\,|x| < |x| =: H_1(x).$$

Покажемо, що умова 3) леми 3.2 також виконується.

$$\mathbb{E}\left(\left(|g(X_{(m)},\gamma)| + |\varepsilon_{(m)}| + H_0\right)H_1(X_{(m)})\right)^{\alpha} \leqslant \mathbb{E}\left(\left(2+|\varepsilon_{(m)}|\right)|X_{(m)}|\right)^{\alpha} =$$

$$= \mathbb{E}\left(2+|\varepsilon_{(m)}|\right)^{\alpha}\mathbb{E}\,|X_{(m)}|^{\alpha} < \infty$$

з умов прикладу, та нерівності Мінковського для першого доданка. □

Окрім функцій $\psi_{j;n}$, $j = 1,...,n$, $n = 1,2,...$ розглянемо функції $h : \mathbb{R}^D \to \mathbb{R}$ і $\rho : \Theta\times\Theta \to \mathbb{R}$.

Будемо казати, що для набору функцій $\psi_{j;n}$ виконується умова $\Psi_{h\rho}$, якщо:

$$(\Psi_{h\rho}) \quad \left|\psi_{j;n}(z,\gamma_1) - \psi_{j;n}(z,\gamma_2)\right| \leqslant h(z)\rho(\gamma_1,\gamma_2),\ \forall\gamma_1,\gamma_2\in\Theta,\ \forall z\in\mathbb{R}^D$$

Тут і далі, символом z позначатимемо пару (x, y).

Лема 3.3. *Нехай $\psi_{j;n}(\xi_{j;n},\gamma) = \left(Y_j - g(X_j,\gamma)\right)\dot{g}(X_j,\gamma)$ і виконуються такі умови:*

1 Для функцій $\dot{g}(x,\gamma)$ виконується умова $\Psi_{h_1,\rho}$, для всіх $\gamma_1,\gamma_2\in\Theta$:

$$\exists h_0,\rho :\ |\dot{g}(x,\gamma_1) - \dot{g}(x,\gamma_2)| < h_1(x)\rho(\gamma_1,\gamma_2)$$

2 Для функцій $g(x,\gamma)\dot{g}(x,\gamma)$ виконується умова $\Psi_{h_{01},\rho}$, для всіх $\gamma_1,\gamma_2\in\Theta$:

$$\exists h_{01},\rho :\ |g(x,\gamma_1)\dot{g}(x,\gamma_1) - g(x,\gamma_2)\dot{g}(x,\gamma_2)| < h_{01}(x)\rho(\gamma_1,\gamma_2)$$

Тоді для функцій $\psi_{j;n}(z,\gamma)$ виконується умова $\Psi_{h\rho}$, де $h(z) = |y|h_1(x) + h_{01}(x)$

Доведення.

$$|\psi_{j;n}(z,\gamma_1) - \psi_{j;n}(z,\gamma_2)| \leqslant |y|\Big|\dot g(x,\gamma_1) - \dot g(x,\gamma_2)\Big| + \Big|g(x,\gamma_1)\dot g(x,\gamma_1) - g(x,\gamma_2)\dot g(x,\gamma_2)\Big| =$$

$$\leqslant \Big(|y|h_1(x) + h_{01}(x)\Big)\rho(\gamma_1,\gamma_2) = h(z)\rho(\gamma_1,\gamma_2)$$

Лему доведено. □

Приклад 3.2. Нехай $\psi_{j;n}$ і g мають вигляд (3.2) та (3.3) відповідно: Тоді для функцій $\psi_{j;n}(z,\gamma)$ виконується умова $\Psi_{h\rho}$.

Доведення. Позначимо $\sigma(\langle x,\gamma\rangle) = g(x,\gamma)$. Оцінимо зверху функції $\dot\sigma(t)$, та $\ddot\sigma(t)$:

$$|\dot\sigma(t)| = |\sigma(t)\,(1-\sigma(t))| \leqslant 1;\ |\ddot\sigma(t)| = |\sigma(t)\,(1-\sigma(t))\,(1-2\sigma(t))| \leqslant 1.$$

Аналогічно,

$$\left|\frac{\partial\sigma(\langle x,\gamma\rangle)}{\partial\gamma}\right| \leqslant |x|;\ \left|\frac{\partial^2\sigma(\langle x,\gamma\rangle)}{\partial\gamma^2}\right| \leqslant |x|^2$$

З теоремою про середнє значення

$$|\dot g(x,\gamma_1) - \dot g(x,\gamma_2)| \leqslant \sup_{\gamma\in\Theta}|\ddot g(x,\gamma)|\Big|\langle x,\gamma_1-\gamma_2\rangle\Big| \leqslant |x|^3|\gamma_1-\gamma_2| := h_1(x)\rho(\gamma_1,\gamma_2).$$

Покажемо виконання умови $\Psi_{h_{01}\rho}$ для функцій $g(x,\gamma)\dot g(x,\gamma)$

$$|g(x,\gamma_1)\dot g(x,\gamma_1) - g(x,\gamma_2)\dot g(x,\gamma_2)| \leqslant \sup_{\gamma\in\Theta}\left|\frac{\partial}{\partial\gamma}(g(x,\gamma)\dot g(x,\gamma))\right||\langle x,\gamma_1-\gamma_2\rangle| \tag{3.4}$$

Оцінимо зверху окремо для множника $\left|\frac{\partial}{\partial\gamma}(g(x,\gamma)\dot g(x,\gamma))\right|$

$$\left|\frac{\partial}{\partial\gamma}(g(x,\gamma)\dot g(x,\gamma))\right| = |\dot g(x,\gamma)\dot g^T(x,\gamma) + g(x,\gamma)\ddot g(x,\gamma)| \leqslant$$
$$\leqslant |g(x,\gamma)^2\big(1-g(x,\gamma)\big)(2-3g(x,\gamma))xx^T| < 2|x|^2.$$

Продовжуючи (3.4), маємо

$$|g(x,\gamma_1)\dot g(x,\gamma_1) - g(x,\gamma_2)\dot g(x,\gamma_2)| \leqslant 2|x|^3|\gamma_1-\gamma_2| =: h_{01}(x)\rho(\gamma_1,\gamma_2).$$

Отже, умови леми 3.3 перевірено.

□

Лема 3.4. *Нехай для функцій $\psi_{j;n}$ виконується умова* $(\Psi_{h\rho})$, *при чому для всіх* $m = 1, ..., M$, $\mathbb{E}\left(h(\xi_{(m)})\right)^2 < \infty$. *Крім того,* $\det \Gamma_\infty \neq 0$.

Тоді існує така послідовність випадкових величин ζ_n, *що* $\zeta_n = o_P(1)$ *і для всіх* $n = 1, 2, ...,$ *всіх* $\gamma_1, \gamma_2 \in \Theta$ *і всіх* $i = 1, ..., n$

$$|\Psi_n^k(\gamma_1) - \Psi_n^k(\gamma_2)| \leqslant \zeta_n \rho(\gamma_1, \gamma_2) \tag{3.5}$$

$$|\Psi_{ni-}^k(\gamma_1) - \Psi_{ni-}^k(\gamma_2)| \leqslant \zeta_n \rho(\gamma_1, \gamma_2) \tag{3.6}$$

Доведення. Почнемо з нерівності (3.5).

$$|\Psi_n^k(\gamma_1) - \Psi_n^k(\gamma_2)| \leqslant \sum_{j=1}^n |a_{j;n}^k| \Big|\psi_{j;n}(\xi_{j;n}, \gamma_1) - \psi_{j;n}(\xi_{j;n}, \gamma_2)\Big| \leqslant$$

$$\leqslant \sum_{j=1}^n |a_{j;n}^k| h(\xi_{j;n}) \rho(\gamma_1, \gamma_2)$$

внаслідок виконання умови $(\Psi_{h\rho})$. За лемою 3.1, для деякого $C_a < \infty$, $|a_{j;n}^k| \leqslant \frac{C_a}{n}$ для всіх j, n. Отже,

$$|\Psi_n^k(\gamma_1) - \Psi_n^k(\gamma_2)| \leqslant \frac{C_a}{n} \rho(\gamma_1, \gamma_2) \sum_{j=1}^n h(\xi_{j;n}).$$

З умов леми випливає, що

$$\max_{j,n} \mathbb{E}\, h(\xi_{j;n}) \leqslant \max_{m=1,\dots,M} \mathbb{E}\, h(\xi_{(m)}) =: A_1 < \infty$$

і

$$\max_{j,n} \mathbb{D}\, h(\xi_{j;n}) \leqslant \max_{m=1,\dots,M} \mathbb{E}\left(h(\xi_{(m)})\right)^2 =: A_2 < \infty.$$

Тому для довільного $\lambda > A_1$, за нерівністю Чебишова отримуємо

$$\mathbb{P}\{\frac{1}{n}\sum_{j=1}^n h(\xi_{j;n}) > \lambda\} \leqslant \frac{\mathbb{D}\,\frac{1}{n}\sum_{j=1}^n h(\xi_{j;n})}{\left(\lambda - \mathbb{E}\,\frac{1}{n}\sum_{j=1}^n h(\xi_{j;n})\right)^2} \leqslant \frac{1}{n}\frac{A_2}{(\lambda - A_1)^2} \to 0$$

при $n \to \infty$. Отже, (3.5) виконується, якщо вибрати $\zeta_n = \frac{C_a}{n}\sum_{j=1}^n h(\xi_{j;n})$.

Щоб довести (3.6), помітимо, що

$$|\Psi_{ni-}^k(\gamma_1) - \Psi_{ni-}^k(\gamma_2)| \leqslant \sum_{j \neq i}^n |a_{ji-;n}^k| h(\xi_{j;n}) \rho(\gamma_1, \gamma_2). \tag{3.7}$$

Оскільки
$$|a^k_{ji-;n}| \leqslant |a^k_{j;n}| + |a^k_{ji-;n} - a^k_{j;n}|,$$
то за лемою 3.1, існує таке $C'_a < \infty$, що $\max_{i,j=1,\dots,n} |a^k_{ji-;n}| \leqslant \frac{C'_a}{n}$.

Тому з (3.7) випливає
$$\max_{i=1,\dots,n} |\Psi^k_{ni-}(\gamma_1) - \Psi^k_{ni-}(\gamma_2)| \leqslant \frac{C'_a}{n} \sum_{j=1}^{n} h(\xi_{j;n}).$$
Отже, (3.6) виконано з $\zeta_n = \frac{C'_a}{n} \sum_{j=1}^{n} h(\xi_{j;n})$ □

Лема 3.5. *Нехай функції $\psi_{j;n}$ мають вигляд (3.2) і виконуються наступні умови:*

1 Для функцій $\dot{g}(x,\gamma)$ і $g(x,\gamma)\dot{g}(x,\gamma)$ виконуються умови $(\Psi_{h_1\rho})$ і $(\Psi_{h_{01}\rho})$ відповідно. В обох випадках функція ρ одна і та ж сама.

2 $\mathbb{E}\left(|Y_{(m)}|h_1(X_{(m)}) + h_{01}(X_{(m)})\right)^2 < \infty$ *для усіх* $m = 1, \dots, M$.

3 $\det \Gamma_\infty \neq 0$.

Тоді має місце лема 3.4 для функцій $\psi_{j;n}$

Доведення. Для доведення леми, перевіримо умови леми 3.4.

За лемою 3.3 і умови 1) леми, для функцій $\psi_{j;n}$ виконується умова $\Psi_{h\rho}$ із $h(z) = |y|h_1(x) + h_{01}(x)$. З цього, (3.2), та умови 2) леми, видно, що умова 2) леми 3.4 також виконується.

Лему доведено □

Приклад 3.3. Нехай функції $\psi_{j;n}$ і g мають вигляд (3.2) і (3.3), та виконуються наступні умови:

1 $\exists\, \mathbb{E}\,|X_{(m)}|^6$ для всіх $m = 1, \dots, M$.

2 $\exists\, \mathbb{E}\,|\varepsilon_{(m)}|^2$ для всіх $m = 1, \dots, M$.

3 $\det \Gamma_\infty \neq 0$.

Тоді виконується твердження леми 3.4 для функцій $\psi_{j;n}$.

Доведення. Для доведення цього твердження, перевіримо умови леми 3.5. З умов прикладу і прикладу (3.2) видно, що для $\psi_{j,n}$ виконується умова $(\Psi_{h\rho})$.

Далі перевіримо умову 2) леми 3.5. Для усіх $m = 1, \dots, M$, з прикладу 3.2 і обмеженості $g(x,\gamma)$ маємо наступне:

$$\mathbb{E}\left(|Y_{(m)}|h_1(X_{(m)}) + h_{01}(X_{(m)})\right)^2 = \mathbb{E}\left((|Y_{(m)}| + 2)^2|X_{(m)}|^6\right) \leqslant$$

$$\leqslant \mathbb{E}\left((3 + \varepsilon_{(m)})^2|X_{(m)}|^6\right) \leqslant \infty$$

з незалежності випадкових величин $\varepsilon_{(m)}$ і $X_{(m)}$, та умов 1) і 2) даного прикладу.

Усі умови перевірено і твердження леми 3.5 виконується.

□

Лема 3.6. *Нехай виконуються наступні умови:*

1 Θ *- компакт в* $\mathbb{R}^d$.

2 Виконується умова $(\Psi_{h\rho})$ *для функцій* $\psi_{j;n}$.

3 Функція $\rho(\gamma_1, \gamma_2)$ *- неперервна на* $\Theta \times \Theta$ *і* $\rho(\gamma, \gamma)=0$ *для всіх* $\gamma \in \Theta$.

4 Для всіх $m = 1, ..., M$, $\mathbb{E}\left(h(\xi_{(m)})\right)^2 < \infty$, $\sup_{j;n} \mathbb{E}\,|\psi_{j;n}(\xi_{(m)}, \theta^{(m)})|^2 < \infty$.

5 $\det \Gamma_\infty \neq 0$.

Тоді

$$\sup_{\gamma\in\Theta} |\Psi^k_n(\gamma) - \mathbb{E}\,\Psi^k_n(\gamma)| \to 0 \tag{3.8}$$

$$\sup_{\gamma\in\Theta} |\Psi^k_{ni-}(\gamma) - \mathbb{E}\,\Psi^k_{ni-}(\gamma)| \to 0 \tag{3.9}$$

при $n \to \infty$ *за ймовірністю.*

Доведення. Розглянемо $\bar{\psi}_{j;n}(\gamma) = \mathbb{E}\,\psi_{j;n}(\xi_{j;n}, \gamma)$. З умов 1-4 випливає, що

$$\max_{j,n} \sup_{\gamma\in\Theta} |\bar{\psi}_{j;n}(\gamma)| < \infty \tag{3.10}$$

і існує така $C_\psi < \infty$, що

$$\max_{j,n} \sup_{\gamma\in\Theta} |\bar{\psi}_{j;n}(\gamma_1) - \bar{\psi}_{j;n}(\gamma_2)| \leqslant C_\psi \rho(\gamma_1, \gamma_2) \tag{3.11}$$

для всіх $\gamma_1, \gamma_2 \in \Theta$.

Визначимо $\tilde{\psi}_{j;n}(\xi_{j;n}, \gamma) = \psi_{j;n}(\xi_{j;n}, \gamma) - \bar{\psi}_{j;n}(\xi_{j;n}, \gamma)$. З (3.10) і (3.11) випливає, що для набору функцій $\tilde{\psi}_{j;n}$ виконується умова $(\Psi_{\tilde{h}\rho})$ з функцією $\tilde{h}(x) = h(x) + C_\psi$. Для цієї функції внаслідок умови 4),

$$\mathbb{E}\left(h(\xi_{(m)})\right)^2 < \infty, \tag{3.12}$$

і враховуючи (3.10), маємо

$$\max_{j,n} \mathbb{E} \left|\tilde{\psi}_{j;n}(\xi_{(m)}, \theta^{(m)})\right|^2 < \infty. \tag{3.13}$$

Оскільки $\Psi_n^k(\gamma) - \mathbb{E}\,\Psi_n^k(\gamma) = \tilde{\Psi}_n(\gamma)$ і $\Psi_{ni-}^k(\gamma) - \mathbb{E}\,\Psi_{ni-}^k(\gamma) = \tilde{\Psi}_{ni-}(\gamma)$, де $\tilde{\Psi}_n(\gamma) = \sum_{j=1}^n a_{j;n}^k \psi_{j;n}(\xi_{k;n}, \gamma)$, $\tilde{\Psi}_{ni-}(\gamma) = \sum_{j\neq i}^n a_{ji-;n}^k \psi_{j;n}(\xi_{k;n}, \gamma)$, То для доведення леми нам досить показати, що

$$\sup_{\gamma\in\Theta} |\tilde{\Psi}_n(\gamma)| \to 0 \text{ за ймовірністю при } n \to \infty \tag{3.14}$$

і

$$\max_{i=1,\dots,n} \sup_{\gamma\in\Theta} |\tilde{\Psi}_{ni-}(\gamma)| \to 0 \text{ за ймовірністю при } n \to \infty. \tag{3.15}$$

Почнемо з доведення (3.14). Спочатку зафіксуємо довільне $\gamma \in \Theta$.

$$\mathbb{E}\,\tilde{\Psi}_n(\gamma) = \sum_{i=1}^n a_{j;n}^k \,\mathbb{E}\,\tilde{\psi}_{j;n}(\xi_{j;n}, \gamma) = 0. \tag{3.16}$$

нагадаємо, що

$$\tilde{\psi}_{j;n}(\xi_{j;n}, \gamma) = \left(\tilde{\psi}_{j;n}^1(\xi_{j;n}, \gamma), ..., \tilde{\psi}_{j;n}^d(\xi_{j;n}, \gamma)\right)^T$$

і, аналогічно,

$$\tilde{\Psi}_n(\gamma) = \left(\tilde{\Psi}_n^1(\gamma), ..., \tilde{\Psi}_n^d(\gamma)\right)^T.$$

Для довільного $l = 1, ..., d$ розглянемо

$$\mathbb{D}\,\tilde{\Psi}_n^l(\gamma) = \sum_{j=1}^n (a_{j;n}^k)^2\,\mathbb{D}\,\tilde{\psi}_{j;n}^l \leqslant \frac{C_a^2}{n} \max_{j=1,\dots,n} \mathbb{D}\,\tilde{\psi}_{j;n}(\xi_{j;n}, \gamma). \tag{3.17}$$

З умови $(\Psi_{\tilde{h}\rho})$ отримуємо

$$\left|\tilde{\psi}_{j;n}(\xi_{j;n}, \gamma)\right| \leqslant \left|\tilde{\psi}_{j;n}(\xi_{j;n}, \theta^{(m)})\right| + \tilde{h}(\xi_{j;n})\rho(\theta^{(m)}, \gamma).$$

Звідси, враховуючи (3.12) і (3.13), отримуємо, що

$$\max_{\substack{j=1,\dots,n\\ n=1,2,\dots}} \mathbb{D}\,\psi_{j;n}^l(\xi_{j;n}, \gamma) < \infty.$$

Отже, з (3.16) і (3.17) отримуємо, що

$$\Psi_n(\gamma) \to 0,\ n \to \infty \text{ за ймовірністю} \tag{3.18}$$

для будь-якого $\gamma \in \Theta$.

Застосовуючи лему 3.4 до функцій $\tilde{\psi}_{j;n}$, отримуємо, що існує така послідовність в.в. $\zeta_n = o_p(1)$, і

$$\left|\tilde{\Psi}_n(\gamma_1) - \tilde{\Psi}_n(\gamma_2)\right| \leqslant \zeta_n \rho(\gamma_1, \gamma_2) \tag{3.19}$$

для всіх $\gamma_1, \gamma_2 \in \Theta$.

Щоб довести (3.14), потрібно показати, що для будь-яких $\delta > 0$ і $\varepsilon > 0$ знайдеться таке n_0, що для всіх $n > n_0$

$$\mathbb{P}\{\sup_{\gamma \in \Theta} |\tilde{\Psi}_n(\gamma)| > \delta\} < \varepsilon. \tag{3.20}$$

Фіксуємо ε і δ. Виберемо невипадкове число $C_\zeta < \infty$ так, щоб $\mathbb{P}\{\zeta_n > C_\zeta\} < \frac{\varepsilon}{2}$ для всіх n.

Оскільки Θ - компакт і виконується умова 3) леми, то можна вибрати скінченний набір $T = \{t_1, ..., t_L\} \subseteq \Theta$ такий, що для кожного $\gamma \in \Theta$ існує номер $l(\gamma) \in \{1, ..., L\}$ для якого

$$\rho(\gamma, t_{l(\gamma)}) < \frac{\delta}{2C_\zeta}. \tag{3.21}$$

Запишемо

$$|\tilde{\Psi}_n(\gamma)| \leqslant |\tilde{\Psi}_n(\gamma) - \tilde{\Psi}_n(t_{l(\gamma)})| + |\tilde{\Psi}_n(t_{l(\gamma)})|.$$

Отже

$$\sup_{\gamma \in \Theta} |\tilde{\Psi}_n(\gamma)| \leqslant \sup_{\gamma \in \Theta} |\tilde{\Psi}_n(\gamma) - \tilde{\Psi}_n(t_{l(\gamma)})| + \max_{l=1,\dots,L} |\tilde{\Psi}_n(t_{l(\gamma)})| \leqslant$$

$$\zeta_n \sup_{\gamma \in \Theta} \rho(\gamma, t_{l(\gamma)}) + \max_{l=1,\dots,L} |\tilde{\Psi}_n(t_l)|.$$

Тому

$$\mathbb{P}\{\sup_{\gamma \in \Theta} |\tilde{\Psi}_n(\gamma)| > \delta\} \leqslant \mathbb{P}\{\zeta_n \sup_{\gamma \in \Theta} \rho(\gamma, t_{l(\gamma)}) > \delta/2\} + \mathbb{P}\{\max_{l=1,\dots,L} |\tilde{\Psi}_n(t_l)| > \delta/2\}. \tag{3.22}$$

Внаслідок (3.18) $\mathbb{P}\{\max_{l=1,\dots,L} |\tilde{\Psi}_n(t_l)| > \delta/2\} \to 0$ при $n \to \infty$, отже, можна вибрати n_0 так, щоб цей доданок у правій частині (3.22) був менший ніж $\varepsilon/2$ для всіх $n > n_0$.

Оцінимо перший доданок. З (3.21) маємо

$$\sup_{\gamma \in \Theta} \rho(\gamma, t_{l(\gamma)}) < \frac{\delta}{2C_\zeta}, \text{ отже}$$

$$\mathbb{P}\{\zeta_n \sup_{\gamma\in\Theta} \rho(\gamma, t_{l(\gamma)}) > \delta/2\} \leqslant \mathbb{P}\{\zeta_n > C_\zeta\} < \frac{\varepsilon}{2}$$

за вибором C_ζ.

Отже, (3.20) виконано і (3.8) доведено.

Доведемо (3.9). Для цього оцінимо

$$\tilde{\Psi}_n(\gamma) - \tilde{\Psi}_{ni-}(\gamma) = a^k_{i;n}\tilde{\psi}_{i;n}(\xi_{i;n}, \gamma) + \sum_{j\neq i} \left(a^k_{j;n} - a^k_{ji-;n}\right)\tilde{\psi}_{j;n}(\xi_{j;n}, \gamma).$$

Для оцінки $\max_{j=1,\dots,n} \sup_{\gamma\in\Theta} |\tilde{\psi}_{j;n}(\xi_{j;n}, \gamma)|$ скористаємося лемою 2.1 з $H(x) = |x|$. Внаслідок умови 4) леми, умова леми 2.1 виконується з $\alpha = 2$. Отже,

$$\max_{i=j,\dots,n} \sup_{\gamma\in\Theta} |\psi_{j;n}(\xi_{j;n}, \gamma)| = o_P\left(\sqrt{n}\right),$$

і враховуючи, що за лемою 3.1, $|a^k_{j;n}| \leqslant \frac{C_a}{n}$, $|a^k_{ji-;n} - a^k_{j;n}| \leqslant \frac{C_a}{n^2}$, отримуємо

$$\max_{i=1,\dots,n} \sup_{\gamma\in\Theta} |\tilde{\Psi}_n(\gamma) - \tilde{\Psi}_{ni-}(\gamma)| = o_P\left(\frac{1}{\sqrt{n}}\right).$$

Звідси, з урахуванням 3.8, отримуємо (3.9).

□

Лема 3.7. *Нехай функції $\psi_{j;n}$ мають вигляд (3.2) і виконуються наступні умови:*

1 Θ *- компакт в* $\mathbb{R}^d$.

2 Для функцій $\dot{g}(x, \gamma), g(x, \gamma)\dot{g}(x, \gamma)$ *виконуються умови* $(\Psi_{h_1,\rho})$, $(\Psi_{h_{01},\rho})$ *відповідно.*

3 Функція $\rho(\gamma_1, \gamma_2)$ *- неперервна на* $\Theta\times\Theta$ *і* $\rho(\gamma, \gamma)$*=0 для всіх* $\gamma \in \Theta$.

4 Для всіх $m = 1, ..., M$, $\mathbb{E}\left(|Y_{(m)}|h_1(\xi_{(m)}) + h_{01}(\xi_{(m)})\right)^2 < \infty$, $\mathbb{E}\,|\varepsilon_{(m)}|^2 < \infty$, $\mathbb{E}\,|\dot{g}(\xi_{(m)}, \theta^{(m)})|^2 < \infty$.

5 $\det\Gamma_\infty \neq 0$.

Тоді для функцій $\psi_{j;n}$ *виконується твердження леми* 3.6

Доведення. Доведення леми полягає у перевірці умов 2) і 4) леми 3.6.

Умова 2) леми 3.6 виконується внаслідок леми 3.3. Покажемо, що умова 4) леми 3.6 також виконується. Дійсно,

$$\mathbb{E}\left|\psi_{j;n}(\xi_{(m)}, \theta^{(m)})|^2 = \mathbb{E}\,|\left(Y_{(m)} - g(X_{(m)}, \theta^{(m)})\right)\dot{g}(X_{(m)}, \theta^{(m)})\right|^2 =$$

$$\mathbb{E}\,\varepsilon_{(m)}^2\,\mathbb{E}\,|\dot{g}(X_{(m)},\theta^{(m)})|^2 < \infty$$

за умовами теореми. За лемою 3.3 і за умовою 4) леми

$$\mathbb{E}\left(h(\xi_{(m)})\right)^2 = \mathbb{E}\left(|Y_{(m)}|h_1(\xi_{(m)}) + h_{01}(\xi_{(m)})\right)^2 < \infty.$$

Лему доведено. □

Приклад 3.4. Нехай функції $\psi_{j;n}$ і g мають вигляд (3.2) і (3.3) відповідно, та виконуються наступні умови:

1 Θ - компакт в $\mathbb{R}^d$.

2 Для всіх $m = 1, ..., M$, $\mathbb{E}\,|\xi_{(m)}|^6 < \infty$, $\mathbb{E}\,|\varepsilon_{(m)}|^2 < \infty$.

3 $\det \Gamma_\infty \neq 0$.

Тоді для функцій $\psi_{j;n}$ виконується твердження леми 3.6

Доведення. Покажемо, що умови 2), 3), 4) леми 3.7 виконуються. Виконання умови 2) леми 3.7 перевірено у прикладі 3.2, із $\rho(\gamma_1, \gamma_2) = |\gamma_1 - \gamma_2|$, яка є неперервною. Це означає, що умова 3) леми 3.7 виконана.

Покажемо, що умова 4) леми 3.7 виконується. Дійсно,

$$\mathbb{E}\left(|Y_{(m)}|h_1(\xi_{(m)}) + h_{01}(\xi_{(m)})\right)^2 \leqslant \mathbb{E}\left(\left(|\varepsilon_{(m)}| + 1\right)|X_{(m)}|^3 + 2|X_{(m)}|^3\right) \leqslant$$

$$\leqslant \mathbb{E}\left(\left(|\varepsilon_{(m)}| + 3\right)^2|X_{(m)}|^6\right) < \infty$$

внаслідок умови 2). Покажемо скінченність математичного сподівання $\mathbb{E}\,|\dot{g}(\xi_{(m)}, \theta^{(m)})|^2$. Дійсно, з умови 2) випливає, що

$$\mathbb{E}\left|\dot{g}(\xi_{(m)}, \theta^{(m)})\right|^2 = \mathbb{E}\,g(\xi_{(m)}, \theta^{(m)})^2\left|1 - g(\xi_{(m)}, \theta^{(m)})\right|^2|X_{(m)}|^2 < \infty.$$

□

3.3 Консистентність УОР-оцінок

У цьому підрозділі ми доведемо теорему про консистентність УОР-оцінок $\hat{\theta}_n^{(k)}$, визначених (3.5). Крім того, ми дослідимо асимптотичну поведінку оцінок $\hat{\theta}_{i-;n}^{(k)}$, заданих (3.9), які використовуються при побудові оцінки методу складаного ножа для асимптотичної коваріації матриці оцінок $\hat{\theta}_n^{(k)}$.

Взагалі кажучи, не очевидно, що рівняння (3.5) і (3.9) мають розв'язки, тобто оцінки $\hat{\theta}^{(k)}_{i-;n}$ і $\hat{\theta}^{(k)}_{n}$ можуть взагалі не існувати при деяких значеннях n та i. У наступній теоремі ми вважаємо, що ймовірність існування розв'язків (3.5) і (3.9) прямує до 1 при $n \to \infty$. Далі у цьому підрозділі доведена теорема, яка дає умови виконання цього припущення. Символом z далі позначатимемо пару (x, y), де $x \in \mathbb{R}^{D-1}, y \in \mathbb{R}$.

Теорема 3.1. *Нехай виконуються наступні умови:*

1 Θ *- компакт в* $\mathbb{R}^d$.

2 Для елементарних оціночних функцій $s(x, \gamma)$ *виконується умова* ($\Psi_{h\rho}$).

3 ρ *- неперервна функція на* $\Theta \times \Theta$ *і* $\rho(\gamma, \gamma) = 0$ *для всіх* $\gamma \in \Theta$.

4 Для всіх $m = 1, ..., M$, $\mathbb{E}\left(s(\xi_{(m)}, \theta^{(m)})\right)^2 < \infty$ *і* $\mathbb{E}\left(h(\xi_{(m)})\right)^2 < \infty$.

5 $\mathbb{E}\, s(\xi_{(k)}, \theta^{(k)}) = 0$ *тоді і тільки тоді, коли* $\gamma = \theta^{(k)}$.

6 $\mathbb{P}\{$*Розв'язок рівняння* $S^{(k)}_n(\gamma) = 0$ *існує*$\} \to 1$ *при* $n \to \infty$.

6' $\mathbb{P}\{$*Для всіх* $i = 1, ..., n$ *розв'язок рівняння* $S^{(k)}_{ni-}(\gamma) = 0$ *існує*$\} \to 1$, *при* $n \to \infty$.

7 $\det \Gamma_\infty \neq 0$.

Тоді при виконанні умов 1)-6) і 7) $\hat{\theta}^{(k)}_n \to \theta^{(k)}$ *за ймовірністю при* $n \to \infty$.

При виконанні умов 1)-5), 6'), 7) $\sup_{i=1,\dots,n} |\hat{\theta}^{(k)}_{i-;n} - \theta^{(k)}| \to 0$ *за ймовірністю, при* $n \to \infty$.

Доведення. 1. Доведемо, що $\hat{\theta}^{(k)}_n \to \theta^{(k)}$ за ймовірністю при $n \to \infty$. Позначимо $S^k_\infty(\gamma) = \mathbb{E}\, s(\xi_{(k)}, \gamma)$. З умов 3) і 4) теореми випливає, що S^k_∞ - неперервна функція на Θ. За умовою 5), $|S^k_\infty(\gamma)| > 0$ для всіх $\gamma \neq \theta^{(k)}$.

Зафіксуємо довільне $\varepsilon > 0$ і розглянемо $N_\varepsilon = \{\gamma \in \Theta : |\gamma - \theta^{(k)}| > \varepsilon\}$. Тоді $s_{min} = \inf_{\gamma \in N_\varepsilon} |S^k_\infty(\gamma)| > 0$.

Оскільки для $\gamma \in N_\varepsilon$,

$$|S^k_n(\gamma)| \geqslant S^k_\infty(\gamma) - |S^k_n(\gamma) - S_\infty(\gamma)| \geqslant$$

$$\geqslant \inf_{\gamma \in N_\varepsilon} S^k_\infty(\gamma) - \sup_{\gamma \in \Theta} |S^k_n(\gamma) - S^k_\infty(\gamma)|,$$

то

$$\mathbb{P}\{\inf_{\gamma \in N_\varepsilon} |S^k_n(\gamma)| < s_{min}/2\} \leqslant \mathbb{P}\{\sup_{\gamma \in \Theta} |S^k_n(\gamma) - S^k_\infty(\gamma)| > s_{min}/2\}.$$

Застосувавши лему 3.6 до функцій $\psi_{j;n} = s$, отримуємо, що

$$\sup_{\gamma\in\Theta} |S_n^k(\gamma) - S_\infty^k(\gamma)| \to \infty,$$

при $n \to \infty$ за ймовірністю.

Отже,

$$\mathbb{P}\{\inf_{\gamma\in N_\varepsilon} |S_n^k(\gamma)| < s_{min}/2\} \to 0 \text{ при } n \to \infty.$$

За означення, $S_n^k(\hat{\theta}_n^{(k)}) = 0$, отже, якщо $\hat{\theta}_n^{(k)} \in N_\varepsilon$, то $\inf_{\gamma\in N_\varepsilon} |S_n^k(\gamma)| = 0$. Умова $\hat{\theta}_n^{(k)} \in N_\varepsilon$ еквівалентна $|\hat{\theta}_n^{(k)} - \theta^{(k)}| > \varepsilon$. Таким чином подія $\{|\hat{\theta}_n^{(k)} - \theta^{(k)}| > \varepsilon\} \subseteq \{\inf_{\gamma\in N_\varepsilon} |S_n^k(\gamma)| < s_{min}/2\}$ і $\mathbb{P}\{|\hat{\theta}_n^{(k)} - \theta^{(k)}| > \varepsilon\} \to \infty$. Внаслідок довільності $\varepsilon > 0$ це еквівалентно $\hat{\theta}_n^{(k)} \to \theta^{(k)}$ за ймовірністю.

2. Доведемо, що

$$\sup_{i=1,\dots,n} |\hat{\theta}_{i-;n}^{(k)} - \theta^{(k)}| \to 0 \text{ за ймовірністю при } n \to \infty.$$

За лемою 3.6, отримуємо

$$\max_{i=1,\dots,n} \sup_{\gamma\in\Theta} |S_{ni-}^k(\gamma) - S_\infty^k(\gamma)| \to 0 \text{ за ймовірністю при } n \to \infty.$$

Відповідно,

$$\mathbb{P}\{\inf_{\gamma\in N_\varepsilon} |S_{ni-}^k(\gamma)| < s_{min}/2\} \leqslant \mathbb{P}\{\sup_{\gamma\in\Theta} |S_{ni-}^k(\gamma) - S_\infty^k(\gamma)| > s_{min}/2\} \to 0 \text{ при } n \to \infty.$$

Звідси, так само, як у першій частині доведення, отримуємо

$$\mathbb{P}\{\max_{i=1,\dots,n} |\hat{\theta}_{i-;n}^{(k)} - \theta^{(k)}| > \varepsilon\} \to 0 \text{ при } n \to \infty.$$

Теорему доведено. □

Теорема 3.2. *Нехай $s(z,\gamma)$ мають вигляд (3.2) і виконуються наступні умови:*

1 Θ *- компакт в* $\mathbb{R}^d$.

2 *Для функцій $\dot{g}(x,\gamma)$ і $g(x,\gamma)\dot{g}(x,\gamma)$ виконуються умова ($\Psi_{h_1\rho}$) і ($\Psi_{h_{01}\rho}$).*

3 *ρ - неперервна функція на $\Theta\times\Theta$ і $\rho(\gamma,\gamma) = 0$ для всіх $\gamma \in \Theta$.*

4 *Для всіх $m = 1,\dots,M$, $\mathbb{E}\,\varepsilon_{(m)}^2 < \infty$, $\mathbb{E}\,|\dot{g}(\xi_{(m)}, \theta^{(m)})|^2 < \infty$, і*

$$\mathbb{E}\left(|Y_{(m)}|h_0(X_{(m)}) + h_{01}(X_{(m)})\right)^2 < \infty.$$

5 $\mathbb{E}\left(Y_{(m)} - g(\xi_{(k)}, \theta^{(k)})\right)\dot{g}(\xi_{(k)}, \theta^{(k)}) = 0$ *тоді і тільки тоді, коли* $\gamma = \theta^{(k)}$.

6 $\mathbb{P}\{$*Розв'язок рівняння* $S_n^{(k)}(\gamma) = 0$ *існує*$\} \to 1$ *при* $n \to \infty$.

6' $\mathbb{P}\{$*Для всіх* $i = 1, ..., n$ *розв'язок рівняння* $S_{ni-}^{(k)}(\gamma) = 0$ *існує*$\} \to 1$ *при* $n \to \infty$.

7 $\det \Gamma_\infty \neq 0$.

Тоді для функцій $s(z,\gamma) = \big(y - g(x,\gamma)\big)\dot{g}(x,\gamma)$ *і оцінок* $\hat{\theta}_n^{(k)}$, $\hat{\theta}_{i-;n}^{(k)}$ *виконується теорема* 3.1

Доведення. Для доведення теореми, використаємо теорему 3.1 і перевіримо її умови 2) і 3).

З умови 2) теореми і леми 3.3, випливатиме умова 2) теореми 3.1 з функцією $h(z) = |y|h_1(x) + h_{01}(x)$.

З попереднього і умови 4) теореми випливатиме умова 4) теореми 3.1. Дійсно,

$$\mathbb{E}\left[\left(Y_{(m)} - g(X_{(m)}, \theta^{(m)})\right)\dot{g}(X_{(m)}, \theta^{(m)})\right]^2 = \mathbb{E}\,\varepsilon_{(m)}^2\,\mathbb{E}\,|\dot{g}(X_{(m)}, \theta^{(m)})|^2 < \infty.$$

Теорему доведено.

□

Приклад 3.5. Нехай $s(z,\gamma) = \big(y - g(x,\gamma)\big)\dot{g}(x,\gamma)$, $g(x,\gamma) = \frac{1}{1+e^{-\langle x,\gamma,\rangle}}$ і виконуються наступні умови:

1 Θ - компакт в $\mathbb{R}^d$.

2 Для всіх $m = 1, ..., M$, $\mathbb{E}\,\varepsilon_{(m)}^2 < \infty$, $\mathbb{E}\,|X_{(m)}|^6 < \infty$.

3 Для довільного $\gamma \in \mathbb{R}^d$, $\gamma \neq 0$, $\mathbb{P}\{\langle X_{(m)}, \gamma, \rangle = 0\} < 1$, для усіх $m = 1, ..., M$.

4 $\mathbb{P}\{$Розв'язок рівняння $S_n^{(k)}(\gamma) = 0$ існує$\} \to 1$ при $n \to \infty$.

4' $\mathbb{P}\{$Для всіх $i = 1, ..., n$ розв'язок рівняння $S_{ni-}^{(k)}(\gamma) = 0$ існує$\} \to 1$ при $n \to \infty$.

5 $\det \Gamma_\infty \neq 0$.

Тоді для функцій s, g і оцінок $\hat{\theta}_n^{(k)}$, $\hat{\theta}_{i-;n}^{(k)}$ виконується твердження теореми 3.1

Доведення. Для доведення, застосуємо теорему 3.2. Для цього, перевіримо її умови 2)-5).

Почнемо із другої умови теореми 3.2. У прикладі 3.2 показано виконання цієї умови, при чому $h(\xi_{(m)}) = \big(|Y_{(m)}| + 2\big)|X_{(m)}|^3$ і $\rho(\gamma_1, \gamma_2) = |\gamma_1 - \gamma_2|$. З останнього випливає умова 3) теореми 3.2

Перевіримо четверту умову теореми 3.2:

$$\mathbb{E}\left(h(\xi_{(m)})\right)^2 < \mathbb{E}(\varepsilon_{(m)})^2\,\mathbb{E}\,|X_{(m)}|^6 < \infty$$

з умови 2).

$$\mathbb{E}\,|\dot g(X_{(m)},\gamma)|^2 = \mathbb{E}\left(g(X_{(m)},\theta^{(m)})^2\big(1-g(X_{(m)},\theta^{(m)})\big)^2\right)^2|X_{(m)}|^2 < \infty$$

з умови 2). Отже, умова 4) теореми 3.2 виконана.

Залишилося показати, що умова 5) теореми 3.2 випливає з умови 3). Перевіримо імплікацію в обидва боки. Спочатку покажемо, що із рівності $\gamma = \theta^{(k)}$ випливає $\mathbb{E}\,s(\xi_{(k)},\theta^{(k)} = 0$. Дійсно,

$$\mathbb{E}\,s(\xi_{(k)},\theta^{(k)} = \mathbb{E}\,\varepsilon_{(k)}\,\mathbb{E}\,\dot g(X_{(k)},\theta^{(k)}) = 0,$$

оскільки $\mathbb{E}\,\varepsilon_{(m)} = 0$ і $\varepsilon_{(m)}$ з $X_{(m)}$ незалежні.

Імплікацію в іншу сторону доведемо від супротивного. Припустимо, що $\gamma \neq \theta^{(k)}$, але $\mathbb{E}\,s(\xi_{(m)},\gamma) = 0$. Помітимо, що функція $g(x,\gamma)$ строго монотонна по параметру $\langle x,\gamma\rangle$. Розглянемо випадок, коли $\langle x,\theta^{(k)}-\gamma\rangle \geqslant 0$ (випадок $\langle x,\theta^{(k)}-\gamma\rangle \leqslant 0$ розглядається аналогічно). Тоді

$$\big(g(X_{(k)},\theta^{(k)}) - g(X_{(k)},\gamma)\big)g(X_{(k)},\gamma)\big(1-g(X_{(k)},\gamma)\big)\langle x,\theta^{(k)}-\gamma\rangle \geqslant 0. \tag{3.2}$$

За початковим припущенням,

$$0 = \mathbb{E}\,s(\xi_{(m)},\gamma) = \mathbb{E}\left(g(X_{(k)},\theta^{(k)}) - g(X_{(k)},\gamma)\right)g(X_{(k)},\gamma)\,\big(1-g(X_{(k)},\gamma)\big)\,\langle x,\theta^{(k)}-\gamma\rangle,$$

а це можливо лише за умови, коли (3.2) дорівнює нулю майже напевно. З цього, та строгої додатності $g(X_{(k)},\gamma)\big(1-g(X_{(k)},\gamma)\big)$ випливає рівність $\langle x,\theta^{(k)}-\gamma\rangle = 0$ майже напевно. Це суперечить умові 3) теореми, а отже і припущення $\gamma \neq \theta^{(k)}$, $\mathbb{E}\,s(\xi_{(m)},\gamma) = 0$ хибне.

Отже, всі умови теореми 3.2 перевірено. □

Перейдемо тепер до перевірки умов 6) і 6') теореми 3.1. Для функцій вигляду $\psi : \mathbb{R}^D \times \theta \to \mathbb{R}^d$ будемо позначати

$$\dot\psi(\xi,\gamma) = \frac{\partial}{\partial\gamma^T}\psi(\xi,\gamma) = \begin{pmatrix} \frac{\partial}{\partial\gamma^1}\psi^1(\xi,\gamma)\dots\frac{\partial}{\partial\gamma^d}\psi^1(\xi,\gamma) \\ \vdots\quad\vdots\quad\vdots \\ \frac{\partial}{\partial\gamma^1}\psi^d(\xi,\gamma)\dots\frac{\partial}{\partial\gamma^d}\psi^d(\xi,\gamma)\end{pmatrix}$$

і, аналогічно, для $\psi : \Theta \to \mathbb{R}^D$, $\dot\psi(\gamma) = \frac{\partial}{\partial\gamma^T}\psi(\gamma)$

Теорема 3.3. *Нехай виконуються умови 1)-5) і 7) теореми* 3.1*,* $\theta^{(k)}$ *- внутрішня точка* Θ*, і крім того, існує* $\dot{S}^k_\infty(\theta^{(k)})$ *і* $\det \dot{S}^k_\infty(\theta^{(k)}) \neq 0$*. Тоді виконуються умови 6) і 6') Теореми* 3.1.

Доведення. Доведення є модифікацією доведення леми 6.А з [7]. Як показано у доведенні Теореми 3.1, при виконанні умов 1)-5) і 7),

$$\mathbb{P}\{\sup_{\gamma\in\Theta} |S^k_n(\gamma) - S^k_\infty(\gamma)| > \varepsilon\} \to 0, \text{ при } n \to \infty \tag{3.3}$$

$$\text{і } \mathbb{P}\{\sup_{\gamma\in\Theta} \max_{i=1,\dots,n} |S^k_{ni-}(\gamma) - S^k_\infty(\gamma)| > \varepsilon\} \to 0, \text{ при } n \to \infty \tag{3.4}$$

Позначимо $\Phi = \dot{S}^k_\infty(\theta^{(k)})$. Задамо функцію

$$f_n(\gamma) = \gamma - \Phi^{-1} S^k_n(\gamma), \ \gamma \in \Theta.$$

Оскільки $S^k_\infty(\gamma)$ диференційовна при $\gamma = \theta^{(k)}$, то

$$S^k_\infty(\gamma) = \Phi(\gamma - \theta^{(k)}) + o(|\gamma - \theta^{(k)}|), \text{ при } \gamma \to \theta^{(k)}.$$

Тому існує таке ε_1, що для всіх γ, таких, що $|\gamma - \theta^{(k)}| \leqslant \varepsilon_1$

$$|S^k_\infty(\gamma) - \Phi(\gamma - \theta^{(k)})| \leqslant \frac{1}{2|\Phi^{-1}|}|\gamma - \theta^{(k)}|.$$

Фіксуємо довільне $\delta > 0$. З (3.3) випливає, що для деякого n_0, при всіх $n \geqslant n_0$,

$$\mathbb{P}\{\sup_{\gamma\in\Theta} |S^k_n(\gamma) - S^k_\infty(\gamma)| > \frac{\varepsilon_1}{2|\Phi^{-1}|}\} < \delta. \tag{3.5}$$

Позначимо $\bar{B}(\theta^{(k)}, \varepsilon_1) = \{\gamma \in \Theta : |\gamma - \theta^{(k)}| \leqslant \varepsilon_1\}$ і A_n - випадкову подію $\{\sup_{\gamma\in\Theta} |S^k_n(\gamma) - S^k_\infty(\gamma)| > \frac{\varepsilon_1}{2|\Phi^{-1}|}\}$. За (3.5), при $n > n_0$, $\mathbb{P}\{A_n\} < \delta$.

Якщо подія A_n не виконана і $\gamma \in \bar{B}(\theta^{(k)}, \varepsilon_1)$, то

$$|f_n(\gamma) - \theta^{(k)}| = |\Phi^{-1}(\Phi(\gamma - \theta^{(k)}) - S^k_n(\gamma))| \leqslant$$

$$|\Phi^{-1}|\Big(|S^k_\infty(\gamma) - \Phi(\gamma - \theta^{(k)})| + |S^k_n(\gamma) - S_\infty(\gamma)|\Big) \leqslant$$

$$\leqslant |\Phi^{-1}|\Big(\frac{|\gamma - \theta^{(k)}|}{2|\Phi^{-1}|} + \frac{\varepsilon_1}{2|\Phi^{-1}|}\Big) \leqslant \varepsilon_1.$$

Отже, $f_n(\gamma) \in \bar{B}(\theta^{(k)}, \varepsilon)$, тобто f_n відображає $\bar{B}(\theta^{(k)}, \varepsilon)$ у $\bar{B}(\theta^{(k)}, \varepsilon)$. Оскільки f_n є неперервним відображенням, то за теоремою Брауера, існує така точка $\hat{\theta}_n^{(k)} \in \bar{B}(\theta^{(k)}, \varepsilon)$, що $\hat{\theta}_n^{(k)} = f_n(\hat{\theta}_n^{(k)})$, тобто $\hat{\theta}_n^{(k)} = \hat{\theta}_n^{(k)} - \Phi^{-1} S_n^k(\hat{\theta}_n^{(k)})$, а отже, $S_n^k(\hat{\theta}_n^{(k)}) = 0$.

Таким чином, якщо не виконано подію A_n, то рівняння $S_n^k(\gamma) = 0$ має розв'язок. Оскільки $\mathbb{P}\{A_n\} < \delta$ для будь-яких δ при достатньо великих n, звідси випливає виконання умови 6) Теореми 3.3.

Виконання умови 6') доводиться аналогічно з використанням (3.4). □

3.4 Асимптотична нормальність УОР-оцінок

У цьому підрозділі сформульовані умови, за яких УОР-оцінки визначені (3.5) є асимптотично нормальними. Оскільки для потреб перевірки гіпотез про параметри $\theta^{(k)}$ може бути корисним спільний розподіл їх оцінок для $k = 1, ..., M$ (тобто для всіх компонент суміші разом) ми розглянемо об'єднаний вектор всіх евклідових параметрів моделі:

$$\theta = \left((\theta^{(1)})^T, (\theta^{(2)})^T, ..., (\theta^{(M)})^T \right)^T,$$

та відповідну оцінку

$$\theta = \left((\hat{\theta}_n^{(1)})^T, (\hat{\theta}_n^{(2)})^T, ..., (\hat{\theta}_n^{(M)})^T \right)^T,$$

і встановимо умови, за яких

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{w} N(0, V),$$

та визначимо вигляд коваріаційної матриці V граничного нормального розподілу. Для опису V нам будуть потрібні додаткові позначення.

Позначимо $M^{(k)}(\gamma) = \mathbb{E}\,\dot{S}(\xi_{(k)}, \gamma)$,

$$M^{(k)} = \mathbb{E}\,\dot{S}(\xi_{(k)}, \theta^{(k)}) = M^{(k)}(\theta^{(k)}),$$

$$\langle a^k a^m p^l p^i \rangle = \lim_{n\to\infty} n \sum_{j=1}^{n} a_{j;n}^k a_{j;n}^m p_{j;n}^l p_{j;n}^i,$$

$$\langle a^k a^m p^l \rangle = \lim_{n\to\infty} n \sum_{j=1}^{n} a_{j;n}^k a_{j;n}^m p_{j;n}^l.$$

(Існування цих математичних сподівань та границь є умовами нашої теореми про асимптотичну нормальність).

Тепер позначимо

$$Z^{(m,l)} = \sum_{i=1}^{M} \langle a^m a^l p^i \rangle \, \mathbb{E}\, s(\xi_{(i)}, \theta^{(m)}) s(\xi_{(i)}, \theta^{(l)})^T -$$

$$\sum_{i_1,i_2=1}^{M} \langle a^m a^l p^{i_1} p^{i_2} \rangle \, \mathbb{E}\, s(\xi_{(i_1)}, \theta^{(m)}) \, \mathbb{E}\, s(\xi_{(i_2)}, \theta^{(l)})^T$$

$$V^{(m,l)} = \left(M^{(m)}\right)^{-1} Z^{(m,l)} \left(M^{(l)}\right)^{-T} \tag{3.2}$$

(тут і далі $V^{-T} = (V^{-1})^T$).

Складемо $d \times d$ матриці $V^{(m,l)}$ в одну матрицю V розміру $(Md) \times (Md)$:

$$V = \begin{pmatrix} V^{(1,1)} & \dots & V^{(1,M)} \\ \vdots & & \vdots \\ V^{(M,1)} & \dots & V^{(M,M)} \end{pmatrix}$$

Теорема 3.4. *Нехай виконуються наступні умови:*

1 *θ є внутрішньою точкою $\Theta \times \ldots \times \Theta$.*

2 *У деякому відкритому околі B точки θ існують і є неперервними похідні*

$$\frac{\partial s^l(x,\gamma)}{\partial \gamma^i, \partial \gamma^j} \text{ для всіх } l, i, j = 1, ..., d$$

і майже всіх (відносно мір F_m, $m = 1, ..., M$).

3 *Існує така функція $h(x)$, що*

$$\max_{l,i,j} \sup_{\gamma \in B} \left| \frac{\partial s^l(x,\gamma)}{\partial \gamma^i, \partial \gamma^j} \right| \leqslant h(x)$$

і $E\Big(h(x_{(l)})\Big)^{\alpha} < \infty$ для деякого $\alpha > 1$ і усіх $l = 1, .., M$.

4 *$\mathbb{E}\, |s(\xi_{(m)}, \theta^{(m)})|^3 < \infty$ для всіх $m = 1, ..., M$.*

5 *Матриці $M^{(m)}$ є скінченними і не виродженими для всіх $m = 1, ..., M$.*

6 *Границі $\langle a^k a^m p^l p^i \rangle$ існують для всіх $k, m, l, i = 1, ..., M$.*

7 *Матриця Γ_∞ існує і є не виродженою.*

8 *$\hat{\theta}_n$ є консистентною оцінкою θ.*

Тоді при $n \to \infty$

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{w} N(0, V).$$

Доведення. Введемо позначення $S_n(\gamma) = \left((S_n^1(\gamma))^T, ..., (S_n^M(\gamma))^T\right)^T$.

Доведення теореми стандартне. Застосувавши розклад Тейлора до (3.5), отримаємо

$$\sqrt{n}(\hat{\theta}_n - \theta) = -\left[\dot{S}_n^{(k)}(\zeta)\right]^{-1}\left(\sqrt{n}S_n(\theta)\right),$$

тут ζ - проміжна точка між $\hat{\theta}_n$ та θ. З припущення 2) випливатиме, що для набору функцій $\dot{s}(x,\gamma)$, виконується умова $\Psi_{h\rho}$. З цього факту, разом з припущеннями 2)-5) і 7) теореми, та лемою 3.6 випливає, що при $n \to \infty$,

$$\dot{S}_n^{(k)}(\zeta) \to \mathbb{E}\,\dot{S}_n^{(k)}(\theta) = M^{(k)}.$$

З (3.3), маємо $\mathbb{E}\,S_n^{(k)}(\theta^{(k)}) = 0$. Отже,

$$\mathbb{C}\mathrm{ov}\left(S_n^{(m)}(\theta^{(m)}), S_n^{(l)}(\theta^{(l)})\right) = \sum_{j=1}^{n} a_{j;n}^m a_{j;n}^l\, \mathbb{C}\mathrm{ov}(s(\xi_j, \theta^{(m)}), s(\xi_j, \theta^{(l)})),$$

$$\mathbb{C}\mathrm{ov}(s(\xi_j, \theta^{(m)}), s(\xi_j, \theta^{(l)})) = \mathbb{E}\,s(\xi_j, \theta^{(m)})s^T(\xi_j, \theta^{(l)})$$

$$-\,\mathbb{E}\,s(\xi_j, \theta^{(m)})\,\mathbb{E}\,s^T(\xi_j, \theta^{(l)}) =$$

$$\sum_{i=1}^{M} p_{j;n}^i\, \mathbb{E}\,s(\xi_{(i)}, \theta^{(m)})s^T(\xi_{(i)}, \theta^{(l)}) - \sum_{i,k=1}^{M} p_{j;n}^i p_{j;n}^k\, \mathbb{E}\,s(\xi_{(i)}, \theta^{(m)})\,\mathbb{E}\,s^T(\xi_{(k)}, \theta^{(l)}),$$

Отже

$$\lim_{n\to\infty} n\,\mathbb{C}\mathrm{ov}\left(S_n^{(m)}(\theta^{(m)}), S_n^{(l)}(\theta^{(l)})\right) = Z^{(m,l)}. \tag{3.3}$$

Використаємо теорему Крамера-Вальда, а потім центральну граничну теорему з умовою Ляпунова, і покажемо, що система векторів $(\sqrt{n}S_n^{(k)}(\theta^{(k)}), k = 1, ..., M)$ збігається за розподілом до гауссової системи векторів $(u^{(k)}, k = 1, ..., M)$, таких, що

$$\mathbb{E}\,u^{(k)} = 0,\ \mathbb{E}\,u^{(k)}(u^{(m)})^T = Z^{(k,m)}.$$

Для простоти, розглянемо $S_n(\theta) = (S_n^{(k)}(\theta^{(k)}), S_n^{(l)}(\theta^{(l)}))$, при різних k і l (випадок $k = l$, аналогічний). За теоремою Крамера-Вальда, достатньо знайти граничний розподіл $\sqrt{n}\langle c, S_n(\theta)\rangle$, $\forall c \in \mathbb{R}^{2d}$. Для цього, перевіримо умову Ляпунова.

Розглянемо випадкові величини

$$\eta_j^{(k,l)} = \left(na_{j;n}^k s(\xi_j, \theta^{(k)})^T, na_{j;n}^l s(\xi_j, \theta^{(l)})^T\right) =$$

$$\left(na_{j;n}^k s_1(\xi_j, \theta^{(k)}), ..., na_{j;n}^k s_1(\xi_j, \theta^{(k)}), na_{j;n}^k s_1(\xi_j, \theta^{(l)}), ..., na_{j;n}^k s_d(\xi_j, \theta^{(l)})\right), j = 1, ..., n.$$

Візьмемо довільне $c \in \mathbb{R}^{2d}$. З властивості навантажень $a_{j;n}^{(k)}$, випливатиме, що $\mathbb{E}\,\eta_j^{(k,l)} = 0$. З Леми 3.1, випливатиме, що $s_n^2 = \mathbb{D}\,\eta_j^{(k,l)} = O(n)$, і $r_n^3 = \sum_{j=1}^n \mathbb{E}\,|\langle c, \eta_j^{(k,l)}\rangle|^3 = O(n)$ за умовою 3) теореми. З цього випливає, що $\frac{r_n}{s_n} = O(\frac{1}{n^{1/6}}) = o(1)$. Отже, за ЦГТ (Ляпунова),

$$\frac{n\langle c, S_n(\theta)\rangle}{\mathbb{D}\,n\langle c, S_n(\theta)\rangle} \xrightarrow{w} N(0,1).$$

Оскільки

$$\mathbb{D}\,n\langle c, S_n(\theta)\rangle = n^2 c^T\,\mathbb{C}\mathrm{ov}\,S_n(\theta)c,$$

і

$$nc^T\,\mathbb{C}\mathrm{ov}\,S_n(\theta)c \to c^T \begin{pmatrix} Z^{(k,k)} & Z^{(k,l)} \\ Z^{(l,k)} & Z^{(l,l)} \end{pmatrix} c, \text{ при } n \to \infty,$$

то з властивостей збіжності за розподілом, та теореми Слуцького

$$\sqrt{n}\langle c, S_n(\theta)\rangle \xrightarrow{w} N\left(0, c^T \begin{pmatrix} Z^{(k,k)} & Z^{(k,l)} \\ Z^{(l,k)} & Z^{(l,l)} \end{pmatrix} c\right)$$

Тому система векторів $\left(\sqrt{n}\left(\hat{\theta}_n^{(k)} - \theta^{(k)}\right), k = 1, ..., M\right)$ збігається слабко до системи $\left(\left(M^{(k)}\right)^{-1} u^{(k)}, k = 1, ..., M\right)$. Це і доводить теорему. □

Введемо позначення для $i, j, l = 1, ..., d$,

$$\dot{g}_i(x,\gamma) = \frac{\partial g(x,\gamma)}{\partial \gamma_i};\ \ddot{g}_{ij}(x,\gamma) = \frac{\partial^2 g(x,\gamma)}{\partial \gamma_i \gamma_j};\ \dddot{g}_{ijl}(x,\gamma) = \frac{\partial^3 g(x,\gamma)}{\partial \gamma_i \gamma_j \gamma_l}.$$

Теорема 3.5. *Нехай $s(z,\gamma) = (y - g(x,\gamma))\,\dot{g}(x,\gamma)$ і виконуються наступні умови:*

1 *θ є внутрішньою точкою $\Theta \times \ldots \times \Theta$.*

2 *У деякому відкритому околі B точки θ існують і є неперервними мішані часткові похідні*

$$\frac{\partial^3 g(x,\gamma)}{\partial \gamma^i, \partial \gamma^j, \partial \gamma^l} \text{ для всіх } l, i, j = 1, ..., d$$

і майже всіх (відносно мір $F_m, m = 1, ..., M$).

3 *Існує така функція $h(x)$, що*

$$\sup_{\gamma\in B}\left|g(x,\gamma)\right| \leqslant h(x),\ \max_i \sup_{\gamma\in B}\left|\dot{g}_i(x,\gamma)\right| \leqslant h(x),$$

$$\max_{i,j} \sup_{\gamma\in B}\left|\ddot{g}_{ij}(x,\gamma)\right| \leqslant h(x),\ \max_{i,j,l} \sup_{\gamma\in B}\left|\dddot{g}_{ijl}(x,\gamma)\right| \leqslant h(x),$$

і $E\Big(h(\xi_{(l)})\Big)^{2\alpha} < \infty$ для деякого $\alpha > 1$ і усіх $l = 1, .., M$.

4 $\mathbb{E}\,\varepsilon_{(m)}^{\max(3,\alpha)} < \infty$, *та* $\mathbb{E}\,|g(X_{(m)}, \theta^{(m)})|^{\max(3,\alpha)} < \infty$ *для всіх* $m = 1, ..., M$.

5 Матриці $M^{(m)}$ *є скінченними і для будь-якого* $c \in \mathbb{R}^d, c \neq 0$

$$\mathbb{P}\{\langle c, \dot{g}(\xi_{(m)}, \theta^{(m)})\rangle = 0\} < 1 \text{ для } m = 1, ..., M.$$

6 Границі $\langle a^k a^m p^l p^i \rangle$ *існують для всіх* $k, m, l, i = 1, ..., M$.

7 Матриця Γ_∞ *існує і є не виродженою.*

8 $\hat{\theta}_n$ *є консистентною оцінкою* θ.

Тоді при $n \to \infty$

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{w} N(0, V).$$

Доведення. Для доведення достатньо показати виконання умов 3) і 4) теореми 3.4.

Покажемо, що умова 3) теореми 3.4 виконується. Дійсно,

$$\max_{l,i,j} \sup_{\gamma \in B} \left| \frac{\partial s^l(z, \gamma)}{\partial \gamma^i, \partial \gamma^j} \right| =$$

$$\max_{l,i,j} \sup_{\gamma \in B} \left| y\ddot{g}_{ijl}(x, \gamma) - \ddot{g}_{ij}(x, \gamma)\dot{g}_l(x, \gamma) - \dot{g}_i(x, \gamma)\ddot{g}_{jl}(x, \gamma) - \dot{g}_j(x, \gamma)\ddot{g}_{il}(x, \gamma) - g(x, \gamma)\ddot{g}_{ijl}(x, \gamma) \right| \leq$$

$$|yh(x)| + 4h^2(x).$$

Внаслідок умов 4) і 3) теореми і нерівності Мінковського, для всіх $l = 1, ..., M$,

$$\mathbb{E}\left(|(g(X_{(l)}, \theta) + \varepsilon_{(l)})h(\xi_{(l)})| + 4h(\xi_{(l)})^2 \right)^\alpha < \infty.$$

З умови 4) теореми, для всіх $m = 1, ..., M$ випливатиме умова 4) теореми 3.4,

$$\mathbb{E}\,|s(\xi_{(m)}, \theta^{(m)})|^3 = \mathbb{E}\,|(Y_{(m)} - g(\xi_{(m)}, \theta^{(m)}))\dot{g}(\xi_{(m)}, \theta^{(m)})|^3 = \mathbb{E}\,|\varepsilon_{(k)}|^3\, \mathbb{E}\,|g(\xi_{(m)}, \theta^{(m)})|^3 < \infty.$$

Для зручного застосування матриці $M^{(k)}$, випишемо її з урахуванням вигляду функції $s(z, \gamma)$.

$$M^{(k)} = \mathbb{E}\,\dot{S}(\xi_{(k)}, \theta^{(k)}) = \mathbb{E} \sum_{j=1}^{n} a_{j;n}^{(k)} \dot{s}(\xi_i, \theta^{(k)}) = \sum_{m=1}^{M} I_{[k=m]}\, \mathbb{E}\,\dot{s}(\xi_{(k)}, \theta^{(k)}) =$$

$$\mathbb{E}\, \frac{\partial}{\partial \gamma}\Big(Y_{(k)} - g(X_{(k)}, \gamma)\Big)\dot{g}(X_{(k)}, \gamma)\Big|_{\gamma = \theta^{(k)}} = -\,\mathbb{E}\,\dot{g}(X_{(k)}, \theta^{(k)})\dot{g}^T(X_{(k)}, \theta^{(k)}).$$

З умови 5), видно, що для довільного $c \in \mathbb{R}^d$,

$$-c^T M^{(k)} c = \mathbb{E}\,|\dot{g}(X_{(k)}, \theta^{(k)})|^2$$

□

Приклад 3.6. Нехай для функцій s і g мають вигляд (3.2) і (3.3) відповідно, виконуються умови 1), 6)-8) теореми 3.5, і наступні умови:

1 $E|X_{(m)}|^{6\alpha} < \infty$ для деякого $\alpha > 1$ і усіх $m = 1, .., M$.

2 $\mathbb{E}\, \varepsilon_{(m)}^{\max(3,\alpha)} < \infty$, та $\mathbb{E}\, |g(X_{(m)}, \theta^{(m)})|^{\max(3,\alpha)} < \infty$ для всіх $m = 1, ..., M$.

3 Для довільного $\gamma \in \mathbb{R}^d$, $\gamma \neq 0$, $\mathbb{P}\{\langle X_{(m)}, \gamma, \rangle = 0\} < 1$, для усіх $m = 1, ..., M$.

Тоді при $n \to \infty$

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{w} N(0, V).$$

Доведення. Для того, щоб довести твердження прикладу, використаємо теорему 3.5. Для цього, потрібно перевірити її умови 3) і 5).

Доведення почнемо із умови 5). Як показано у теоремі 3.5,

$$M^{(k)} = -\,\mathbb{E}\, \dot{g}(X_{(m)}, \gamma)\dot{g}(X_{(m)}, \gamma)^T,$$

а отже, $-M^{(m)}$ є невід'ємно визначеною. Оскільки, за умови 3) прикладу, для довільного $\gamma \in \mathbb{R}^d$ такого, що $\gamma \neq 0$,

$$-\gamma M^{(m)} \gamma = \mathbb{E}\, |g(X_{(m)}, \gamma)(1 - g(X_{(m)}, \gamma))\langle X_{(m)}, \gamma\rangle|^2 > 0 \text{ м.н.}$$

Тепер покажемо, що виконується умова 3) теореми 3.5. Випишемо усі обмеження функцій із умови 3). Для усіх $x \in \mathbb{R}^D$

$$\max_{\gamma \in \Theta} |g(x, \gamma)| \leqslant 1; \ \max_{\gamma \in \Theta} |\dot{g}(x, \gamma)| \leqslant |x|,$$

$$\max_{\gamma \in \Theta} |\ddot{g}(x, \gamma)| \leqslant |x|^2; \ \max_{\gamma \in \Theta, j; t = 1, \ldots, d} \left| \frac{\partial \ddot{g}_{j,t}(x, \gamma)}{\partial \gamma} \right| \leqslant 4|X|^3.$$

Отже, у нашому випадку, потрібно показати, що для $h(x) = 4\sum_{i=0}^{3} |x|^i$ і $\alpha > 1$,

$$\mathbb{E}\left(4 \sum_{i=0}^{3} |X_{(m)}|^i\right)^{2\alpha} < \infty.$$

Це дійсно так, оскільки старший степінь цього полінома буде обмежений за умовою 1) прикладу. □

3.5 Оцінка методу складаного ножа для асимптотичної коваріації УОР-оцінок

Для $k, m = 1, ...M$ позначимо

$$\hat{V}_n^{(k,m)} = n \sum_{i=1}^{n} \left(\hat{\theta}_{i-;n}^{(k)} - \hat{\theta}_n^{(k)}\right) \left(\hat{\theta}_{i-;n}^{(m)} - \hat{\theta}_n^{(m)}\right)^T.$$

Будемо називати $\hat{V}_n^{(k,m)}$ оцінкою методом складаного ножа для $V^{(k,m)}$.

Теорема 3.6. *Нехай виконуються наступні умови:*

1 Θ *- компакт,* θ *- внутрішня точка* Θ.

2 Існує така функція $h : \mathbb{R}^D \to \mathbb{R}$, *що*

$$\sup_{\gamma\in\Theta} |s(x,\gamma)| \leqslant h(x), \ \sup_{\gamma\in\Theta} |\dot{s}(x,\gamma)| \leqslant h(x),$$

$$\max_{i,j,l=1,\dots,d} \sup_{\gamma\in\Theta} \left| \frac{\partial s^l(x,\gamma)}{\partial\gamma^i \partial\gamma^j} \right| \leqslant h(x) \text{ і, для деякого } \alpha \geqslant 4,$$

$$\mathbb{E}\left(h(\xi_{(l)})\right)^\alpha < \infty \text{ для всіх } l = 1, \dots, M.$$

3 Матриця Γ_∞ *існує і* $\det \Gamma_\infty \neq 0$.

4 Існують границі $\langle a^k a^m p^l p^i \rangle$ *для всіх* $i, l = 1, \dots, M$..

5 Матриці $M^{(k)} = \mathbb{E}\,\dot{s}(\xi_{(k)}, \theta^{(k)})$, $M^{(m)} = \mathbb{E}\,\dot{s}(\xi_{(m)}, \theta^{(m)})$ *є невиродженими.*

6 $\hat{\theta}_n^{(k)}, \hat{\theta}_n^{(m)}$ *є* $\sqrt{n}$ *- консистентними оцінками* $\theta^{(k)}, \theta^{(m)}$.

7 $\sup_{i=1,\dots,m} |\hat{\theta}_{i-;n}^{(l)} - \theta^{(l)}| \to 0$, *для* $l = k, m$ *за ймовірністю при* $n \to \infty$.

Тоді $\hat{V}_n^{(k,m)} \to V^{(k,m)}$ *за ймовірністю.*

Зауваження. Для перевірки виконання умови 7, можна скористатися Теоремою 3.1, для умови 6), Теоремою 3.4.

Доведення. Не обмежуючи загальності, будемо вважати, що $k = 1, m = 2$. (Доведення у випадку $k = m$ не відрізняється принципово від доведення у випадку, коли k і m різні).

Позначимо $s_j^{(l)}(\xi_j, \gamma) = s(\xi_j, \gamma) - \mathbb{E}\,s(\xi_j, \theta^{(l)})$, $(l = 1, 2)$.

Тоді

$$\sum_{j=1}^n a_{j;n}^l s_j^{(l)}(\xi_j, \gamma) = \sum_{j=1}^n a_{j;n}^l s(\xi_j, \gamma) - \sum_{m=1}^M \sum_{j=1}^n a_{j;n}^l p_{j;n}^m \,\mathbb{E}\,s(\xi_{(m)}, \theta^{(l)}) =$$

$$= \sum_{j=1}^n a_{j;n}^l s(\xi_j, \gamma) = S_n^{(l)}(\gamma).$$

Отже,

$$S_n^{(l)}(\hat{\theta}_n^{(l)}) = \sum_{j=1}^n a_{j;n}^l s_j^{(l)}(\xi_j, \hat{\theta}_n^{(l)}) = 0 \tag{3.2}$$

за означенням $\hat{\theta}_n^{(l)}$. Аналогічно

$$S_{ni-}^{(l)}(\hat{\theta}_{i-;n}^{(l)}) = \sum_{j\neq i} a_{ji-;n}^{l} s_j^{(l)}(\xi_j, \hat{\theta}_{i-;n}^{(l)}) = 0. \tag{3.3}$$

Тому за теоремою Лагранжа, існує таке $t_{ni}^l \in [0,1]$, що для $\zeta_{ni}^l = (1-t_{ni}^l)\hat{\theta}_{i-;n}^{(l)} + t_{ni}^l \hat{\theta}_n^{(l)}$ виконано

$$-S_{ni-}^{(l)}(\hat{\theta}_n^{(l)}) = S_{ni-}^{(l)}(\hat{\theta}_{i-;n}^{(l)}) - S_{ni-}^{(l)}(\hat{\theta}_n^{(l)}) = \dot{S}_{ni-}^{(l)}(\zeta_{ni}^l)(\hat{\theta}_{i-;n}^{(l)} - \hat{\theta}_n^{(l)}). \tag{3.4}$$

Помітимо, що внаслідок умови 2),

$$|\dot{s}_j^{(l)}(x,\gamma_1) - \dot{s}_j^{(l)}(x,\gamma_2)| \leqslant h(x)|\gamma_1 - \gamma_2|.$$

Отже, застосовуючи Лему 3.6 з $\psi_j(x,\gamma) = \dot{s}_j^{(l)}(x,\gamma)$, $\rho(\gamma_1,\gamma_2) = |\gamma_1 - \gamma_2|$, отримуємо, що

$$\sup_{\gamma\in\Theta} |\dot{S}_{ni-}^{(l)}(\gamma) - M^{(l)}(\gamma)| \to 0 \text{ за ймовірністю} \tag{3.5}$$

(нагадаємо, що $M^{(l)}(\gamma) = \mathbb{E}\,\dot{s}(\xi_{(l)},\gamma)$, оскільки $\frac{\partial s(\xi_j,\theta^{(k)})}{\partial\gamma} = 0$).

Легко бачити, що внаслідок умови 2), $M^{(l)}(\gamma) \to M^{(l)}(\theta^{(l)})$ при $\gamma \to \theta^{(l)}$. За умовами 6) і 7), $\max_{i=1,\dots,n} |\hat{\theta}_{i-;n}^{(l)} - \hat{\theta}_n^{(l)}| \to 0$, при $n\to\infty$ за ймовірністю, отже

$$\max_{i=1,\dots,n} |\dot{S}_{ni-}^{(l)}(\zeta_{ni}^l) - M^{(l)}(\theta^{(l)})| \xrightarrow{P} 0,\ n\to\infty.$$

Враховуючи, що $M^{(l)}(\theta^{(l)}) = M^{(l)}$ - не вироджена матриця, звідки отримаємо, що

$$\mathbb{P}\{\det \dot{S}_{ni}^{(l)}(\zeta_{ni}^l) \neq 0 \text{ для всіх } i = 1,\dots,n\} \to 1 \text{ при } n\to\infty,$$

і

$$\Lambda_n^l = \max_{i=1,\dots,n} |(\dot{S}_{ni}^{(l)}(\zeta_{ni}^l))^{-1}| = O_p(1). \tag{3.6}$$

Отже, з ймовірністю, що прямує до одиниці,

$$|\hat{\theta}_{i-;n}^{(l)} - \hat{\theta}_n^{(l)}| = |(\dot{S}_{ni-}^{(l)}(\zeta_{ni}^l))^{-1}(S_{ni-}^{(l)}(\hat{\theta}_n)| \leqslant \Lambda_n^l |S_{ni-}^{(l)}(\hat{\theta}_n^{(l)})| =$$

$$\Lambda_n^l \left| S_{ni-}^{(l)}(\hat{\theta}_n^{(l)}) + \dot{S}_{ni}(\tilde{\zeta}_{ni}^{(l)})(\hat{\theta}_{i-;n}^{(l)} - \hat{\theta}_n^{(l)}) \right|, \tag{3.7}$$

де $\tilde{\zeta}_{ni}^{(l)} = (1-\tilde{t}_{ni}^{l})\hat{\theta}_{n}^{(l)} + \tilde{t}_{ni-}^{l}\theta^{(l)}$, $\tilde{t}_{ni}^{l} \in [0,1]$ - деяка проміжна точка. Оскільки, за умовою 6), $\hat{\theta}_{n}^{(l)} - \theta^{(l)} = o_P(\frac{1}{\sqrt{n}})$, а за 3.5, $\max_{i=1,\ldots,n} \sup_{\gamma\in\Theta} |\dot{S}_{ni-}^{(l)}(\gamma)| = O_P(1)$, отримуємо

$$\dot{S}_{ni-}^{(l)}(\tilde{\zeta}_{ni}^{l})(\hat{\theta}_{i-;n}^{(l)} - \hat{\theta}_{n}^{(l)}) = O_P(n^{-0.5}) \tag{3.8}$$

Далі

$$S_{ni-}^{(l)}(\theta^{(l)}) = S_{n}^{(l)}(\theta^{(l)}) - a_{i;n}^{l} s_{i}^{(l)}(\xi_i, \theta^{(l)}) - \sum_{j\neq i}(a_{j;n}^{l} - a_{ji;n}^{l}) s_{j}^{(l)}(\xi_j, \theta^{(l)}).$$

$\mathbb{E}\, S_{n}^{(l)}(\theta^{(l)}) = 0$, а $\mathbb{D}\, S_{n}^{(l)}(\theta^{(l)}) = \sum_{j=1}^{n}(a_{j;n}^{l})^2\, \mathbb{D}\, s_{j}^{(l)}(\xi_j, \theta^{(l)}) = o(\frac{1}{n})$, тому $S_{n}^{(l)}(\theta^{(l)}) = O_P(\frac{1}{\sqrt{n}})$. За лемами 3.1 і 2.1, з урахуванням умови 2),

$$\sup_{i=1,\ldots,n} |a_{i;n}^{l} s_{i}^{(l)}(\xi_i, \theta^{(l)}) + \sum_{j\neq i}(a_{j;n}^{l} - a_{ji;n}^{l}) s_{j}^{(l)}(\xi_j, \theta^{(l)})| = O_P(n^{-0.5}). \tag{3.9}$$

Отже, отримуємо,

$$\sup_{i=1,\ldots n} |S_{ni-}^{(l)}(\theta^{(l)})| = O_P(n^{-0.5}).$$

Звідси, та з (3.6)-(3.8), отримуємо

$$\sup_{i=1,\ldots n} |\hat{\theta}_{i-;n}^{(l)} - \hat{\theta}_{n}^{(l)}| = O_P(n^{-0.5}). \tag{3.10}$$

Позначимо $M_{ni}^{(l)} = \dot{S}_{ni}^{(l)}(\zeta_{ni}^{(l)})$. За (3.2) і (3.4), отримуємо

$$\hat{\theta}_{i-;n}^{(l)} - \hat{\theta}_{n}^{(l)} = (M_{ni}^{(l)})^{-1}(S_{n}^{(l)}(\hat{\theta}_{n}^{(l)}) - S_{ni-}^{(l)}(\hat{\theta}_{n}^{(l)})) =$$

$$(M_{ni}^{(l)})^{-1}(a_{i;n}^{l} s_{i}^{(l)}(\xi_i, \hat{\theta}_{n}^{(l)}) + \sum_{j\neq i}(a_{j;n}^{l} - a_{ji-;n}^{l}) s_{j}^{(l)}(\xi_j, \hat{\theta}_{n}^{(l)})) \tag{3.11}$$

позначимо

$$U_{i}^{(l)} = a_{i;n}^{l} s_{i}^{(l)}(\xi_i \theta^{(l)}) + \sum_{j\neq i}(a_{j;n}^{l} - a_{ji-;n}^{l}) s_{j}^{(l)}(\xi_j, \theta^{(l)}), \tag{3.12}$$

$$\Delta_{i}^{(l)V} = a_{i;n}^{l}(s_{i}^{(l)}(\xi_i, \hat{\theta}_{n}^{(l)}) - s_{i}^{(l)}(\xi_i, \theta^{(l)})) + \sum_{j\neq i}(a_{j;n}^{l} - a_{ji-;n}^{l})(s_{j}^{(l)}(\xi_j, \hat{\theta}_{n}^{(l)}) - s_{j}^{(l)}(\xi_j, \theta^{(l)})), \tag{3.13}$$

$$\Delta_{i}^{(l)M} = (M_{ni}^{(l)})^{-1} - (M^{(l)})^{-1}. \tag{3.14}$$

Тоді за (3.11),

$$\hat{\theta}^{(l)}_{i-;n} - \hat{\theta}^{(l)}_n = \Big((M^{(l)})^{-1} + \Delta_i^{(l)M}\Big)\Big(U_i^{(l)} + \Delta_i^{(l)}V\Big).$$

Таким чином,

$$\hat{V}_n^{(1,2)} = n\sum_{i=1}^n \big(\hat{\theta}^{(1)}_{i-;n} - \hat{\theta}^{(1)}_n\big)\big(\hat{\theta}^{(2)}_{i-;n} - \hat{\theta}^{(2)}_n\big)^T = n\sum_{i=1}^n v_i, \tag{3.15}$$

де

$$v_i = \Big((M^{(1)})^{-1} + \Delta_i^{(1)M}\Big)\Big(U_i^{(1)} - \Delta_i^{(1)}V\Big)\Big(U_i^{(2)} - \Delta_i^{(2)}V\Big)^T\Big((M^{(2)})^{-1} + \Delta_i^{(2)M}\Big)^T \tag{3.16}$$

Розглянемо

$$\tilde{V}_n^{(1,2)} = n\sum_{i=1}^n \tilde{v}_i, \text{ де } \tilde{v}_i = (M^{(1)})^{-1}U_i^{(1)}{U_i^{(2)}}^2(M^{(2)})^{-T}. \tag{3.17}$$

Ми покажемо, що

$$\left|\hat{V}_n^{(1,2)} - \tilde{V}_n^{(1,2)}\right| = o_p(1) \tag{3.18}$$

$$\text{і } \tilde{V}_n^{(1,2)} \to V^{(1,2)} \text{ за ймовірністю при } n\to\infty. \tag{3.19}$$

Звідси випливає твердження теореми. Ідея доведення (3.19) належить О.В. Сугаковій.

Для доведення (3.18) скористаємося розкладом

$$v_i - \tilde{v}_i = v_i^1 + v_i^2 + v_i^3 + v_i^4, \text{ де}$$

$$v_i^1 = \Delta_i^{(1)M}\Big(U_i^{(1)} + \Delta_i^{(1)V}\Big)\Big(U_i^{(2)} + \Delta_i^{(2)V}\Big)^T\Big((M^{(2)})^{-1} + \Delta_i^{(2)M}\Big)^T,$$

$$v_i^2 = (M^{(1)})^{-1}\Delta_i^{(1)V}\Big(U_i^{(2)} + \Delta_i^{(2)V}\Big)^T\Big((M^{(2)})^{-1} + \Delta_i^{(2)M}\Big)^T, \tag{3.20}$$

$$v_i^3 = (M^{(1)})^{-1}U_i^{(1)}(\Delta_i^{(2)V})^T((M^{(2)})^{-1} + \Delta_i^{(2)M})^T,$$

$$v_i^4 = M^{(1)})^{-1}U_i^{(1)}(U_i^{(2)})^T(\Delta_i^{(2)M})^T.$$

Оцінимо окремо кожне $v_i^l, l = 1, ..., 4$.

Почнемо з оцінки $\Delta^{(1)M}$.

Застосовуючи лему 3.6 до функцій $\psi_{j;n}(x,\gamma) = \frac{\partial}{\partial\gamma^k}\dot{s}^{(l)}(x,\gamma)$ при $l = 1, 2$; $k = 1, ..., d$ з урахуванням умов 1), 2) теореми, отримуємо, що для

$$\frac{\partial}{\partial\gamma^k}\dot{S}^{(l)}_{ni-}(\gamma) = \sum_{j\neq i}\frac{\partial}{\partial\gamma^k}s^{(l)}_j(\xi_j,\gamma) = \sum_{j\neq i}\psi_{j;n}(\xi_j,\gamma),$$

виконано

$$\max_{i=1,\dots,n}\sup_{\gamma\in\Theta}\left|\frac{\partial}{\partial\gamma^k}\dot{S}^{(l)}_{ni-}(\gamma) - \mathbb{E}\,\frac{\partial}{\partial\gamma^k}\dot{s}(\xi_{(k)},\gamma)\right| = o_P(1),$$

а отже і

$$\max_{i=1,\dots,n}\sup_{\gamma\in\Theta}\left|\frac{\partial}{\partial\gamma^k}\dot{S}^{(l)}_{ni-}(\gamma)\right| = O_P(1). \tag{3.21}$$

Враховуючи, що

$$\dot{S}^{(l)}_{ni-}(\gamma) = \dot{S}^{(l)}_{ni-}(\theta^{(l)}) + \sum_{k=1}^{d}\frac{\partial}{\partial\gamma^k}\dot{S}^{(l)}_{ni-}(\tilde{\zeta}^l_{ni})(\gamma^k - \theta^{k(l)}),$$

де $\tilde{\zeta}^k_{ni}$ - деяка проміжна точка між γ і $\theta^{k(l)}$, отримуємо з (3.21) і умови 6) теореми, що

$$\sup_{i=1,\dots,n}\left|\dot{S}^{(l)}_{ni-}(\tilde{\zeta}^l_{ni}) - \dot{S}^{(l)}_{ni}(\theta)\right| = O_P(1)O_P(\frac{1}{\sqrt{n}}) = O_P(\frac{1}{\sqrt{n}}). \tag{3.22}$$

Далі,

$$\dot{S}^{(l)}_{ni-}(\theta) = \dot{S}^{(l)}_{n}(\theta) - a^l_{i;n}\dot{s}^{(l)}_i(\xi_j,\theta) + \sum_{j\neq i}(a^l_{j;n} - a^l_{ji-;n})\dot{S}^{(l)}_j(\xi_j,\theta).$$

Так само, як для оцінки $S^{(l)}_{ni-}$ у формулі (3.9), використовуючи леми 3.1 і 2.1, отримуємо, що для $\beta \geqslant 1/\alpha$ (далі саме таке β усюди)

$$\max_{i=1,\dots,n}|\dot{S}^{(l)}_{ni-}(\theta) - \dot{S}^{(l)}_n(\theta)| = O_P(n^{\beta-1}). \tag{3.23}$$

Оцінюючи дисперсії кожного елемента матриці $\dot{S}^{(l)}_n(\theta)$, як $O(n^{-1})$,

$$\text{Отримуємо, що } |\dot{S}^{(l)}_n(\theta) - M^{(l)}| = O_P\left(\frac{1}{\sqrt{n}}\right). \tag{3.24}$$

Згадавши, що $M^{(l)}_{ni} = \dot{S}^{(l)}_{ni-}(\zeta^{(l)}_{ni})$, з (3.22)-(3.24) отримуємо

$$\max_{i=1,\dots,n}|M^{(l)}_{ni} - M^{(l)}| = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Оскільки $\det M^{(l)} \neq 0$ за умовою 5) теореми, то звідси випливає, що

$$\max_{i=1,\dots,n} |\Delta_i^{(l)M}| = O_P(n^{-0.5}). \tag{3.25}$$

Внаслідок лем 3.1 і 2.1 і умови 2) теореми,

$$\max_{i=1,\dots,n} |U_i^{(l)}| = O_P(n^{\beta-1}). \tag{3.26}$$

Оцінимо

$$\max_{i=1,\dots,n} |\Delta_i^{(l)V}| \leqslant \max_{i=1,\dots,n} (|a_{i;n}^l| \dot{s}_i^{(l)}(\tilde{\zeta}_i^{(l)}) |\hat{\theta}_n^{(l)} - \theta^{(l)}|) +$$

$$\sum_{j \neq i} |a_{j;n}^l - a_{ji-;n}^l| \dot{s}_i^{(l)}(\tilde{\zeta}_i^{(l)}) |\hat{\theta}_n^{(l)} - \theta^{(l)}| \tag{3.27}$$

(тут $\tilde{\zeta}_i^{(l)}$, як за звичай, проміжні точки між $\hat{\theta}_n^{(l)}$ і $\theta^{(l)}$). Знову застосовуючи лему 2.1, отримуємо

$$\sup_{\gamma \in \Theta} |s_i^{(l)}(\gamma)| = O_P(n^{\beta}),$$

за умовою 6) теореми, $|\hat{\theta}_n^{(l)} - \theta^{(l)}| = o_P(\sqrt{n})$, отже з урахуванням леми 3.1, отримуємо з (3.27)

$$\max_{i=1,\dots,n} |\Delta_i^{(l)V}| \leqslant O_P(n^{\beta-3/2}). \tag{3.28}$$

Тепер оцінимо доданки v_i^m, визначені (3.20).

Враховуючи (3.25), (3.26) і (3.28), маємо

$$\max_{i=1,\dots,n} |v_i^1| \leqslant \max_{i=1,\dots,n} |\Delta_i^{(1)M}| \Big(|U_i^{(1)}| + |\Delta_i^{(1)V}| \Big) \Big(|U_i^{(2)}| + |\Delta_i^{(2)V}| \Big) \Big(|(M^{(2)})^{-1}| + |\Delta_i^{(2)M}| \Big) =$$

$$= O_P(n^{-0.5})(O_P(n^{\beta-1}))^2 O_P(1) = O_P(n^{2\beta-5/2}).$$

Аналогічно,

$$\max_{i=1,\dots,n} |v_i^4| = O_P(n^{2\beta-5/2}).$$

Для v_i^2 (і аналогічно, для v_i^3) отримуємо

$$\max_{i=1,\dots,n} |v_i^2| \leqslant \max_{i=1,\dots,n} |(M^{(1)})^{-1} |\Delta_i^{(1)}| (|U_i^{(2)}| + |\Delta_i^{(2)V}|)(|(M^{(2)})^{-1}| + |\Delta_i^{(2)M}|) =$$

$$= O_P(n^{\beta-3/2}) O_P(n^{\beta-1}) O_P(1) = O_P(n^{n^{2\beta-5/2}}).$$

Отже

$$|\hat{V}_n^{(1,2)} - \tilde{V}_n^{(1,2)}| \leqslant n \sum_{i=1}^{n} |v_i - \tilde{v}_i| \leqslant n^2 \max_{i=1,\dots,n} \sum_{k=1}^{4} |v_i^k| = O_P(n^{2\beta-1/2}) = o_P(1),$$

при $\beta \leqslant 1/4$. (нагадаємо, що можна взяти будь-яке $\beta \geqslant 1/\alpha$), а $\alpha \geqslant 4$.

Отже, (3.18) доведено. Для доведення теоерми залишилось перевірити (3.19).

Розглянемо $\hat{z}_n^{(1,2)} = n \sum_{i=1}^{n} U_i^{(1)} \big(U_i^{(2)}\big)^T$.

$$\mathbb{E}\, \hat{z}_n^{(1,2)} = \mathbb{E}\, n \sum_{i=1}^{n} U_i^{(1)} \big(U_i^{(2)}\big)^T = \bar{z}_{1,n} + \bar{z}_{2,n},$$

де

$$\bar{z}_{1,n} = n \sum_{i=1}^{n} a_{i;n}^1 a_{i;n}^2\, \mathbb{E}\, s_i^{(1)}(\xi_i, \theta^{(1)}) s_i^{(2)}(\xi_i, \theta^{(2)}),$$

$$\bar{z}_{2,n} = n \sum_{i=1}^{n} \sum_{j\neq i} (a_{j;n}^1 - a_{ji;n}^1)(a_{j;n}^2 - a_{ji;n}^2)\, \mathbb{E}\, s_i^{(1)}(\xi_i, \theta^{(1)}) s_i^{(2)}(\xi_i, \theta^{(2)}).$$

За лемою 3.1, отримуємо

$$\max_{i=1,\dots,n} \Big| \sum_{j\neq i} (a_{j;n}^1 - a_{ji;n}^1)(a_{j;n}^2 - a_{ji;n}^2)\, \mathbb{E}\, s_i^{(1)}(\xi_i, \theta^{(1)}) s_i^{(2)}(\xi_i, \theta^{(2)}) \Big| = O(n^{-3}),$$

тому $\bar{z}_{2,n} = n^2 O(n^{-3}) = O(n^{-1})$. За означенням $z^{(1,2)}$, $\bar{z}_{1,n} \to z^{(1,2)}$, при $n \to \infty$. Отже,

$$E\hat{z}_n^{(1,2)} \to z^{(1,2)}. \tag{3.29}$$

Оцінимо

$$\mathbb{E}\, |\hat{z}_n^{(1,2)} - \mathbb{E}\, \hat{z}_n^{(1,2)}| = \sum_{l,k=1}^{d} \mathbb{V}\mathrm{ar}(\hat{z}_n^{lk(1,2)}) \tag{3.30}$$

(тут $\hat{z}_n^{lk(1,2)}$ - (lk)-тий елемент матриці $\hat{z}_n^{(1,2)}$). Оцінимо окремо lk-тий доданок у цій сумі ($U_i^{l(1)}$-l-тий елемент вектора $U_i^{(1)}$):

$$\mathbb{V}\mathrm{ar}(\hat{z}_n^{lk(1,2)}) = n^2 \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(U_i^{l(1)} U_i^{k(2)}) \leqslant n^3 \max_{i,l,k} \mathbb{E}(U_i^{l(1)} U_i^{k(2)})^2 \leqslant$$

$$n^3 \sqrt{\max_{i,l,k} \mathbb{E}(U_i^{l(1)})^4\, \mathbb{E}(U_i^{k(2)})^4} \leqslant n^3 \max_{i,j,k} \mathbb{E}(U_i^{k(l)})^4 \tag{3.31}$$

Для спрощення позначимо $\eta_{i;n} = s_i^{k(l)}(\xi_{i;n}, \theta^{(l)})$. За означенням s_i, $\mathbb{E}\,\eta_{i;n} = 0$. Отже

$$\mathbb{E}(U_i^{k(l)})^4 = \mathbb{E}(a_{i;n}^{(l)}\eta_{i;n} + \sum_{j\neq i}(a_{i;n}^l - a_{ji-;n}^l)\eta_{j;n})^4 =$$

$$= (a_{i;n}^l)^4\,\mathbb{E}(\eta_{i;n})^4 + 6(a_{i;n}^l)^2\,\mathbb{E}(\eta_{i;n})^2\,\mathbb{E}(\sum_{j\neq i}(a_{i;n}^l - a_{ji-;n}^l)\eta_{j;n})^2 + \mathbb{E}(\sum_{j\neq i}(a_{i;n}^l - a_{ji-;n}^l)\eta_{j;n})^4. \tag{3.32}$$

З умови 2 теореми випливає, що $\sup_{j=1,\dots,n;n=1,\dots}\mathbb{E}(\eta_{j;n})^4 \leqslant C_\eta < \infty$.

За лемою 3.1 отримуємо, що перший доданок у правій частині (3.32) є $O(n^{-4})$. Другий доданок не перевищує

$$\frac{C}{n^2}\sum_{j\neq i}(a_{i;n}^l - a_{ji-n}^l)^2\,\mathbb{E}(\eta_{j;n})^2 \leqslant \frac{C}{n^2}\sum_{j=1}^{n}\frac{1}{n^4} = O(n^{-5}).$$

Позначимо $a_{i;n}^l - a_{ji-;n}^l = b_{ij} = O(n^{-2})$.

Для третього доданка маємо

$$\mathbb{E}\left(\sum_{j\neq i}(b_{ij}\eta_j)\right)^4 \leqslant \sum_{j\neq i}(b_{ij})^4\,\mathbb{E}(\eta_{j;n})^4 + 6\sum_{j_1,j_2\neq i}(b_{ij_1})^2(b_{ij_2})^2\,\mathbb{E}(\eta_{j_1,n})^2(\eta_{j_2,n})^2$$

$$= O(n^{-7}) + n^2 O(n^{-8}) = O(n^{-6}).$$

Враховуючи ці оцінки, отримуємо з (3.32)

$$\sup_{i=1,\dots,n;n=1,2,\dots}\mathbb{E}(U_i^{k(l)})^2 = O(n^{-4}).$$

Тому за (3.31), (3.30) отримуємо

$$\mathbb{E}\,|\hat{z}_n^{(1,2)} - \mathbb{E}\,\hat{z}_n^{(1,2)}| = O(n^{-4})n^3 = O(n^{-1}),$$

що разом за (3.29) дає (3.19). Теорема доведена. □

Теорема 3.7. *Нехай $s(z,\gamma) = (y - g(x,\gamma))\dot{g}(x,\gamma)$ і виконуються наступні умови:*

1 Θ - компакт, θ - внутрішня точка Θ.

2 Нехай існують такі функції $h_i : \mathbb{R}^D \to \mathbb{R}$, $i = \overline{0,3}$, що для всіх $x \in \mathbb{R}^D$

$$\max_{\gamma\in\Theta}|g(x,\gamma)| \leqslant h_0(x),\ \max_{\gamma\in\Theta}|\dot{g}(x,\gamma)| \leqslant h_1(x),$$

$$\max_{\gamma\in\Theta}|\ddot{g}(x,\gamma)|\leqslant h_2(x),\quad \max_{\gamma\in\Theta,j;t=1,\dots,d}\left|\frac{\partial\ddot{g}_{j,t}(x,\gamma)}{\partial\gamma}\right|\leqslant h_3(x),$$

і для деякого $\alpha\geqslant 4$ *і* $l=k,m$

$$\mathbb{E}\left(\left(|Y_{(l)}|+h_0(X_{(l)})\right)\Big(\sum_{i=1}^{3}h_i(X_{(l)})\Big)+3h_1(X_{(l)})h_2(X_{(l)})+h_1^2(X_{(l)})\right)^{\alpha}<\infty$$

3 Матриця Γ_∞ *існує і* $\det\Gamma_\infty\neq 0$.

4 Існують границі $\langle a^k a^m p^l p^i\rangle$ *для всіх* $i,l=1,...,M$..

5 Для будь-якого $c\in\mathbb{R}^d, c\neq 0$ $\mathbb{P}\{\langle c,\dot{g}(\xi_{(l)},\theta^{(l)})\rangle=0\}<1$ *для* $l=k,m$.

6 $\hat{\theta}_n^{(k)},\hat{\theta}_n^{(m)}$ *є* $\sqrt{n}$ *- консистентними оцінками* $\theta^{(k)},\theta^{(m)}$.

7 $\sup_{i=1,\dots,m}|\hat{\theta}_{i-;n}^{(l)}-\theta^{(k)}|\to 0$, *для* $l=k,m$ *за ймовірністю при* $n\to\infty$.

Тоді $\hat{V}_n^{(k,m)}\to V^{(k,m)}$ *за ймовірністю.*

Доведення. Для доведення теореми, скористаємось теоремою 3.6 і достатньо лише перевірити її умови 2) і 5).

Почнемо з умови 5). Покажемо, що матриці $-\mathbb{E}\,\dot{s}(\xi_{(l)},\theta^{(l)}),\ l=k,m$ за умови 5) теореми, будуть додатно визначеними. Дійсно, для усіх $c\in\mathbb{R}^d,\ c\neq 0$,

$$-c\,\mathbb{E}\,\dot{s}(\xi_{(l)},\theta^{(l)})c^T=\mathbb{E}\left(\langle c,\dot{g}(\xi_{(l)},\theta^{(l)})\rangle\right)^2>0.$$

Покажемо, що умова 2) теореми 3.6 виконується. Оцінимо далі $s(z,\gamma)$ і $\dot{s}(z,\gamma)$

$$\max_{\gamma\in\Theta}|s(z,\gamma)|\leqslant\left(|y|+h_0(x)\right)h_1(x)=H_0(z),$$

$$\max_{\gamma\in\Theta}|\dot{s}(z,\gamma)|=\max_{\gamma\in\Theta}|-\dot{g}(x,\gamma)\dot{g}(x,\gamma)^T+(y-g(x,\gamma))\ddot{g}(x,\gamma)|\leqslant$$

$$h_1^2(x)+\left(|y|+h_0(x)\right)h_2(x)=H_1(z),$$

і

$$\max_{i,j,l=1,\dots,d}\sup_{\gamma\in\Theta}\left|\frac{\partial s^l(z,\gamma)}{\partial\gamma^i\partial\gamma^j}\right|\leqslant 3h_1(x)h_2(x)+\left(|y|+h_0(x)\right)h_3(x)=H_2(z).$$

З відношення $\mathbb{E}[\sum_{j=1}^3 H_j(\xi_{(l)})]^\alpha<\infty$, умови 2) теореми випливає умова 2) теореми 3.6.

□

Приклад 3.7. Нехай $\psi_{j;n}(z,\gamma)=(y-g(x,\gamma))\dot{g}(x,\gamma)$; $g(x,\gamma)=\frac{1}{1+e^{-\langle x,\gamma\rangle}}$, виконуються умови 1), 3), 4), 6), 7) теореми 3.7 і наступні умови:

1 Для $l = k, m$ і деякого $\alpha \geqslant 4$, $\mathbb{E}\, \varepsilon_{(l)}^{\alpha} < \infty$, та $\mathbb{E}\, |X_{(l)}|^{3\alpha} < \infty$

2 Для будь-якого $c \in \mathbb{R}^d, c \neq 0$ $\mathbb{P}\{\langle c, X_{(l)} \rangle = 0\} < 1$ для $l = k, m$.

Тоді $\hat{V}_n^{(k,m)} \to V^{(k,m)}$ за ймовірністю.

Доведення. Скористаємось теоремою 3.7. Для цього лише достатньо перевірити її умови 2) і 5).

Почнемо з умови 5). Візьмемо довільне $c \in \mathbb{R}^d, c \neq 0$. Тоді

$$\mathbb{P}\{\langle c, \dot{g}(\xi_{(l)}, \theta^{(l)}) \rangle = 0\} = \mathbb{P}\{g(\xi_{(l)}, \theta^{(l)})\big(1 - g(\xi_{(l)}, \theta^{(l)})\langle c, X_{(l)}\big)\rangle = 0\} < 1,$$

за умовою 2) теореми.

Покажемо, що умова 2) теореми 3.7 виконується. Випишемо усі обмеження функцій із умови 2) теореми 3.7. Для усіх $x \in \mathbb{R}^D$

$$\max_{\gamma \in \Theta} |g(x, \gamma)| \leqslant 1, \ \max_{\gamma \in \Theta} |\dot{g}(x, \gamma)| \leqslant |x|,$$

$$\max_{\gamma \in \Theta} |\ddot{g}(x, \gamma)| \leqslant |x|^2, 7 \max_{\gamma \in \Theta, j; t = 1, \dots, d} \left| \frac{\partial \ddot{g}_{j,t}(x, \gamma)}{\partial \gamma} \right| \leqslant 4|x|^3.$$

Тоді

$$\mathbb{E}\left(\big(|Y_{(l)}| + h_0(X_{(l)})\big)\big(\sum_{i=1}^{3} h_i(X_{(l)})\big) + 3h_1(X_{(l)})h_2(X_{(l)}) + h_1^2(X_{(l)}) \right)^{\alpha} \leqslant$$

$$\leqslant \mathbb{E}\left(\big(|\varepsilon_{(l)}| + 2\big)\big(\sum_{i=1}^{3} |X_{(l)}|^i\big) + 3|X_{(l)}|^3 + |X_{(l)}|^2 \right)^{\alpha}.$$

З нерівності Мінковського, незалежності $\varepsilon_{(l)}$ і $X_{(l)}$ і умови 1), випливає виконання умови 2) теореми 3.7.

□

3.6 Довірчі еліпсоїди для параметрів регресії

Для побудови асимптотичних довірчих множин параметрів $\theta^{(k)} \in \mathbb{R}^d$, використаємо результати теореми 3.4 про асимптотичну нормальність оцінок $\hat{\theta}_n^{(k)}$. Для $t \in \mathbb{R}^d$ та будь-якої невиродженої матриці $S \in \mathbb{R}^{d \times d}$, визначимо

$$T^{(k)}(t, S) = n(\hat{\theta}_n^{(k)} - t)^T S^{-1} (\hat{\theta}_n^{(k)} - t).$$

Зрозуміло, якщо теорема 3.4 виконується, і матриця $V^{(k,k)}$ - невироджена, то

$$T^{(k)}(\theta^{(k)}, V^{(k,k)}) \xrightarrow{w} \chi_d^2, \tag{3.2}$$

тут χ_d^2 - це χ^2-розподіл з d ступенями вільності. Зауважимо, що (3.2) виконується, якщо $V^{(k,k)}$ замінити на її деяку консистентну оцінку. Квантиль рівня α розподілу χ_d^2 позначимо χ_α. Тоді множина

$$B_n^k(\alpha) = \{t \in \mathbb{R}^d : T^{(k)}(\theta^{(k)}, \hat{V}_n^{(k,k)}) < \chi_\alpha\}$$

є асимптотичною довірчою множиною вектору параметрів $\theta^{(k)}$ рівня α у тому сенсі, що

$$\mathbb{P}\{\theta^{(k)} \in B_n^k(\alpha)\} \to 1 - \alpha, \text{ при } n \to \infty.$$

Для завершення побудови довірчих множин, нам потрібно навести умови за яких $V^{(k,k)}$ буде невиродженою. Розглянемо ці умови у випадку, коли $s(\xi_j, \gamma) = \left(Y_j - g(X_j, \gamma)\right) \dot{g}(X_j, \gamma)$

3.6.1 Невиродженість $V^{(k,k)}$

Оскільки

$$V^{(k,k)} = (M^{(k)})^{-1} Z^{(k,k)} (M^{(k)})^{-T},$$

потрібно знайти умови невиродженості матриць $M^{(k)} = \mathbb{E}\, \dot{g}(X_{(k)}, \theta^{(k)}) \dot{g}(X_{(k)}, \theta^{(k)})^T$, та $Z^{(k,k)}$.

Теорема 3.5 показує, що за умови

$$\mathbb{P}\{\langle c, \dot{g}(\xi_{(m)}, \theta^{(m)}) \rangle = 0\} < 1 \text{ для } m = 1, ..., M, \text{ і } c \in \mathbb{R}^d, c \neq 0,$$

матриці $-M^{(k)}$ будуть невиродженими і додатно визначеними.

Теорема 3.8. *Припустимо, що матриця $Z^{(k,k)}$ існує, скінченна, та*

$$\mathbb{P}\{\langle c, \dot{g}(\xi_{(m)}, \theta^{(m)}) \rangle = 0\} < 1 \text{ для } m = 1, ..., M,$$

для $c \in \mathbb{R}^d, c \neq 0$, і $\mathbb{D}\, \varepsilon_{(k)} > 0$. Тоді $V^{(k,k)}$ існує і невироджена.

Доведення. Далі символом $\geqslant$ позначатимемо порівняння Льовнера для матриць, тобто, $A \geqslant Z$ означатиме те, що $A - Z$ - додатно визначена матриця.

Помітимо, що

$$\mathbb{C}\mathrm{ov}\, s(\xi_j, \theta^{(k)}) = \mathbb{E}[\mathbb{C}\mathrm{ov}[s(\xi_j, \theta^{(k)})]|\kappa_j] + \mathbb{C}\mathrm{ov}[E[s(\xi_j, \theta^{(k)})|\kappa_j]] \geqslant$$

$$\mathbb{E}[\mathbb{C}\mathrm{ov}(s(\xi_j, \theta^{(k)}))|\kappa_j] = \sum_{l=1}^{M} p^l_{j;n}\, \mathbb{C}\mathrm{ov}[s(\xi_{(l)}, \theta^{(k)})] \geqslant$$

$$p^k_{j;n}\, \mathbb{C}\mathrm{ov}[s(\xi_j, \theta^{(k)})] = p^k_{j;n}\, \mathbb{D}\, \varepsilon_{(k)}(-M^{(k)}).$$

З (3.3) маємо, що

$$Z^{(k,k)} \geqslant \langle (a^{(k)})^2 p^k, \rangle\, \mathbb{D}\, \varepsilon_{(k)}(-M^{(k)}).$$

Оскільки $(-M^{(k)}) \geqslant 0$, та $\det M^{(k)} \neq 0$, то для доведення теореми достатньо показати, що $\langle (a^{(k)})^2 p^k, \rangle > 0$. Для того, щоб це довести, згадаємо, що

$$\sum_{j=1}^{n} a^{(k)}_{j;n} p^k_{j;n} = 1.$$

Тоді

$$\sum_{j=1}^{n} a^{(k)}_{j;n} p^k_{j;n} I[a^k_{j;n} > 1/(2n)] = 1 - \sum_{j=1}^{n} a^{(k)}_{j;n} p^k_{j;n} I[a^k_{j;n} \leqslant 1/(2n)] \geqslant 1/2,$$

з умови $0 \leqslant p^k_{j;n} \leqslant 1$, та

$$n \sum_{j=1}^{n} (a^k_{j;n})^2 p^k_{j;n} \geqslant n \sum_{j=1}^{n} (a^k_{j;n} p^k_{j;n})(a^k_{j;n} I[a^k_{j;n} > 1(2n)]) \geqslant 1/4.$$

Отже, і

$$\langle (a^{(k)})^2 p^k, \rangle = \lim_{n\to\infty} n \sum_{j=1}^{n} (a^k_{j;n})^2 p^k_{j;n} \geqslant 1/4.$$

Теорему доведено. □

3.6.2 Консистентні оцінки $V^{(k,k)}$

Є принаймні дві способи оцінювання $V^{(k,k)}$. Перший базується на техніці підстановки. Для цього будуються емпіричні аналоги матриць $M^{(k)}$ і $Z^{(k,k)}$, та підставляємо їх у (3.2) щоб отримати оцінку $V^{(k,k)}$. Матрицю $M^{(k)} = -\mathbb{E}\, \dot g(X_{(k)}, \theta^{(k)}) \dot g(X_{(k)}, \theta^{(k)})^T$ оцінимо матрицею

$$\hat M^{(i,k)}_n = -\sum_{j=1}^{n} a^i_{j;n} \dot g(X_j, \hat\theta^{(k)}_n) \dot g(X_j, \hat\theta^{(k)}_n)^T.$$

Оцінимо тепер матрицю $Z^{(k,k)}$. За виразом (3.3) випишемо її емпіричний аналог. Для цього запишемо наступні оцінки:

$$\hat{S}_n^{(i,k)} = \sum_{j=1}^{n} a_{j;n}^{i} s(\xi_j, \theta_n^{(k)})$$

$$\hat{D}_n^{(i,k)} = \sum_{j=1}^{n} a_{j;n}^{i} s(\xi_j, \hat{\theta}_n^{(k)}) s(\xi_j, \hat{\theta}_n^{(k)})^T,$$

та замінимо наступні границі $\langle (a^{(i)})^2, p^k \rangle$ і $\langle (a^{(i)})^2, p^k p^l \rangle$ виразами

$$\alpha(i,k) = \sum_{j=1}^{n} (a_{j;n}^{i})^2 p_{j;n}^{k}, \; \alpha(k,l,m) = \sum_{j=1}^{n} (a_{j;n}^{i})^2 p_{j;n}^{k} p_{j;n}^{l}.$$

З цього маємо, що оцінка матриці $Z^{(k,k)}$ матиме наступний вигляд

$$\hat{Z}_n^{(k,k)} = \sum_{i=1}^{M} \alpha(k,i) \hat{D}_n^{(i,k)} - \sum_{i,l=1}^{M} \alpha(k,i,l) \hat{S}_n^{(i,k)} (\hat{S}_n^{(l,k)})^T.$$

Отже, оцінка $V^{(k,k)}$, що отримана методом підстановки, наступна:

$$^{plug}\hat{V}_n^{(k,k)} = (\hat{M}_n^{(k,k)})^{-1} \hat{Z}_n^{(k,k)} (\hat{M}_n^{(k,k)})^{-T}. \tag{3.3}$$

Другий спосіб оцінювання $V^{(k,k)}$ - використання методу складаного ножа. Зокрема, якщо теорема 3.6 виконується, то $V^{(k,k)}$ можна оцінити виразом $\hat{V}_n^{(k,k)}$, визначеним у (3.10), яка буде консистентною оцінкою

$$^{JK}\hat{V}_n^{(k)} = \sum_{i=1}^{n} \left(\hat{\theta}_{i-;n}^{(k)} - \hat{\theta}_n^{(k)}\right)\left(\hat{\theta}_{i-;n}^{(k)} - \hat{\theta}_n^{(k)}\right)^T.$$

3.7 Алгоритмічні питання і результати моделювання

Якість побудованих довірчих еліпсоїдів перевірено імітаційними моделюваннями на $N = 1000$ симульованих вибірках для кожного експерименту. В експериментах було побудовано довірчі еліпсоїди параметрів регресії з номінальним довірчим рівнем 95%, та пораховано частоти покриття довірчими множинами справжніх значень параметрів.

Дані було створено за моделлю суміші з двох компонентів ($M = 2$), та ймовірностей концентрацій, які також отримані випадково за наступним правилом:

$$p_{j;n}^k = \frac{u_j^k}{u_j^1 + u_j^2},\ k = 1, 2,$$

де $u_j^i,\ i = 1, 2,\ j = 1, ..., n$ - незалежні, рівномірно розподілені на [0, 1] випадкові величини.

Кожне спостереження складається з двох змінних (Y, X), а їх розподіл для m-го компонента відповідає моделі логістичної регресії з неперервним відгуком:

$$Y_{(m)} = g(X_{(m)}, \theta^{(m)}) + \varepsilon_{(m)},\ X_{(m)} \simeq N(\mu^{(m)}, \Sigma^{(m)}),$$

$$g(X, \gamma) = \frac{1}{1 + exp(\gamma_0 + \gamma_1 X)}.$$

Тут $\theta^{(m)} = (\theta_0^{(m)}, \theta_1^{(m)})^T$ - вектор невідомих параметрів регресії m-го компонента, які потрібно оцінити. Справжні значення параметрів з якими було створено вибірки виписані у таблиці 3.1.

m	1	2
$\mu^{(m)}$	0.0	1.0
$\Sigma^{(m)}$	2.0	2.0
$b_0^{(m)}$	0.5	0.5
$b_1^{(m)}$	2	-1/3

Табл. 3.1. Параметри експериментів

$\varepsilon_{(m)}$ - центрована регресійна похибка, що незалежна із $X_{(m)}$. Її розподіл різний для різних експериментів.

Довірчі множини були побудовані за $N = 1000$ генерованих вибірках та пораховано ймовірності покриття. Ці ймовірності записані у таблицях для кожного експерименту.

3.7.1 Пришвидшений розрахунок навантажень

Розрахунок оцінки матриці $V^{(k,k)}$ потребує знаходження оцінок $\hat{\theta}_{i-;n}^{(k)}, i = 1, ..., n,\ k = 1, ..., M$, кожна з яких потребує розрахунок вагових коефіцієнтів $a_{ji-;n}^k$. Позначимо матрицю навантажень $A_n = (a_{j;n}^k)_{k=1,j=1}^{M,n}$ і $A_{i-;n} = (a_{ji-;n}^k)_{k=1,j=1}^{M,n}$.

Виразом $P_{i-;n}$ позначимо матрицю концентрацій компонент P_n, з якої вилучили концентрації i-го спостереження.

Матриці $A_{i-;n}$ можна зручно (і довго) рахувати за наступними формулами:

$$A_{i-;n} = P_{i-;n}\Gamma^{-1}_{i-;n};\ \Gamma_{i-;n} = P^T_{i-;n}P_{i-;n}$$

Для швидкого перерахунку $\Gamma^{-1}_{i-;n}$ можна використати наступні формули:

$$\Gamma^{-1}_{i-;n} = \Gamma^{-1}_n + \frac{1}{1-h_i}\Gamma^{-1}_n p_i p_i^T \Gamma^{-1}_n;\ \Gamma_n = P^T_n P_n,$$

тут $p_i = (p_i^1, ..., p_i^M)$, та $h_i = p_i^T\Gamma_n^{-1}p_i$.

3.7.2 Консистентність оцінки коваріації

У кожному експерименті було пораховано:

1 Квазі-довірчі множини зі справжнім значенням матриці $V^{(k,k)}$;

2 Довірчі еліпсоїди за оцінкою методу підстановки ${}^{plug}\hat{V}_n^{(k,k)}$;

3 Еліпсоїди методом складаного ножа ${}^{JK}\hat{V}_n^{(k,k)}$;

n	Перший компонент			Другий компонент		
	$V^{(k,k)}$	${}^{plug}\hat{V}_n^{(k,k)}$	${}^{JK}\hat{V}_n^{(k,k)}$	$V^{(k,k)}$	${}^{plug}\hat{V}_n^{(k,k)}$	${}^{JK}\hat{V}_n^{(k,k)}$
100	0.668	0.942	0.955	0.729	0.939	0.953
500	0.929	0.931	0.956	0.948	0.934	0.939
1 000	0.954	0.951	0.95	0.944	0.937	0.939
5 000	0.959	0.951	0.943	0.952	0.940	0.931
7 500	0.961	0.942	0.933	0.951	0.938	0.957
10 000	0.954	0.949	0.944	0.944	0.947	0.954

Табл. 3.2. Ймовірності покриття експерименту 1

3.7.2.1 Експеримент 1

Тут похибки регресії були центровані із дисперсією $\mathbb{D}\,\varepsilon_{(k)}$=0.25, для $k = 1, 2$. Ймовірності покриття показано у таблиці 3.2. Видно, що точність методу підстановки у цьому експерименті невисока, але достатня для практичних застосувань на вибірках, розмір яких більший за 1000. Точність цих еліпсоїдів приблизно однакова із точністю справжніх довірчих множин, а отже, спостережувані відхилення ймовірностей покриття від номінального рівня 95% не може бути пояснена похибками оцінювання $V^{(k,k)}$. Отже, оцінки складаного ножа більш точні ніж, методом підстановки.

n	Перший компонент			Другий компонент		
	$V^{(k,k)}$	${}^{plug}\hat{V}_n^{(k,k)}$	${}^{JK}\hat{V}_n^{(k,k)}$	$V^{(k,k)}$	${}^{plug}\hat{V}_n^{(k,k)}$	${}^{JK}\hat{V}_n^{(k,k)}$
100	0.593	0.952	0.909	0.684	0.964	0.939
500	0.9907	0.929	0.938	0.924	0.929	0.934
1 000	0.917	0.959	0.946	0.951	0.944	0.939
5 000	0.938	0.941	0.934	0.947	0.959	0.933
7 500	0.934	0.948	0.948	0.958	0.956	0.944
10 000	0.937	0.947	0.943	0.955	0.945	0.950

Табл. 3.3. Ймовірності покриття експерименту 2

3.7.2.2 Експеримент 2

Тут ми розглядаємо обмежені похибки регресії, а саме рівномірно розподілені на інтервалі [-0.25, 0.25]. Результати записано у таблиці 3.3. Видно, що точність методу підстановки та складаного ножа приблизно однакові, як і у експерименті 1. Парадоксально, проте квазі-довірчі еліпсоїди з використанням справжніх значень $V^{(k,k)}$ показують гіршу точність, ніж та, що була у експерименті 1.

n	Перший компонент			Другий компонент		
	$V^{(k,k)}$	${}^{plug}\hat{V}_n^{(k,k)}$	${}^{JK}\hat{V}_n^{(k,k)}$	$V^{(k,k)}$	${}^{plug}\hat{V}_n^{(k,k)}$	${}^{JK}\hat{V}_n^{(k,k)}$
100	0.568	0.942	0.903	0.701	0.932	0.931
500	0.900	0.938	0.944	0.917	0.929	0.956
1 000	0.936	0.932	0.926	0.944	0.931	0.919
5 000	0.930	0.949	0.945	0.942	0.936	0.948
7 500	0.946	0.939	0.953	0.941	0.953	0.926
10 000	0.955	0.947	0.935	0.925	0.944	0.936

Табл. 3.4. Ймовірності покриття експерименту 3

3.7.2.3 Експеримент 3

Тут ми порівняємо точності еліпсоїдів регресії для похибок із важкими хвостами. Похибки генеровані розподілом $\nu/10$, де ν - t розподіл Стьюдента із 4-ма ступенями вільності. Результати показано у таблиці 3.4. У цьому випадку, точність методу складаного ножа здається значно гіршою, ніж у експериментах 1 і 2. Метод підстановки демонструє приблизно ту ж саму якість.

3.8 Висновки

У цьому розділі було розглянуто модель суміші нелінійних регресій зі змінними концентраціями компонентів. Побудовано семі-параметричні оцінки невідомих параметрів регресії для випадку нелінійної регресії методом оціночних рівнянь. Для оцінок параметрів цієї моделі доведено консистентність і асимптотичну нормальність, а для оцінок матриць коваріацій методом СН доведено консистентність, та побудовано довірчі еліпсоїди для параметрів регресії. Окрім цього, асимптотичні властивості конкретизовано для випадку оцінок МНК, та додатково для логістичної функції регресії.

Отримані результати було перевірено імітаційним моделюванням логістичної функції регресії, та різних розподілів похибок. З результатів моделювання видно, що оцінки методом СН поступаються у якості відповідним оцінкам методом підстановки.

Розділ 4

Аналіз залишків

Проблема аналізу залишків у регресійних моделях виникає під час регресійного аналізу, та вивчення розподілу похибок регресії і його параметрів. Це важливо для побудови статистичних тестів. Наприклад, для тесту Фішера у поясненні результатів, чи відкидання зайвих регресорів. У нашому випадку, ми будуємо техніку регресійного аналізу, щоб побудувати: оцінки розподілів залишків, оцінки квантилів, та оцінки дисперсії залишків для застосування на соціологічних даних.

Розглянемо далі модель регресійної суміші зі змінними концентраціями із вибіркою $\Xi_n = \left(X_j, Y_j\right)_{j=1}^{n}$. Зв'язок між регресорами X і відгуками Y, у випадку, коли об'єкт належить k-му компоненту має наступний вигляд

$$Y_{(k)} = g(X_{(k)}, \theta^{(k)}) + \varepsilon_{(k)}.$$

Спостереження $\xi_{j;n} = \left(X_j, Y_j\right) \in \mathbb{R}^D,\ j = 1, ..., n$ є незалежними при різних j і мають розподіл

$$\mathbb{P}\{\xi_{j;n} \in A\} = \sum_{m=1}^{M} p_{j;n}^{m} F_m(A), \tag{4.1}$$

і для усіх $1 \leqslant j \leqslant n$, $\sum_{m=1}^{M} p_{j;n}^{m} = 1$.

Припустимо, що похибки регресії k - го компонента $\varepsilon_{(k)}$ мають розподіл із наступними властивостями: $\mathbb{E}\,\varepsilon_{(k)} = 0$, та $\mathbb{D}\,\varepsilon_{(k)} = \sigma_k^2 > 0$. Далі параметр σ_k^2 невідомий і перед нами стоїть задача побудувати для нього оцінку за вибіркою Ξ_n.

Результати цього розділу надруковані в статтях [66], [67],

4.1 Консистентність оцінки дисперсії залишків лінійної моделі

У цьому підрозділі розглянемо випадок простої лінійної регресійної суміші, у якій

$$Y_j = \langle X_j, \theta^{(\kappa_j)} \rangle + \varepsilon_j.$$

Символом $\theta^{(k)}$ позначимо вектор параметрів регресії, що невідомий.

Позначимо наступну функцію

$$\hat{\sigma}^2_{k,n}(\gamma) = \sum_{j=1}^{n} a^k_{j;n} \left(Y_j - \langle X_j, \gamma \rangle \right)^2 .$$

Далі, $\hat{\theta}^{(k)}_n = \left(X^T A^{(k)}_n X \right)^{-1} X^T A^{(k)}_n Y$ - оцінка параметрів регресії $\theta^{(k)}$ (2.1) з розділу 3. На роль оцінки параметра дисперсії залишків σ^2_k виберемо оцінку $\hat{\sigma}^2_{k,n}(\hat{\theta}^{(k)}_n)$.

Наступна теорема показує умови, за яких оцінка $\hat{\sigma}^2_{k,n}(\hat{\theta}^{(k)}_n)$ буде консистентною.

Теорема 4.1. *Нехай виконуються такі умови:*

1 Навантаження $a^k_{j;n}$ задовольняють умові (Г).

2 Оцінка $\hat{\theta}^{(k)}_n \xrightarrow{\mathbb{P}} \theta^{(k)}$, $n \to \infty$.

3 $\exists\, \mathbb{D}\, \varepsilon_{(k)} = \sigma^2_k > 0$

4 $\exists\, \mathbb{E}\, |X_{(k)}|^2 < \infty$ для $1 \leqslant k \leqslant M$.

Тоді

$$\hat{\sigma}^2_{k,n}(\hat{\theta}^{(k)}_n) \xrightarrow{\mathbb{P}} \sigma^2_k,\ n \to \infty.$$

Доведення. З теореми 1.3 маємо, що $\hat{\sigma}^2_{k,n}(\theta^{(k)}) \xrightarrow{\mathbb{P}} \sigma^2_k$, при $n \to \infty$. Це означає, що достатньо довести наступну збіжність для доведення теореми:

$$\hat{\sigma}^2_{k,n}(\hat{\theta}^{(k)}_n) - \hat{\sigma}^2_{k,n}(\theta^{(k)}) \xrightarrow{\mathbb{P}} 0, n \to \infty.$$

Для цього покажемо, що права частина наступного виразу також консистентна.

$$\hat{\sigma}^2_{k,n}(\hat{\theta}^{(k)}_n) - \hat{\sigma}^2_{k,n}(\theta^{(k)}) = \sum_{j=1} a^k_{j;n} \Big(-2Y_j \big(g(X_j, \hat{\theta}^{(k)}_n) - g(X_j, \theta^{(k)}) \big) + g^2(X_j, \hat{\theta}^{(k)}_n) - g^2(X_j, \theta^{(k)}) \Big)$$

Покажемо, що обидва доданки консистентні:

1) За умов 2, 3, 4 і леми 1.3

$$\sum_{j=1} a^k_{j;n} Y_j \big(g(X_j, \hat{\theta}^{(k)}_n) - g(X_j, \theta^{(k)})\big) = \langle 2\sum_{j=1}^{n} a^k_{j;n} Y_j X_j; \hat{\theta}^{(k)}_n - \theta^{(k)}\rangle \xrightarrow{\mathbb{P}} 0,\ n \to \infty.$$

2) За умов 2, 3 і леми 1.3

$$\sum_{j=1} a^k_{j;n}\Big(g^2(X_j, \hat{\theta}^{(k)}_n) - g^2(X_j, \theta^{(k)})\Big) = \sum_{j=1} a^k_{j;n}\langle X_j, \hat{\theta}^{(k)}_n + \theta^{(k)}\rangle\langle X_j, \hat{\theta}^{(k)}_n - \theta^{(k)}\rangle =$$

$$= (\hat{\theta}^{(k)}_n + \theta^{(k)})^T \big(\sum_{j=1} a_{j;n} X_j X_j^T\big)(\hat{\theta}^{(k)}_n - \theta^{(k)}) \xrightarrow{\mathbb{P}} 0,\ n \to \infty.$$

□

4.2 Консистентність оцінки дисперсії залишків нелінійної моделі

Використаємо введену раніше функцію для побудови оцінки дисперсії залишків нелінійної регресії:

$$\hat{\sigma}^2_{k,n}(\gamma) = \sum_{j=1}^{n} a^k_{j;n}\left(Y_j - g(X_j, \gamma)\right)^2.$$

Тут $g(x, \gamma) : \mathbb{R}^{D-1} \times \Theta \to R$ - деяка нелінійна функція, $\Theta \subset \mathbb{R}^d$. Тут і далі $\hat{\theta}^{(k)}_n$ - оцінка параметрів регресії $\theta^{(k)}$ знайдена методом оціночних рівнянь (УОР-оцінка). Тобто, $S^k_n(\hat{\theta}^{(k)}_n) = 0$ м.н.

Теорема 4.2. *Нехай виконуються наступні умови:*

1 $\det \Gamma_\infty > 0$

2 $\hat{\theta}^{(k)}_n \xrightarrow{\mathbb{P}} \theta^{(k)}$, *при* $n \to \infty$

3 Для функцій $g(x, \gamma)$ *і* $g(x, \gamma)^2$ *виконуються умови* $\Psi_{h\rho}$ *і* $\Psi_{h_2\rho}$ *відповідно*

4 $\mathbb{E}\,|g(X_{(t)}, \theta^{(t)})||h(X_{(t)})| < \infty$ *для всіх* $t = 1, ..., M$

5 $\mathbb{E}\,|h_2(X_{(t)})| < \infty$ *для всіх* $t = 1, ..., M$

6 $\rho(\gamma_1, \gamma_2)$ *- неперервна по двом аргументам, і* $\rho(\gamma, \gamma) = 0$.

Тоді $\hat{\sigma}^2_{k,n}(\hat{\theta}^{(k)}_n) \xrightarrow{\mathbb{P}} \sigma^2_k$

Доведення. За теоремою 1.3, $\sigma_k^2(\theta^{(k)}) \xrightarrow{\mathbb{P}} \sigma_k^2$, при $n \to \infty$. Це означає, що достатньо показати збіжність

$$\sigma_k^2(\theta^{(k)}) - \sigma_k^2(\hat{\theta}_n^{(k)}) \xrightarrow{\mathbb{P}} 0, \text{ при } n \to \infty.$$

Отже, за означенням

$$\sigma_k^2(\theta^{(k)}) - \sigma_k^2(\hat{\theta}_n^{(k)}) = -2\sum_{j=1}^{n} a_{j;n}^k Y_j(g(X_j, \theta^{(k)}) - g(X_j, \hat{\theta}_n^{(k)}))+$$

$$+\sum_{j=1}^{n} a_{j;n}^k (g(X_j, \theta^{(k)})^2 - g(X_j, \hat{\theta}_n^{(k)})^2) =$$

Покажемо, що обидві суми збігаються до 0.

1. За лемою 3.1, і умовами 1, 3 теореми,

$$\Big|\sum_{j=1}^{n} a_{j;n}^k Y_j(g(X_j, \theta^{(k)}) - g(X_j, \hat{\theta}_n^{(k)}))\Big| \leqslant \frac{c}{n}\sum_{j=1}^{n} |Y_j| h(X_j)\rho(\theta^{(k)}, \hat{\theta}_n^{(k)})$$

Цей добуток збігається до 0 із умов 2, 4, 6 теореми і теореми 1.3. 2. За лемою 3.1, і умовами 1, 3 теореми,

$$\Big|\sum_{j=1}^{n} a_{j;n}^k (g(X_j, \theta^{(k)})^2 - g(X_j, \hat{\theta}_n^{(k)})^2)\Big| \leqslant \frac{c}{n}\sum_{j=1}^{n} h_2(X_j)\rho(\theta^{(k)}, \hat{\theta}_n^{(k)})$$

Цей добуток збігається до 0 із умов 2, 5, 6 теореми 1.3.

Теорему доведено □

Теорема 4.2 показує консистентність $\hat{\sigma}_k^2(\hat{\theta}_n^{(k)})$ для довільної нелінійної функції $g(x, \gamma)$. Наступний приклад розглянуто для випадку, коли

$$g(x, \gamma) = \frac{1}{1 + e^{-\langle x, \gamma\rangle}} \tag{4.2}$$

Приклад 4.1. Нехай функція $g(x, \gamma)$ має вигляд (4.2), і виконуються такі умови:

1 $\det \Gamma_\infty > 0$

2 $\hat{\theta}_n^{(k)} \xrightarrow{\mathbb{P}} \theta^{(k)}$, при $n \to \infty$

3 $\mathbb{E}|Y_{(t)}||X_{(t)}| < \infty$, $\mathbb{E}|X_{(t)}| < \infty$ для всіх $t = 1, ..., M$

Тоді

$$\hat{\sigma}_{k,n}^2(\hat{\theta}_n^{(k)}) \xrightarrow{\mathbb{P}} \sigma_k^2, n \to \infty.$$

Доведення. Аналогічно, до Теореми 4.2, Достатньо показати, що консистентно збіжні наступні дві суми:

1) $S_1 = \left|\sum_{j=1}^{n} a^k_{j;n} Y_j (g(X_j, \theta^{(k)}) - g(X_j, \hat{\theta}^{(k)}_n))\right|$. За теоремою про середнє значення, і теоремою 3.1, умови 3 прикладу і факту $\dot{g}(x,\gamma) = g(x,\gamma)(1 - g(x,\gamma))x^T$

$$S_1 \leqslant \frac{C}{n} \sum_{j=1}^{n} |Y_j| \sup_{\gamma \in \Theta} |\dot{g}(X_j, \gamma)| |\hat{\theta}^{(k)}_n - \theta^{(k)}| \leqslant \frac{C}{n} \Big(\sum_{j=1}^{n} |Y_j||X_j| \Big) |\hat{\theta}^{(k)}_n - \theta^{(k)}| \xrightarrow{\mathbb{P}} 0,$$

при $n \to \infty$.

2) $S_2 = \left|\sum_{j=1}^{n} a^k_{j;n} (g(X_j, \theta^{(k)})^2 - g(X_j, \hat{\theta}^{(k)}_n)^2)\right|$. За теоремою про середнє значення, і теоремою 3.1, умови 3 прикладу

$$S_2 \leqslant \frac{C}{n} \sum_{j=1}^{n} |g(X_j, \hat{\theta}^{(k)}_n) + g(X_j, \theta^{(k)})||g(X_j, \hat{\theta}^{(k)}_n) - g(X_j, \theta^{(k)})| \leqslant$$

$$\leqslant \frac{2C}{n} \sum_{j=1}^{n} |X_j||\hat{\theta}^{(k)}_n - \theta^{(k)}| \xrightarrow{\mathbb{P}} 0,$$

при $n \to \infty$. □

4.3 Модифікація діаграм квантиль проти квантиля для моделей регресійних сумішей

Для використання підігнаної регресійної моделі важливо вміти аналізувати розподіл похибок. У випадку однорідних спостережень для візуального порівняння розподілу похибок регресії з деяким теоретичним розподілом, прийнято застосовувати діаграми квантиль проти квантиля (QQ-діаграми) залишків регресії.

Нагадаємо, як будується стандартна QQ-діаграма залишків. Нехай регресійна модель задається формулою

$$Y_j = g(X_j, \theta) + \varepsilon_j, \ \ j = 1, ..., n,$$

де (X_j, Y_j) - спостережувані дані, θ - невідомий параметр, g - відома регресійна функція, а ε_j - похибки регресії, які вважаються незалежними, однаково розподіленими випадковими величинами. На роль спостережуваного замінника неспостережуваних похибок використовують залишки регресії

$$U_j = Y_j - g(X_j, \hat{\theta}),$$

де $\hat{\theta}$ - деяка оцінка θ (зазвичай, оцінка МНК).

На QQ-діаграмі $j-$тому спостереженню ставиться у відповідність точка з координатами $(U_{[j]}, Q^F(\alpha_j))$, де $U_{[j]}$ - порядкові статистики набору U_j, а $Q^F(\alpha_j)$ - квантилі рівня α_j для того розподілу F, який є гіпотетичним розподілом похибок ε_j. Рівні α_j підбираються так, що U_j можна було розглядати як оцінку для $Q^F(\alpha_j)$ (Насправді, популярним є вибір $\alpha_j = \frac{j-1/2}{n}$).

Якщо на такій діаграмі точки розташовані вздовж бісектриси першого координатного кута, то це свідчить на користь гіпотези, що F є розподілом ε. Наявність значущих відхилень суперечить цій гіпотезі.

Для моделі суміші регресій поняття залишка ускладнюється, оскільки для кожного компонента суміші є свої власні оцінки невідомого параметра $\hat{\theta}_n^{(k)}$ і залишки потрібно визначити окремо для кожного компонента.

Нехай дані задані моделлю суміші регресій (1.5)

$$Y_j = g(X_j, \theta^{(\kappa_j)}) + \varepsilon_j,$$

де похибки регресій мають розподіл, що залежить від κ_j - номера компонента суміші, якому належить спостереження. Позначимо $F_\varepsilon^{(k)}(x) = P\{\varepsilon_j < x | \kappa_j = k\}$ - функція розподілу похибки для k-го компонента суміші. Для того, щоб побудувати QQ-діаграму, залишки потрібно визначити для кожного компонента окремо. Отже, позначимо залишок $k-$го компонента

$$U_j^{(k)} = Y_j - g(X_j, \hat{\theta}_n^{(k)}). \tag{4.3}$$

Якщо $j-$те спостереження належить $k-$му компоненту $(\kappa_j = k)$, а оцінка $\hat{\theta}_n^{(k)}$ є близькою до справжнього $\theta^{(k)}$, то можна сподіватись, що $U_j^{(k)}$ буде близьким до $\varepsilon_j = Y_{j;n} - g(X_j, \theta^{(k)})$ і, отже, за значенням таких $U_j^{(k)}$ можна було б спробувати оцінити $F_\varepsilon^{(k)}(x)$. Відобразити на QQ-діаграмі значення залишків $k-$го компонента, розташованих у порядку зростання, як це робиться для однорідних вибірок, є недоцільним, оскільки неможливо визначити для яких j $\kappa_j = k$.

Ми розглянемо QQ-діаграми, в яких використовуються оцінки теоретичних квантилів розподілу $F_\varepsilon^{(k)}$ на основі залишків (4.3). Для визначення таких оцінок розглянемо спочатку оцінки для $F_\varepsilon^{(k)}$:

$$\tilde{F}_n^{(k)}(x) = \sum_{j=1}^{n} a_{j;n}^k I[U_j^{(k)} < x], \tag{4.4}$$

а потім як оцінку теоретичних квантилів використаємо квантилі $\tilde{F}_n^{(k)}$.

Теорема 4.3. *Нехай виконуються умови:*

1 $\det \Gamma_\infty > 0$.

2 Для функції g *виконується умова* $\Psi_{h,\rho}$ *з* $\rho(x_1, x_2) = |x_1 - x_2|$ *і функцією* h *такою, що для деякого* $\alpha > 2$

$$\mathbb{E}\left(h(X_{(k)})\right)^\alpha < \infty, \text{ при всіх } k = 1, \dots, M.$$

3 Розподіли $F_\varepsilon^{(k)}$ *мають щільність* $f_\varepsilon^{(k)}(x)$*, які є обмеженими функціями на* $\mathbb{R}$.

4 $\hat{\theta}_n^{(k)} - \sqrt{n}$ *консистентна оцінка* $\theta^{(k)}$.

Тоді

$$\sup_x |\tilde{F}_n^{(k)}(x) - F_\varepsilon^{(k)}(x)| \xrightarrow{\mathbb{P}} 0, \ n \to \infty.$$

Доведення. Розглянемо

$$z_j = \varepsilon_j + g(X_j, \theta^{(\kappa_j)}) - g(X_j, \theta^{(k)}); \quad \delta_j = g(X_j, \theta^{(k)}) - g(X_j, \hat{\theta}_n^{(k)}).$$

Тоді

$$U_j^{(k)} = Y_j - g(X_j, \hat{\theta}_n^{(k)}) = g(X_j, \theta^{(\kappa_j)}) + \varepsilon_j - g(X_j, \hat{\theta}_n^{(k)}) = z_j + \delta_j.$$

Нехай Δ_n - деяка числова послідовність, $\Delta_n \to 0, \ n \to \infty$. Задамо випадкову подію $A_n = \{\omega : \sup_{j=1,\dots,n} |\delta_j| < \Delta_n\}$. Припустимо, що подія A_n виконується. Тоді

$$I[z_j < x - \Delta_n] \leqslant I[U_j^{(k)} < x] \leqslant I[z_j < x - \delta_j] \leqslant I[z_j < x + \Delta_n].$$

Отже,

$$|I[U_j^{(k)} < x] - I[z_j^{(k)} < x]| \leqslant I[z_j^{(k)} < x + \Delta_n] - I[z_j^{(k)} < x - \Delta_n] =$$

$$I[x - \Delta_n < z_j^{(k)} < x + \Delta_n] \tag{4.5}$$

позначимо $\hat{F}_{z;n}^{(k)}(x) = \sum_{j=1}^n a_{j;n}^k I[z_j < x]$. З (4.5) отримуємо, що при виконанні події Δ_n

$$\sup_x |\tilde{F}_n^{(k)}(x) - \hat{F}_{z;n}^{(k)}(x)| \leqslant \sup_x \sum_{j=1}^n |a_{j;n}^k| I[x - \Delta_n < z_j < x + \Delta_n] \leqslant$$

$$\frac{C_a}{n} \sup_x \sum_{j=1}^{n} I[x - \Delta_n < z_j < x + \Delta_n]. \tag{4.6}$$

позначимо $\bar{F}_n(x) = \sum_{k=1}^{M} \langle p^k \rangle_n F_z^{(k)}(x)$, де

$$F_z^{(k)}(x) = P\{z_j < x | \kappa_j = k\}, \ \langle p^k \rangle_n = \frac{1}{n} \sum_{j=1}^{n} p_{j;n}^k; \ \ \check{F}_n(x) = \frac{1}{n} \sum_{j=1}^{n} I[z_j < x]$$

За Твердженням 1.7 отримуємо, що для деякої майже напевно скінченної випадкової величини Λ

$$\sup_x |\check{F}_n(x) - \bar{F}_n(x)| \leqslant \Lambda \sqrt{\frac{\ln n}{n}} \text{ для всіх } n = 1, 2, ...$$

Тому з (4.6) отримуємо, що при виконанні A_n,

$$|\check{F}_n(x) - \bar{F}_n(x)| \leqslant C_a(\check{F}_n(x+\Delta_n) - \check{F}_n(x-\Delta_n)) \leqslant C_a \left(\bar{F}_n(x + \Delta_n) - \bar{F}_n(x - \Delta_n) + 2\Lambda \frac{\ln n}{n} \right) \leqslant$$

$$C_a \left(\sum_{m=1}^{M} \left(F_z^{(m)}(X + \Delta_n) - F_z^{(m)}(X - \Delta_n) \right) + 2\Lambda \frac{\ln n}{n} \right).$$

Оцінимо $J(x, \Delta_n) = F_z^{(m)}(x + \Delta_n) - F_z^{(m)}(x - \Delta_n)$. Помітимо, що

$$F_z^{(m)}(x) = P\{z_j < x | \kappa_j = m\} =$$

$$P\{\varepsilon_j + g(X_j, \theta^{\kappa_j}) - g(X_j, \theta^{(m)}) < x | \kappa_j = m\} = P\{\varepsilon_j < x | \kappa_j = m\} = P\{\varepsilon_{(m)} < x\}.$$

$$\text{Тоді } J(x, \Delta_n) \leqslant \sup_t f_\varepsilon^{(m)}(t) 2\Delta_n \to 0, \text{ при } n \to \infty \tag{4.7}$$

З (4.6) і (4.7) отримуємо

$$\sup_x |\tilde{F}_n^{(k)}(x) - \hat{F}_{z;n}^{(k)}(x)| \leqslant C_a \left(\Delta_n c_p + 2\Lambda \sqrt{\frac{\ln n}{n}} \right), \tag{4.8}$$

де $c_p = \max_{k=1,...m} \sup_t f_\varepsilon^{(k)}(t)$.

$$\text{Виберемо } \Delta_n \text{ так, щоб } P(A_n) \to 1, \text{ при } n \to \infty. \tag{4.9}$$

З умови 2) отримуємо $|\delta_j| \leqslant h(X_j)|\hat{\theta}_n^{(k)} - \theta^{(k)}|$, при чому за Лемою 4.2 (2.1)

$$\max_{j=1,...,n} h(X_j) = o_P(n^{1/\alpha}),$$

а за умовою 4), $|\hat{\theta}_n^{(k)} - \theta^{(k)}| = O_P(n^{-1/2})$.

Отже, $|\delta_j| = o_P(n^{1/\alpha-1/2})$ і обравши $\Delta_n = n^{1/\alpha-1/2}$, отримуємо (4.9). Використавши Твердження 1.7, отримуємо

$$\sup_x |\hat{F}_{z;n}^{(k)}(x) - F_\varepsilon^{(k)}(x)| \xrightarrow{\mathbb{P}} 0, \ n \to \infty. \tag{4.10}$$

Об'єднавши (4.8), (4.9) (4.10), отримуємо твердження теореми. □

Розглянемо тепер оцінки для квантилів $F_\varepsilon^{(k)}$. Ці квантилі $Q^{F_\varepsilon^{(k)}}(\alpha) = Q_\varepsilon^{(k)}(\alpha)$ визначаються як

$$Q_\varepsilon^{(k)}(\alpha) = \inf\{x : F_\varepsilon^{(k)}(x) > \alpha\}. \tag{4.11}$$

Нехай $N_k = \{x : 0 < F_\varepsilon^{(k)}(x) < 1\}$ - носій розподілу $F_\varepsilon^{(k)}$. Ми обмежимось випадком, коли F_ε - строго зростаюча, неперервна на N_k функція. У цьому випадку $Q_\varepsilon^{(k)}(*)$ - функція обернена до функції $F_\varepsilon^{(k)}(*)$ для всіх $\alpha \in (0,1)$, і всіх $x \in N_k$.

$$F_\varepsilon^{(k)}\left(Q_\varepsilon^{(k)}(\alpha)\right) = \alpha, \ Q_\varepsilon^{(k)}\left(F_\varepsilon^{(k)}(x)\right) = x.$$

Для оцінювання $Q_\varepsilon^{(k)}(\alpha)$ можна замінити у (4.11) $F_\varepsilon^{(k)}$ на її оцінку $\tilde{F}_n^{(k)}(x)$, визначену у (4.4)

$$Q_{k;n}^{-}(\alpha) = \inf\{x : \tilde{F}_n^{(k)}(x) > \alpha\}$$

- це крайня ліва точка, в якій графік $\tilde{F}_n^{(k)}(x)$ перестрибує через рівень α (або досягає цього рівня).

Крім того, можна розглянути крайню праву точку, в якій $\tilde{F}_n^{(k)}(x)$ перестрибує α:

$$Q_{k;n}^{+}(\alpha) = \sup\{x : \tilde{F}_n^{(k)}(x) < \alpha\}.$$

На роль оцінки для $Q_\varepsilon^{(k)}(\alpha)$ можна розглянути будь-яку точку між $Q_{k;n}^{-}(\alpha)$ і $Q_{k;n}^{+}(\alpha)$.

Теорема 4.4. *Нехай виконані умови Теореми 4.3 і функція $F_\varepsilon^{(k)}(x)$ є строго зростаючою у деякому околі $Q_\varepsilon^{(k)}(\alpha)$. Тоді*

$$Q_{k;n}^{+}(\alpha) \xrightarrow{\mathbb{P}} Q_\varepsilon^{(k)}(\alpha) \text{ і } Q_{k;n}^{-}(\alpha) \xrightarrow{\mathbb{P}} Q_\varepsilon^{(k)}(\alpha) \text{ при } n \to \infty.$$

Доведення. Проведемо для $Q^-_{k;n}(\alpha)$. Для $Q^+_{k;n}(\alpha)$ доведення аналогічне. Розглянемо довільне $x > Q^{(k)}_\varepsilon(\alpha)$. За означенням $Q^-_{k;n}(\alpha)$, якщо $\tilde{F}^{(k)}_n(x) > \alpha$, то $Q^-_{k;n}(\alpha) < x$. Отже

$$P\{Q^-_{k;n}(\alpha) < x\} \geqslant P\{\tilde{F}^{(k)}_n(x) > \alpha\}. \tag{4.12}$$

За Теоремою 4.3, $\forall\delta > 0$ $P\{\tilde{F}^{(k)}_n(x) > F^{(k)}_\varepsilon(x) - \delta\} \to 1$, при $n \to \infty$, а за умовою теореми, $F^{(k)}_\varepsilon(x) > \alpha$. Враховуючи (4.12), поклавши $\delta = F^{(k)}_\varepsilon(x) - \alpha$, маємо $P\{\tilde{F}^{(k)}_n > \alpha\} \to 1$. Отже,

$$P\{Q^-_{k;n}(\alpha) < x\} \to 1, \text{ при } x > Q^{(k)}_\varepsilon(\alpha). \tag{4.13}$$

Тепер розглянемо довільне $x < Q^{(k)}_\varepsilon(\alpha)$. Якщо $Q^-_{k;n}(\alpha) < x$, то

$$S_{x;n} = \sup_{z<x} \tilde{F}^{(k)}_n(x) > \alpha.$$

Покладемо $\delta = \alpha - F^{(k)}_\varepsilon(x)$. Позначимо $J_n = \sup_z |\tilde{F}^{(k)}_n(z) - F^{(k)}_\varepsilon(z)|$.

Тоді

$$S_{x;n} < \sup_{z<x} F^{(k)}_\varepsilon(x) + J_n.$$

$P\{|J_n| > \delta\} \to 0$, отже

$$P\{Q^-_{k;n}(\alpha) < x\} \leqslant P\{S_{x;n} > \alpha\} \leqslant P\{F^{(k)}_\varepsilon(x) + J_n > \alpha\} =$$

$$P\{J_n > \delta\} \to 0 \text{ при } n \to \infty. \tag{4.14}$$

З (4.13) і (4.14) випливає $Q^-_{k;n}(\alpha) \xrightarrow{\mathbb{P}} Q^{(k)}_\varepsilon(\alpha)$ при $n \to \infty$. □

4.3.1 Діаграми квантиль проти квантилю

Для того, щоб побудувати QQ-діаграму у моделі нелінійної регресійної суміші, потрібно вибрати кількість рівнів K, і визначити самі рівні. Наприклад, їх можна обрати наступним чином $\alpha_k = \frac{k}{K+1}$, $k = 1, ..., K$. Кожному рівню на QQ-діаграмі ставиться у відповідність пара з квантилів: теоретичного, та оціненого розподілу похибок регресії для відповідного компонента $\left(Q^{(k)}_\varepsilon(\alpha_i), Q^+_{k;n}(\alpha_i)\right)$ (або $Q^-_{k;n}$), $i = 1, ..., K$.

Якщо на такій діаграмі точки розташовані вздовж бісектриси першого координатного кута, то це свідчить на користь гіпотези, що $F^{(k)}_\varepsilon$ є розподілом похибок у $k-$му компоненті. Наявність значущих відхилень суперечить цій гіпотезі.

	компонент	
	1	2
$\mu^{(m)}$	0.0	1.0
$\Sigma^{(m)}$	2.0	2.0
$\sigma^{(m)}$	0.05	0.05
$\theta_0^{(m)}$	0.5	0.5
$\theta_1^{(m)}$	2	-1/3

	компонент	
	1	2
$\mu^{(m)}$	-2.5	2.5
$\Sigma^{(m)}$	2.0	2.0
$\theta_0^{(m)}$	0.5	-0.5
$\theta_1^{(m)}$	2	-2

Табл. 4.1. Таблиця параметрів моделювання експериментів 1 (ліворуч), 3 (праворуч)

4.4 Алгоритмічні питання та моделювання

Було проведено серію із $H = 1000$ експериментів для різних розмірів вибірки двокомпонентної суміші. Цими експериментами перевірялася консистентність оцінки $\hat{\sigma}_k^2(\hat{\theta}_n^{(k)})$ для нелінійної функції $g(x, \gamma)$ вигляду (4.2). Для цього було пораховано середнє абсолютне зміщення справжніх параметрів від їх оцінок. Додатково рахувалася і дисперсія отриманих оцінок. В наступних трьох експериментах похибка регресії мала нормальний розподіл з різною дисперсією. Квадратами на рисунках позначені статистики оцінок першого компонента, кружечками - другого.

Вибірки генерувалися наступним чином: номери компонент $\kappa_i = \kappa(O_i), i = 1, ..., n$ обиралися відповідно до розподілів $(p_{j;n}^1, ..., p_{j;n}^M)_{j=1}^n$, де $p_{j;n}^k$ згенеровані процедурою (2.26). Характеристики, які спостерігаються для m-го компонента суміші $\xi_i^{(m)} = (Y_i^{(m)}, X_i^{(m)})$ генерувалися наступним чином:

$$Y_i^{(m)} = g(X_i^m, \theta^{(m)}) + \varepsilon_i^{(m)};\ X_i^{(m)} \simeq N(\mu^{(m)}, \Sigma^{(m)}),$$

де

$$g(X, \gamma) = \frac{1}{1 + e^{-\sum_{d=1}^D X_d \gamma_d}} = \sigma(\langle X, \gamma \rangle),$$

і розподіл похибки $\varepsilon_i^{(m)}$ залежить від номера експерименту.

Параметри розподілів та регресії, що описані у таблиці 4.1 (ліворуч) і будуть використані за замовчуванням.

Експеримент 1

Значення параметрів для цього експерименту зазначено у Таблиці 4.1 (ліворуч). Похибки компонент ε^k мали нормальний розподіл з параметрами: $\varepsilon_{1,i}^{(m)} \simeq N(0, 0.05^2),\ m = 1, ..., M$.

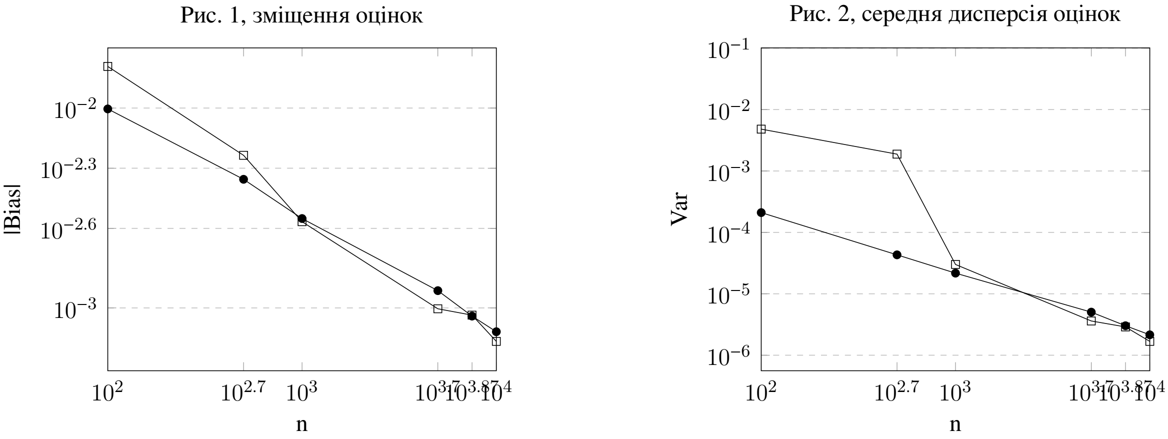


Рис. 4.1. Експеримент оцінювання $\hat{\sigma}_k(\hat{\theta}_n^{(k)})$ 1: 1 - зміщення оцінок, 2 - середня дисперсія оцінок

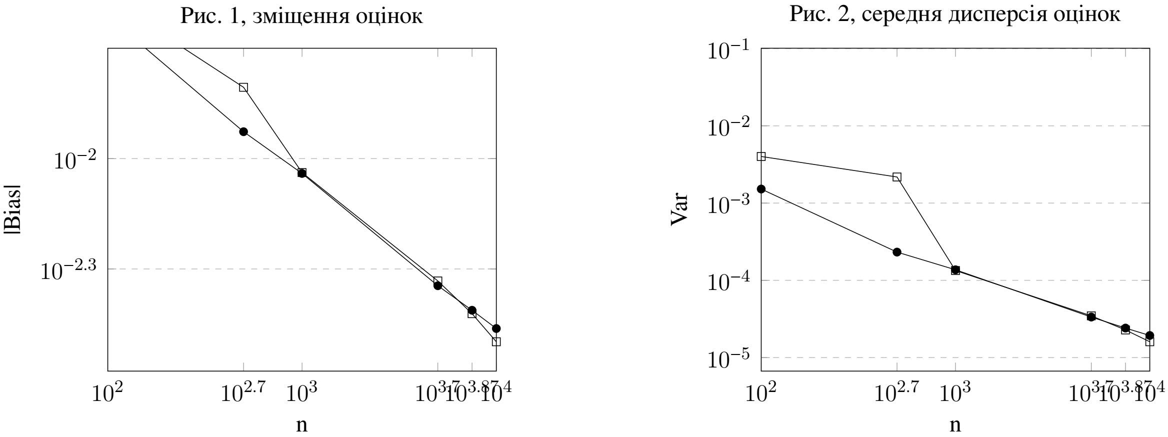


Рис. 4.2. Експеримент оцінювання $\hat{\sigma}_k(\hat{\theta}_n^{(k)})$ 2: 1 - зміщення оцінок, 2 - середня дисперсія оцінок

Цей експеримент описується моделлю “повної роздільності”, коли візуально спостереження можуть бути розділені на дві групи відповідно до різних компонент суміші. Результати моделювання наведено на рисунку 4.1.

Експеримент 2

Для того, щоб подивитися вплив розкиду похибок регресії на якість методу, параметр дисперсії похибки було збільшено $\varepsilon_{1,i}^{(m)} \simeq N(0, 0.25^2)$, $m = 1, ..., M$. Всі інші параметри було залишено як є. Результати моделювання наведено на рисунку 4.2.

Експеримент 3

У даному експерименті параметри регресії підібрані таким чином, щоб залишки регресії об’єктів, які належать іншим компонентам, впливали на значення функціо-

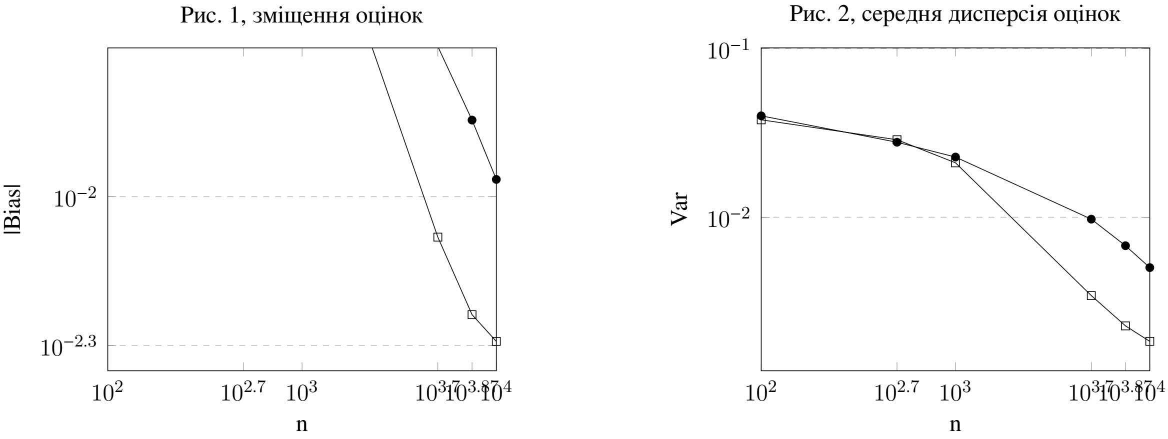


Рис. 4.3. Експеримент оцінювання $\hat{\sigma}_k(\hat{\theta}_n^{(k)})$ 3: 1 - зміщення оцінок, 2 - середня дисперсія оцінок

налу $S_n^{(k)}$ сильніше за навантаження $a_{i:n}^k$. Тобто, ці залишки регресії відіграватимуть роль викидів. Параметри похибок регресії ті ж самі, що і у експерименті 2. Дані параметри підібрані таким чином, щоб продемонструвати роботу моделі у найгіршому випадку. У таблиці 4.1 (праворуч) наведені параметри розподілів і регресії цього експерименту. Розподіли похибок аналогічні розподілу похибок у другому експерименті. Діаграма розсіювання пар схожа на діаграму експерименту 2.

Результати моделювання наведено на рисунку 4.3.

Висновки моделювання Результати усіх трьох експериментів показують консистентність оцінок дисперсії похибки регресії для різних компонент. У найскладнішому випадку видно, що і зміщення і дисперсії зменшуються зі зростанням об'єму вибірки, але значення статистик явно вищі за аналогічні статистики перших двох експериментів.

4.5 Висновки

У даному розділі було розглянуто оцінки дисперсії залишків регресій, та оцінки квантилів функції розподілу залишків для моделі суміші зі змінними концентраціями. Для цих оцінок було отримано умови їх консистентності.

Для оцінок дисперсій залишків було проведено серію імітаційних моделювань, щоб перевірити отримані результати.

Розділ 5

Застосування до соціологічних даних

У цьому розділі описано застосування регресійної моделі суміші зі змінними концентраціями до соціологічних даних. Регресійний аналіз проведено на прикладі об'єднаних даних, що створені на базі: результатів зовнішнього незалежного оцінювання (ЗНО) за 2016 рік [49], та результатами виборів до Верховної ради у 2014-му році. У статті [49] описано результати регресійного аналізу у припущенні про лінійність функції залежності. У статті [67] наведено результати нелінійного регресійного аналізу цих даних. У статті [50] застосовано модель нелінійної суміші для опису соціологічних даних, та побудовані довірчі множини оцінок параметрів регресії з використанням оцінок матриць коваріації методом складаного ножа.

5.1 Опис соціологічних даних

Для демонстрації можливостей розробленої регресійної техніки, розглянемо оцінювання функцій регресії, та приклад побудови довірчих еліпсоїдів для параметрів моделей у статистичному аналізі залежності між шкільними успіхами абітурієнтів та політичними уподобаннями у їх дорослому середовищі. Аналіз базується на двох наборах даних. Перший набір складається із результатів Зовнішнього незалежного оцінювання в Україні 2016-го року - ЗНО-2016. Ці дані було узято з сайту Українського Центру Оцінювання Якості Освіти (https://zno.testportal.com.ua/stat/2016). ЗНО - це набір екзаменів, які проходять випускники старшої школи для вступу до вищих навчальних закладів. Дані ЗНО-2016 містять індивідуальні оцінки абітурієнтів із деякою додатковою інформацією, що включає назву регіону України, у якому розташована відвідувана абітурієнтом школа. Оцінки усіх іспитів варіюються у межах від 100 до 200 балів.

Ми розглянемо інформацію про оцінки із двох основних предметів: Українська мова та література (UKR) і Математика (MATH). ЗНО-2016 містить дані цих оцінок для близько 246 000 абітурієнтів. Очевидно, що оцінки UKR та MATH пов'язані. У загальному випадку припускаємо, що цю залежність можна описати регресійною моделлю:

$$UKR = g(MAT, \theta) + \varepsilon,$$

в якій параметри θ залежать від політичних уподобань дорослого оточення, у якому учні зростали. Скажемо, у сім'ї прихильників незалежності України, очікується більша зацікавленість до вивчення української мови, ніж у середовищі, у якому переважає скептичне ставлення до української незалежності. Оскільки дані ЗНО не містять інформації про політичні уподобання, для оцінки параметрів потрібно залучити додаткову інформацію. Для цього було обрано другий набір даних із офіційними результатами Українських парламентських виборів 2014 року. Він дає змогу отримати приблизні пропорції прихильників різних політичних течій у різних областях України.

29 політичних партій та блоків приймали участь у виборах. Виборці також мали змогу голосувати проти усіх або не приймати участі у голосуванні. Ми розділили всіх виборців на три компонента:

(1) Ті виборці, які голосували за партії, які створили правлячу коаліцію (БПП, Народний фронт, Батьківщина, Радикальна партія, Самопоміч). Цей компонент складається з людей із позитивним ставленням до про-європейського курсу української держави.
(2) Ті виборці, які голосували за: Опозиційний блок, голосували проти усіх, малі партії, що не пройшли 5%-й виборчий бар'єр. Ці виборці мають критичне ставлення до про-європейського курсу України, але вони беруть участь у політичному житті держави.
(3) Виборці, які не брали участі у парламентських виборах. Це ті, хто не розглядає Українську державу як власну або взагалі не цікавляться політикою.

Ми використали результати виборів для розрахунку пропорцій кожного компонента суміші у кожному регіоні України, в яких проводилося голосування. Ці пропорції були узяті за оцінки ймовірностей того, що студент із конкретного регіону, зростав у оточенні відповідного компонента. Тобто ми розглядали їх як ймовірності змішування. Такої інформації достатньо, щоб побудувати мінімаксні навантаження, і оцінити параметри регресії. Регресійних аналіз проведено для двох функцій регресії окремо: лінійної та логістичної.

Очевидно, що більш точні оцінки можна отримати, якщо ймовірності змішування абітурієнтів взяти із відповідних їм виборчих округів, а не областям. Проте у даних результатів ЗНО недостатньо даних щоб отримати зведення на рівні дільниць.

5.2 Модель лінійної регресійної суміші

У даній частині аналізу буде розглянуто одну з найпростіших моделей регресії, - лінійну:

$$UKR = \theta_0 + MATH \times \theta_1 + \varepsilon.$$

Результати іспитів - числа в інтервалі від 100 до 200, то вони були переведені в інтервал від 0 до 1 відніманням і діленням на 100. Це зроблено для того, щоб області значень відгуків лінійної та нелінійної функції були однаковими.

	1	2	3
θ_0	0.359	0.534	0.183
θ_1	0.638	0.410	0.509
σ_k^2	0.0298	0.0237	0.0096

Табл. 5.1. Параметри регресії всередині різних політичних компонент

Параметри регресії та дисперсія похибки наведені у таблиці 5.1.

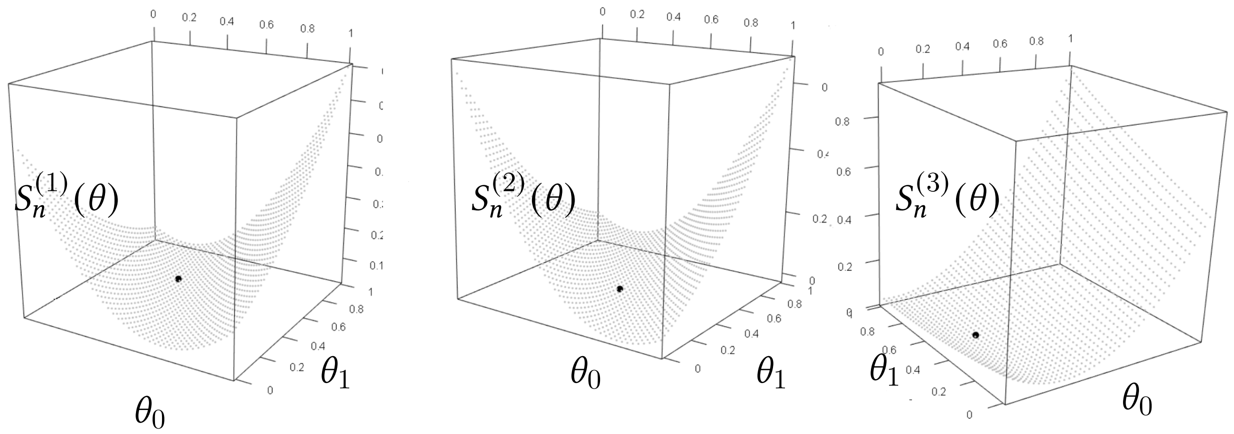
Вигляд навантаженого функціоналу МНК $S_n(\gamma)$ для лінійної функції регресії наведено на рисунку 5.1 для різних компонент. Чорною точкою на графіку позначено точки (значення оцінок), які мінімізують його значення.

Питання про те, чи непараметрична модель відповідає даним, знаходиться поза аналізом, як і питання про соціологічну інтерпретацію отриманих результатів.

5.3 Модель нелінійної регресійної суміші

Ми розглянули випадок, коли функція залежності - логістична:

$$g(X,\gamma) = \frac{1}{1+e^{-\gamma_0-\gamma_1 X_1}}.$$



(а) $S_n^{(1)}(\theta)$. (б) $S_n^{(2)}(\theta)$. (в) $S_n^{(3)}(\theta)$.

Рис. 5.1. Вигляд функціоналів МНК $S_n(\gamma)$ різних компонент для даних ЗНО

Тут, як і для лінійної моделі, відгуки були нормовані відніманням і діленням на 100.

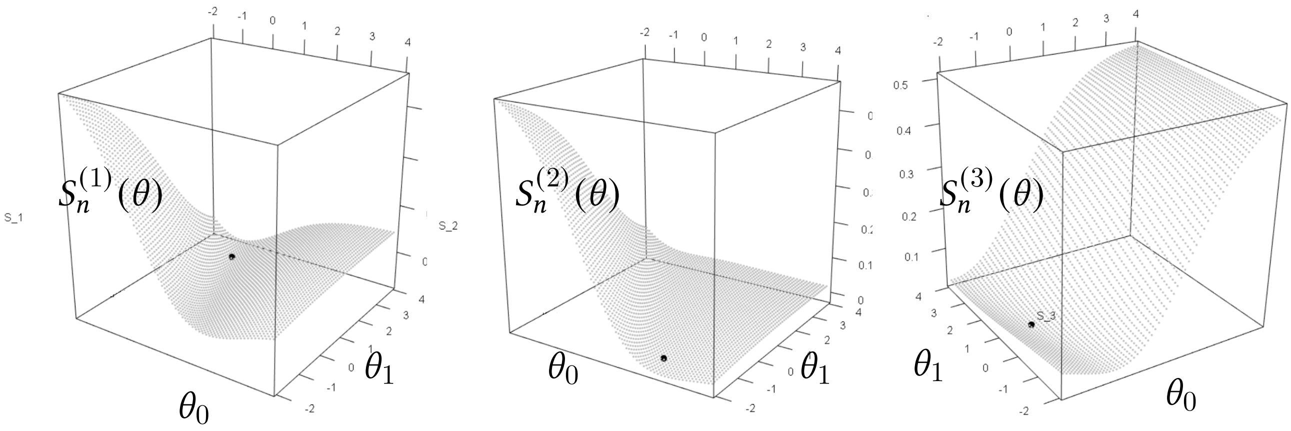
Оцінки параметрів регресії і оцінка дисперсії залишків наведені у Таблиці 3.

Компонент	1	2	3
$\theta_0^{(k)}$	-0.615	2.174	-1.151
$\theta_1^{(k)}$	2.970	-1.242	0.999
σ_k^2	0.0264	0.0252	0.0233

Табл. 5.2. Параметри регресії всередині різних політичних компонент

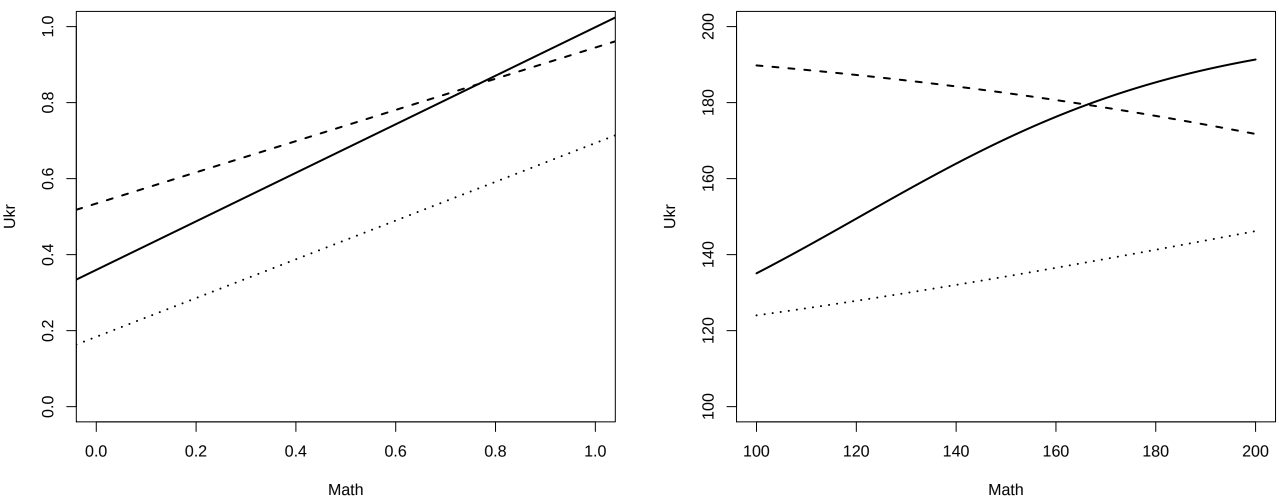
Вигляд навантаженого функціоналу МНК $S_n(\gamma)$ наведено на рисунку 5.2 для різних компонент. Чорною точкою на графіку позначено точки (значення оцінок), які мінімізують його значення.

На відміну від симульованих даних, при аналізі соціологічних даних, справжні параметри регресії невідомі й в такому випадку неможливо точно сказати якою буде справжня матриця коваріації з теореми 3.5. Її доведеться оцінювати іншими способами. Було побудовано дві оцінки матриці коваріації $\hat{V}_{plug-in}$, ${}^{JK}\hat{V}_n^{(k,k)}$. Оцінка $\hat{V}_{plug-in}$ отримана підставленням оцінених параметрів регресії у матриці із теореми 3.5, а оцінка ${}^{JK}\hat{V}_n^{(k,k)}$ порахована за означенням (3.10).



(а) $S_n^{(1)}(\theta)$. (б) $S_n^{(2)}(\theta)$. (в) $S_n^{(3)}(\theta)$.

Рис. 5.2. Вигляд функціоналів МНК $S_n(\gamma)$ різних компонент для даних ЗНО

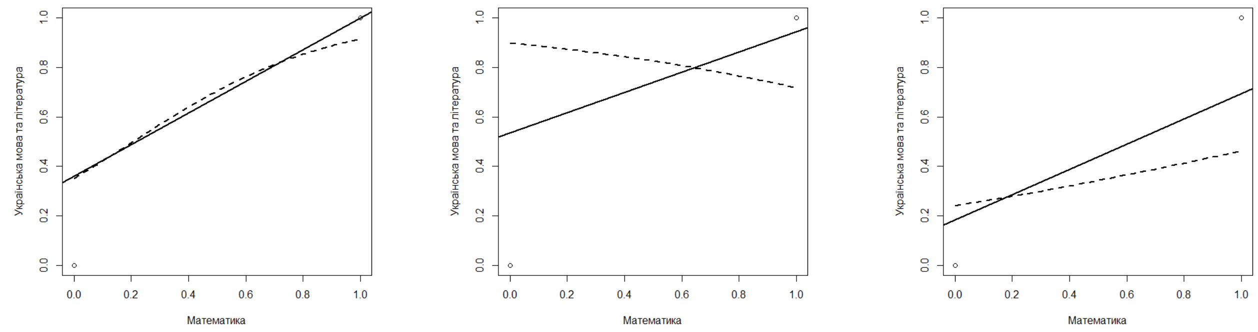


(а) Лінійна регресійна залежність (б) Нелінійна регресійна залежність

Рис. 5.3. Зображення кривих залежності різних компонент. Пряма - коаліція, розривна пряма - опозиція, пунктир - інші

5.4 Порівняння результатів лінійної та нелінійної моделі

Порівняння результатів почнемо із оцінок дисперсій. У таблиціях 5.1 та 5.2 приведені ці оцінки σ_k^2. Майже усі оцінки дисперсії залишків у нелінійній моделі схожі на оцінки у лінійній моделі. Лише значення оцінки для компонента “інші” значно менша у лінійній моделі.



(а) Коаліція (б) Опозиція (в) Інші

Рис. 5.4. Лінійні та нелінійна функції залежності окремо для різних компонент

Обидві частини рисунку 5.3 показують, що у різних випадках моделі (лінійна і нелінійна), нелінійні криві знаходять у тих самих областях, що і відповідні їм лінії. Бачимо, що для "опозиційного" компонента лінія нелінійної залежності спадає, на відміну від лінійної моделі.

	Компонент		
Модель	1	2	3
Лінійна	0.0264	-0.00121	0.01134
Нелінійна	0.0263	-0.0222	0.00940

Табл. 5.3. Значення функції $S_n^{(k)}(\hat{\theta}_n^{(k)})$

Три графіка на рисунку 5.4 показують, що лінії залежностей нелінійної моделі досить близькі до відповідних їм ліній у більш простої лінійній моделі. Таблиця 5.3 показує, що значення функції $S_n^{(k)}$ трошки менші у випадку нелінійної моделі. Це означає, що доречно використовувати саме її, проте різниця не надто велика. Потрібно зауважити, що від'ємні значення у таблиці не є помилкою, оскільки навантаження $a_{j;n}^k$ можуть бути від'ємними.

5.5 Квантиль проти квантиля

За параметрами із таблиць 5.1 і 5.2 та виправленими мінімаксними навантаженнями A_n були оцінені функції розподілу залишків і побудовані діаграми квантиль проти квантиля у припущенні що апріорний розподіл залишків має гаусів розподіл.

На обох графіках перші колонки - порівняння оцінки функції розподілу компонент (суцільна лінія) і нормального розподілу (пунктиром) із відповідною

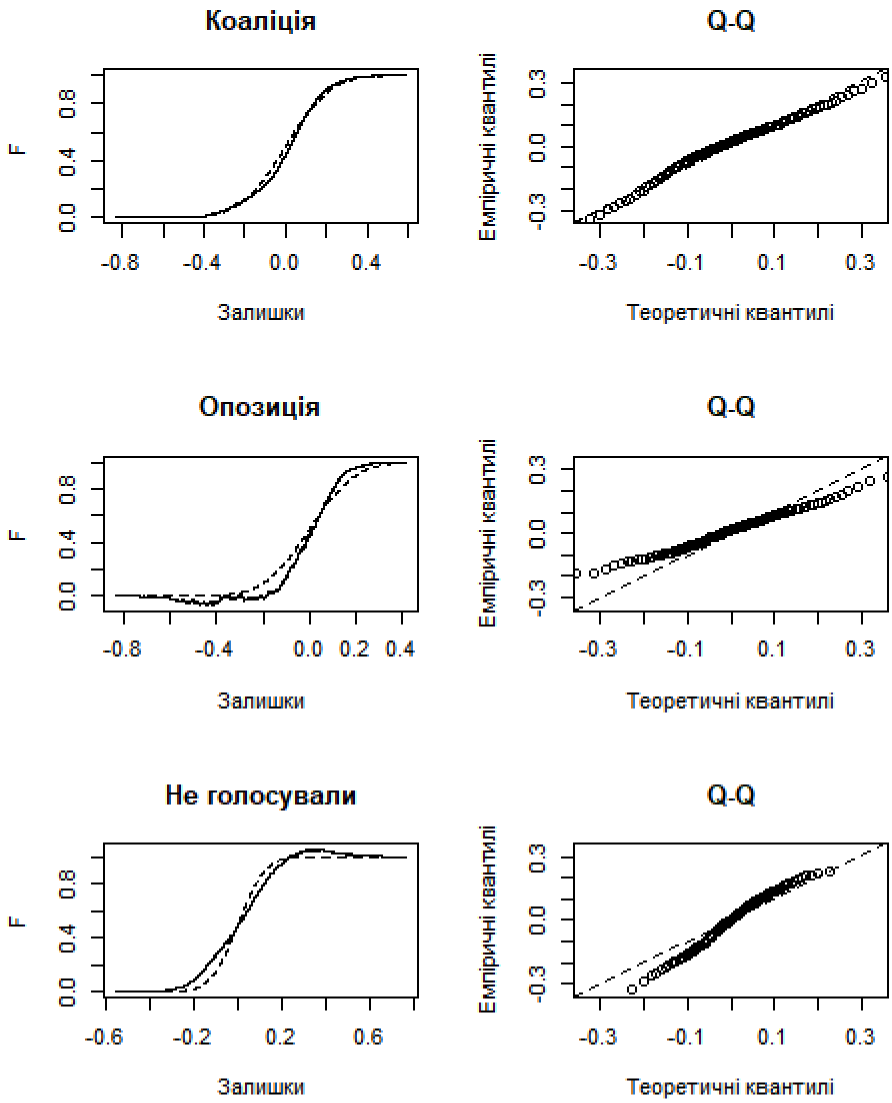


Рис. 5.5. Розподіл похибок для лінійної моделі. Оцінка функції розподілу залишків - ліва колонка, діаграма квантиль проти квантилю - права колонка

дисперсією σ_k^2 (значення оцінок, що отримані у лінійній та нелінійній моделях). У другій колонці зображені діаграми квантиль проти квантиля для оцінок розподілів і нормального розподілу.

Як видно з діаграм, квантилі гаусової функції розподілу з оціненою дисперсією схожі на квантилі оцінених розподілів для "Коаліційного" компонента. Це каже нам про те, що при припущеній логістичній залежності, залишки можуть бути розподілені нормально. Візуально, для "Опозиційного" компонента лінійна модель дає більш обережні передбачення, а для компонента "Інші", краще використовувати нелінійну модель.

Для впевненості слід розробити статистичний тест перевірки гіпотез про

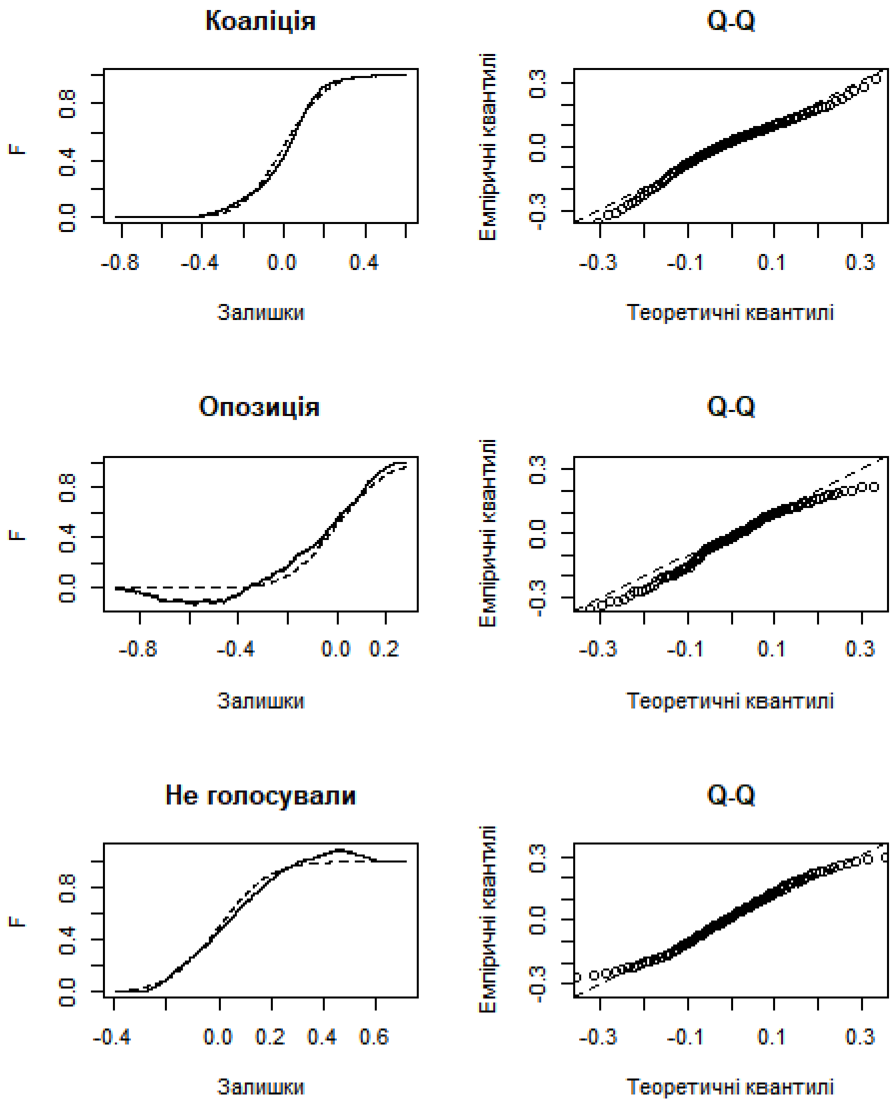


Рис. 5.6. Розподіл похибок для нелінійної моделі. Оцінка функції розподілу залишків - ліва колонка, діаграма квантиль проти квантилю - права колонка

рівність розподілів.

5.6 Довірчі еліпсоїди для параметрів

Лінійна модель За даними ЗНО-2016, та Українських парламентських виборів було отримано EM- та LS-еліпсоїди для вектору параметрів (θ_0, θ_1) лінійної регресії більше детально ці довірчі множини описано у розділі 3. Вони зображені на Рис. 4. Еліпсоїди побудовані із рівнем значущості $\alpha = 0.05/3 \approx 0.01667$, утворюють набір одночасних довірчих множин з рівнем значущості $\alpha_0 = 0.05$. Оскільки еліпсоїди не перетинаються в обох випадках, можемо зробити висновок, що коефіцієнти

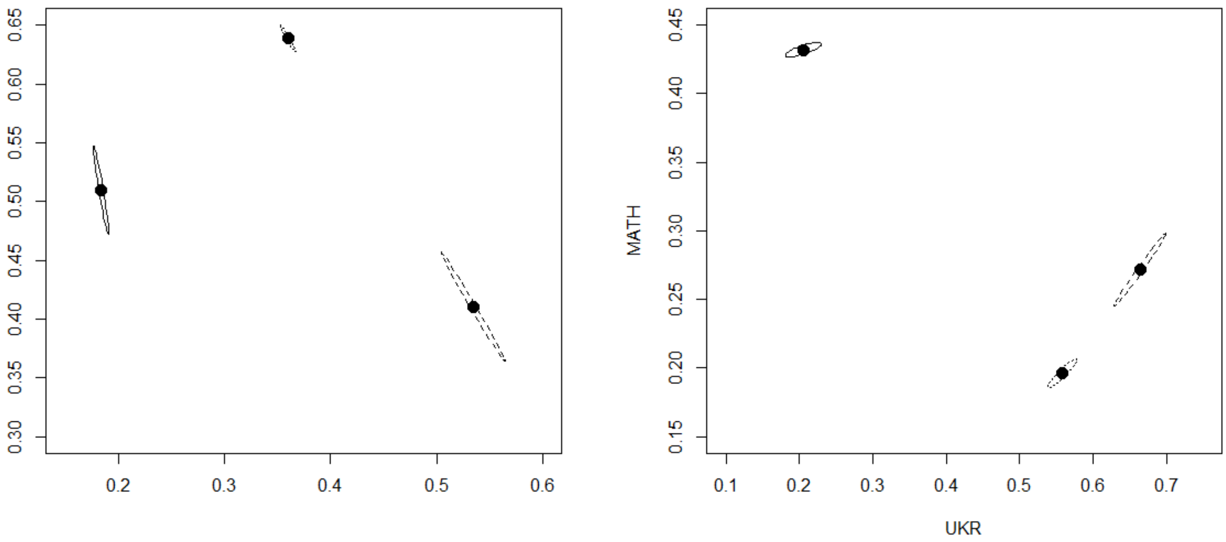


Рис. 5.7. LS(ліворуч) та EM (праворуч) довірчі еліпсоїди для даних ЗНО. Компоненти: (1) точкова лінія,(2) пунктир, (3) цільна лінія. θ_0 відкладено по горизонталі, θ_1 по вертикалі.

регресії (θ_0^i, θ_1^i), $i = 1, ..., 3$ істотно відрізняються для різних компонент.

Нелінійна модель. На рисунку 5.8 зображені довірчі еліпсоїди із центрами, які відповідають знайденим параметрам нелінійної регресії, що оцінені на основі соціологічних даних ЗНО-2016, та Українських парламентських виборів 2014 року. Відрізняються ці рисунки еліпсоїдами, матриці коваріацій яких оцінено різними способами. Перший графік рисунку 5.8 зображає довірчі еліпси, матриці коваріацій яких побудовані на основі формул із теореми 3.4 ($\hat{V}_{plug-in}$). Правий графік зображає довірчі еліпсоїди, коваріації яких були пораховані методом складаного ножа (${}^{JK}\hat{V}_n^{(k,k)}$). Візуально еліпсоїди на обох графіках рисунку 5.8 відрізняються шириною та об'ємами, проте кут нахилу однаковий у відповідних еліпсоїдів. Об'єми довірчих множин методом $\hat{V}_{plug-in}$ менші, ніж об'єми довірчих множин за оціненими матрицями ${}^{JK}\hat{V}_n^{(k,k)}$ методом складаного ножа. В цих двох випадках довірчі множини не перетинаються між собою. Даний факт є аргументом того, що є сенс розглядати ці соціологічні дані як суміш із трьох різних компонент, і їх число краще не зменшувати.

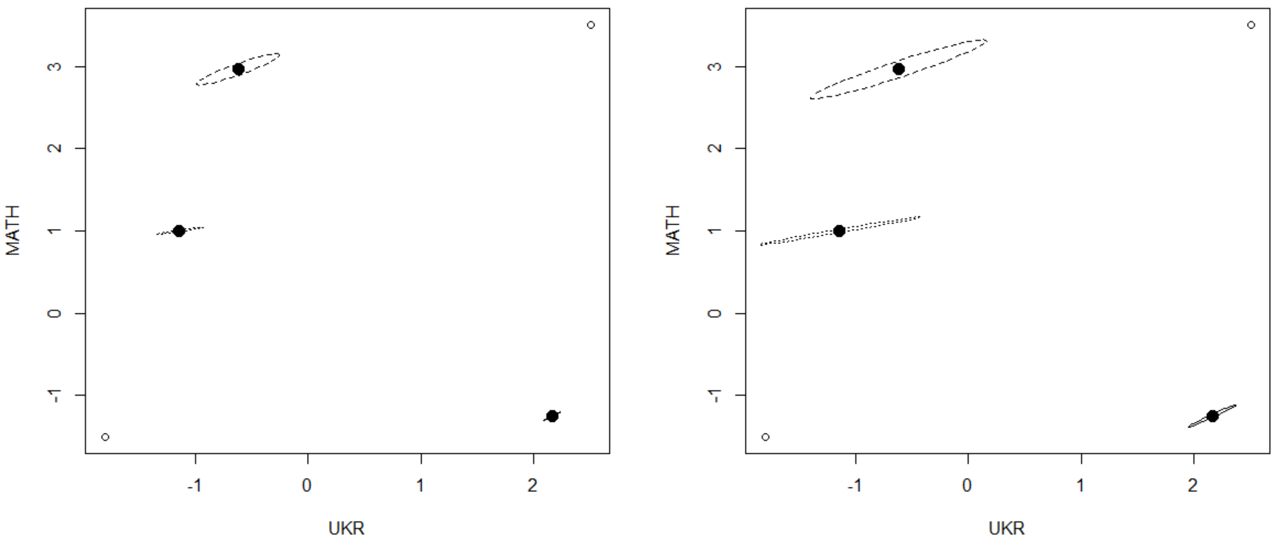


Рис. 5.8. Довірчі еліпсоїди методами $\hat{V}_{plug-in}$, ${}^{JK}\hat{V}_n^{(k,k)}$. Еліпсоїд першого компонента - розривна лінія, для другого - неперервна лінія, третій - пунктир крапками

5.7 Висновки

Розробленні методи регресійного аналізу сумішей зі змінними концентраціями було застосовано до прикладної задачі, а саме вивчення інтересу до української мови та літератури в залежності від інтересу до математики абітурієнтів в умовах, коли вибірка цих абітурієнтів є сумішшю осіб, які росли у середовищах із різними політичними настроями. Цей інтерес між дисциплінами було описано лінійною та нелінійною функцією. З вигляду діаграм квантиль проти квантилю можна зробити висновок, що принаймні для двох компонент (Коаліція, та “Не голосували”) похибки нелінійної регресії мають нормальний розподіл. Додатково було отримано довірчі еліпсоїди оцінок залежностей декількома модифікаціями оцінок коваріацій. Той факт, що довірчі інтервали різних компонент не перетинаються, каже що різну функцію залежності результатів ЗНО між української мовою та математикою всередині усіх компонент суміші. Інтерпретація отриманих результатів лежить поза цим дослідженням.

Висновки

У дисертаційній роботі одержано наступні результати.

1 Для лінійної моделі регресійної суміші доведено консистентність оцінок методу підстановки та методу складаного ножа для матриць коваріацій оцінок параметрів лінійної регресії. Ці оцінки матриць застосовано для побудови довірчих множин.

2 Доведено консистентність, та асимптотичну нормальність оцінок коефіцієнтів у нелінійних моделях регресійних сумішей.

3 Наведено умови збіжності ймовірності існування розв'язку оціночних рівнянь.

4 Для асимптотичної коваріаційної матриці оцінок регресійних параметрів наведено умови невиродженості цих матриць та консистентність оцінки за сендвіч-формулою.

5 Для оцінки коваріаційної матриці методом складаного ножа для оцінок регресійних параметрів доведено теорему про її консистентність. Ці оцінки матриць застосовано для побудови довірчих множин.

6 Доведено теореми про консистентність оцінок дисперсій залишків різних регресійних моделей суміші. Також, отримано умови консистентності непараметричної оцінки розподілу залишків регресії для різних компонент, та консистентність оцінок квантилів розподілу залишків.

7 Розроблені алгоритми застосовано до аналізу соціологічних даних.

Бібліографія

1. Іванов О.В., Орловський І.В. Конзистентність М-оцінок у нелінійних моделях регресії з неперервним часом // Наукові вісті НТУУ “КПІ”. — 2005. — № 4. — С. 140—147.
2. Іванов О.В., Орловський І. В. Асимптотична нормальність М-оцінок у класичній нелінійній моделі регресії // Український математичний журнал. — 2008. — Т. 60, № 11. — С. 1470–1488. — Бібліогр.: 44 назв. — укр.
3. Іванов О.В., Орловський І. В. Про єдиність М-оцінок параметрів нелінійних моделей регресії // Наукові вісті НТУУ “КПІ”: науково-технічний журнал, № 4(66), (2009)
4. Майборода Р.Є. "Оцінка середніх положень та концентрацій по спостереженнях двокомпонентної суміші симетричних розподілів". Теорія ймов. та матем. статистика, № 78 , - 2008
5. Майборода Р. Є., Сугакова О. В. Оцінювання та класифікація за спостереженнями із суміші // Київський Університет, (2009)
6. Майборода Р. Є., Сугакова О. В. Тести для гіпотез про квантилі розподілів компонентів суміші // Теорія ймовірностей та математична статистика, Vol.101, Iss. pp. 157 - 168, - 2019
7. Масюк С. В. , Кукуш О. Г. , Шкляр С. В. , Чепурний М. І. , Ліхтарьов І. А. Моделі регресії з похибками вимірювання та їх застосування до оцінювання радіаційних ризиків // ДІА, 288 p. - 2015
8. Autin F., Pouet Ch. Test on the components of mixture densities // Statistics $ Risk Modeling 28(4), 389–410 (2011).
9. Bilancia M., Pollice A. “Bayesian estimation of finite mixtures of gaussian mixtures” // February Metron LVII(1-2):181-208, 1999
10. Benaglia T., Chauveau D., Hunter D. R. An R Package for Analyzing Finite MixtureModels // Journal of Statistical Software 32(6), 1–29 (2009).

11. Bishop Y. M. , Fienberg S. E., Holland P. W. Discrete Multivariate Analysis Theory and Practice // Springer, 2007
12. Bilmes J. A. A gentle tutorial of the EM algorithm and its Application to Parameter Estimation for Gaussian mixture and Hidden Markov Models // U.C Berkeley 1998
13. Blei D.M., Ng A.Y., Jordan M.I.: Latent dirichlet allocation. Journal of Machine Learning, Research 3 (2003) 993–1022
14. Bordes L., Delmas C., Vandekerkhove P. Semiparametric estimation of a two-component mixture model when a component is known. Scandinavian Journal of Statistics, 33, p. 733-752 (2006)
15. Borovkov A. A. Mathematical statistics // Gordon and Breach Science Publishers, Amsterdam (1998)
16. Breiman L. Stacked Regressions // Machine Learning, 24, 49-64 (1996)
17. Butucea C. , Vandekerkhove P. Semiparametric mixtures of symmetric distributions. Scandinavian Journal of Statistics, 41, p. 227-239 (2014)
18. Deming, W. E. Statistical adjustment of data // Wiley, NY (Dover Publications edition, 1985), (1943)
19. DeSarbo W., Wedel M. Mixture Regression Models // Article in Journal of Classification · February 1988
20. DeSarbo W., Cron W. A Maximum Likelihood Methodology for Clusterwise Linear Regression // Applied Latent Class Analysis in Honor of C. Clogg (pp.366- 382) Chapter: 13, Cambridge University Press, (2002)
21. DeVeaux, R.D Parameter estimation for a mixture of linear regressions, Technical Report No. 247, and unpublished Ph.D. dissertation, Statistics Department, Stanford University, 1986
22. DeVeaux, R.D. Mixtures of linear regressions, Comput. Statist. Data Anal. 8 (1989), pp. 227–245
23. Dempster A. P., Laird N. M., Rubin D. B Maximum-likelihood from incomplete data via the em algorithm // J. Royal Statist. Soc. Ser. B., 39, 1977.
24. Efron B. The Jackknife, the Bootstrap, and Other Resampling Plans // Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 19103. 1982
25. Faria S., Soromenho G. Fitting mixtures of linear regressions // Journal of Statistical Computation and Simulation 80, No 2, 201:225 (2010)
26. Fazekas I., Kukush A. Infill asymptotics inside increasing domain for the least

squares estimator in linear models . Statistical Inference for Stochastic Processes, 2000, 3, N3, 199—223.

27. Golub G. H., van Loan C. F. Matrix Computations // The Johns Hopkins University Press, Baltimore, 3rd edition, 1996.
28. Grun B., Leisch F. Fitting finite mixtures of linear regression models with varying & fixed effects in R // In Alfredo Rizzi and Maurizio Vichi, editors, Compstat 2006 - Proceedings in Computational Statistics, pages 853-860. Physica Verlag, Heidelberg, Germany, 2006.
29. Grun B., Leisch F. Finite Mixtures of Generalized Linear Regression Models // Technical Report Number 013, Department of Statistics, University of Munich, 2007
30. Ivanov A. V. Leonenko N.N.: Statistical Analysis of Random Fields. Kluwer Academic Publishers, Dordecht/Boston/London (1989).
31. Ivanov A. V. Orlovskyi I. V. Large deviations of regression parameter estimator in continuous-time models with sub-Gaussian noise // Modern Stochastics: Theory and Applications 5 (2) (2018) 191–206
32. Hansena B., Racine J. Jackknife model averaging / Journal of Econometrics Volume 167, Issue 1, March 2012, Pages 38-46
33. Hall P., Zhou X. H. Nonparametric estimation of component distributions in a multivariate mixture. Annals of Statistics, 31:201–224 (2003)
34. Hoffman M., Blei D., Bach F. Online Learning for Latent Dirichlet Allocation // Advances in Neural Information Processing Systems 23 (NIPS 2010)
35. Hoffman M., Blei D., Wang C., Paisley J. “Stochastic Variational Inference” // Journal of Machine Learning Research 14 (2013) 1303-1347
36. Hunter D. R., Wang S., Hettmansperger T. P. “Inference for mixtures of symmetric distributions.” Ann. Statist. 35 (1) 224 - 251, February 2007
37. Jordan M. and Jacobs. R. Hierarchical mixtures of experts and the em algorithm // Neural Computation, 6:181–214, 199
38. Jordan M. and Xu L. Convergence results for the em approach to mixtures of experts, architectures. Neural Networks, 8:1409–1431, 1996
39. Kariya T., Takeaki H. Generalized Least Squares. // Chichester Wiley 2004
40. Kariya T., Toyooka Y. Nonlinear versions of the Gauss-Markov theorem and GLSE. In Multivariate Analysis VI (ed. Krishnaiah PR), pp. 345–354. Elsevier Science Publishers B.V. (North Holland). 1985

41. Kariya T., Kurata H. A Maximal Extension of the Gauss–Markov Theorem and Its Nonlinear Version. Journal of Multivariate Analysis 83, 37–55 (2002)
42. Kenney J. F., Keeping E. S. Linear Regression and Correlation // Ch. 15 in Mathematics of Statistics, Pt. 1, 3rd ed. Princeton, NJ: Van Nostrand, pp. 252–285 (1962)
43. Kessy S., Sherris M., Villegas A., Ziveyi J. Mortality Forecasting Using Stacked Regression Ensembles // Article in SSRN Electronic Journal · January 2021
44. Kukush A. and Van Huffel S., Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model AX=B. Metrika 59 (2004)
45. Kwon J., Caramanis C. EM Converges for a Mixture of Many Linear Regressions, arXiv:1905.12106, 2019
46. Lenk P., DeSarbo W. “Bayesian inference for finite mixtures of generalized linear models with random effects.” Psychometrika 65, no. 1 (2000): 93–119.
47. Liubashenko D., Maiboroda R. Linear regression by observatons from mixture with varying concentrations // Modern Stochastics: Theory and Applications, 2 ,No 4, 343 – 353, (2015)
48. Maiboroda R., Kubaichuk O. Asymptotic normality of improved weighted empirical distribution functions // Theor. Probab. Math. Stat. 69, 95–102 (2004).
49. Maiboroda R. E., Miroshnichenko V. O. Confidence ellipsoids for regression coefficients by observations from a mixture // Modern Stochastics: Theory and Applications, Vol.5, Iss.2 pp. 225 - 245, - 2018
50. Maiboroda R., Miroshnichenko V., Sugakova O. Jackknife for nonlinear estimating equations.- Modern Stochastics: Theory and Applications, 2022 https://doi.org/10.15559/22-VMSTA208
51. Maiboroda R. E., Sugakova O. V. Statistics of mixtures with varying concentrations with application to DNA microarray data analysis // Journal of nonparametric statistics. 24 , No 1 201–205 (2012)
52. Maiboroda R. E., Sugakova O. V. Jackknife covariance matrix estimation for observations from mixture // Modern Stochastics: Theory and Applications, 2019
53. Maiboroda R. E., Sugakova O. V. Sampling bias correction in the model of mixtures with varying concentrations // Methodology and Computing in Applied Probability. 17(1 (2015)

54. Maiboroda R. E., Navara H., Sugakova O. V. Orthogonal regression for observations from mixtures // Teoriya imovirnostey ta matematychna statystyka, Vol.99, Iss. pp. 152 - 167, - 2018
55. Maiboroda R. E., Sugakova, O. V. Jackknife covariance matrix estimation for observations from mixture // Modern Stochastics: Theory and Applications, 2019
56. Maiboroda R. E., Sugakova O. V. “Generalized estimating equations for symmetric distributions observed with admixtur”. Communications in Statistics -Theory and Methods, V. 40: 1, pp. 96 - 116, - 2011
57. Maiboroda R. E., Sugakova O. V., Doronin A. V. Generalized estimating equations for mixtures with varying concentrations // The Canadian Journal of Statistics 41(2), 217–236 (2013)
58. Keribin, C. Consistent estimation of the order of mixture models. Sankhya A, 62:49–65, 2000
59. Masiuk S., Kukush A., Shklyar S., Chepurny M., Likhtarov I. (ed.): Radiation Risk Estimation: Based on Measurement Error Models (2nd ed) // de Gruyter series in Mathematics and Life
60. Mclachlan, G. J., Basford K. Mixture models. Inference and Application to Clustering // Marcel Dekker inc., New York and Basel, 1987
61. McLachlan, G. J. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. Journal of the Royal Statistical Society Series C, 36:318–324., 1987
62. McLachlan, G.J., Peel, D. Robust cluster analysis via mixtures of multivariate t-distributions, SSPR /SPR 1998: Advances in Pattern Recognition pp 658–666 (1998)
63. Mclachlan, G. J., Peel D. Finite mixture models // Wiley-Interscience (2000)
64. McLachlan G., Thriyambakam K. The EM Algorithm and Extensions, 2nd Edition // Wiley, (2008)
65. McIntosh A. I. The Jackknife Estimation Method // Boston University Retrieved 2016-04-30.: p. 3.
66. Miroshnychenko V. O. Generalized least squares estimates for mixture of nonlinear regressions // Bulletin of Taras Shevchenko National University of Kyiv; Series: Physics Mathematics, 2018, 5
67. Miroshnychenko V. O. Residual analysis in regression mixture model // Bulletin of

Taras Shevchenko National University of Kyiv; Series: Physics Mathematics, 2019, 3

68. Miroshnychenko V. O., Maiboroda R. E. Asymptotic normality of modified LS estimator for mixture of nonlinear regressions // Modern Stochastics: Theory and Applications, Vol.7, Iss.7 pp. 435 - 448, - 2020
69. Miroshnichenko V. O Fast calculations of Jackknife covariance matrix estimator // Bulletin of Taras Shevchenko National University of Kyiv, pp. 27-36, №1, 2021
70. Hall P., Titterington D. M. The Use of Uncategorized Data to Improve the Performance of a Nonparametric Estimator of a Mixture Density // Journal of the Royal Statistical Society. Series B (Methodological), Vol. 47, No. 1 (1985), pp. 155-163
71. Quenouille M. H. Notes on bias in estimation // Biometrika, 43, 353-60, (1956).
72. Redner R.A., Walker H.F. Mixture densities, maximum likelihood and the EM algorithm. SIAM Review 26, 195–239. (1984)
73. Richardson S. Green P. J. On Bayesian analysis of mixtures with an unknown number of components (with discussion). Journal of the Royal Statistical Society Series B, 59:731–792., 197
74. Seber G. A. F , Lee A. J. Linear Regression Analysys // Wiley (2003)
75. Seber G. A. F., Wild C. J. Nonlinear Regression // Wiley, (2003)
76. Shao J. Mathematical Statistics // Springer, 2007.
77. Shao J. One-step jackknife for M-estimators computed using Newton's method // Ann. Inst. Statist. Math. Vol. 44, No. 4, 687-701 (1992)
78. Shao J. Consistency of least-squares estimator and its jackknife variance estimator in nonlinear models // Canadian Joumal of Statistics Vol. 20. No. 4. 1992, Pages 415-428
79. Shao J., Tu D. Jackknife and bootstrap // Springer, 1995.
80. Shao J., Tu D. The Jackknife and Bootstrap // Springer, NY, (1995)
81. Shao J, Wu C. F. J. A General Theory for Jackknife Variance Estimation // The Annals of Statistics, Vol. 17, No. 3 (Sep., 1989), pp. 1176-1197.
82. Shao, J., An asymptotic theory for linear model selection. Statistica Sinica 7, 221–264 1997.
83. Shklyar S. Consistency of the total least squares estimator in the linear errors-in-variables regression // Modern Stochastics: Theory and Applications, Vol.5, Iss.3

pp. 247–295, - 2018

84. Schlattmann P. Medical Applications of Finite Mixture Models. Berlin: Springer, 2009
85. Stephens M. Bayesian Methods for Mixtures of Normal Distributions // Michaelmas Term (1997).
86. Stephens M. Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. Annals of Statistics, 28:40–74, 2000
87. Tatarinova T., Schumitzky A. Nonlinear Mixture Models : A Bayesian Approach // Imperial College Press (2015).
88. Titterington, D. M., Smith, A. F., Makov, U. E. Analysis of Finite Mixture Distributions. Wiley, New York
89. Titterington D. M., Smith A. F, Makov U. E. Statistical analysis of finite mixture distributions // Wiley series in probability and mathematical statistics. Applied probability and statistics, New York, 1985. –364p.
90. Tukey J. W., Bias and confidence in not quite large samples // The Annals of Mathematical Statistics. 29 (2): 614 (1958)
91. Wolpert D. Stacked Generalization // Neural Networks. — 1992. — T. 5, vol. 2.
92. Yao W., Wei Y., Yu C. Robust mixture regression using the t-distribution. Computational Statistics and Data Analysis, 71:116–127. 2014

Додаток

Список публікацій здобувача за темою дисертації

Публікації, в яких опубліковано основні наукові результати дисертації

1 Maiboroda R. E., Miroshnichenko V. O. Confidence ellipsoids for regression coefficients by observations from a mixture. Modern Stochastics: Theory and Applications. 2018. Vol.5, Iss.2, pp. 225-245
2 Miroshnichenko V. O. Generalized least squares estimates for mixture of nonlinear regressions. Bulletin of Taras Shevchenko National University of Kyiv. 2018. Vol. 3, pp. 25-29
3 Miroshnichenko V. O. Residual analysis in regression mixture model. Bulletin of Taras Shevchenko National University of Kyiv. 2019. Vol. 3, pp. 8-16
4 Miroshnychenko V. O., Maiboroda R. E. Asymptotic normality of modified LS estimator for mixture of nonlinear regressions. Modern Stochastics: Theory and Applications. 2020. Vol.7 Iss.4, pp. 435–448
5 Miroshnychenko V. O. Fast calculations of Jackknife covariance matrix estimator. Bulletin of Taras Shevchenko National University of Kyiv. 2021. Vol. 1, pp. 27-36
6 Maiboroda R., Miroshnichenko V., Sugakova O. Jackknife for nonlinear estimating equations. Modern Stochastics: Theory and Applications. 2022. https://doi.org/10.15559/22-VMSTA208

Публікації, які засвідчують апробацію матеріалів дисертації

1 Miroshnychenko V.O. EM алгоритм для аналізу регресійної суміші. XV International Scientific-Practical Conference "Shevchenkivska Vesna - 2018". Kyiv, Ukraine, 2017.

2 Maiboroda R.E., Miroshnychenko V.O. Confidence Sets For Regression Coefficients By Observations From a Mixture. Modern Stochastics: Theory And Applications IV. Kyiv, Ukraine, 2018

3 Miroshnychenko V.O. Asymptotics of generalized least squares estimates for mixture of nonlinear regressions. International Conference of Young Mathematicians. Kyiv, Ukraine, 2019

4 Miroshnychenko V.O. Оцінки параметрів нелінійної регресії за спостереженнями з суміші. XVII International Scientific-Practical Conference "Shevchenkivska Vesna - 2019". Kyiv, Ukraine, 2019

5 Miroshnychenko V.O. Асимптотична нормальність коефіцієнтів у суміші регресій. XVIII International Scientific-Practical Conference "Shevchenkivska Vesna - 2020". Kyiv, Ukraine, 2020

6 Miroshnychenko V.O., Maiboroda R.E. Asymptotic normality of modified LS estimator for mixture of non-linear regressions. Scientific Conference "Actual Problems Of Stochastic Analysis" dedicated to the 80-th anniversary of the birth of academician Sh.K.Formanov. Tashkent, Uzbekistan, 2021

7 Maiboroda R.E., Miroshnychenko V.O. Consistency of jackknife covariance estimator for mixture of nonlinear regressions. International conference "Modern stochastics: theory and applications V". Kyiv, Ukraine, June 1-4, 2021

8 Семінар з математичної статистики механіко-математичного факультету Київського Національного університету імені Тараса Шевченка під керівництвом проф. О.Г. Кукуша та проф. Р.Є. Майбороди. Київ, Україна, 2021

9 Maiboroda R.E., Miroshnychenko V.O. Метод складаного ножа для суміші нелінійних регресій. XX International Scientific-Practical Conference "Shevchenkivska Vesna - 2022". Kyiv, Ukraine 2022

10 Miroshnychenko V.O. Jackknife for nonlinear estimating equations. The 4-th International Conference on Statistics: Theory and Applications, Virtual conference. Prague, Czech Republic, 2022
https://avestia.com/ICSTA2022_ Proceedings/files/paper/ICSTA_ 149.pdf

Відомості про апробацію результатів дисертації

Конференції

1 XV International Scientific-Practical Conference “Shevchenkivska Vesna - 2018”, Kyiv, Ukraine (2017)
2 Modern Stochastics: Theory And Applications. IV, Kyiv, Ukraine (2018)
3 International Conference of Young Mathematicians, Kyiv, Ukraine (2019)
4 XVII International Scientific-Practical Conference “Shevchenkivska Vesna - 2019”, Kyiv, Ukraine (2019)
5 XVIII International Scientific-Practical Conference “Shevchenkivska Vesna - 2020”, Kyiv, Ukraine (2020)
6 Scientific Conference “Actual Problems Of Stochastic Analysis” dedicated to the 80-th anniversary of the birth of academician Sh.K.Formanov Tashkent, Uzbekistan (2021)
7 International conference “Modern stochastics: theory and applications V”, Kyiv, Ukraine, June 1-4 (2021)
8 XX International Scientific-Practical Conference “Shevchenkivska Vesna - 2022”, Kyiv, Ukraine (2022)
9 The 4-th International Conference on Statistics: Theory and Applications, Virtual conference (2022), Prague, Czech Republic
https://avestia.com/ICSTA2022_ Proceedings/files/paper/ICSTA_ 149.pdf

Наукові семінари

1. Семінар з математичної статистики механіко-математичного факультету Київського Національного університету імені Тараса Шевченка під керівництвом проф. О. Г. Кукуша та проф. Р. Є. Майбороди (м. Київ, 2021 р.)
2. Семінар “Статистичні проблеми для випадкових процесів і полів” кафедри математичного аналізу та теорії ймовірностей Національного технічного університету України “Київський політехнічний інститут імені Ігоря Сікорського” під керівництвом проф. О. І. Клесова, проф. О. В. Іванова (м. Київ, 2022 р.)
3. Семінар кафедри теорії ймовірностей і математичного аналізу Ужгородського Національного університету під керівництвом проф. Г. І. Сливки-Тилищак (м. Ужгород, 2022 р.)

Документ підписано у сервісі Вчасно (продовження)
thesis_miroshnychenko.pdf

Документ відправлено: 18:04 29.10.2022

Власник документу

Електронний підпис
18:04 29.10.2022
Ідентифікаційний код: 3439009251
Мірошниченко Віталій Олегович
Власник ключа: Мірошниченко Віталій Олегович
Час перевірки КЕП/ЕЦП: 18:04 29.10.2022
Статус перевірки сертифікату: Сертифікат діє
Серійний номер: 4454624186E4B18704000000032990100194F0A00
Тип підпису: удосконалений