Київський національний університет імені Тараса Шевченка
Кафедра теорії ймовірностей, статистики та актуарної математики

Р. Майборода

Робочі матеріали по курсу лекцій

Регресійний аналіз та асимптотична статистика

Поточна версія від 27 травня 2020 р.

Цей документ має на меті допомогти студентам, що навчаються дистанційно. У ньому зібрані конспекти частини лекцій з курсу “Регресійний аналіз та асимптотична статистика” для студентів четвертого курсу бакалаврату. Документ буде поповнюватись і редагуватись, тому рекомендую регулярно користуватись останньою версією. Дата версії зазначена на титульній сторінці. Отримати останню версію можна за адресою:

`probability.univ.kiev.ua/userfiles/mre/regrasympb.pdf`

Тут вміщені лише теоретичні відомості по кусу. Завдання для самостійного виконання з регресійного аналізу (1й семестр) знаходяться за адресою:

`probability.univ.kiev.ua/userfiles/mre/regr.pdf`

Завдання для самостійного виконання з асимптотичної статистики (2й семестр) знаходяться за адресою:

`probability.univ.kiev.ua/userfiles/mre/asympt.pdf`

Для успішного складання іспиту потрібно виконати і здати всі 10 завдань - 5 завдань першого і 5 другого семестру.

У першому розділі даного документа вміщено перелік тем, що вже були розглянуті у даному курсі та чергова тема, запланована до розгляду. Рекомендую ознайомлюватись із матеріалами по черговій темі до її обговорення онлайн. Після обговорення потрібно надіслати мені на пошту звіт з відповідями на запитання, поставлені до теми у документі.

Розділ 1

Організаційні питання

1.1 Умови і організація іспиту

Для допуску на іспит потрібно здати не менше чотирьох індивідуальних завдань першого семестру (регресійний аналіз) і не менше чотирьох завдань другого семестру (асимптотична статистика). Кожна вчасно здана без зауважень робота оцінюється у 8 балів. За невчасну здачу оцінка знижується на два бали.

Додаткові бали можна отримати за відповіді на теоретичні питання другого семестру, оголошені при вивченні онлайн відповідних тем, а також за активну роботу під час очного навчання (по 5 балів за кожний семестр, максимум - 10 балів).

Студенти, що здали не менше чотирьох завдань першого і чотирьох завдань другого семестру і вважають себе готовими до складання іспиту можуть звернутись до мене е-мейлом у зручний для них час для отримання теоретичних питань іспиту (двва питання у білеті). На отримані питання студент дає письмову відповідь у формі екзаменаційної роботи, як правило, протягом 12 годин. Я перевіряю цю відповідь і призначаю час для співбесіди також протягом 12 годин. На основі балів, отриманих протягом семестру, екзаменаційної роботи (макс. 10 балів за кожне запитання) та по результатах співбесіди (макс. 10 балів) виставляється остаточна оцінка.

За розкладом іспит призначений:

— у групі **"Комп'ютерна статистика та аналіз даних"** на **01.06.2020**. Здати роботи і отримати запитання для підготовки потрібно не пізніше

12:00 31.05.2020, надіслати письмову роботу - не пізніше 24:00 31.05.2020.

— у групі **“актуарна та фінансова математика”** на **04.06.2020**. Отримати запитання для підготовки потрібно не пізніше 12:00 03.06.2020, надіслати письмову роботу - не пізніше 24:00 03.06.2020.

Студентам, які не складуть іспит у встановлений термін, оцінки будуть виставлені за результатами перескладання. Графік перескладань визначає деканат.

1.2 Запитання на іспит

Запитання на іспит:

1. Оцінки методу найменших квадратів для коефіцієнтів лінійної регресії. Їх геометричний зміст та алгебраїчне зображення.

2. Оптимальність оцінок МНК в класі всіх лінійних незміщених оцінок. Коваріаційна матриця оцінок МНК.

3. Незміщена оцінка дисперсії похибок у моделі лінійної регресії.

4. Проста лінійна регресія, формули для оцінок та їх дисперсій, їх отримання з формул загального випадку.

5. Неодночасні довірчі інтервали для коефіцієнтів лінійної регресії. метод Бонферроні отримання одночасних довірчих інтервалів.

6. Метод Шеффе отримання одночасних довірчих інтервалів для коефіцієнтів лінійної регресії.

7. Загальна лінійна гіпотеза для коефіцієнтів лінійної регресії. Тест Фішера у загальному випадку.

8. Перевірка гіпотези про відсутність залежності між двома змінними. Коефіцієнт кореляції Пірсона. Його використання у тесті для перевірки залежності між двома змінними.

9. Перевірка гіпотези про відсутність залежності від хоча б одного з регресорів. Коефіцієнт детермінації лінійної регресійної моделі. Його використання для перевірки залежності від хоча б одного з регресорів.

10. Перевірка однорідності математичних сподівань (середніх) в однофакторному дисперсійному аналізі. Тест Фішера та графічний тест для перевірки однорідності середніх.

11. Розшарована вибірка. Перевірка гіпотез про однорідність розшарованої вибірки. Тест Чоу.

12. Графічні засоби аналізу залишків лінійної регресії: діаграма прогноз-залишки, гістограма, QQ-діаграма залишків.

13. Поняття баєсового тесту. Баєсів тест для перевірки простої гіпотези проти простої альтернативи.

14. Поняття мінімаксного тесту. Мінімаксний тест для перевірки простої гіпотези проти простої альтернативи.

15. Поняття найбільш потужного тесту. Найбільш потужний тест із заданим рівнем значущості для перевірки простої гіпотези проти простої альтернативи.

16. Асимптотичне дослідження тесту відношення вірогідності для альтернатив, що зближуються (випадок простих гіпотез). Вибір порога тесту та оцінка його потужності.

17. Визначення мінімального обсягу вибірки, що гарантує задані ймовірності помилок першого і другого роду для тесту відношення вірогідності на основі асимптотики альтернатив, що зближуються.

18. Тест хі-квадрат для перевірки узгодженості розподілу.

19. Тест хі-квадрат для перевірки незалежності двох змінних. Таблиця спряженості ознак. Мозаїчна діаграма.

20. Структурна модель нелінійної регресії. Консиситентність та асимптотична нормальність оцінок методу найменших квадратів у нелінійних регресійних моделях.

21. Асимптотичні довірчі інтервали для коефіцієнтів нелінійної регресії.

22. Модель логістичної регресії з бінарним відгуком. Оцінки найбільшої вірогідності для коефіцієнтів логістичної регресії. Застосування логістичної регресії у задачі класифікації.

1.3 Пройдені теми для вивчення

У курсі були розглянуті такі теми.

Семестр 1. Регресійний аналіз:

1. Оцінки методу найменших квадратів для коефіцієнтів лінійної регресії. Їх геометричний зміст та алгебраїчне зображення. ([2], п.2.1, 2.2, теорема 2.2.1)

2. Оптимальність оцінок МНК в класі всіх лінійних незміщених оцінок. Коваріаційна матриця оцінок МНК. ([2], 2.2, теорема 2.2.2)

3. Незміщена оцінка дисперсії похибок.([2], 2.2, виправлена вибіркова дисперсія)

4. Проста лінійна регресія, формули для оцінок та їх дисперсій.([2], 2.2, приклад 2.2.1)

5. Довірчі інтервали для коефіцієнтів регресії. Одночасні довірчі інтервали – методи Бонферроні і Шеффе.([2], 2.3, надійні проміжки для коефіцієнтів регресії, теорема 2.3.2; п.2.4.2, теорема 2.4.2)

6. Поняття про тест для перевірки статистичної гіпотези. Помилки першого і другого роду. Потужність тесту. досягнутий рівень значущості. Тест відношення вірогідності для складних гіпотез.([3], п.9.1, 9.3.1)

7. Загальна лінійна гіпотеза. Тест Фішера у загальному випадку.([2], п. 2.4, теорема 2.4.1)

8. Перевірка гіпотези про відсутність залежності між двома змінними. Коефіцієнт кореляції Пірсона.([2], п. 2.4, приклад 2.4.1)

9. Перевірка гіпотези про відсутність залежності від хоча б одного з регресорів. Коефіцієнт детермінації.([2], п. 2.4, приклад 2.4.3)

10. Однофакторний дисперсійний аналіз з фіксованими ефектами. Перевірка однорідності середніх тестом Фішера і графічним тестом. ([2], п. 2.5)

11. Методи графічного аналізу залишків. Дисперсії залишків. Стьюдентизовані залишки. Діаграма розсіювання залишків, гістограма, Q-Q і P-P діаграми.([2], додаток D.2, [3], п. 10.1, [4], роботи 2,3)

12. Розшарована вибірка. Перевірка гіпотез про однорідність розшарованої вибірки. Тест Чоу.([2], 2.4, теорема 2.4.4)

Семестр 2. Асимптотична статистика:

13. Перевірка статистичних гіпотез. Баєсові, мінімаксні та нейманові тести. Потужність тесту, ймовірності помилок першого і другого роду, досягнутий рівень значущості.([1],п.41)

14. Баєсів тест для перевірки простої гіпотези проти простої альтернативи.([1],п.42)

15. Найбільш потужний тест для перевірки простої гіпотези проти простої альтернативи. Лема Неймана-Пірсона.([1],п.42)

16. Асимптотичне дослідження тесту відношення вірогідності для альтернатив, що зближуються. Вибір порога тесту та оцінка його потужності (помилки другого роду).(п. 2.1)

17. Визначення мінімального обсягу вибірки, що гарантує задані ймовірності помилок першого і другого роду для тесту відношення вірогідності. (п. 2.2)

Теми дистанційного навчання за період 12.03.2020 – 3.04.2020:

Тема 1.Тест хі-квадрат для перевірки узгодженості розподілу.

Теорія. ознайомтесь з підрозділами 9.6.1 – 9.6.3 книги [3].

Дайте відповіді на такі питання:

1. Як визначаються і як підраховуються в R емпіричні частоти у тесті хі-квадрат для перевірки узгодженості, якщо розподіл даних є неперервним?

2. Як визначаються і як підраховуються в R теоретичні частоти у тесті хі-квадрат для перевірки узгодженості якщо розподіл має невідомі параметри?

3. Як визначається статистика тесту хі-квадрат?

4. Як визначається поріг тесту? Як підрахувати кількість ступенів вільності для визначення порогу, якщо у теоретичного розподілу є невідомі параметри?

5. Які умови на обсяг вибірки повинні виконуватись, щоб тест хі-квадрат можна було застосовувати на практиці?

Тема 2. Таблиці спряженості і перевірка незалежності змінних.

Теорія. Теоретична частина. ознайомтесь з підрозділом 9.6.4 книги [3].

Дайте відповіді на такі питання:

1. Дайте означення незалежності двох випадкових подій, незалежності двох випадкових величин.

2. Як перевірити, чи будуть незалежними дві дискретні випадкові величини, якщо задано їх спільний розподіл?

3. Що таке таблиця спряженості ознак?

4. Що таке мозаїчна діаграма і як перевірити незалежність двох ознак за цією діаграмою?

5. Як визначаються теоретичні ймовірності у тесті хі-квадрат для перевірки незалежності?

6. Як визначається кількість ступенів вільності для хі-квадрат теоретичного у тесті хі-квадрат перевірки незалежності?

Обговорення онлайн 1.04.2020 о 14-00. — обговорювалась тема 1 — тест хі-квадрат для перевірки узгодженості розподілу. Виконання першої самостійної роботи по асимптотичній статистиці.

Теми дистанційного навчання з 6.04.2020:

Обговорення онлайн: 10.04.2020 о 14-00.

Тема: Виконання другої і третьої самостійних робіт по асимптотичній статистиці. Див. п. 2.1, 2.2, а також [1], п. 42, 43.3.

Дайте відповіді на такі питання:

1. Дайте означення інформації за Фішером, що міститься у одному спостереженні.

2. Дайте означення мінімаксного тесту. Який тест буде мінімаксним для перевірки простої гіпотези проти простої альтернативи?

3. Поясніть, чому у мінімаксного тесту для перевірки простих гіпотез ймовірності помилок першого і другого роду повинні бути рівні.

Обговорення онлайн: 17.04.2020 о 14-00.

Тема: Асимптотичне дослідження оцінок методу найменших квадратів для параметрів нелінійних регресійних моделей. Див. п. 3.1.

Запитання до теми:

1. Сформулюйте означення слабкої збіжності послідовності випадкових векторів.

2. Сформулюйте центральну граничну теорему, яка була використана при доведенні асимптотичної нормальності оцінок методу найменших квадратів.

3. Дайте означення матриці розсіювання оцінки.

4. Нехай функція регресії є лінійною: $g(\mathbf{X},\vartheta) = \sum_{i=1}^{d} X_j^i \vartheta^i$. Який вигляд матиме в цьому випадку матриця $\mathbf{A}_\infty$?

Обговорення онлайн: 1.05.2020 о 14-00. Тести та довірчі інтервали для параметрів нелінійних регресійних моделей (п. 3.2).

Запитання до теми:

1. Дайте означення асимптотичного довірчого інтервалу.

2. Який вигляд матиме асимптотичний довірчий інтервал (3.12), якщо використовувати його для коефіцієнта b_1 простої лінійної регресії?

3. Дайте означення досягнутого рівня значущості (довільного) тесту.

4. Поясніть, як можна отримати формулу (3.13) для досягнутого рівня значущості асимптотичного тесту Стьюдента.

Обговорення онлайн: 4.05.2020 о 14-00. Регресія з бінарним відгуком. Логістична і пробіт-регресія (п. 3.3).

Запитання до теми:

1. Об'єкти вашого дослідження — студенти механко-математичного факультету. Ви хочете побудувати прогнозну формулу для передбачення того, з якою ймовірністю студент здасть залік з статистики. Для цього як регресори ви можете використати оцінки, виставлені студенту за окремі індивідуальні завдання, які він виконував протягом семестру. Яку регресійну модель ви використаєте?

2. Розгляньте задачу класифікації для двох класів, в якій вектор спостережуваних характеристик одновимірний, розподіл цієї характеристи-

ки гауссів для обох класів, причому для різних класів різними є як математичні сподівання, так і дисперсії. Опишіть модель регресії з бінарним відгуком для цієї задачі.

Розділ 2

Асимптотичне дослідження статистичних тестів

2.1 Тести відношення вірогідності для простих альтернатив, що зближуються

Розглянемо кратну вибірку $\mathbf{X} = (\xi_1, \dots, \xi_n)$, де ξ_j незалежні, однаково розподілені спостереження[1] причому розподіл одного спостереження має щільність $f^\xi(x)$ відносно деякої міри[2] μ на просторі можливих значень ξ_j.

Нехай відомо, що $f^\xi(x)$ належить деякій параметричній сім'ї розподілів $f(x;\vartheta)$, $\vartheta \in \Theta$, де ϑ — одновимірний невідомий параметр, $\Theta \subseteq \mathbb{R}$ — множина можливих значень θ. Тобто існує таке "справжнє значення" ϑ, що $f^\xi(x) = f(x;\vartheta)$ для всіх x.

Задача полягає в тому, щоб побудувати оптимальний у певному розумінні тест для перевірки гіпотези $\vartheta = \vartheta_0$ проти альтернативи $\vartheta = \vartheta_1$, де ϑ_i, $i = 0, 1$ — відомі значення.

Як ми знаємо, найкращі тести (в усіх звичайних розуміннях: баєсові, мінімаксні, найбільш потужні) треба шукати серед тестів відношення вірогідності.

---

[1]Це можуть бути випадкові величини або вектори, або випадкові значення з фіксованого набору, тощо.

[2]Традиційні приклади: якщо ξ_j абсолютно неперервні, то μ це міра Лебега і f^ξ — традиційна ймовірнісна щільність розподілу. Якщо ξ_j — дискретні, μ — лічильна міра, $f^\xi(x) = \mathsf{P}\{\xi_j = x\}$.

Функція вірогідності у даному випадку має вигляд:

$$L(\vartheta) = \prod_{j=1}^{n} f(\xi_j; \vartheta).$$

Відповідно, логарифмічне відношення вірогідності є:

$$\mathsf{lr}_n(\mathbf{X}) = \log\left(\frac{L(\vartheta_1)}{L(\vartheta_0)}\right) = \sum_{j=1}^{n} \log\left(\frac{f(\xi_j; \vartheta_1)}{f(\xi_j; \vartheta_0)}\right) = \sum_{j=1}^{n} \eta_j, \qquad (2.1)$$

де

$$\eta_j = \log\left(\frac{f(\xi_j; \vartheta_1)}{f(\xi_j; \vartheta_0)}\right).$$

Тест відношення вірогідності використовує статистику $\mathsf{lr}_n(\mathbf{X})$ і поріг C. Якщо $\mathsf{lr}_n(\mathbf{X}) \le C$, тест приймає основну гіпотезу, інакше — альтернативу. Для того, щоб тест був оптимальним, треба правильно обрати поріг C. Скажімо, коли потрібен тест, який був би найбільш потужним у класі тестів з рівнем значущості α_0, то C слід обрати з умови $C = C_{\alpha_0}$,

$$\mathsf{P}_{\vartheta_0}\{\mathsf{lr}_n(\mathbf{X}) > C_{\alpha_0}\} = \alpha_0. \qquad (2.2)$$

Якщо намагатись знайти таке C_α, що задовольняє (2.2) при фіксованому обсязі вибірки, то потрібно явно виписати розподіл статистики відношення вірогідності при виконанні основної гіпотези і шукати його квантиль рівня $1-\alpha$. Для більшості задач перевірки гіпотез це не просто. Але коли перейти до $n \to \infty$, задача спрощується. А отримані результати можна використовувати при великих обсягах вибірки як хороше наближення. Це ми і зробимо.

Одразу помітимо, що, коли розглядати нашу задачу перевірки гіпотез з фіксованими ϑ_0 і ϑ_1 при $n \to \infty$, то, як правило, у всіх розумних тестів ймовірності помилок і першого і другого роду будуть дуже швидко прямувати до 0. У цій ситуації всі тести виявляються "чудовими" і не зрозуміло, як обрати найкращий. Тому зручніше розглядати випадок, коли із зростанням обсягу вибірки альтернативна гіпотеза наближається до основної: $\vartheta_1 \to \vartheta_0$ при $n \to \infty$. Тоді із зростанням об'єму інформації у даних зростає і складність задачі оцінювання і ймовірності похибок найкращого тесту не прямують до 0. Швидкість збіжності $\vartheta_1 \to \vartheta_0$ підбирають так, щоб ці ймовірності прямували до якихось фіксованих чисел більших 0 і менших 1. Ці числа і будуть характеризувати асимптотичну якість тесту. Це зветься дослідженням тесту на альтернативах, що зближуються (локальних альтернативах).

Виявляється, що у нашій задачі така правильна збіжність задається як

$$\vartheta_1 = \vartheta_0 + \frac{v}{\sqrt{n}}, \tag{2.3}$$

тобто альтернатива наближається до основної гіпотези по порядку величини як $1/\sqrt{n}$, а v характеризує швидкість цього наближення.

Тепер дослідимо асимптотичну поведінку логарифмічного відношення вірогідності (2.1), вважаючи, що виконана основна гіпотеза $H_0 : \vartheta = \vartheta_0$, а альтернатива $H_1 : \vartheta = \vartheta_1$ наближається до основної гіпотези відповідно до (2.3). Дослідження проводиться неформально, насправді, потрібно іще обґрунтовувати збіжність до 0 залишкових доданків, за певних додаткових умов це можна зробити, але ми не будемо.

Будемо позначати

$$f'(x;\vartheta) = \frac{\partial f(x;\vartheta)}{\partial \vartheta},\ f''(x;\vartheta) = \frac{\partial^2 f(x;\vartheta)}{\partial \vartheta^2}.$$

Використаємо розклади за формулою Тейлора:

$$f(x;\vartheta_1) = f(x;\vartheta_0) + f'(x;\vartheta_0)\frac{v}{\sqrt{n}} + \frac{1}{2}f''(x;\vartheta_0)\frac{v^2}{n} + o(1/n)$$

і

$$\log(1+z) = z - \frac{1}{2}z^2 + o(z^2).$$

Ми розглядаємо асимптотику при $n \to \infty$. При цьому ми враховуємо лише доданки, у які n входить з порядком $n^{-1/2}$, або n^{-1}. Доданки, які по n прямують до 0 швидше, ніж n^{-1} не враховуються (про це далі). Отримуємо:

$$\eta_j = \log\left(1 + \frac{f(\xi_j;\vartheta_0 + v/\sqrt{n}) - f(\xi_j;\vartheta_0)}{f(\xi_j;\vartheta_0)}\right)$$

$$= \frac{1}{f(\xi_j;\vartheta_0)}\left(f'(\xi_j;\vartheta_0)\frac{v}{\sqrt{n}} + f''(\xi_j;\vartheta_0)\frac{v^2}{n} + \dots\right) - \frac{(f'(\xi_j;\vartheta_0))^2}{2(f(\xi_j;\vartheta_0))^2}\frac{v^2}{n} + \dots$$

Отже

$$\mathsf{lr}_n(\mathbf{X}) = \sum_{j=1}^{n} \eta_j = vS_1 + \frac{v^2}{2}S_2 - \frac{v^2}{2}S_3 + \dots, \tag{2.4}$$

$$S_1 = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{f'(\xi_j;\vartheta_0)}{f(\xi_j;\vartheta_0)},$$

$$S_2 = \frac{1}{n} \sum_{j=1}^{n} \frac{f''(\xi_j;\vartheta_0)}{f(\xi_j;\vartheta_0)},$$

$$S_3 = \frac{1}{n} \sum_{j=1}^{n} \left(\frac{f'(\xi_j;\vartheta_0)}{f(\xi_j;\vartheta_0)} \right)^2,$$

A ... у цьому розкладі відповідає доданкам вигляду $\sum_{j=1}^{n} o(1/n)$. Вони прямують до 0, оскільки кожен доданок прямує до 0 швидше, ніж зростає кількість доданків у сумі (насправді, це треба обґрунтовувати акуратніше).

Помітимо, що, за підсиленим законом великих чисел, при $n \to \infty$ майже напевне,

$$S_2 \to \mathsf{E}_{\vartheta_0} \frac{f''(\xi_1;\vartheta_0)}{f(\xi_1;\vartheta_0)} = \int f''(x;\vartheta_0)\mu(dx)$$

$$= \frac{\partial^2}{\partial \vartheta^2} \int f(x;\vartheta_0)\mu(dx) = \frac{\partial^2}{\partial \vartheta^2} 1 = 0,$$

і, аналогічно,

$$S_3 \to \mathsf{E}_{\vartheta_0} \left(\frac{f'(\xi_1;\vartheta_0)}{f(\xi_1;\vartheta_0)} \right)^2 = I^{\xi}(\vartheta_0),$$

де $I^{\xi}(\vartheta_0)$ — інформація за Фішером, що міститься в одному спостереженні ξ_j про параметр ϑ при $\vartheta = \vartheta_0$.

Нарешті, за центральною граничною теоремою,

$$S_1 \xrightarrow{W} N(0, D),$$

де

$$D = \mathsf{E}_{\vartheta_0} \left(\frac{f'(\xi_1;\vartheta_0)}{f(\xi_1;\vartheta_0)} \right)^2 = I^{\xi}(\vartheta_0).$$

Об'єднуючи ці результати, отримуємо, що, при виконанні H_0,

$$\mathsf{lr}_n(\mathbf{X}) \xrightarrow{W} |v|\zeta - \frac{v^2}{2} I^{\xi}(\vartheta_0), \tag{2.5}$$

де $\zeta \sim N(0, I^{\xi}(\vartheta_0))$.

Розглянемо тепер ймовірність помилки першого роду для нашого тесту відношення вірогідності з порогом C:

$$\alpha(C) = \mathsf{P}_{\vartheta_0}\{\mathsf{lr}_n(\mathbf{X}) > C\} \to \mathsf{P}\{|v|\zeta - \frac{v^2}{2} I^\xi(\vartheta_0) > C\}$$

$$= \mathsf{P}\left\{\frac{\zeta}{\sqrt{I^\xi(\vartheta_0)}} > \frac{v^2 I^\xi(\vartheta_0)/2 + C}{|v|\sqrt{I^\xi(\vartheta_0)}}\right\}$$

Отже

$$\alpha(C) \approx \alpha_\infty(C) = 1 - \Phi\left(\frac{v^2 I^\xi(\vartheta_0)/2 + C}{|v|\sqrt{I^\xi(\vartheta_0)}}\right) \text{ при великих } n. \tag{2.6}$$

Тут $\Phi(x)$ — функція розподілу стандартного нормального розподілу. Це — асимптотична формула для ймовірності помилки першого роду для тесту відношення вірогідності з порогом C.

Ми обираємо поріг C так, щоб $\alpha_\infty(C) = \alpha_0$, де α_0 — стандартний рівень значущості. Отримуємо

$$C_{\alpha_0} = \lambda_{\alpha_0}|v|\sqrt{I^\xi(\vartheta_0)} - v^2 I^\xi(\vartheta_0)/2. \tag{2.7}$$

Це і є поріг тесту відношення вірогідності, який забезпечує заданий асимптотичний рівень значущості.

Тепер оцінимо ймовірність помилки другого роду для нашого тесту. Для цього треба дослідити асимптотику відношення вірогідності вже в припущенні, що виконана альтернатива: $\vartheta = \vartheta_1$. Відповідно, розклади за формулою Тейлора треба брати в околі ϑ_1 і математичні сподівання теж брати в припущенні, що справжнє значення параметра — це ϑ_1. В результаті отримуємо, що при виконанні H_1:

$$\mathsf{lr}_n(\mathbf{X}) \xrightarrow{W} |v|\zeta + \frac{v^2}{2} I^\xi(\vartheta_0) \text{ при } n \to \infty.$$

Таким чином, для ймовірності помилки другого роду тесту відношення вірогідності з порогом C маємо такий асимптотичний результат:

$$\beta(C) = \mathsf{P}_{\vartheta_1}\{\mathsf{lr}_n(\mathbf{X}) < C\} \approx \beta_\infty(C) = \Phi\left(\frac{-v^2 I^\xi(\vartheta_0)/2 + C}{|v|\sqrt{I^\xi(\vartheta_0)}}\right). \tag{2.8}$$

Тут $\beta_\infty(C)$ — асимптотична помилка другого роду для тесту відношення вірогідності з порогом C.

Якщо у цю формулу підставити поріг C_{α_0} з (2.2) то отримаємо

$$\beta_\infty(C_{\alpha_0}) = \Phi(-|v|\sqrt{I^\xi(\vartheta_0)} + \lambda_\alpha).$$

Це найкраще (найменше) можливе значення ймовірності помилки другого роду для будь-якого тесту, що перевіряє гіпотезу $\vartheta = \vartheta_0$ проти локальної альтернативи і забезпечує рівень значущості α_0. Легко бачити, що похибка тим менша, чим більше $I^\xi(\vartheta_0)$ — інформація, що міститься в одному спостереженні і $|v|$ — величина, що характеризує відстань від основної гіпотези до альтернативи. Крім того, із зменшенням α_0, $\beta_\infty(C_{\alpha_0})$ зростає.

Отримані асимптотичні формули (2.6) і (2.8) дозволяють також визначати пороги для баєсових та мінімаксних тестів. Наприклад, якщо $C = 0$, отримуємо

$$\alpha_\infty(0) = \beta_\infty(0) = \Phi(-|v|\sqrt{I^\xi(\vartheta_0)}/2),$$

отже тест відношення вірогідності з порогом $C = 0$ є асимптотично мінімаксним тестом.

2.2 Визначення обсягу вибірки для забезпечення заданої якості тестування

Оскільки тест відношення вірогідності є найкращим можливим, поліпшити ймовірності його помилок першого і другого роду одночасно можна лише збільшивши обсяг вибірки, за якою проводиться тестування.

Нехай за кратною вибіркою $\mathbf{X} = (\xi_1, \dots, \xi_n)$ з щільністю спостереження $f^\xi(x) = f(x;\vartheta)$ проводиться перевірка основної гіпотези $H_0 :\ \vartheta = \vartheta_0$ проти альтернативи $H_1 :\ \vartheta = \vartheta_1$. Який обсяг вибірки потрібен, щоб забезпечити одночасно ймовірності помилки першого роду[3] α_0 і другого роду — β_0?

Для відповіді на це питання скористаємось асимптотичними формулами (2.6) і (2.8). Обравши $C = C_{\alpha_0}$ ми забезпечуємо $\alpha_\infty(C_{\alpha_0}) = \alpha_0$. Тоді

[3]На практиці α_0 і β_0 — фіксовані числа, що визначаються замовником статистичного дослідження.

для ймовірності помилки другого роду маємо:

$$\beta_\infty(C_{\alpha_0}) = \Phi(-|v|\sqrt{I^\xi(\vartheta_0)} + \lambda_{\alpha_0}) = \beta_0.$$

Звідси отримуємо

$$v = \frac{\lambda_{\alpha_0} + \lambda_{\beta_0}}{\sqrt{I^\xi(\vartheta_0)}}.$$

Згадаємо, що за (2.3)

$$v = \sqrt{n}(\vartheta_1 - \vartheta_0).$$

Прирівнюючи праві частини двох останніх рівностей, отримуємо

$$n = \frac{(\lambda_{\alpha_0} + \lambda_{\beta_0})^2}{I^\xi(\vartheta_0)(\vartheta_1 - \vartheta_0)^2}. \tag{2.9}$$

Приклад. Нехай дані є кратною вибіркою з розподілу Бернуллі: $\mathbf{X} = (\xi_1, \ldots, \xi_n)$,

$$\xi_j = \begin{cases} 1 & \text{з ймовірністю } p, \\ 0 & \text{з ймовірністю } 1-p. \end{cases}$$

Ймовірність успіху p невідома, щодо неї є дві гіпотези. Основна $H_0: \ p = p_0 = 0.05$, альтернативна — $H_1: \ p = p_1 = 0.1$. Потрібно визначити необхідний обсяг вибірки і побудувати тест, який матиме ймовірності помилок першого і другого роду не більше 0.05.

Для перевірки цих гіпотез скористаємось тестом відношення вірогідності. Оскільки спостереження є дискретними, на роль міри μ, відносно якої визначається щільність, беремо лічильну міру на множині $\{0, 1\}$. Тоді щільність одного спостереження

$$f(x;p) = \begin{cases} p & \text{якщо } x = 1, \\ 1-p & \text{якщо } x = 0, \\ 0 & \text{в усіх інших випадках.} \end{cases}$$

Логарифмічна функція вірогідності

$$\mathsf{lr}_n(\mathbf{X}) = \sum_{j=1}^n \log\left(\frac{f(\xi_j;p_1)}{f(\xi_j;p_0)}\right) = S_n \log(p_1/p_0) + (n - S_n)\log((1-p_1)/(1-p_0)),$$

де $S_n = \sum_{j=1}^n \xi_j$.

Щоб визначити обсяг вибірки скористаємось формулою (2.9). Для цього потрібно визначити $I^{\xi}(p_0)$. Інформація за Фішером на одне спостереження визначається як

$$I^{\xi}(p_0) = \mathsf{E}_{p_0}\left(\frac{f'(\xi_1;p_0)}{f(\xi_1;p_0)}\right)^2$$

$$= p_0(1/p_0)^2 + (1-p_0)(1/(1-p_0))^2 = \frac{1}{p_0(1-p_0)} = 21.05263.$$

У нашому випадку $\lambda_{\alpha_0} = \lambda_{\beta_0} = Q^{N(0,1)}(1-0.05) \approx 1.64$. Отже, за (2.9) отримуємо:

$$n = \frac{(2\times 1.64)^2}{(0.05\times 0.95)^{-1}\times(0.1-0.05)^2} \approx 204.4096.$$

Округлюючи з надлишком, отримуємо, що для заданої точності тестування потрібно $n = 205$ спостережень. Це значення n відповідає параметру

$$v = \sqrt{n}(p_1 - p_0) = \sqrt{205}(0.1-0.05) = 0.7158911.$$

Тепер за заданим v і рівнем значущості $\alpha_0 = 0.05$ можна визначити поріг тесту за формулою (2.7):

$$C_{0.05} = 1.64\times 0.716\times\sqrt{(}21.05) - 0.716^2\times 21.05/2 \approx -0.0078$$

— практично 0. (Можна було б і не рахувати, оскільки за побудовою у нашого тесту ймовірності помилок обох родів рівні між собою. Отже — це мінімаксний тест, у якого $C = 0$).

Розділ 3

Асимптотичне дослідження оцінок

3.1 Консистентність та асимптотична нормальність оцінок методу найменших квадратів

Розглянемо структурну модель нелінійної регресії, в якій спостереження $\mathbf{Z}_j = (\mathbf{X}_j, Y_j)$ є незалежними, однаково розподіленими випадковими векторами. Тут $\mathbf{X}_j = (X_j^1, \dots, X_j^m)^T \in \mathbb{R}^m$ — вектор регресорів[1] а Y_j — значення відгуку, що відповідає цим регресорам у j-тому спостереженні. Відгук і регресори пов'язані між собою регресійним зв'язком:

$$Y_j = g(\mathbf{X}_j, \vartheta) + \varepsilon_j, \tag{3.1}$$

де

$g(\mathbf{x}, \mathbf{t})$ — функція регресії, яка вважається відомою з точністю до значення невідомого параметра $\mathbf{t} \in \Theta \subseteq \mathbb{R}^d$;

$\vartheta = (\vartheta^1, \dots, \vartheta^d)^T$ — невідоме справжнє значення параметра;

ε_j — випадкова похибка регресії для j-го спостереження.

Далі ми всюди будемо вважати, що виконані такі **умови структурної L_2 регресії**:

[1] Їх іще називають "незалежними змінними" (independent variables) але це не означає, що регресори мають бути незалежними між собою як випадкові величини. Це просто назва, яка протиставляє їх "залежній змінній" Y, тобто змінній, залежність якої від регресорів ми вивчаємо.

1. Вектори спостережень $\mathbf{Z}_j$, $j = 1, \ldots, n$ є незалежними, однаково розподіленими при різних j.

2. $\mathsf{E}\,\varepsilon_j = 0$, $\mathrm{D}\,\varepsilon_j = \sigma^2 < \infty$.

3. $\mathbf{X}_j$ і ε_j незалежні між собою при всіх j.

Наше завдання полягає в тому, щоб за спостереженнями $\mathbf{Z}_j$, $j = 1, \ldots, n$, оцінити невідоме значення параметра $\vartheta \in \Theta$.

Для оцінювання скористаємось методом найменших квадратів. Складемо **функціонал найменших квадратів**:

$$J_n(\mathbf{t}) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - g(\mathbf{X}_j, \mathbf{t}))^2. \tag{3.2}$$

Тут $\mathbf{t}$ пробігає всі можливі значення невідомого параметра: $\mathbf{t} \in \Theta$.

Оцінкою методу найменших квадратів (МНК, LSE — least squares estimator) $\hat{\vartheta}_n$ невідомого параметра ϑ називають те значення $\mathbf{t}$, при якому $J_n(\mathbf{t})$ досягає свого мінімуму:

$$J_n(\hat{\vartheta}_n) = \min_{\mathbf{t}\in\Theta} J_n(\mathbf{t}). \tag{3.3}$$

Якщо таких значень декілька, то на роль оцінки можна обрати будь-яке з них, але при цьому $\hat{\vartheta}_n$ повинна бути вимірною функцією від даних. Якщо мінімум у (3.3) не досягається, оцінка МНК не існує[2].

У загальному випадку задача знаходження оцінок МНК не розв'язується у явному вигляді, оцінки шукають чисельними методами. Нас буде цікавити асимптотична поведінка $\hat{\vartheta}_n$ коли обсяг вибірки $n \to \infty$.

Перше, чого вимагають від оцінки — це **консистентність**, тобто збіжність до величини, що оцінюється: $\hat{\vartheta}_n \to \vartheta$ за ймовірністю.

Розглянемо неформальні міркування, які дозволяють зрозуміти, за яких умов оцінки МНК будуть консистентними.

Помітимо, що

$$\mathsf{E}_\vartheta\, J_n(\mathbf{t}) = \frac{1}{n}\sum_{j=1}^{n}\mathsf{E}_\vartheta(Y_j - g(\mathbf{X}_j, \mathbf{t}))^2$$

$$= \mathsf{E}_\vartheta(g(\mathbf{X}_j, \vartheta) + \varepsilon_j - g(\mathbf{X}_j, \mathbf{t}))^2$$

[2] В цьому випадку можна ввести узагальнені або наближені оцінки МНК, але ми їх тут не розглядатимем.

$$= \mathsf{E}_\vartheta(g(\mathbf{X}_j,\vartheta)) - g(\mathbf{X}_j,\mathbf{t}))^2 + 2\,\mathsf{E}_\vartheta(g(\mathbf{X}_j,\vartheta)) - g(\mathbf{X}_j,\mathbf{t}))\varepsilon_j + \mathsf{E}\,\varepsilon_j^2.$$

Другий доданок у цій сумі дорівнює 0, оскільки $\mathsf{E}\,\varepsilon_j = 0$ і ε_j та $\mathbf{X}_j$ — незалежні. Отже

$$\mathsf{E}_\vartheta\, J_n(\mathbf{t}) = J_\infty(\mathbf{t}),$$

де

$$J_\infty(\mathbf{t}) = \mathsf{E}_\vartheta(g(\mathbf{X}_j,\vartheta)) - g(\mathbf{X}_j,\mathbf{t}))^2 + \sigma^2.$$

За законом великих чисел

$$J_n(\mathbf{t}) \to J_\infty(\mathbf{t}),\ \text{при } n \to \infty$$

майже напевно, а отже, і за ймовірністю.

Помітимо, що перший доданок у $J_\infty(\mathbf{t})$ завжди невід'ємний, а при $\mathbf{t} = \vartheta$ він обертається у 0. Другий доданок — σ^2 від $\mathbf{t}$ не залежить, отже $\mathbf{t} = \vartheta$ — точка мінімуму $J_\infty(\mathbf{t})$. Оскільки $J_n(\mathbf{t})$ прямує до $J_\infty(\mathbf{t})$, можна сподіватись, що і точка мінімуму $J_n(\mathbf{t})$, тобто $\hat\vartheta_n$ прямує до ϑ. Це і є консистентність оцінки МНК.

Для того, щоб ця логіка працювала, потрібні додаткові умови.

По-перше, потрібно, щоб ϑ була єдиною точкою глобального мінімуму $J_\infty(\mathbf{t})$. Якщо для якогось $\mathbf{t}' \neq \vartheta$ виконано $g(\mathbf{X}_j,\mathbf{t}') = g(\mathbf{X}_j,\vartheta)$ м.н., то $J_\infty(\mathbf{t}') = J_\infty(\vartheta)$ і $\mathbf{t}'$ — така сама точка точка мінімуму, як і ϑ. Відповідно, оцінка $\hat\vartheta_n$ може збігатись до "хибного мінімуму" $\mathbf{t}'$. Тоді вона, зрозуміло, не буде консистентною.

Отже, першою умовою консистентності буде

$$\text{Для всіх } \mathbf{t}' \neq \mathbf{t}'',\ \mathsf{P}\{g(\mathbf{X}_1,\mathbf{t}') \neq g(\mathbf{X}_1,\mathbf{t}'')\} > 0. \tag{3.4}$$

Цю умову називають умовою контрасту.

По-друге, множина можливих значень невідомого параметра Θ повинна бути компактом (замкненою обмеженою множиною в $\mathbb{R}^d$). Інакше можливо, що інфінум $J_\infty(\mathbf{t})$ досягається на нескінченності і, відповідно, оцінка МНК буде "тікати на нескінченність" замість того, щоб збігатись до ϑ.

Нарешті, по-третє, збіжність функціоналу МНК до границі має бути рівномірною:

$$\sup_{\mathbf{t}\in\Theta} |J_n(\mathbf{t}) - J_\infty(\mathbf{t})| \to 0,\ \text{при } n \to \infty.$$

Саме ця умова разом з умовою контрасту забезпечує збіжність точок мінімуму.

Підсумовуючи, сформулюємо умови консистентності у вигляді теореми.

Теорема 3.1.1 *Нехай виконуються такі умови.*

1. Виконуються умови структурної L_2 регресії.

2. Θ є компактом.

3. Виконується умова контрасту (3.4).

4. Функції $g(\mathbf{x},\mathbf{t})$ є неперервними по $\mathbf{t}\in\Theta$ рівноступенево по $\mathbf{x}$, тобто

$$\sup_{\mathbf{x}} |g(\mathbf{x},\mathbf{t}) - g(\mathbf{x},\mathbf{t}_0)| \to 0, \ \textit{при } \mathbf{t}\to\mathbf{t}_0 \textit{ для всіх } \mathbf{t}\in\Theta.$$

5. $\mathsf{E}\sup_{\mathbf{t}\in\Theta}(g(\mathbf{X}_1,\mathbf{t}))^2 < \infty$.

Тоді $\hat{\vartheta}_n \xrightarrow{P} \vartheta$ *при* $n\to\infty$.

Доведення цієї теореми можна провести аналогічно доведенню твердження 5.4 у книзі [6]. Умови 4 і 5 цієї теореми забезпечують рівномірну збіжність функціонала МНК до граничного значення.

Перейдемо тепер до асимптотичної нормальності оцінок МНК. Ми будемо досліджувати граничний розподіл нормованих різниць між оцінкою та справжнім значенням параметра:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{W} N(0,\mathbf{V}). \tag{3.5}$$

Матриця $\mathbf{V}$ зветься матрицею розсіювання оцінки $\hat{\vartheta}$.

Далі неформально описано спосіб такого дослідження і виведена формула для $\mathbf{V}$. Формальний виклад відповідних результатів можна знайти у книзі [6], п. 5.4.3.

Далі крапка над символом функції позначає диференціювання за змінною $\mathbf{t}$. Оскільки $\mathbf{t} = (t^1,\dots,t^d)$ є вектором, то йдеться про набір похідних за всіма координатами цього ветора, отже

$$\dot{g}(\mathbf{x},\mathbf{t}) = \begin{pmatrix} \frac{\partial g(\mathbf{x},\mathbf{t})}{\partial t^1} \\ \frac{\partial g(\mathbf{x},\mathbf{t})}{\partial t^2} \\ \vdots \\ \frac{\partial g(\mathbf{x},\mathbf{t})}{\partial t^d} \end{pmatrix}$$

це градієнт функції g (по змінній $\mathbf{t}$), а

$$\ddot{g}(\mathbf{x},\mathbf{t}) = \begin{pmatrix} \frac{\partial^2 g(\mathbf{x},\mathbf{t})}{\partial t^1 \partial t^1} & \frac{\partial^2 g(\mathbf{x},\mathbf{t})}{\partial t^1 \partial t^2} & \dots & \frac{\partial^2 g(\mathbf{x},\mathbf{t})}{\partial t^1 \partial t^d} \\ \frac{\partial^2 g(\mathbf{x},\mathbf{t})}{\partial t^2 \partial t^1} & \frac{\partial^2 g(\mathbf{x},\mathbf{t})}{\partial t^2 \partial t^2} & \dots & \frac{\partial^2 g(\mathbf{x},\mathbf{t})}{\partial t^2 \partial t^d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 g(\mathbf{x},\mathbf{t})}{\partial t^d \partial t^1} & \frac{\partial^2 g(\mathbf{x},\mathbf{t})}{\partial t^d \partial t^2} & \dots & \frac{\partial^2 g(\mathbf{x},\mathbf{t})}{\partial t^d \partial t^d} \end{pmatrix}$$

— гессіан функції g. Надалі ми вважатимем, що всі ці похідні існують і є неперервними функціями $\mathbf{t}$. Аналогічні позначення використовуються і для функції $J_n(\mathbf{t})$.

Ми будемо припускати, що оцінка МНК $\hat{\vartheta}_n$ є консистентною. Оскільки у наших умовах $J_n(\mathbf{t})$ — неперервно диференційовна функція, для її точки мінімуму $\hat{\vartheta}_n$ повинна виконуватись достатня умова екстремуму:

$$\dot{J}_n(\hat{\vartheta}) = 0.$$

Розкладаючи $\dot{J}_n(\mathbf{t})$ за формулою Тейлора у околі точки ϑ, отримуємо

$$\dot{J}_n(\mathbf{t}) = \dot{J}_n(\vartheta) + \ddot{J}_n(\zeta)(\mathbf{t} - \vartheta),$$

де ζ — проміжна точка між $\mathbf{t}$ і ϑ. Підставивши $\mathbf{t} = \hat{\vartheta}_n$, отримуємо:

$$0 = \dot{J}_n(\vartheta) + \ddot{J}_n(\zeta)(\hat{\vartheta}_n - \vartheta),$$

звідки

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = -[\ddot{J}_n(\zeta)]^{-1}(\sqrt{n}\dot{J}_n(\vartheta)). \tag{3.6}$$

Тепер дослідимо окремо поведінку $\ddot{J}_n(\zeta)$ і $\sqrt{n}\dot{J}_n(\vartheta)$ при $n \to \infty$.

$$\sqrt{n}\dot{J}_n(\vartheta) = \frac{-2}{\sqrt{n}} \sum_{j=1}^{n} (Y_j - g(\mathbf{X}_j, \vartheta))\dot{g}(\mathbf{X}_j, \vartheta),$$

отже

$$-\sqrt{n}\dot{J}_n(\vartheta)/2 = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varepsilon_j \dot{g}(\mathbf{X}_j, \vartheta).$$

Доданки у цій сумі незалежні, однаково розподілені і

$$\mathsf{E}_\vartheta\, \varepsilon_j \dot{g}(\mathbf{X}_j, \vartheta) = \mathsf{E}_\vartheta\, \varepsilon_j\, \mathsf{E}_\vartheta\, \dot{g}(\mathbf{X}_j, \vartheta) = 0,$$

$$\mathrm{Cov}_\vartheta\, \varepsilon_j \dot{g}(\mathbf{X}_j, \vartheta) = \mathsf{E}_\vartheta\, \varepsilon_j \dot{g}(\mathbf{X}_j, \vartheta)(\varepsilon_j \dot{g}(\mathbf{X}_j, \vartheta))^T$$

$$= \sigma^2 \mathbf{A}_\infty,$$

де

$$\mathbf{A}_\infty = \mathsf{E}_\vartheta\, \dot{g}(\mathbf{X}_j, \vartheta)(\dot{g}(\mathbf{X}_j, \vartheta))^T. \tag{3.7}$$

(Ми припускаємо, що всі елементи цієї матриці скінченні).

Отже, за центральною граничною теоремою,

$$\eta_n = -\sqrt{n} J_n(\vartheta) \xrightarrow{W} \zeta_\infty \sim N(0, \sigma \mathbf{A}_\infty). \tag{3.8}$$

Тепер розглянемо

$$-\ddot{J}_n(\mathbf{t})/2 = -\frac{\partial}{\partial \mathbf{t}}(-\dot{J}_n(\mathbf{t})/2) = \frac{1}{n}\sum_{j=1}^n \frac{\partial}{\partial \mathbf{t}}(Y_j - g(\mathbf{X}_j, \mathbf{t}))\dot{g}(\mathbf{X}_j, \mathbf{t})$$

$$= \frac{1}{n}\sum_{j=1}^n \left[-\dot{g}(\mathbf{X}_j, \mathbf{t}))(\dot{g}(\mathbf{X}_j, \mathbf{t}))^T + (Y_j - g(\mathbf{X}_j, \mathbf{t}))\ddot{g}(\mathbf{X}_j, \mathbf{t})\right].$$

Підставляючи у цей вираз $\mathbf{t} = \vartheta$, отримуємо

$$-\ddot{J}_n(\vartheta)/2 = \frac{1}{n}\sum_{j=1}^n \left[-\dot{g}(\mathbf{X}_j, \vartheta))(\dot{g}(\mathbf{X}_j, \vartheta))^T + \varepsilon_j \ddot{g}(\mathbf{X}_j, \vartheta)\right]$$

Доданки у цій сумі є незалежними однаково розподіленими випадковими матрицями. Отже, за підсиленим законом великих чисел, $-\ddot{J}_n(\vartheta)/2$ прямує майже непевно до математичного сподівання одного доданка. Легко бачити, що це математичне сподівання дорівнює $-\mathbf{A}_\infty$. Отже

$$-\ddot{J}_n(\vartheta)/2 \to -\mathbf{A}_\infty \text{ майже напевно при } n \to \infty.$$

Враховуючи, що у (3.6) ζ — проміжна точка між $\hat{\vartheta}_n$ і ϑ прямує до ϑ при $n \to \infty$, можна показати (при виконанні додаткових умов на гладкість g), що

$$\ddot{J}_n(\vartheta) - \ddot{J}_n(\zeta) \xrightarrow{P} 0.$$

Тому

$$-\ddot{J}_n(\zeta)/2 \xrightarrow{P} -\mathbf{A}_\infty \text{ при } n \to \infty.$$

Об'єднуючи наші результати, отримуємо:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = -[\ddot{J}_n(\zeta)]^{-1}(\sqrt{n}\dot{J}_n(\vartheta)) \xrightarrow{W} \mathbf{A}_\infty^{-1}\zeta_\infty \sim N(0, \sigma^2 \mathbf{A}_\infty^{-1}).$$

Таким чином, якщо функція регресії є достатьо гладенькою, оцінка МНК — консистентною і всі елементи матриці $\mathbf{A}_\infty$ скінченні, а сама матриця невироджена, то

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{W} N(0, \sigma^2 \mathbf{A}_\infty^{-1}). \tag{3.9}$$

Це і є твердження про асимптотичну нормальність МНК оцінок параметрів нелінійної регресії, а $\sigma^2 \mathbf{A}_\infty^{-1}$ — матриця розсіювання цих оцінок.

3.2 Довірчі інтервали і тести для коефіцієнтів нелінійної регресії

Розглянемо задачу побудови довірчих інтервалів для коефіцієнтів (параметрів) нелінійної регресійної моделі (3.1) — ϑ^i, $i = 1, \ldots, d$. Для цього скористаємось асимптотичною нормальністю оцінок найменших квадратів (3.9).

Позначимо $\mathbf{A}_\infty^{-1} = (\bar{a}_\infty^{im})_{im=1}^d$. За (3.9),

$$\frac{\sqrt{n}(\hat{\vartheta}^i - \vartheta^i)}{\sigma \sqrt{\bar{a}_\infty^{ii}}} \xrightarrow{W} N(0, 1), \tag{3.10}$$

для всіх $i = 1, \ldots, d$. Безпосередньо скористатись цією збіжністю для побудови довірчого інтервалу не можна, оскільки ми не знаємо σ і $\bar{a}_\infty^{ii}$. Знайдемо оцінки для цих характеристик.

Звичайною оцінкою для дисперсії похибок є вибіркова дисперсія залишків:

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{j=1}^n (Y_j - g(\mathbf{X}_j, \hat{\vartheta}_n))^2.$$

Для того, щоб оцінити $\mathbf{A}_\infty$, замінимо у формулі (3.7) математичне сподівання вибірковим середнім, а справжнє значення невідомого параметра ϑ — його оцінкою:

$$\mathbf{A}_n = \frac{1}{n} \sum_{j=1}^n \dot{g}(\mathbf{X}_j, \hat{\vartheta}_n)(\dot{g}(\mathbf{X}_j, \hat{\vartheta}_n))^T.$$

При виконанні досить широких умов, використовуючи закон великих чисел, можна показати, що $\mathbf{A}_n$ є консистентною оцінкою для $\mathbf{A}_\infty$, а $\hat{\sigma}_n^2$

— для σ^2. Тому

$$\frac{\sqrt{n}(\hat{\vartheta}^i - \vartheta^i)}{\hat{\sigma}_n\sqrt{\bar{a}_n^{ii}}} \xrightarrow{W} N(0,1), \tag{3.11}$$

де $\bar{a}_n^{ii}$ — i-й діагональний елемент матриці$\mathbf{A}_n^{-1}$.

З (3.11) випливає, що

$$\leq \mathsf{P}\left\{-\lambda_{\alpha/2} \leq \frac{\sqrt{n}(\hat{\vartheta}^i - \vartheta^i)}{\hat{\sigma}_n\sqrt{\bar{a}_n^{ii}}} \leq \lambda_{\alpha/2}\right\} \to 1-\alpha, \text{ при } n \to \infty.$$

(Тут, як і раніше, $\lambda_\alpha = Q^{N(0,1)}(1-\alpha)$).

Розв'язуючи ці нерівності відносно ϑ^i, знаходимо межі асимптотичного довірчого інтервалу для ϑ^i:

$$\mathsf{P}\{\vartheta^i \in [\vartheta^i_-, \vartheta^i_+]\} \to 1-\alpha,$$

де

$$\vartheta^i_\pm = \hat{\vartheta}^i_n \pm \frac{\lambda_{\alpha/2}\hat{\sigma}_n\sqrt{\bar{a}_n^{ii}}}{\sqrt{n}}. \tag{3.12}$$

Такими довірчими інтервалами можна користуватись для графічної перевірки різних гіпотез про невідомі коефіцієнти регресії.

Наприклад, нехай потрібно перевірити, чи відрізняється коефіцієнт (параметр) ϑ^i значущо від 0, тобто

$$H_0 : \vartheta^i = 0 \text{ проти альтернативи } H_1 : \vartheta^i \neq 0.$$

Для цього можна побудувати довірчий інтервал для ϑ^i використовуючи (3.12) і перевірити, чи потрапляє 0 у цей інтервал.

Якщо $0 \in [\vartheta^i_-, \vartheta^i_+]$ — приймаємо H_0, інакше — відхиляємо.

Зрозуміло, що цей тест можна організувати по іншому: обчислюємо статистику

$$T^i = \frac{\sqrt{n}|\hat{\vartheta}^i_n|}{\hat{\sigma}_n\sqrt{\bar{a}_n^{ii}}},$$

і порівнюємо її з порогом $\lambda_{\alpha/2}$.

Якщо $T^i < \lambda_{\alpha/2}$ — приймаємо H_0, інакше — відхиляємо.

Статистику T^i називають статистикою Стьюдента, а тест - асимптотичним тестом Стюдента для перевірки значущості коефіцієнта ϑ^i. (По аналогії з класичним тестом Стюдента для гауссової лінійної регресії у цьому тесті прийнято використовувати як поріг не $\lambda_{\alpha/2} = Q^{N(0,1)}(1-\alpha/2)$, а $t^{n-d}_{\alpha/2} = Q^{T_{n-d}}(1-\alpha/2)$. Асимптотичний тест застосовується лише при великих обсягах вибірки n, а в цьому випадку квантилі T-розподілу мало відрізняються від квантилів нормального розподілу. Тому використання того чи іншого порогу не створює великих відмінностей).

Досягнутий рівень значущості цього тесту можна підрахувати за формулою

$$\text{p-value} = 2(1 - \Phi(T_i)), \tag{3.13}$$

де Φ — функція розподілу стандартного нормального розподілу.

Інколи буває потрібно перевірити гіпотезу про декілька параметрів одразу. У цьому випадку застосовують асимптотичні аналоги тесту Фішера. Наприклад, нехай треба перевірити гіпотезу

$$H_0 : \vartheta^i = 0 \text{ для всіх } i = 1, \ldots, d,$$

проти альтернативи

$$H_1 : \vartheta^i \neq 0 \text{ для деякого } i.$$

Розглянемо статистику

$$F_n = \frac{n}{\hat{\sigma}^2} \hat{\vartheta}_n^T \mathbf{A}_n^{-1} \hat{\vartheta}_n.$$

Покажемо, що при виконанні H_0, розподіл цієї статистики при $n \to \infty$ прямує до розподілу χ^2_d (хі-квадрат з d ступенями вільності).

Оскільки ми вважаємо, що $\mathbf{A}_n \to \mathbf{A}_\infty$ і $\hat{\sigma}_n^2 \to \sigma^2$, то асимптотична поведінка F_n буде така ж, як і у величин

$$\tilde{F}_n = (\sqrt{n}(\hat{\vartheta}_n - \vartheta))^T \frac{1}{\sigma^2} \mathbf{A}_\infty^{-1} (\sqrt{n}(\hat{\vartheta}_n - \vartheta)).$$

(Тут $\vartheta = 0$ внаслідок виконання H_0). Оскільки

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{W} \eta \sim N(0, \mathbf{V}),$$

де $\mathbf{V} = \sigma^2 \mathbf{A}$, то

$$F_n \xrightarrow{W} F_\infty = \eta^T \mathbf{V}^{-1} \eta.$$

Позначимо $\mathbf{V}^{-1/2}$ таку симетричну $d \times d$ матрицю, що $\mathbf{V}^{-1/2}\mathbf{V}^{-1/2} = \mathbf{V}^{-1}$. Тоді

$$F_n = (\mathbf{V}^{-1/2}\eta)^T (\mathbf{V}^{-1/2}\eta) = \|\mathbf{V}^{-1/2}\eta\|.$$

Випадковий вектор $\mathbf{V}^{-1/2}\eta$ є гауссовим з нульовим математичним сподіванням. Його коваріаційна матриця

$$\mathrm{Cov}(\mathbf{V}^{-1/2}\eta) = \mathsf{E}(\mathbf{V}^{-1/2}\eta)(\mathbf{V}^{-1/2}\eta)^T = \mathbf{V}^{-1/2}\,\mathsf{E}\,\eta\eta^T\mathbf{V}^{-1/2} = \mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2} = \mathbf{E},$$

де $\mathbf{E}$ — одинична матриця. Отже координати η^i цього вектора є незалежними стандартними випадковими величами. Тому $\|\mathbf{V}^{-1/2}\eta\| = \sum_{i=1}^{d}(\eta^i)^2 \sim \chi^2_d$.

Таким чином, при виконанні H_0, $F_n \xrightarrow{W} \chi^2_d$. Тому

$$\mathsf{P}_{H_0}\{F_n > Q^{\chi^2_d}(1-\alpha)\} \to \alpha.$$

Це дає нам можливість побудувати тест для перевірки H_0, який приймає H_0, якщо $F_n \le Q^{\chi^2_d}(1-\alpha)$ і відхиляє в іншому випадку. Цей тест матиме асимптотичний рівень значущості α.

3.3 Логістична та пробіт регресія з бінарним відгуком

У прикладній статистиці часто виникають задачі дослідження впливу деяких дійснозначних змінних $X^1,\dots X^m$ на змінну Y, що приймає тільки одне з двох значень (так/ні, успіх/невдача, хворий/здоровий, склав залік/не склав). Далі, для зручності позначень ми будемо вважати, що ці значення є 0 або 1. Для опису таких залежностей можна використовувати моделі, у яких описується умовна ймовірність того, що змінна Y приймає те чи інше значення, при фіксованих регресорах X^i, $i = 1\dots,m$. Ми одразу будемо розглядати набір (вибірку) з n спостережень $Z_j = (Y_j, X_j^1,\dots,X_j^m)^T$, де j — номер об'єкта, для якого вимірюються відповідні змінні. Як і раніше, позначимо $\mathbf{X} = (X^1,\dots,X^m)^T$ — вектор значень всіх регресорів.

Модель може мати такий вигляд:

$$\mathsf{P}\{Y_j = 1\ |\mathbf{X}\} = g(b_0 + \mathbf{b}^T\mathbf{X}_j), \tag{3.14}$$

де g — фіксована відома функція (функція зв'язку), b_0, $\mathbf{b} = (\mathbf{b}^1,\dots,\mathbf{b}^m)^T$ — невідомі коефіцієнти, які потрібно оцінити за спостереженнями.

Для того, щоб оцінити ці коефіцієнти, можна скористатись методом найбільшої вірогідності. У даній моделі розподіл регресорів не залежить від коефіцієнітв, які ми оцінюємо. Тому складемо функцію вірогідності, розглядаючи як випадкові лише значення відгуку Y_j. Оскільки відгук є дискретною випадковою величиною, використовуємо щільності відносно лічильної міри, тобто функція вірогідності буде дорівнювати ймовірності події, що спостерігається у вибірці:

$$L(b_0, \mathbf{b}) = \prod_{j=1}^{n} (g(b_0 + \mathbf{b}^T \mathbf{X}_j))^{Y_j} (1 - g(b_0 + \mathbf{b}^T \mathbf{X}_j))^{1-Y_j}.$$

Оцінка методу найбільшої вірогідності визначається як точка мінімуму цієї функції по b_0 і $\mathbf{b}$:

$$(\hat{b}_0, \hat{\mathbf{b}}) = \operatorname{argmax}_{b_0, \mathbf{b}} L(b_0, \mathbf{b}).$$

Як правило, знайти цю оцінку аналітично не можна, її підраховують чисельними методами для конкретних даних. Використовуючи методи асимптотичної статистики, можна на основі цих оцінок будувати довірчі інтервали для справжніх значень коефіцієнтів та тести для перевірки гіпотез про них.

Розглянемо приклади конкретних функції регресії g для таких моделей з бінарним відгуком.

3.3.1 Задача класифікації і логістична регресія

Задачі діагностики часто виникають у прикладних статистичних дослідженнях, наприклад, у медичній діагностиці.

Нехай спостережувані об'єкти (наприклад, пацієнти) можуть відноситись до одного з двох класів, які ми позначимо 0 (здорові, не хворі даною хворобою) і 1 (хворі даною хворобою). У кожного об'єкта спостерігається набір деяких характеристик $\mathbf{X} = (X^1, \ldots, X^m)$ (скажімо, результати аналізу крові пацієнта). Потрібно за цим набором визначити, до якого класу належить об'єкт. Цей номер позначимо Y.

Це — задача класифікації об'єктів за їх спостережуваними характеристиками.

Будемо вважати, що для об'єктів кожного класу спостережувані характеристики мають нормальний розподіл, тобто їх щільність для i-го класу

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{m/2} \sqrt{\det \mathbf{S}}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mu_i)^T \mathbf{S}^{-1} (\mathbf{x} - \mu_i) \right),$$

де $\mu_i = (\mu_i^1, \dots, \mu_i^m)$ — вектор математичних сподівань характеристик i-того класу; $\mathbf{S}$ — коваріаційна матриця векторів характеристик, яка вважається однаковою для обох класів.

Ми також будемо вважати, що існують апріорні ймовірності того, що спостережуваний об'єкт належить i-тому класу:

$$\pi_i = \mathsf{P}\{Y = i\}.$$

(У нашому прикладі, π_1 — ймовірність того, що навмання вибраний пацієнт хворий хворобою, що діагностується, $\pi_0 = 1 - \pi_1$ — ймовірність того, що цієї хвороби у нього немає).

Ці ймовірності називають апріорними, вони не залежать від спостережуваних характеристик об'єкта і характеризують загальну інформацію про розподіл розглядуваних об'єктів по класах. Після того, як отримані значення спостережуваних характеристик даного об'єкта, ми можемо використати їх для уточнення наших уявлень про можливості для нього належати тому чи іншому класу. Ці уточнення описуються апостеріорними ймовірностями:

$$p_i(\mathbf{x}) = \mathsf{P}\{Y = i \mid \mathbf{X} = \mathbf{x}\}$$

— умовна ймовірність того, що об'єкт належить i-му класу, за умови, що його характеристики дорівнюють $\mathbf{x}$.

Якщо ми хочемо визначити, як класифікувати об'єкт, то це зручно робити саме за апостеріорними ймовірностями: якщо $p_1(\mathbf{x}) > p_0(\mathbf{x})$ об'єкт відносять до першого класу, інакше — до другого. Такий спосіб називають баєсовою класифікацією. Оскільки $p_0(\mathbf{x}) = 1 - p_1(\mathbf{x})$, цей метод класифікації вибирає перший клас, якщо $p_1(\mathbf{x}) > 1/2$, і нульовий клас, якщо $p_1(\mathbf{x}) < 1/2$. У випадку, коли $p_1(\mathbf{x}) = 1/2$ можна віднести об'єкт до будь-якого класу, наприклад, підкинувши монетку.

Підрахуємо, чому дорівнюють апостеріорні ймовірності у нашому випадку:

$$\begin{aligned} p_1(\mathbf{x}) &= \frac{\pi_1 f_1(\mathbf{x})}{\pi_1 f_1(\mathbf{x}) + \pi_0 f_0(\mathbf{x})} \\ &= \frac{1}{1 + \pi_0 f_0(\mathbf{x})/(\pi_1 f_1(\mathbf{x}))} \\ &= \frac{1}{1 + \exp(-(b_0 + \mathbf{b}^T\mathbf{x}))}, \end{aligned}$$

де

$$b_0 = -\log\frac{\pi_0}{\pi_1} + \frac{1}{2}(\mu_0^T\mathbf{S}^{-1}\mu_0 - \mu_1^T\mathbf{S}^{-1}\mu_1),\ \mathbf{b} = \mathbf{S}^{-1}(\mu_0 - \mu_1).$$

Таким чином, ми отримали для залежності між Y та $\mathbf{X}$ формулу виду (3.14) з функцією зв'язку

$$g(t) = \mathrm{Logist}(t) = \frac{1}{1+\exp(-t)}.$$

Функцію $\mathrm{Logist}(t)$ називають логістичною функцією, а модель (3.14) з цією функцією зв'язку — моделлю логістичної регресії з бінарним відгуком.

Модель логістичної регресії часто використовують і до даних, які не пов'язані із задачею класифікації.

3.3.2 Пробіт-регресія

Розглянемо тепер інший механізм породження спостережуваних даних, який також приводить до моделі регресії з бінарним відгуком.

Нехай регресори $\mathbf{x} = (X^1, \dots, X^m)$ пов'язані з деякою величиною η моделлю гауссової регресії:

$$\eta = \beta_0 + \sum_{i=1}^m \beta_i X^i + \varepsilon,$$

де ε — гауссова випадкова величина з нульовим математичним сподіванням і дисперсією σ^2.

Ми спостерігаємо регресори X^i, але замість η спостерігається тільки $Y = \mathbb{1}\{\eta < c\}$, де c — деякий поріг.

Скажімо, η може бути температурою поверхні землі на дворі у даний момент спостережень, але ми не маємо термометра, щоб її виміряти, тому відмічаємо лише, чи замерзла вода у калюжі. Якщо замерзла, то $\eta < 0$ і $Y = 1$, якщо ні — $\eta > 0$ і $Y = 0$.

Підрахуємо

$$\mathsf{P}\{Y = 1 \mid \mathbf{X} = \mathbf{x}\} = \mathsf{P}\{\beta_0 + \sum_{i=1}^m \beta_i x^i + \varepsilon < c\}$$

$$= \mathsf{P}\left\{\frac{\varepsilon}{\sigma} < \frac{c-\beta_0}{\sigma} + \sum_{i=1}^m \frac{-\beta_i}{\sigma} x^i\right\}.$$

Отже

$$\mathsf{P}\{Y = 1 \mid \mathbf{X} = \mathbf{x}\} = \Phi(b_0 + \sum_{i=1}^{m} b_i x^i), \tag{3.15}$$

де

$\Phi(x)$ — функція розподілу стандартного нормального розподілу,

$$b_0 = \frac{c - \beta_0}{\sigma}, \; b_i = \frac{-\beta_i}{\sigma}.$$

Модель (3.15) називають моделлю пробіт регресії. Це — частковий випадок моделі (3.14) з функцією зв'язку $g(t) = \Phi(t)$. Оцінюючи коефіцієнти у цій моделі за методом найбільшої вірогідності можна будувати прогнозні формули для прогнозування ймовірності того, що $Y = 1$, в залежності від значень регресорів.

Бібліоґрафія

[1] Боровков А.А. (1997) Математическая статистика.

[2] Майборода Р.Є. "Регресія: лінійні моделі". ВПЦ “Київський університет”, 296 p. - 2007

http://probability.univ.kiev.ua/userfiles/mre/ora0.pdf

[3] Майборода Р. Є. Комп’ютерна статистика/ 2017.- 405с. Режим доступу:

http://probability.univ.kiev.ua/userfiles/mre/compsta.pdf

[4] Майборода Р.Є. Самостійна робота по курсу “Регресійний аналіз” . Індивідуальні завдання та рекомендації по їх виконанню.

http://probability.univ.kiev.ua/userfiles/mre/regr.pdf

[5] Майборода Р.Є. Самостійна робота по курсу “Асимптотична статистика”.

http://probability.univ.kiev.ua/userfiles/mre/asympt.pdf

[6] Shao J. Mathematical statistics.- Springer-Verlag: New York, 1998. - 530 p.