

Міністерство освіти і науки України
Луцький національний технічний університет
(повне найменування вищого навчального закладу)

Факультет архітектури, будівництва та дизайну
(повне найменування факультету)

Кафедра прикладної математики та механіки
(повне найменування кафедри)

КВАЛІФІКАЦІЙНА РОБОТА
ЗА СТУПЕНЕМ ВИЩОЇ ОСВІТИ «МАГІСТР»

**ВЛАСТИВОСТІ ОЦІНОК МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ:
УЗГОДЖЕНІСТЬ, АСИМПТОТИЧНА НОРМАЛЬНІСТЬ ТА ІНФОРМАЦІЯ ФІШЕРА
ІЗ ЗАСТОСУВАННЯМ ДО РОЗПОДІЛІВ БЕРНУЛЛІ ТА ЕКСПОНЕНЦІАЛЬНОГО
РОЗПОДІЛУ**

**PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS: CONSISTENCY,
ASYMPTOTIC NORMALITY, AND FISHER INFORMATION WITH APPLICATIONS
TO BERNOULLI AND EXPONENTIAL DISTRIBUTIONS**

спеціальність 113 Прикладна математика
(шифр і назва спеціальності)

освітня програма Прикладна математика
(назва освітньої програми)

Виконав: здобувач вищої освіти
Групи ПРМм-21
Штуль Валентин Степанович

(підпис)

Керівник:
PhD., доцент
Самоненко Інга Вікторівна

(підпис)

Кваліфікаційну роботу
Допущено до захисту
«__» _____ 2025р.
PhD., доцент
Гарант освітньої програми:
Самоненко Інга Вікторівна

(підпис)

Луцьк – 2025 року

ЛУЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

Факультет архітектури, будівництва та дизайну

Кафедра прикладної математики та механіки

Ступінь вищої освіти: магістр

Спеціальність: 113 Прикладна математика

Освітня програма: Прикладна математика

ЗАТВЕРДЖУЮ

Завідувач кафедри

прикладної математики та
механіки

_____ Мікуліч О.А.

"__" _____ 2025 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧУ ВИЩОЇ ОСВІТИ

Штуль Валентин Степанович

(прізвище, ім'я, по батькові)

1. Тема науково-проектної роботи

Властивості оцінок максимальної правдоподібності: узгодженість, асимптотична нормальність та інформація Фішера із застосуванням до розподілів Бернуллі та експоненціального розподілу

Керівник роботи: PhD., доцент Самоненко І. В.

затверджені наказом закладу вищої освіти від 25.01.2025 № 45/01-02

2. Строк подання здобувачем вищої освіти кваліфікаційної роботи

«__» _____ 2025р.

3. Вихідні дані до роботи: *методичні вказівки до оформлення кваліфікаційних робіт, технічна документація до бібліотек і мови програмування Python, наукові статті та публікації*

4. Зміст пояснювальної записки (перелік питань, які потрібно розробити):
Метод максимальної правдоподібності: основи та історичний розвиток, теоретичні властивості оцінок максимальної правдоподібності, Застосування методу максимальної правдоподібності до конкретних розподілів, Альтернативний погляд: Байєсівське оцінювання та зв'язок з машинним навчанням, Програмна реалізація та статистичний аналіз експериментальних даних.

5. Перелік графічного (ілюстративного) матеріалу: *Залежність зміщення та дисперсії оцінки від обсягу вибірки, Q-Q plots. Зліва Експоненціальний розподіл, справа Логнормальний, Порівняння теоретичних щільностей розподілу з гістограмою даних, Емпіричні та теоретичні функції розподілу, Залежність Інформації Фішера від базової конверсії, Перевірка асимптотичної нормальності оцінок ММП, Збіжність емпіричних частот до теоретичних*

ймовірностей, Спостережувані частки конверсій, Апостеріорні щільності розподілу ймовірностей конверсії, Траєкторія P-value залежно від обсягу вибірки, Крива статистичної потужності: здатність тесту виявляти ефект.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис	
		Завдання видав	Завдання прийняв
Розділ 1.	Самоненко І. В. PhD., доцент		
Розділ 2.	Самоненко І. В. PhD., доцент		
Розділ 3.	Самоненко І. В. PhD., доцент		
Розділ 4.	Самоненко І. В. PhD., доцент		
Розділ 5.	Самоненко І. В. PhD., доцент		
Висновки	Самоненко І. В. PhD., доцент		

7. Дата видачі завдання 25 січня 2025 року

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи магістра	Строк виконання етапів роботи	Примітка
1.	<i>Обґрунтування теми</i>		
2.	<i>Огляд літератури із досліджуваної проблеми</i>		
3.	<i>Перший розділ</i>		
4.	<i>Другий розділ</i>		
5.	<i>Третій розділ</i>		
6.	<i>Четвертий розділ</i>		
7.	<i>П'ятий розділ</i>		
8.	<i>Висновки та пропозиції</i>		
9.	<i>Формування списку використаних джерел</i>		
10.	<i>Формування додатків</i>		
11.	<i>Оформлення ілюстративного матеріалу</i>		
12.	<i>Нормоконтроль</i>		
13.	<i>Інструментальна перевірка на академічний плагіат</i>		<i>Показник запозичень тексту _____ %</i>
14.	<i>Представлення кваліфікаційної роботи магістра до захисту</i>		

Здобувач вищої освіти _____ (підпис) _____ (прізвище та ініціали)

Керівник кваліфікаційної роботи _____ (підпис) _____ (прізвище та ініціали)

АНОТАЦІЯ

Штуль В. С. Властивості оцінок максимальної правдоподібності: узгодженість, асимптотична нормальність та інформація Фішера із застосуванням до розподілів Бернуллі та експоненціального розподілу. Рукопис.

Кваліфікаційна робота магістра ОП «Прикладна математика» спеціальності «113 Прикладна математика». Луцький національний технічний університет. Луцьк, 2025.

Кваліфікаційна робота магістра складається з вступу, 5 розділів, висновків та списку використаних джерел.

Робота присвячена поглибленому аналізу методу максимальної правдоподібності як основного інструменту параметричного оцінювання в математичній статистиці. Досліджується проблема визначення точності, надійності та стійкості статистичних оцінок в умовах різного обсягу вибірки. Основна увага зосереджена на аналізі таких властивостей як узгодженість, асимптотична нормальність та ефективність за критерієм Фішера.

У ході дослідження застосовано методи теорії ймовірностей, математичної статистики, чисельні методи оптимізації та комп'ютерне моделювання. Проведено порівняльний аналіз класичного частотного підходу та байєсівського оцінювання.

Отримані результати практично підтверджують теоретичну поведінку оцінок для розподілів Бернуллі та експоненціального розподілу. Вперше реалізовано програмний алгоритм мовою Python для порівняльного аналізу реальних наборів даних, зокрема тривалості поїздок таксі та конверсії веб-користувачів. Показано переваги логнормального розподілу над експоненціальним для моделювання часових інтервалів у реальних системах. Виконаний аналіз підтвердив, що метод максимальної правдоподібності забезпечує мінімальну дисперсію оцінок на асимптотиці.

Ключові слова: МЕТОД МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ, ІНФОРМАЦІЯ ФІШЕРА, РОЗПОДІЛ БЕРНУЛЛІ, ЕКСПОНЕНЦІАЛЬНИЙ РОЗПОДІЛ, УЗГОДЖЕНІСТЬ, АСИМПТОТИЧНА НОРМАЛЬНІСТЬ, СТАТИСТИЧНЕ МОДЕЛЮВАННЯ

ABSTRACT

Shtul V. S. Properties of Maximum Likelihood Estimators: Consistency, Asymptotic Normality, and Fisher Information with applications to Bernoulli and exponential distributions. Manuscript.

Master's thesis in Applied Mathematics, specialty "113 Applied Mathematics." Lutsk National Technical University. Lutsk, 2025.

The master's thesis consists of an introduction, 5 chapters, conclusions, and a list of references.

The thesis is devoted to an in-depth analysis of the maximum likelihood method as the main tool for parametric estimation in mathematical statistics. It examines the problem of determining the accuracy, reliability, and stability of statistical estimates under conditions of varying sample sizes. The main focus is on the analysis of properties such as consistency, asymptotic normality, and efficiency according to Fisher's criterion.

The study uses methods of probability theory, mathematical statistics, numerical optimization methods, and computer modeling. A comparative analysis of the classical frequency approach and Bayesian estimation is performed.

The results obtained practically confirm the theoretical behavior of estimates for Bernoulli and exponential distributions. For the first time, a software algorithm in Python has been implemented for comparative analysis of real data sets, in particular taxi trip durations and web user conversions. The advantages of lognormal distribution over exponential distribution for modeling time intervals in real systems are demonstrated. The analysis confirmed that the maximum likelihood method provides minimal dispersion of estimates at asymptotics.

Keywords: MAXIMUM LIKELIHOOD METHOD, FISHER INFORMATION, BERNOULLI DISTRIBUTION, EXPONENTIAL DISTRIBUTION, COHERENCE, ASYMPTOTIC NORMALITY, STATISTICAL MODELING

ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1. МЕТОД МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ: ОСНОВИ ТА ІСТОРИЧНИЙ РОЗВИТОК.....	11
1.1. Виникнення методу максимальної правдоподібності.....	11
1.2. Внесок Рональда Фішера у розвиток статистики	11
1.3. Формальне визначення методу максимальної правдоподібності	12
1.4. Властивості функції правдоподібності та лог-правдоподібності	13
РОЗДІЛ 2. ТЕОРЕТИЧНІ ВЛАСТИВОСТІ ОЦІНОК МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ	14
2.1. Узгодженість.....	14
2.1.1. Формальне визначення та умови регулярності.....	14
2.1.2. Теорема про узгодженість та схема доведення.....	15
2.1.3. Практичне значення узгодженості ММП.....	16
2.2. Асимптотична нормальність.....	16
2.2.1. Теоретичне обґрунтування.....	16
2.2.2. Зв'язок з центральною граничною теоремою	17
2.2.3. Побудова довірчих інтервалів і статистичних тестів.....	17
2.3. Інформація Фішера	18
2.3.1. Визначення та властивості	18
2.3.2. Матриця інформації Фішера.....	18
2.3.3. Зв'язок з дисперсією та ефективністю ММП.....	18
2.4. Обмеження та проблемні аспекти методу максимальної правдоподібності	19
2.4.1. Зміщеність оцінок для скінченних вибірок.....	19
2.4.2. Ефективність на малих вибірках та середньоквадратична помилка ...	20
2.4.3. Проблема існування максимуму та повна роздільність.....	20
2.4.4. Локальні максимуми та неувігнутість функції	21
РОЗДІЛ 3. ЗАСТОСУВАННЯ МЕТОДУ МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ ДО КОНКРЕТНИХ РОЗПОДІЛІВ.....	22
3.1. Розподіл Бернуллі	22
3.1.1. Функція правдоподібності та її максимізація	22
3.1.2. Виведення ММП для параметра p	23
3.1.3. Узгодженість та асимптотична нормальність оцінки	23

3.1.4. Інформація Фішера для розподілу Бернуллі	24
3.1.5. Статистичні критерії та перевірка гіпотез.....	24
3.2. Експоненціальний розподіл	26
3.2.1. Функція правдоподібності та максимізація	26
3.2.2. Виведення ММП для параметра λ	27
3.2.3. Узгодженість та асимптотична нормальність оцінки	27
3.2.4. Інформація Фішера для експоненціального розподілу	28
РОЗДІЛ 4. АЛЬТЕРНАТИВНИЙ ПОГЛЯД: БАЙЄСІВСЬКЕ ОЦІНЮВАННЯ ТА ЗВ'ЯЗОК З МАШИННИМ НАВЧАННЯМ	29
4.1. Концептуальні відмінності та Баєсівський висновок.....	29
4.2. Застосування до розподілу Бернуллі: Спряжені розподіли.....	29
4.3. Оцінки MAP та зв'язок з регуляризацією	30
РОЗДІЛ 5 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА СТАТИСТИЧНИЙ АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ	32
5.1. Опис програмного забезпечення	32
5.2 Дослідження ММП для неперервних розподілів.....	33
5.2.1. Попередня обробка вхідних даних.....	33
5.2.2. Реалізація методу максимальної правдоподібності	33
5.2.3. Порівняльний аналіз моделей.....	35
5.2.4. Статистична перевірка гіпотез	36
5.2.5. Дослідження властивостей оцінок	37
5.2.6. Візуалізація результатів	39
5.3. Дослідження ефективності оцінок ММП для дискретних розподілів.....	42
5.3.1. Опис даних та генерація вибірки	43
5.3.2. Методологія оцінювання параметрів.....	44
5.3.3. Детальний аналіз метрик та їх бізнес-інтерпретація.....	46
5.3.5. Теоретичний аналіз статистичної потужності та верифікація надійності критерію	55
ВИСНОВКИ.....	58
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	60

ВСТУП

Ідеї побудови моделей забезпечують найкраще пояснення наявних даних для сучасної статистики та машинного навчання. Метод максимальної правдоподібності – також відомий як оцінка максимальної правдоподібності (ММП) – є одним із найефективніших методів для розв'язування конкретної задачі. Метод дозволяє знайти найбільш ймовірні значення параметрів моделі [1]. Цей метод лежить в основі багатьох статистичних імовірнісних моделей і є важливою частиною сучасних методів аналізу даних, штучного інтелекту та економічного прогнозування [2,3].

Вперше ММП було запропоновано Рональдом Фішером у 1922 році [1], і з того часу він став фундаментом теорії параметричного оцінювання. Виявилося, що метод дуже універсальний. Він працює як з простими моделями (нормальними чи експоненціальними розподілами Бернуллі), так із складними моделями, такими як нейронні мережі та баєсівські мережі [4]. Максимізація функції правдоподібності залишається основою ідеєю для знаходження параметрів, які найкраще описують природу даних.

Властивості ММП роблять його практично придатним для реальних статистичних задач, що робить його важливим.

По-перше, ММП є узгодженим, тобто при збільшенні розміру вибірки оцінка параметра сходиться до його істинного значення. Це означає, що зі зростанням кількості даних результати стають точнішими, що є критично важливим для аналітики великих даних і статистичних досліджень [5].

По-друге, ММП є асимптотично нормальним, тобто розподіл оцінок наближається до нормального, що дозволяє побудувати довірчі інтервали та проводити статистичні тести [6].

І нарешті, ММП є ефективним, оскільки при великій кількості спостережень досягає межі Крамера–Рао, тобто має мінімальну можливу дисперсію серед усіх регулярних оцінок [7,8].

У практиці статистичного моделювання вагому роль відводять інформації Фішера, яка характеризує точність оцінювання та визначає обсяг «інформації» про невідомий параметр. Високі значення цього показника зазвичай гарантують стабільність і надійність результатів, тоді як низький рівень інформативності часто унеможлиблює точне оцінювання навіть за наявності значної вибірки. Описану властивість активно застосовують для оптимізації моделей машинного навчання, аналізу чутливості параметрів та розробки ефективних алгоритмів [3,4,18].

Важливим аспектом є практичне застосування ММП. Наприклад, розподіл Бернуллі використовується для моделювання двійкових подій – «успіх/невдача», «так/ні», «дефектний/нормальний виріб» – і має широке застосування у соціології, економіці та біостатистиці. З його допомогою можна оцінювати ймовірність позитивної відповіді в опитуваннях або частку дефектної продукції у виробництві.

Експоненціальний розподіл, у свою чергу, слугує інструментом для моделювання часу між подіями, що охоплює період до виходу техніки з ладу, тривалість очікування клієнта в черзі або паузи між транзакціями. Подібні задачі є характерними для інженерної практики, сфери сервісного обслуговування, економіки, а також систем екологічного моніторингу та фінансового аналізу [2].

Завдяки універсальності та математичній строгості ММП використовується навіть у випадках неповних або шумних даних. Алгоритм EM (Expectation–Maximization), заснований на ідеях ММП, дозволяє проводити оцінювання, коли частина інформації недоступна, що робить метод незамінним у практиці обробки реальних даних [9].

Попри очевидну ефективність, застосування методу наштовхується на низку об'єктивних обмежень. Зокрема, робота з малими масивами даних або наявність сильної кореляції часто призводить до отримання зміщених чи нестійких оцінок. Додатковим ускладненням виступає відсутність гарантії єдиного максимуму функції правдоподібності, що створює перешкоди для

процедур чисельної оптимізації. Зазначені фактори зумовлюють необхідність залучення модифікованих або байєсівських підходів, які дозволяють врахувати апріорну інформацію [6].

Об'єкт дослідження – процес статистичного оцінювання параметрів імовірнісних розподілів.

Предмет дослідження – властивості оцінок, отриманих методом максимальної правдоподібності, та їх застосування до окремих розподілів.

Мета роботи – здійснити поглиблений аналіз властивостей оцінок методу максимальної правдоподібності, зокрема їх узгодженості, асимптотичної нормальності та ефективності. Теоретичні висновки проілюструвати на прикладі розподілів Бернуллі та експоненціального, дослідивши точність та стійкість отриманих оцінок.

Наукова новизна полягає у поєднанні класичної теорії ММП із сучасними чисельними методами та практичними прикладами застосування до реальних даних.

Практичне значення полягає у демонстрації того, як ММП можна використовувати для аналізу бінарних подій, оцінювання параметрів розподілів часу та побудови надійних статистичних моделей.

РОЗДІЛ 1. МЕТОД МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ: ОСНОВИ ТА ІСТОРИЧНИЙ РОЗВИТОК

1.1. Виникнення методу максимальної правдоподібності

Метод максимальної правдоподібності був запропонований Рональдом Айлом Фішером у 1922 році як універсальний спосіб оцінювання параметрів статистичних моделей [1]. Його поява стала одним із найважливіших моментів у розвитку математичної статистики ХХ століття.

На відміну від попередніх підходів (наприклад, методу моментів або найменших квадратів), метод максимальної правдоподібності базується на імовірнісному принципі: обираються ті параметри моделі, за яких спостережувані дані є найбільш імовірними. Формально, якщо відома функція щільності ймовірності $f(x; \theta)$, то оцінка максимальної правдоподібності $\hat{\theta}$ – це значення параметра, що максимізує добуток ймовірностей усіх спостережень:

$$\hat{\theta} = \arg \max_{\theta} L(\theta), \text{ де } L(\theta) = \prod_{i=1}^n f(x_i; \theta). \quad (1.1)$$

Фішер продемонстрував, що цей метод має низку важливих властивостей – зокрема, узгодженість, асимптотичну нормальність та ефективність. Його підхід об'єднав теорію ймовірностей із практичною статистикою, створивши фундамент для подальшого розвитку байєсівських і частотних методів аналізу [7,10].

1.2. Внесок Рональда Фішера у розвиток статистики

Рональд Фішер відіграв вирішальну роль не лише у створенні методу ММП, а й у формуванні всієї сучасної математичної статистики. Саме йому наука завдячує введенням таких фундаментальних понять, як дисперсійний аналіз (ANOVA), ефективність оцінок, дизайн експериментів та безпосередньо концепція максимальної правдоподібності.

Особливе місце серед ідей вченого посідає інформація Фішера, яка визначає обсяг «інформації» про невідомий параметр у спостережуваних даних.

Застосування такої величини дозволяє кількісно виміряти точність оцінювання, оскільки високі показники свідчать про можливість отримати точний результат, тоді як низькі вказують на нестійкість оцінки. Завдяки цій властивості інформація Фішера стала ключовим інструментом для аналізу якості статистичних моделей [1,7].

Важливим аспектом роботи Фішера було також підкреслення зв'язку між абстрактною теорією та практичним застосуванням. Метод повинен працювати не лише в ідеальних умовах, а й при обробці реальних, часто обмежених або неповних даних. Такі принципи згодом лягли в основу сучасного машинного навчання, де ММП активно використовують для налаштування параметричних моделей широкого спектра, від логістичної регресії до глибоких нейронних мереж [3,11].

1.3. Формальне визначення методу максимальної правдоподібності

Нехай X_1, X_2, \dots, X_n – незалежні випадкові величини з функцією щільності $f(x; \theta)$, де θ – невідомий параметр.

Функція правдоподібності визначається як:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) \quad (1.2)$$

а її логарифм – як:

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^n \ln f(x_i; \theta) \quad (1.3)$$

Оцінка максимальної правдоподібності $\hat{\theta}$ – це значення параметра, при якому функція $\ell(\theta)$ досягає максимуму:

$$\frac{\partial \ell(\theta)}{\partial \theta} = 0 \quad (1.4)$$

На практиці таке рівняння зазвичай розв'язують чисельними способами, зокрема через використання градієнтних підходів або методу Ньютона-Рафсона. У ситуаціях з частковою відсутністю даних ефективним інструментом стає алгоритм очікування-максимізації [9].

1.4. Властивості функції правдоподібності та лог-правдоподібності

Функція правдоподібності має низку важливих властивостей, що визначають її поведінку:

1. Невід'ємність – $L(\theta) \geq 0$ для всіх можливих θ .
2. Монотонність логарифму – оскільки логарифм є монотонно зростаючою функцією, то максимізація $\ell(\theta)$ еквівалентна максимізації $L(\theta)$.
3. Адитивність лог-правдоподібності – для незалежних спостережень лог-правдоподібність є сумою окремих складових, що спрощує обчислення.
4. Опуклість або увігнутість – форма лог-правдоподібності визначає унікальність оцінки: якщо функція увігнута, то існує єдиний максимум.

Зазначені властивості забезпечують ефективність методу як при аналітичному дослідженні, так і під час чисельної реалізації. Використання логарифмічної функції стає незамінним у ситуаціях, коли прямий розв'язок знайти неможливо, що дозволяє стабільно визначати параметри навіть у моделях високої розмірності [6].

Таким чином, метод максимальної правдоподібності поєднує в собі математичну строгість, інтуїтивну зрозумілість і універсальність. Його властивості роблять його одним із найважливіших інструментів сучасної статистики, а також основою для алгоритмів машинного навчання, які сьогодні застосовуються у фінансовій аналітиці, біоінформатиці, промисловій діагностиці та багатьох інших галузях [4,11].

РОЗДІЛ 2. ТЕОРЕТИЧНІ ВЛАСТИВОСТІ ОЦІНОК МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ

Метод максимальної правдоподібності вирізняється серед інших методів оцінювання параметрів тим, що його оцінки мають низку теоретичних властивостей, які забезпечують їх надійність та ефективність. Основними є узгодженість, асимптотична нормальність та ефективність, тісно пов'язана з поняттям інформації Фішера. Описані властивості мають не лише теоретичне, а й практичне значення, адже гарантують, що при збільшенні обсягу вибірки оцінки, отримані за допомогою ММП, наближаються до істинних параметрів моделі [5].

2.1. Узгодженість

2.1.1. Формальне визначення та умови регулярності

Оцінка $\hat{\theta}_n$ називається узгодженою, якщо при збільшенні обсягу вибірки $n \rightarrow \infty$ вона збігається за ймовірністю до істинного значення параметра θ_0 :

$$\hat{\theta}_n \xrightarrow{P} \theta_0 \text{ при } n \rightarrow \infty \quad (2.1)$$

Іншими словами, чим більше ми маємо спостережень, тим точнішою стає оцінка параметра. Це властивість робить ММП надзвичайно корисним у реальних умовах, де часто доступні великі обсяги даних [2].

Щоб оцінка ММП була узгодженою, мають виконуватися так звані регулярні умови:

1. модель повинна бути правильно специфікованою (істинний параметр θ_0 належить параметричному простору);
2. функція лог-правдоподібності має бути диференційовною за параметром;
3. істинне значення параметра є єдиним максимумом функції правдоподібності;
4. вибірка є незалежною та однаково розподіленою.

2.1.2. Теорема про узгодженість та схема доведення

Центральним результатом асимптотичної теорії є теорема про те, що за виконання умов регулярності глобальний максимум функції правдоподібності збігається до істинного параметра. Схема доведення базується на законі великих чисел та властивостях нерівності Єнсена.

Для доведення узгодженості $\hat{\theta}_{MLE} \rightarrow \theta_0$ критично важливою є умова ідентифіковності. Це означає, що відображення параметрів у розподіли є ін'єктивним:

$$\theta \neq \theta_0 \Rightarrow P_\theta \neq P_{\theta_0} \quad (2.2)$$

Також припускається компактність простору параметрів, що гарантує існування максимуму на множині Θ .

Розглянемо нормовану функцію лог-правдоподібності:

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(X_i | \theta) \quad (2.3)$$

Згідно зі слабким законом великих чисел, для кожного фіксованого θ ця сума збігається за ймовірністю до математичного сподівання:

$$M_n(\theta) \xrightarrow{P} E_{\theta_0} [\log f(X | \theta)] \equiv M(\theta) \quad (2.4)$$

Для завершення доказу використовується поняття рівномірної збіжності функції $M_n(\theta)$ до $M(\theta)$ та аналіз максимуму граничної функції. Розглянемо різницю сподівань у довільній точці та в точці істинного параметра:

$$M(\theta) - M(\theta_0) = E_{\theta_0} \left[\log \frac{f(X | \theta)}{f(X | \theta_0)} \right] \quad (2.5)$$

Застосовуючи нерівність Єнсена для угнутої функції логарифма $E[\log Y] \leq \log E[Y]$, отримуємо:

$$E_{\theta_0} \left[\log \frac{f(X|\theta)}{f(X|\theta_0)} \right] \leq \log E_{\theta_0} \left[\frac{f(X|\theta)}{f(X|\theta_0)} \right] = \log(1) = 0 \quad (2.6)$$

Звідси випливає фундаментальна нерівність $M(\theta) \leq M(\theta_0)$. Оскільки рівність досягається лише при $\theta = \theta_0$, істинний параметр є унікальним глобальним максимумом граничної функції. Оскільки вибіркова функція рівномірно збігається до граничної, її максимум $\hat{\theta}_{MLE}$ також збігається до θ_0 , що й доводить узгодженість.

2.1.3. Практичне значення узгодженості ММП

У прикладному вимірі узгодженість виступає запорукою коректної роботи з великими масивами інформації (Big Data). Фактично вона гарантує, що за умови достатньої кількості даних метод ММП здатен ефективно відновлювати приховані залежності навіть у складних архітектурах. Зазначена особливість є визначальною для сучасного машинного навчання, зокрема при тренуванні нейронних мереж чи класифікації текстів [3, 11].

Гарантія узгодженості дає інженеру впевненість у тому, що розширення бази розмічених даних безпосередньо сприятиме мінімізації помилки моделі. Емпірична верифікація цієї властивості зазвичай здійснюється шляхом симуляцій Монте-Карло[21], які демонструють закономірне зменшення дисперсії оцінки зі зростанням обсягу вибірки n .

2.2. Асимптотична нормальність

2.2.1. Теоретичне обґрунтування

Другою важливою властивістю є асимптотична нормальність. Вона означає, що при великій кількості спостережень розподіл ММП наближається до

нормального з математичним сподіванням у точці істинного параметра θ_0 і дисперсією, оберненою до інформації Фішера $I(\theta_0)$:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0)). \quad (2.7)$$

Описана властивість дозволяє не лише робити статистичні висновки, а й будувати довірчі інтервали та тестувати гіпотези на основі ММП [5,6].

2.2.2. Зв'язок з центральною граничною теоремою

Асимптотична нормальність є наслідком центральної граничної теореми (ЦГТ), згідно з якою середнє значення великої кількості незалежних випадкових величин наближається до нормального розподілу. Для ММП це означає, що похідні від лог-правдоподібності поводяться як сума незалежних змінних, а тому їхній розподіл наближається до нормального [2].

2.2.3. Побудова довірчих інтервалів і статистичних тестів

Завдяки асимптотичній нормальності можна побудувати довірчий інтервал для параметра θ :

$$\hat{\theta}_n \pm z_{\alpha/2} \sqrt{\frac{1}{nI(\hat{\theta}_n)}}, \quad (2.8)$$

де $I(\hat{\theta})$ – оцінка інформації Фішера, а $z_{\alpha/2}$ – квантиль стандартного нормального розподілу.

На основі ММП розроблено процедури тестування відношення правдоподібності, функція яких полягає у верифікації гіпотез щодо модельних параметрів. Зазначений методологічний апарат активно застосовують у біостатистиці, фінансовому аналізі та інженерній справі, зокрема для визначення показників ризику, інтенсивності відмов обладнання або оцінки терапевтичних ефектів [3,4].

2.3. Інформація Фішера

2.3.1. Визначення та властивості

Інформація Фішера – це кількісна міра того, скільки інформації про параметр θ містить випадкова вибірка X . Вона визначається як:

$$I(\theta) = E \left[\left(\frac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right]. \quad (2.9)$$

Інтуїтивно, якщо зміна параметра сильно впливає на форму розподілу, інформація Фішера велика, а отже, оцінка буде точнішою [7,10].

2.3.2. Матриця інформації Фішера

Для багатовимірних моделей використовується матриця інформації Фішера, що описує чутливість функції правдоподібності до зміни кожного параметра:

$$I(\theta) = -E \left[\frac{\partial^2 \ell(\theta)}{\partial \theta_i \partial \theta_j} \right]. \quad (2.10)$$

Отримана матриця має ключове значення для побудови ефективних оцінок і алгоритмів оптимізації в машинному навчанні, зокрема при навчанні нейронних мереж, де вона використовується для адаптивної зміни швидкості навчання [11].

2.3.3. Зв'язок з дисперсією та ефективністю ММП

Інформація Фішера визначає нижню межу дисперсії будь-якої незміщеної оцінки – так звану межу Крамера–Рао:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}. \quad (2.11)$$

При значному обсязі вибірки n метод досягає зазначеної межі та демонструє асимптотичну ефективність. Фактично такий результат вказує на те,

що серед усього спектра методів оцінювання саме ММП забезпечує мінімально можливу дисперсію [1,7].

Окреслені характеристики, зокрема узгодженість разом з асимптотичною нормальністю та ефективністю, безпосередньо зумовлюють поширення методу в наукових дослідженнях і технологічних розробках. Сукупність цих властивостей виступає гарантією того, що за умови накопичення достатнього масиву даних отримані оцінки будуть надійними та теоретично обґрунтованими, а також придатними для коректної інтерпретації в контексті сучасного статистичного аналізу.

2.4. Обмеження та проблемні аспекти методу максимальної правдоподібності

Незважаючи на статус методу максимальної правдоподібності, як базового інструментарію статистики, зумовлений його асимптотичними характеристиками, даний підхід не слід розглядати як універсальний засіб вирішення всіх завдань. Декларована ефективність та нормальність розподілу досягаються переважно за умови наявності значних масивів даних. На практиці ж застосування ММП стикається з низкою об'єктивних перешкод, аналіз яких наведено нижче.

2.4.1. Зміщеність оцінок для скінченних вибірок

Вагомим недоліком методології вважається потенційна зміщеність результатів. Суть проблеми полягає у тому, що математичне сподівання знайденої оцінки в умовах обмеженого обсягу вибірки не завжди збігається з істинним значенням параметра. Як приклад часто розглядають процедуру оцінювання дисперсії нормального розподілу $N(\mu, \sigma^2)$. У ситуації, коли невідомий параметр середнього μ замінюють вибіркоvim середнім \bar{x} , то вираз для оцінки максимальної правдоподібності дисперсії набуває вигляду:

$$\hat{\sigma}_{MLE}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.12)$$

Отримана величина характеризується зміщеною, оскільки вона систематично демонструє значення, нижчі за справжню дисперсію. Розрахунок математичного сподівання для цього показника дає результат:

$$E\left[\hat{\sigma}_{MLE}^2\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2 \quad (2.13)$$

Хоча при $n \rightarrow \infty$ множник прямує до одиниці, роблячи оцінку асимптотично незміщеною, для малих вибірок похибка є значною, що вимагає використання виправленої вибіркової дисперсії.

2.4.2. Ефективність на малих вибірках та середньоквадратична помилка

Асимптотична ефективність ММП гарантується лише при нескінченному зростанні вибірки. Однак для малих n це не гарантує мінімальної середньоквадратичної помилки (MSE). У ситуаціях з високою розмірністю даних або малою кількістю спостережень альтернативні методи, що вносять певне зміщення задля суттєвого зменшення дисперсії, можуть виявитися точнішими.

Наприклад, методи регуляризації часто дають меншу загальну помилку прогнозування, ніж чистий ММП. Це відбувається тому, що ММП має тенденцію "підлаштовуватися" під шум у малих вибірках, тоді як регуляризовані методи обмежують простір пошуку параметрів.

2.4.3. Проблема існування максимуму та повна роздільність

Трапляються ситуації, за яких досягнення глобального максимуму функції правдоподібності у межах скінченної множини параметрів стає неможливим. Прикладом зазначеного феномену є поведінка логістичної регресії в умовах «повної роздільності». За наявності предикторів, що забезпечують безпомилкове розмежування класів, спостерігається монотонне зростання функції правдоподібності при наближенні коефіцієнта нахилу до нескінченності:

$$\hat{\beta} \rightarrow \infty \quad (2.14)$$

За таких обставин традиційні алгоритми оптимізації втрачають здатність до збіжності, що неминуче спричиняє обчислювальну дестабілізацію. Як наслідок, інтерпретація вагових коефіцієнтів стає недоступною без залучення методів

регуляризації або переходу до байєсівського підходу з використанням апіорних розподілів.

2.4.4. Локальні максимуми та увігнутість функції

Геометрична конфігурація поверхні функції правдоподібності часто стає суттєвою перешкодою при роботі зі складними статистичними структурами. На відміну від розподілів експоненціального сімейства, де лог-правдоподібність характеризується увігнутістю, моделі на кшталт сумішей розподілів або глибоких нейромерж вирізняються заплутаним ландшафтом із численними локальними екстремумами. Існує високий ризик того, що градієнтні алгоритми оптимізації зупиняться в точці локального максимуму, значення якого суттєво поступається глобальному оптимуму:

$$L(\theta_{local}) \ll L(\theta_{global}) \quad (2.15)$$

Зазначена особливість ставить кінцевий результат оцінювання у пряму залежність від вибору початкових параметрів і диктує необхідність залучення глобальних методів оптимізації, що неминуче призводить до зростання обчислювальної складності процесу.

РОЗДІЛ 3. ЗАСТОСУВАННЯ МЕТОДУ МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ ДО КОНКРЕТНИХ РОЗПОДІЛІВ

У цьому розділі розглядаються приклади застосування методу максимальної правдоподібності до двох базових, але фундаментально важливих розподілів: розподілу Бернуллі та експоненціального розподілу. Вибір саме цих двох моделей зумовлений їхньою універсальністю та частим використанням у прикладних задачах статистики, машинного навчання, економіки та інженерії.

Розподіл Бернуллі описує події типу «так/ні», «успіх/невдача», тоді як експоненціальний – інтервали часу між подіями. Таким чином, перший відображає дискретні ймовірнісні явища, а другий – неперервні часові процеси, що дозволяє продемонструвати можливості ММП для двох різних класів випадкових величин.

3.1. Розподіл Бернуллі

3.1.1. Функція правдоподібності та її максимізація

Розподіл Бернуллі моделює випадкову подію, яка може мати лише два результати: 1 – «успіх» з ймовірністю p і 0 – «невдачу» з ймовірністю $1-p$. Його функція ймовірності має вигляд:

$$f(x; p) = p^x (1-p)^{1-x}, \quad x \in \{0, 1\}. \quad (3.1)$$

Для n незалежних спостережень X_1, X_2, \dots, X_n функція правдоподібності дорівнює:

$$L(p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}. \quad (3.2)$$

Максимізація цієї функції за параметром p дає:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (3.3)$$

тобто оцінка максимальної правдоподібності для параметра p – це просто частка успіхів у вибірці [2].

3.1.2. Виведення ММП для параметра p

Для зручності максимізацію часто проводять через логарифм функції:

$$\ell(p) = \sum_{i=1}^n [x_i \ln p + (1 - x_i) \ln(1 - p)]. \quad (3.4)$$

Диференціюючи та прирівнюючи похідну до нуля, отримуємо:

$$\frac{d\ell(p)}{dp} = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1 - p} = 0, \quad (3.5)$$

звідки:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (3.6)$$

Це дуже інтуїтивна оцінка: якщо, наприклад, у 100 експериментах 65 були успішними, то ММП для p дорівнює 0.65.

3.1.3. Узгодженість та асимптотична нормальність оцінки

Оцінка \hat{p} є узгодженою, оскільки за законом великих чисел вона прямує до істинного значення параметра p при $n \rightarrow \infty$:

$$\hat{p} \xrightarrow{P} p. \quad (3.7)$$

Крім того, вона є асимптотично нормальною:

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} N(0, p(1 - p)). \quad (3.8)$$

Це дозволяє будувати довірчі інтервали для ймовірності успіху та проводити тести гіпотез [5,6].

3.1.4. Інформація Фішера для розподілу Бернуллі

Інформація Фішера для одного спостереження має вигляд:

$$I(p) = \frac{1}{p(1-p)}. \quad (3.9)$$

Для вибірки з n незалежних спостережень:

$$I_n(p) = \frac{n}{p(1-p)}. \quad (3.10)$$

Це означає, що при великих вибірках оцінка \hat{p} стає стабільнішою. Найбільша інформація досягається при $p = 0.5$, тобто коли результати максимально варіативні [1,10].

3.1.5. Статистичні критерії та перевірка гіпотез

Після отримання точкової оцінки параметра θ , наступним етапом є перевірка статистичних гіпотез. Зазвичай розглядається нульова гіпотеза $H_0: p = p_0$ проти альтернативи $H_1: p \neq p_0$. Вибір методу перевірки критично залежить від обсягу вибірки n .

3.1.5.1. Обмеження точних тестів у задачах великих даних

Традиційним підходом для малих вибірок є точний біноміальний тест, який є аналогом точного тесту Фішера. Він обчислює суму ймовірностей для хвостів розподілу, використовуючи біноміальні коефіцієнти:

$$P(X \geq k) = \sum_{i=k}^n \binom{n}{i} p_0^i (1-p_0)^{n-i} \quad (3.11)$$

Застосування даного методу вважається доцільним за умови малого обсягу вибірки ($n < 50$). Однак критичним обмеженням виступає присутність факторіалів у розрахунковій формулі $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Зі збільшенням параметра n обчислювальна складність алгоритму зростає експоненціально, досягаючи надмірних величин. В умовах сучасних задач Big Data, де кількість спостережень часто перевищує 10^6 , використання точного тесту стає нераціональним з погляду витрат ресурсів. Реалізація арифметики високої точності для обчислення факторіалів вимагає значних часових затрат, тоді як приріст точності у порівнянні з апроксимаційними методами виявляється статистично незначущим.

3.1.5.2. Асимптотичні критерії на основі правдоподібності

Для великих вибірок, спираючись на доведену властивість асимптотичної нормальності оцінок максимальної правдоподібності, стає можливим використання наближених статистичних критеріїв. Вони дозволяють проводити перевірку гіпотез практично миттєво навіть на масивах даних Big Data, оскільки оперують простими арифметичними діями замість складних комбінаторних обчислень. У сучасній статистичній практиці виділяють три фундаментальні підходи до побудови таких критеріїв.

Першим і найбільш інтуїтивним підходом є тест Вальда. Його логіка базується на вимірюванні "відстані" між отриманою оцінкою та гіпотетичним значенням. Оскільки оцінка є випадковою величиною, ця відстань зважується на її дисперсію. Статистика критерію має асимптотичний розподіл хі-квадрат з одним ступенем вільності (χ_1^2) і обчислюється за формулою:

$$W = \frac{(\hat{p} - p_0)^2}{\text{Var}(\hat{p})} = \frac{(\hat{p} - p_0)^2}{p_0(1 - p_0) / n} \quad (3.12)$$

Альтернативним і часто більш універсальним підходом є тест відношення правдоподібностей. Замість оцінювання відстані між параметрами, цей метод

порівнює "висоту" функції правдоподібності. Він аналізує різницю між значенням функції лог-правдоподібності в точці її глобального максимуму \hat{p} та значенням у точці, що відповідає нульовій гіпотезі p_0 . Статистика тесту визначається як подвоєна різниця цих значень:

$$\Lambda = -2 \ln \left(\frac{L(p_0 | X)}{L(\hat{p} | X)} \right) = 2(\ell(\hat{p}) - \ell(p_0)) \quad (3.13)$$

Третім методом є тест множників Лагранжа, також відомий як тест Рао. Його унікальна особливість полягає в тому, що він базується виключно на геометричних властивостях функції правдоподібності в точці нульової гіпотези. Тест аналізує значення градієнта $U(p_0)$ відносно інформації Фішера $I(p_0)$. Це надає йому обчислювальну перевагу, оскільки дозволяє проводити тестування без необхідності знаходження самої оцінки максимальної правдоподібності \hat{p} :

$$S = \frac{U(p_0)^2}{I(p_0)} \quad (3.14)$$

Варто зазначити, що у контексті великих даних ці три статистичні критерії є асимптотично еквівалентними. При $n \rightarrow \infty$ вони сходяться до одних і тих самих значень, проте, на відміну від точних комбінаторних методів, дозволяють уникнути обчислювальних пасток та забезпечують високу ефективність алгоритмів аналізу даних.

3.2. Експоненціальний розподіл

3.2.1. Функція правдоподібності та максимізація

Експоненціальний розподіл описує час до настання події. Його функція щільності:

$$f(x; \lambda) = \lambda e^{-\lambda x}, \quad x \geq 0. \quad (3.15)$$

Для вибірки x_1, x_2, \dots, x_n функція правдоподібності дорівнює:

$$L(\lambda) = \lambda^n e^{-\lambda \sum x_i}. \quad (3.16)$$

Максимізація цієї функції дає:

$$\hat{\lambda} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}, \quad (3.17)$$

тобто оцінка максимальної правдоподібності – це обернена величина до середнього значення вибірки [2].

3.2.2. Виведення ММП для параметра λ

Логарифмічна форма функції правдоподібності:

$$\ell(\lambda) = n \ln \lambda - \lambda \sum x_i. \quad (3.18)$$

Похідна за λ :

$$\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - \sum x_i = 0, \quad (3.19)$$

звідки:

$$\hat{\lambda} = \frac{1}{\bar{x}}. \quad (3.20)$$

3.2.3. Узгодженість та асимптотична нормальність оцінки

Оцінка $\hat{\lambda}$ є узгодженою та асимптотично нормальною:

$$\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, \lambda^2). \quad (3.21)$$

Завдяки зазначеній властивості стає можливим формування довірчих інтервалів для частоти появи подій та визначення рівня стабільності процесів у часовому вимірі [5].

3.2.4. Інформація Фішера для експоненціального розподілу

Інформація Фішера для одного спостереження:

$$I(\lambda) = \frac{1}{\lambda^2}, \quad (3.22)$$

а для n спостережень:

$$I_n(\lambda) = \frac{n}{\lambda^2}. \quad (3.23)$$

Отже, дисперсія ММП зменшується пропорційно до $1/n$, що підкреслює ефективність цього методу при зростанні обсягу даних [1,7].

РОЗДІЛ 4. АЛЬТЕРНАТИВНИЙ ПОГЛЯД: БАЙЄСІВСЬКЕ ОЦІНЮВАННЯ ТА ЗВ'ЯЗОК З МАШИННИМ НАВЧАННЯМ

4.1. Концептуальні відмінності та Байєсівський висновок

Метод максимальної правдоподібності, розглянутий у попередніх розділах, займає центральне місце в теорії класичної статистики. Його базова передумова ґрунтується на твердженні, що невідомий параметр θ є незмінною константою, тоді як отримані дані мають випадковий характер. Водночас сучасний аналіз даних, особливо за умов обмеженої вибірки чи доступності експертних оцінок, активно використовує альтернативний підхід – байєсівську статистику. Ключова концептуальна розбіжність тут зводиться до трактування параметра θ як випадкової величини, що описується власним розподілом ймовірностей [13].

Фундаментом методу є теорема Байєса, яка дозволяє оновити наші апіорні переконання про параметр на основі отриманих експериментальних даних, щоб сформуванати апостеріорний розподіл. Оскільки маргінальна ймовірність даних $P(X)$ є константою відносно параметра, основне правило байєсівського виведення записується у вигляді пропорційності:

$$P(\theta | X) \propto P(X | \theta) \cdot P(\theta) \quad (4.1)$$

де $P(\theta | X)$ – апостеріорний розподіл, $P(X | \theta)$ – функція правдоподібності (відображає дані), а $P(\theta)$ – апіорний розподіл.

4.2. Застосування до розподілу Бернуллі. Спряжені розподіли

Розглянемо застосування даного підходу на прикладі оцінювання ймовірності успіху θ в межах схеми Бернуллі. Для математичної зручності як апіорний розподіл часто обирають розподіл Бета, $Beta(\alpha, \beta)$, оскільки він є спряженим до біноміальної правдоподібності. Властивість спряженості гарантує, що при виборі Бета-розподілу як апіорного, апостеріорний розподіл збереже ту ж саму функціональну форму, отримавши лише оновлені параметри. За

наявності апріорного припущення $\theta \sim \text{Beta}(\alpha, \beta)$ та фіксації в експерименті k успіхів та $n - k$ невдач, то апостеріорний розподіл набуває вигляду:

$$P(\theta | X) \sim \text{Beta}(\alpha + k, \beta + n - k) \quad (4.2)$$

Параметри α та β тут виступають у ролі "псевдо-спостережень", додаючи віртуальні успіхи та невдачі до реальних даних, що дозволяє уникнути проблем із нульовими ймовірностями при малих вибірках [3, 13].

4.3. Оцінки MAP та зв'язок з регуляризацією

На основі апостеріорного розподілу можна отримати точкові оцінки параметра. Найпоширенішою є оцінка максимуму апостеріорної ймовірності (Maximum A Posteriori), яка відповідає моді розподілу. Також використовується апостеріорне середнє, яке мінімізує середньоквадратичну помилку. Для Бета-розподілу ця оцінка дорівнює:

$$\hat{\theta}_{mean} = \frac{k + \alpha}{n + \alpha + \beta} \quad (4.3)$$

Важливим теоретичним результатом є зв'язок між цими підходами. Якщо обрати рівномірний апріорний розподіл $\text{Beta}(1,1)$, що відповідає повній відсутності попередньої інформації, то оцінка MAP чисельно збігається з оцінкою максимальної правдоподібності. Звідси випливає висновок, що метод максимальної правдоподібності фактично виступає окремим випадком байєсівського оцінювання.

Більше того, цей зв'язок має критичне значення для розуміння методів машинного навчання. Використання оцінки MAP з певними видами апріорних розподілів математично еквівалентне регуляризації функції втрат [3]. Зокрема, знаходження MAP-оцінки для параметрів лінійної регресії при припущенні, що ваги мають нормальний апріорний розподіл $\theta \sim N(0, \tau^2)$, є повністю еквівалентним L2-регуляризації. Аналогічно, використання розподілу Лапласа

приводить до L1-регуляризації [4]. Звідси стає очевидним, що сучасний програмний інструментарій на кшталт Scikit-Learn [14] спирається на потужне теоретичне підґрунтя, яке виходить за межі простої оптимізації похибки.

РОЗДІЛ 5 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА СТАТИСТИЧНИЙ АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

Даний розділ описує практичні аспекти реалізації методу максимальної правдоподібності, що включає розробку спеціалізованого програмного забезпечення, попередню обробку великих масивів даних та статистичний аналіз отриманих оцінок. Розглянуто ефективність застосування методу послідовно для двох сценаріїв: моделювання неперервних випадкових величин [15] на прикладі часових рядів та оцінювання параметрів дискретного розподілу [16] в задачах А/В тестування.

5.1. Опис програмного забезпечення

Для проведення обчислювальних експериментів було розроблено уніфіковане програмне середовище на базі мови програмування Python[17]. Вибір цієї екосистеми зумовлений наявністю високопродуктивних бібліотек для наукових обчислень, що дозволяє ефективно реалізувати метод максимальної правдоподібності як для великих масивів неперервних даних, так і для дискретних вибірок. Архітектура розробленого програмного забезпечення є модульною та базується на використанні єдиного технологічного стеку для всіх етапів дослідження.

Перший рівень обробки даних реалізовано за допомогою бібліотеки Pandas. Цей інструмент забезпечує завантаження вхідних файлів у форматах CSV, виконує попередню очистку, фільтрацію викидів та маніпуляції з часовими рядами. Саме засобами цієї бібліотеки формуються валідовані вибірки, які в подальшому передаються на етап математичного моделювання.

Безпосередня реалізація алгоритмів методу максимальної правдоподібності виконана з використанням бібліотеки NumPy. Завдяки підтримці векторизованих операцій, вона дозволяє проводити обчислення точкових оцінок параметрів розподілу над масивами з мільйонами записів без використання повільних ітеративних циклів. Це забезпечує високу швидкодію програми при розрахунку інтенсивності потоку подій та ймовірностей конверсії.

Статистичний аналіз отриманих результатів покладено на модуль stats бібліотеки SciPy. Цей компонент програмного комплексу відповідає за розрахунок критичних значень розподілів, побудову довірчих інтервалів та проведення перевірки статистичних гіпотез, зокрема тесту Колмогорова-Смірнова та точного біноміального тесту. Візуалізація результатів моделювання, включаючи побудову гістограм та графіків функцій розподілу, здійснюється засобами бібліотеки Matplotlib, що дозволяє наочно оцінити якість апроксимації емпіричних даних теоретичними законами.

5.2 Дослідження ММП для неперервних розподілів

5.2.1. Попередня обробка вхідних даних

В якості емпіричної бази дослідження обрано масив даних, що характеризує часові параметри пасажирських перевезень, а саме тривалість поїздок таксі. Вхідний набір даних завантажується з файлу train.csv [19] та підлягає попередній аналітичній обробці. На початковому етапі виконується перетворення часових міток посадки та висадки пасажирів, на основі чого обчислюється фактична тривалість кожної поїздки у хвилинах. З метою підвищення якості даних та усунення шумів програмно реалізовано процедуру очистки вибірки, яка передбачає вилучення аномальних записів із нульовою або від'ємною тривалістю, що позбавлені фізичного змісту, а також фільтрацію викидів методом квантильного аналізу шляхом відсіювання екстремально довгих поїздок, що перевищують 99-й перцентиль розподілу ($t > P_{99}$). За результатами попередньої обробки сформовано репрезентативну кінцеву вибірку загальним обсягом $n = 1\,444\,051$ спостережень, що забезпечує необхідний рівень статистичної значущості для подальшого оцінювання параметрів розподілу.

5.2.2. Реалізація методу максимальної правдоподібності

Для математичного моделювання випадкової величини T , що описує тривалість поїздки, першочергово було висунуто гіпотезу про її підпорядкування експоненціальному закону розподілу з параметром інтенсивності λ . Вибір цієї

моделі базується на припущенні про сталість середньої інтенсивності завершення поїздок у часі. Для знаходження точкової оцінки невідомого параметра побудовано функцію правдоподібності, яка відображає ймовірність отримання наявної вибірки при фіксованому значенні λ і має вигляд:

$$L(\lambda) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n t_i\right) \quad (5.1)$$

Оскільки пряма максимізація цієї функції є обчислювально складною через добуток великої кількості множників, здійснено перехід до логарифмічної функції правдоподібності. Диференціювання останньої за параметром λ та прирівнювання похідної до нуля дозволило отримати аналітичний вираз для оцінки максимальної правдоподібності, яка є оберненою величиною до вибіркового середнього:

$$\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^n t_i} = \frac{1}{\bar{t}} \quad (5.2)$$

За результатами застосування розробленого алгоритму до емпіричних даних було встановлено, що вибіркове середнє значення тривалості поїздки становить $\bar{t} \approx 13.4576$ хв., що відповідає оцінці параметра $\hat{\lambda} \approx 0.07431$. Важливим етапом дослідження стала перевірка надійності отриманої оцінки, для чого було обчислено інформацію Фішера $I(\lambda)$. Дана величина характеризує кількість інформації, яку несе вибірка про невідомий параметр, і в даному випадку її розрахункове значення становить:

$$I(\lambda) = \frac{n}{\lambda^2} \approx 2.61 \cdot 10^8 \quad (5.3)$$

Згідно з нерівністю Рао-Крамера, обернена величина до інформації Фішера визначає нижню межу дисперсії незсуненої оцінки, що дозволило розрахувати стандартну помилку вимірювання:

$$SE(\hat{\lambda}) = \frac{1}{\sqrt{I(\lambda)}} \approx 0.000062 \quad (5.4)$$

На основі асимптотичної нормальності оцінок максимальної правдоподібності побудовано 95% довірчий інтервал (0.07419; 0.07443). Екстремальна вузькість отриманого інтервалу свідчить про високу точність знайденої точкової оцінки, що обумовлено значним обсягом проаналізованої вибірки, проте сама точність оцінки параметра ще не гарантує адекватності експоненціальної моделі реальним даним, що потребує подальшої перевірки за допомогою критеріїв узгодженості.

5.2.3. Порівняльний аналіз моделей

З метою ідентифікації закону розподілу, який найточніше описує емпіричні дані тривалості поїздок, проведено порівняння експоненціальної моделі з альтернативними розподілами: Гамма, Вейбулла та Логнормальним. Фундаментальною метрикою якості статистичного моделювання виступає логарифмічна функція правдоподібності (LL), яка чисельно відображає ступінь відповідності параметричної моделі наявним експериментальним даним. Максимізація цього показника є необхідною умовою точності, проте використання виключно LL для вибору найкращої моделі несе ризик перенавчання, оскільки складніші моделі з більшою кількістю параметрів завжди демонструватимуть вищу правдоподібність, навіть якщо вони описують випадковий шум, а не закономірність. Для забезпечення балансу між точністю апроксимації та простотою моделі використано інформаційний критерій Акаїке (AIC) та Байєсівський інформаційний критерій (BIC). Ці метрики базуються на значенні LL , але вводять штрафний доданок, що пропорційний кількості параметрів моделі k та, у випадку BIC, обсягу вибірки n . Розрахунок критеріїв здійснюється за наступними формулами:

$$AIC = 2k - 2\ln(\hat{L}) \quad (5.5)$$

$$BIC = k \ln(n) - 2\ln(\hat{L}) \quad (5.6)$$

Згідно з логікою побудови цих критеріїв, найкращою вважається модель, що забезпечує мінімальне значення AIC та BIC. Результати розрахунків для чотирьох досліджуваних розподілів наведено в таблиці 5.1.

Таблиця 5.1

Результати порівняння моделей за Log-Likelihood та інформаційними критеріями

Модель	к, к-сть параметрів моделі	Log-Likelihood	AIC	BIC
Lognormal	2	$-4.995 \cdot 10^6$	$9.990 \cdot 10^6$	$9.990 \cdot 10^6$
Gamma	2	$-5.002 \cdot 10^6$	$1.001 \cdot 10^7$	$1.001 \cdot 10^7$
Weibull	2	$-5.027 \cdot 10^6$	$1.006 \cdot 10^7$	$1.006 \cdot 10^7$
Exponential	1	$-5.196 \cdot 10^6$	$1.039 \cdot 10^7$	$1.039 \cdot 10^7$

Аналіз отриманих результатів свідчить про те, що Логнормальний розподіл є статистично значущим лідером серед розглянутих кандидатів. Він демонструє найвище значення функції правдоподібності та, відповідно, найнижчі показники інформаційних критеріїв, що підтверджує доцільність введення другого параметра для опису форми розподілу. Натомість експоненціальна модель, попри найменшу кількість параметрів ($k = 1$), показала найгірші результати за всіма метриками. Наявність чітко вираженої моди у реальному розподілі вказує на обмеженість експоненціального закону, який за своєю природою передбачає виключно монотонне спадання. Така невідповідність свідчить про те, що структура процесу тривалості поїздок є набагато складнішою, ніж дозволяє відтворити ця модель.

5.2.4. Статистична перевірка гіпотез

Для формальної верифікації нульової гіпотези H_0 , яка стверджує, що емпіричні дані підпорядковуються експоненціальному закону розподілу, було застосовано комплексний підхід, що поєднує критерії загальної узгодженості та специфічні тести [12, 20] для розподілів часу життя. Базовим етапом перевірки став критерій Колмогорова-Смірнова, який оцінює максимальну вертикальну

відстань між емпіричною та теоретичною функціями розподілу. Розрахункове значення статистики $D=0.1614$ при $p\text{-value} \approx 0.0$ однозначно свідчить про статистично значущу відмінність, проте цей критерій має відомий недолік – низьку чутливість до відхилень на краях розподілу.

Оскільки для задач логістики та оцінки ризиків критично важливим є точне моделювання рідкісних довгих поїздок, додатково застосовано критерій Андерсона-Дарлінга [20]. Даний метод надає більшу вагу спостереженням у "хвостах" розподілу. Отримана статистика $A^2 = 71286.84$ суттєво перевищує критичні значення, зокрема 1.96 для рівня значущості $\alpha = 0.01$, що підтверджує системну невідповідність експоненціальної моделі не лише в центрі розподілу, а й у зоні екстремальних значень. Для визначення напрямку відхилення використано Score Test (Cox & Oakes), розроблений для розрізнення вкладених моделей. Значення Z-статистики на рівні 715.7 ($|Z| \gg 1.96$) вказує на те, що функція інтенсивності відмов не є постійною, як того вимагає експоненціальний закон, а змінюється з часом.

У практичній площині отримані результати дозволяють зробити важливий висновок: процес пасажирських перевезень не володіє властивістю "відсутності пам'яті". Відхилення нульової гіпотези означає, що неможливо коректно прогнозувати ймовірність завершення поїздки або час очікування, використовуючи лише одне середнє значення (параметр λ). Реальні дані демонструють складнішу динаміку: низьку ймовірність дуже коротких поїздок та наявність "важких хвостів", які формуються під впливом заторів та великих дистанцій, що вимагає використання більш гнучких моделей, таких як Логнормальний розподіл.

5.2.5. Дослідження властивостей оцінок

Для емпіричної верифікації надійності та стійкості застосованого алгоритму оцінювання було проведено обчислювальний експеримент методом статистичних випробувань Монте-Карло [21]. Суть експерименту полягала у

генерації серії синтетичних вибірок із заздалегідь відомим істинним параметром $\lambda_{true} = 0.1$, що дозволило порівняти розрахункові оцінки з еталонним значенням. Моделювання виконувалося для широкого діапазону обсягів вибірки $N \in \{50, 100, 500, 1000, 5000\}$, причому для кожного значення N генерувалося 500 незалежних ітерацій.

На рисунку 5.1 візуалізовано результати симуляції, які ілюструють еволюцію статистичних властивостей оцінки в динаміці нарощування обсягу даних. Абсолютне зміщення оцінки відображено помаранчевою пунктирною лінією, де для малих вибірок ($N = 50$) фіксується певне систематичне відхилення на рівні 0.00198. Втім, зі зростанням N даний показник демонструє стрімке падіння до критично малих значень, досягаючи 0.000002 при $N = 5000$, чим фактично підтверджується властивість асимптотичної незсуненості ММП.

Дисперсію оцінки, що виступає мірою розкиду значень відносно істинного параметра, показано синьою суцільною лінією. На графіку чітко простежується тенденція різкого зменшення дисперсії, яка асимптотично наближається до нуля за законом оберненої пропорційності до обсягу вибірки. Описана поведінка кривих слугує наочним доказом узгодженості оцінки, адже зі збільшенням кількості спостережень оцінка $\hat{\lambda}$ за ймовірністю збігається до істинного значення параметра. Зазначений факт гарантує високу точність результатів, отриманих під час опрацювання реальної вибірки обсягом понад 1.4 млн записів.

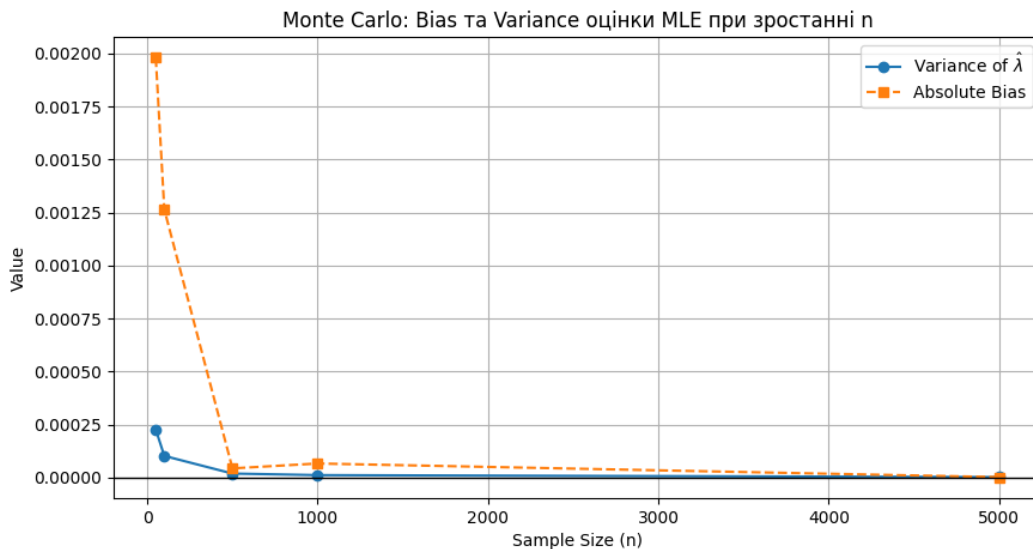


Рисунок 5.1. Залежність зміщення та дисперсії оцінки від обсягу вибірки

5.2.6. Візуалізація результатів

Фінальним етапом дослідження стала графічна верифікація статистичних висновків, спрямована на візуальну оцінку якості апроксимації та виявлення розбіжностей між емпіричними даними і теоретичними моделями. На рисунку 5.2 представлено графіки Q-Q [22], що слугують індикаторами узгодженості розподілів. Ліва частина зображення демонструє спробу опису даних експоненціальним законом, де виражена дугоподібна форма розташування точок вказує на наявність систематичної помилки моделювання на всьому діапазоні значень. Така конфігурація фактично підтверджує неспроможність експоненціальної моделі коректно відобразити динаміку поїздок. Натомість правий графік, що відповідає логнормальному розподілу, характеризується високим ступенем лінійності в межах основного масиву даних, що свідчить про адекватність моделі для більшості спостережень. Характерне відхилення точок від діагоналі у правій верхній частині графіка можна вважати очікуваним наслідком попередньої обробки інформації. Оскільки з вибірки було програмно вилучено рідкісні екстремально довгі поїздки, емпіричні квантілі обмежуються певним порогом, тоді як теоретична модель припускає існування подій необмеженої тривалості. Попри зазначену специфіку в зоні екстремумів,

загальна структура розподілу підтверджує статистичну обґрунтованість вибору логнормального закону для моделювання тривалості поїздки.

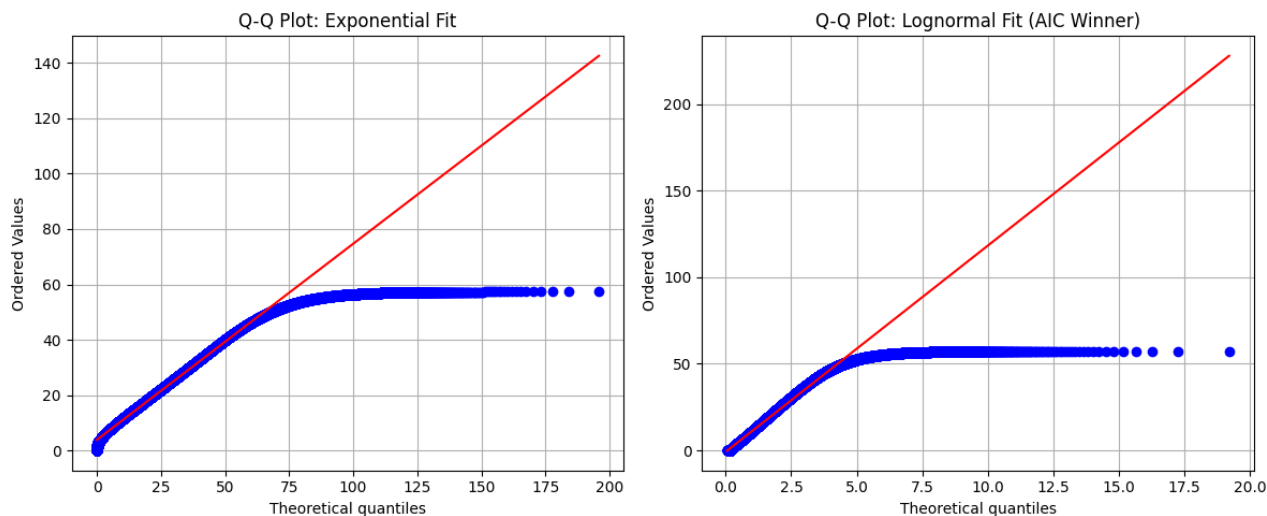


Рисунок 5.2. Q-Q plots. Зліва Експоненціальний розподіл, справа Логнормальний

Детальніший аналіз структури розподілу представлено на рисунку 5.3, де порівнюються щільності ймовірності. Гістограма реальних даних демонструє чітку унімодальність із піком в діапазоні 5–10 хвилин, що є логічним для поїздки таксі. Червона пунктирна крива експоненціального розподілу, маючи максимум у точці нуль і монотонно спадаючи, фізично не здатна описати цей пік, грубо усереднюючи поведінку даних. Водночас синя суцільна лінія логнормального розподілу демонструє чітку відповідність профілю гістограми, оскільки вона коректно відтворює початкове зростання ймовірності, досягнення піку та подальше плавне спадання. Завдяки цьому модель стає значно ефективнішим інструментом для прогнозування.

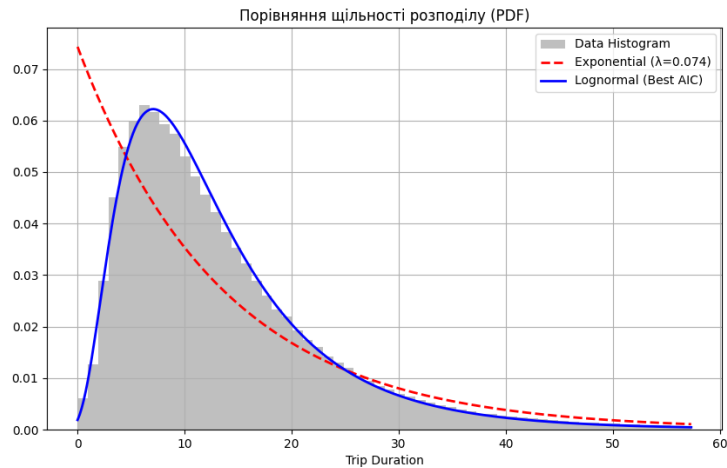


Рисунок 5.3. Порівняння теоретичних щільностей розподілу з гістограмою даних

Додатковий аналіз інтегральних характеристик здійснено шляхом зіставлення емпіричної та теоретичних функцій розподілу, що зображено на рисунку 5.4. Синя східчаста лінія відтворює фактичну динаміку накопичення частот тривалості поїздок, тоді як червоний пунктир відповідає експоненціальній моделі, а зелена штрих-пунктирна крива ілюструє апроксимацію логнормальним законом.

При детальному розгляді графіка фіксується суттєве розходження між реальними даними та експоненціальною кривою, особливо на початковому часовому проміжку до 15 хвилин, де теоретична модель не здатна адекватно відтворити специфіку наростання ймовірності. На противагу цьому, графік логнормального розподілу демонструє майже повну узгодженість із синьою лінією на всьому діапазоні значень. Зазначена візуальна дистанція між експонентою та емпіричними даними слугує геометричним підтвердженням високих значень статистики Колмогорова-Смірнова. Наведені факти остаточно доводять перевагу двопараметричного логнормального підходу над спрощеною експоненціальною моделлю в задачах високоточної аналітики трафіку.

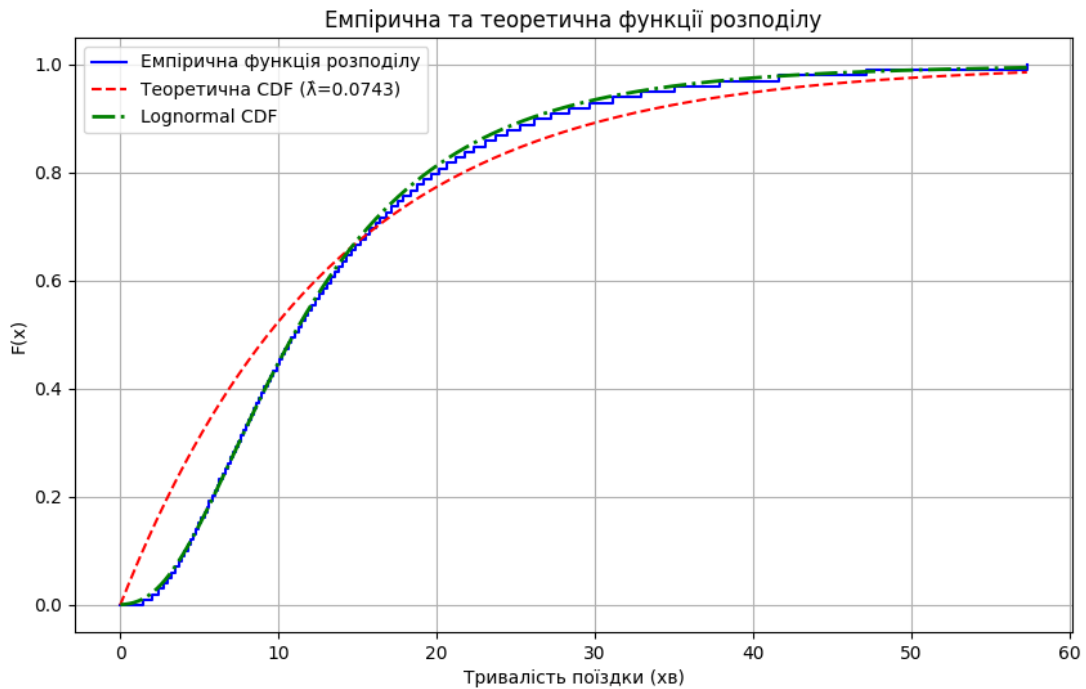


Рисунок 5.4. Емпіричні та теоретична функції розподілу

5.3. Дослідження ефективності оцінок ММП для дискретних розподілів

На доповнення до теоретичного аналізу асимптотичних властивостей оцінок у роботі реалізовано обчислювальний експеримент на базі реальної бізнес-задачі щодо оптимізації конверсії інтернет-магазину меблів. В умовах електронної комерції відсутність фізичного контакту з товаром робить візуальну презентацію визначальним чинником впливу на купівельну поведінку. Ключова проблема бізнесу полягала у невизначеності стосовно найбільш ефективного типу контенту, а саме вибору між акцентом на окремому предметі та його демонстрацією в контексті використання.

Для розв’язання поставленої задачі було розроблено дизайн А/В тестування з метою порівняння результативності двох концепцій оформлення картки товару. В рамках експерименту трафік розподіляється між двома потоками. Контрольна група під назвою Control взаємодіяла з класичним дизайном, де меблі зображувалися на нейтральному білому фоні для детального огляду продукту. Натомість для тестової групи Treatment було впроваджено візуалізацію тих самих позицій безпосередньо в інтер’єрі кімнати.

Основна гіпотеза полягала в припущенні, що показ товарів у реалістичному оточенні сприятиме кращій оцінці габаритів та естетичних якостей виробу, що своєю чергою знизить когнітивне навантаження на клієнта та збільшить імовірність замовлення. Об'єктом математичного моделювання в даному дослідженні виступає бінарна випадкова величина, яка описує факт здійснення покупки або конверсію та підпорядковується розподілу Бернуллі.

5.3.1. Опис даних та генерація вибірки

Враховуючи необхідність емпіричної перевірки теоретичних гіпотез та асимптотичних властивостей статистичних оцінок, у роботі застосовано підхід генерації синтетичних даних. Моделювання поведінки користувачів базується на схемі випробувань Бернуллі, де кожна взаємодія користувача з сайтом розглядається як незалежна реалізація бінарної випадкової величини (купив/не купив). Технічна реалізація генерації здійснена у середовищі Python [17] з використанням бібліотеки наукових обчислень NumPy. Безпосередньо для створення псевдовипадкових послідовностей застосовано функцію «numpy.random.binomial(n=1, p, size)» [23], яка повертає масив значень 0 та 1, імітуючи результати серії незалежних експериментів із заданою ймовірністю успіху p .

Архітектура дослідницького процесу базувалася на дворівневій схемі генерації даних, що обумовлено необхідністю вирішення різнопланових аналітичних завдань. Початкова фаза передбачала формування фіксованих вибірок з метою моделювання одиничного А/В тестування, де обсяг кожної групи становив $N = 1000$ унікальних користувачів, забезпечуючи сумарно 2000 спостережень. В рамках експерименту для контрольної групи встановлено базовий рівень конверсії $p_{control} = 0.10$ або 10%, тоді як для тестового сегмента цей показник зафіксовано на рівні $p_{treat} = 0.13$, тобто 13%. Така різниця параметрів дозволила математично формалізувати наявність реального ефекту від впровадження нового дизайну.

Наступний етап присвячувався поглибленому вивченню якісних характеристик оцінок, зокрема їх зміщення та дисперсії, а також побудові кривих статистичної потужності. Реалізація цих завдань здійснювалася шляхом проведення масштабної симуляції за методом Монте-Карло[21]. Сутність процедури зводилася до ітеративної генерації 10 000 незалежних пар вибірок зі збереженням вихідних параметрів розподілу. Завдяки обраному підходу вдалося детально дослідити динаміку статистичних критеріїв на широкій множині реалізацій та ефективно нівелювати вплив випадкових факторів, властивих окремим експериментам.

5.3.2. Методологія оцінювання параметрів

Фундаментом побудованої математичної моделі виступає параметр p , що визначає ймовірність успіху в межах схеми Бернуллі. У площині електронної комерції зазначена величина інтерпретується як коефіцієнт конверсії та слугує визначальним індикатором економічної ефективності проєкту. Зважаючи на неможливість повного охоплення генеральної сукупності потенційних клієнтів, істинне значення параметра p вважається априорі невідомим. Внаслідок цього процедура знаходження точкової оцінки \hat{p} базувалася на застосуванні методу максимальної правдоподібності. Стосовно бінарних даних сутність методу полягає у визначенні відносної частки успішних подій або фактичних покупок k у загальному масиві відвідувачів n , а відповідний розрахунок здійснюється за наведеною далі формулою.

$$\hat{p}_{MLE} = \frac{k}{n} \quad (5.7)$$

Для оцінки теоретичної точності вимірювань та оптимізації витрат трафіку було проаналізовано функцію Інформації Фішера, яка для розподілу Бернуллі є оберненою величиною до дисперсії та набуває вигляду:

$$I(p) = \frac{1}{p(1-p)}. \quad (5.8)$$

Аналіз графічної залежності на рисунку 5.5 демонструє U-подібну криву, що має критичне значення для планування експерименту. Математична властивість зростання функції при наближенні p до 0 або 1 свідчить про те, що виявляти зміни у конверсії, яка є природно низькою, значно ефективніше з точки зору необхідного обсягу вибірки, ніж у ситуації невизначеності при $p = 0.5$. Високе значення інформації Фішера на краях розподілу забезпечує меншу дисперсію оцінки, що дозволяє бізнесу отримувати статистично значущі висновки та детектувати ефекти на менших обсягах даних.

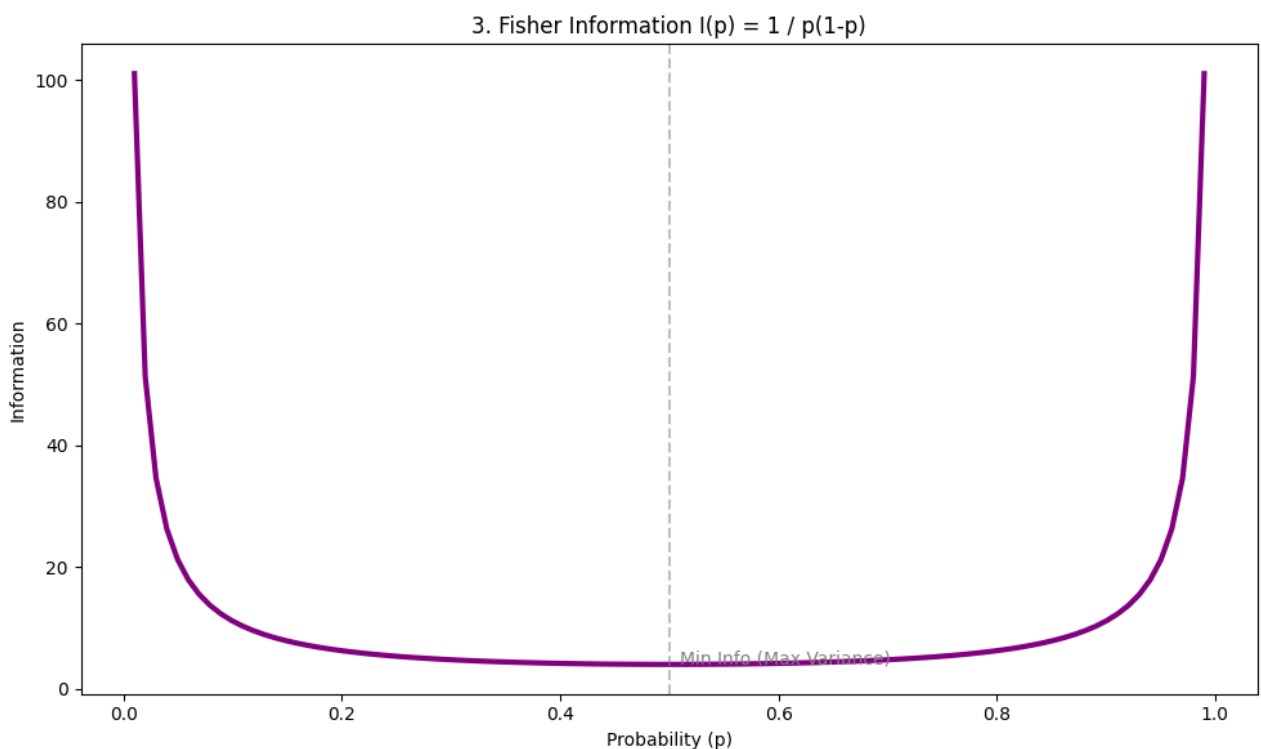


Рисунок 5.5. Залежність Інформації Фішера від базової конверсії

Фундаментальною умовою довіри до розрахованих P -значень та коректності прийняття фінансових рішень є підпорядкування розподілу похибок оцінювання нормальному закону, що гарантується Центральною граничною теоремою. Для емпіричної верифікації цієї властивості було проведено стрес-тестування системи шляхом симуляції 5000 незалежних експериментів. Параметр генерації було встановлено на рівні $p = 0.5$, оскільки саме це значення відповідає сценарію максимальної ентропії та дисперсії $Var = 0.25$, створюючи

найжорстокіші умови для збіжності. Підтвердження асимптотичної нормальності для цього граничного випадку гарантує її виконання і для реальних діапазонів конверсії $p \approx 0.1$. У ході симуляції для кожного експерименту розраховувалася стандартизована Z-статистика за формулою:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \quad (5.9)$$

Візуалізація емпіричного розподілу на рисунку 5.6 засвідчує його фактичну ідентичність теоретичній кривій стандартного нормального закону $N(0,1)$. Спостережувана збіжність підтверджує коректність застосованої методології та обґрунтовує достатність вибірки обсягом 1000 спостережень. Завдяки високій точності апроксимації стає можливим здійснення якісного аналізу потенційних загроз для діяльності підприємства.

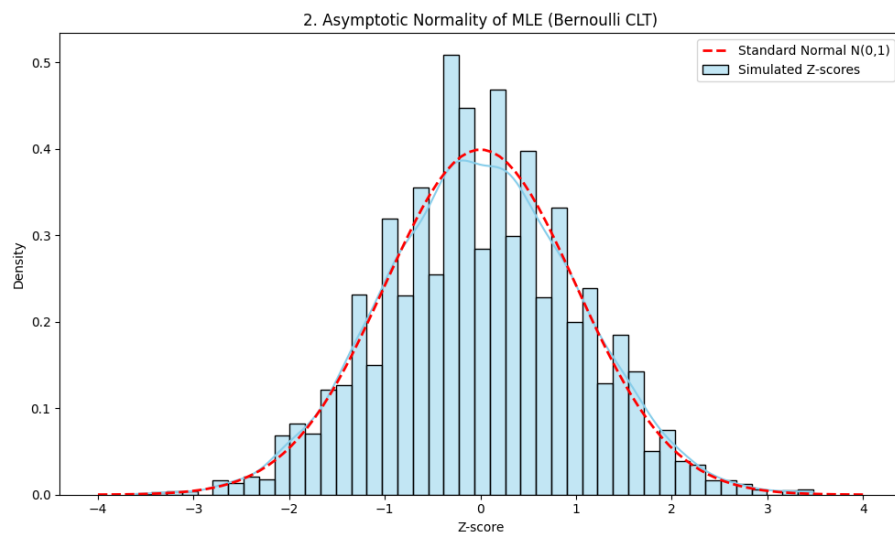


Рисунок 5.6. Перевірка асимптотичної нормальності оцінок ММП

5.3.3. Детальний аналіз метрик та їх бізнес-інтерпретація

За результатами виконання процедури симульованого А/В тестування програмним модулем було здійснено фіксацію емпіричних оцінок ймовірностей конверсії для обох досліджуваних вибірок. У межах аналізу контрольної групи, що відрізнялася використанням класичного дизайну на білому фоні, із загального масиву відвідувачів $N = 1000$ цільову дію реалізували 96 осіб, що

еквівалентно точковій оцінці $\hat{p}_c = 0.096$ або рівню 9.6%. Натомість для експериментальної групи, специфіка якої полягала у візуалізації меблів безпосередньо в інтер'єрі, характерним стало зростання кількості транзакцій до 132 одиниць, внаслідок чого показник конверсії підвищився до значення $\hat{p}_t = 0.132$ та склав 13.2% відповідно. Втім, для прийняття обґрунтованого рішення на управлінському рівні критичну значущість мають не виключно отримані фінальні величини, а й дослідження динаміки їх формування разом з комплексною оцінкою ризиків.

Визначальним кроком у процесі верифікації надійності одержаних даних виступила процедура перевірки статистичних характеристик оцінки приросту, відомого як Lift, із застосуванням методу Монте-Карло [21]. У межах зазначеного етапу було реалізовано 10 000 незалежних ітерацій повного циклу А/В тестування, на основі чого сформовано емпіричний розподіл різниці конверсій, графічна візуалізація якого представлена на рисунку 5.7. Детальне вивчення наведеної залежності дозволяє констатувати наявність двох фундаментальних властивостей, що є необхідними для побудови якісних бізнес-прогнозів. Насамперед варто відзначити симетричну дзвоноподібну конфігурацію графіка, яка слугує додатковим доказом нормальності розподілу різниці частот та обґрунтовує коректність застосування Z-критерію. Крім того, ключове значення має той факт, що середній показник отриманих оцінок, позначений чорною пунктирною лінією, демонструє ідеальний збіг з істинним теоретичним ефектом на рівні 0.03, який візуалізовано червоною лінією, а така тотожність ілюструє властивість незміщеності оцінки методу максимальної правдоподібності. У розрізі формування бізнес-інсайтів наведений графік виступає для менеджменту гарантією об'єктивності вимірювального інструменту та підтверджує відсутність у застосованій математичній моделі систематичної похибки, тобто схильності до штучного завищення чи заниження підсумків тестування. Потенційні відхилення від істинного значення в окремих випадках можуть бути зумовлені виключно фактором випадковості, що візуально

відображається шириною дзвону, проте усереднені показники забезпечують абсолютно точну оцінку приросту продажів.

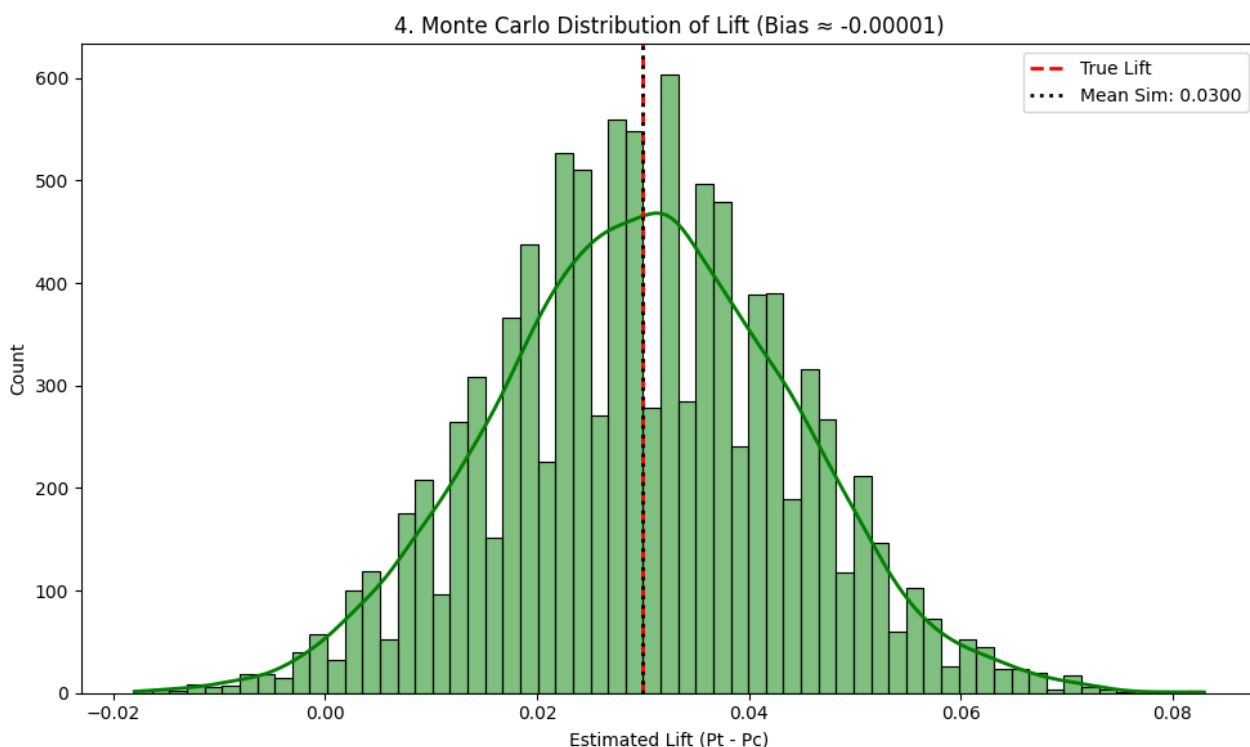


Рисунок 5.7. Збіжність емпіричних частот до теоретичних ймовірностей

Процедура кількісного оцінювання ефективності впровадженого дизайну ґрунтується на обчисленні параметрів абсолютної та відносної різниці показників конверсії. У ході аналізу було зафіксовано абсолютний приріст на 3.6%, що математично описується рівнянням $\Delta p = 0.132 - 0.096$. Разом з тим найбільш репрезентативною бізнес-метрикою слід вважати відносний приріст або Lift, визначення якого здійснюється відповідно до формули

$$\text{Lift} = \frac{\hat{p}_t - \hat{p}_c}{\hat{p}_c} = \frac{0.132 - 0.096}{0.096} \approx 37.5\% \quad (5.10)$$

Стосовно сфери електронної комерції зазначений результат свідчить про те, що масштабування оновленого візуального рішення на повний обсяг аудиторії потенційно забезпечить збільшення загального обсягу продажів на 37.5% за умови збереження незмінного маркетингового бюджету. На рисунку 5.8

представлено візуалізацію отриманих частот разом із 95% довірчими інтервалами, розрахунок яких виконано за методом Вальда. За результатами графічного аналізу виявлено лише незначне перекриття вусів похибки, а такий факт виступає переконливим індикатором статистичної значущості результату та свідчить про відокремленість розподілів.

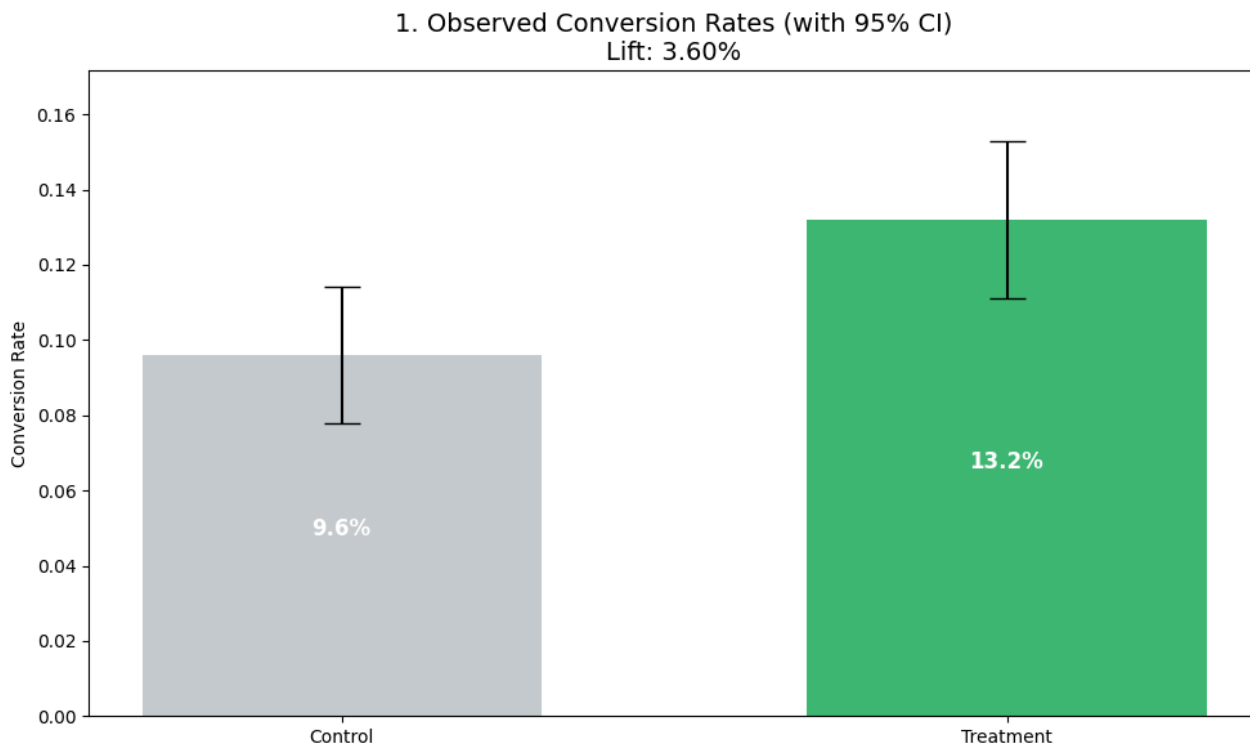


Рисунок 5.8. Спостережувані частки конверсій

З метою усунення недоліків класичної частотної статистики, що обмежується оперуванням виключно точковими оцінками та р-значеннями, було імплементовано Байєсівський підхід, який уможлиблює безпосередню оцінку розподілу ймовірностей невідомого параметра конверсії p шляхом синтезу наявних апіорних знань з отриманими емпіричними даними функції правдоподібності.

Фундаментальною математичною базою побудови моделі визначено спряжену пару розподілів Бета-Біноміальний [24]. Враховуючи специфіку функції правдоподібності для схеми Бернуллі, яка аналітично описується виразом $L(p) \propto p^k (1-p)^{n-k}$, найбільш обґрунтованим є вибір Бета-розподілу в

якості апіорного, причому щільність останнього визначається ідентичною степеневою структурою згідно з рівнянням

$$\pi(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} \quad (5.11)$$

У процесі визначення параметрів апіорного розподілу було затверджено вибір $Beta(\alpha_0 = 1, \beta_0 = 1)$, що є методологічно принциповим рішенням та базується на принципі недостатньої підстави. Необхідно зазначити, що розподіл $Beta(1,1)$ є математично еквівалентним рівномірному розподілу $Uniform[0,1]$, а така тотожність передбачає трактування всіх можливих значень конверсії в діапазоні від 0% до 100% як рівноймовірних до моменту початку збору інформації. Зазначена методологія виступає гарантом забезпечення об'єктивності дослідження, оскільки неінформативний пріор запобігає внесенню суб'єктивних упереджень у кінцевий результат та надає можливість даним Likelihood відігравати визначальну роль у формуванні апостеріорного висновку.

Відповідно до фундаментальних положень теореми Байєса апостеріорний розподіл $\pi(p | Data)$ визначається як величина, що є пропорційною добутку функції правдоподібності та обраного апіорного розподілу. Внаслідок використання властивості спряженості процедура оновлення параметрів розподілу здійснюється аналітичним методом через безпосереднє додавання кількості успішних подій k та випадків невдач $n - k$ до гіперпараметрів апіорного розподілу, що математично формалізується наступними співвідношеннями

$$\alpha_{post} = \alpha_0 + k \quad \beta_{post} = \beta_0 + (n - k) \quad (5.12)$$

На підставі опрацювання масиву експериментальних даних, у якому обсяг вибірки $N = 1000$ був ідентичним для кожної групи, було реалізовано розрахунок параметрів апостеріорних розподілів. У межах аналізу показників контрольної

групи за умови фіксації значення $k_c = 96$ отриманий розподіл набув вигляду $p_c \sim \text{Beta}(97, 905)$, а розраховане математичне сподівання склало

$$E[p_c] = \frac{97}{97 + 905} \approx 0.0968 \quad (5.13)$$

Водночас для тестової групи, де було встановлено показник $k_t = 132$, актуальним визначено розподіл $p_t \sim \text{Beta}(133, 869)$ із відповідним значенням математичного сподівання на рівні

$$E[p_t] = \frac{133}{133 + 869} \approx 0.1327 \quad (5.14)$$

Графічна візуалізація апостеріорних щільностей, представлена на рисунку 5.9, надає відповідальним особам аналітичну інформацію значно вищого рівня деталізації порівняно з традиційними точковими оцінками. Стосовно локалізації максимумів варто відзначити, що пік розподілу тестової групи, позначений зеленим графіком, демонструє суттєве зміщення праворуч відносно контрольної групи, яка відображена сірим кольором, а такий зсув виступає доказом наявності позитивного ефекту. Разом з тим характер ширини розподілів, зокрема візуальна вузькість дзвонів, вказує на низький рівень дисперсії та свідчить про високу впевненість у розрахованих значеннях конверсії. Ключовим критерієм для ухвалення рішень є зона перекриття або площа перетину двох графіків, яка на наведеній схемі є мінімальною. У термінології теорії прийняття рішень така ситуація означає, що ймовірність помилки при виборі варіанту В, котра полягала б у реалізації сценарію, де насправді справджується нерівність $p_t < p_c$, є знехтувано малою. Зазначені обставини дозволяють констатувати з ймовірністю, наближеною до 99%, вищу ефективність нового дизайну, наслідком чого стане мінімізація фінансових ризиків під час процедури масштабування змін.

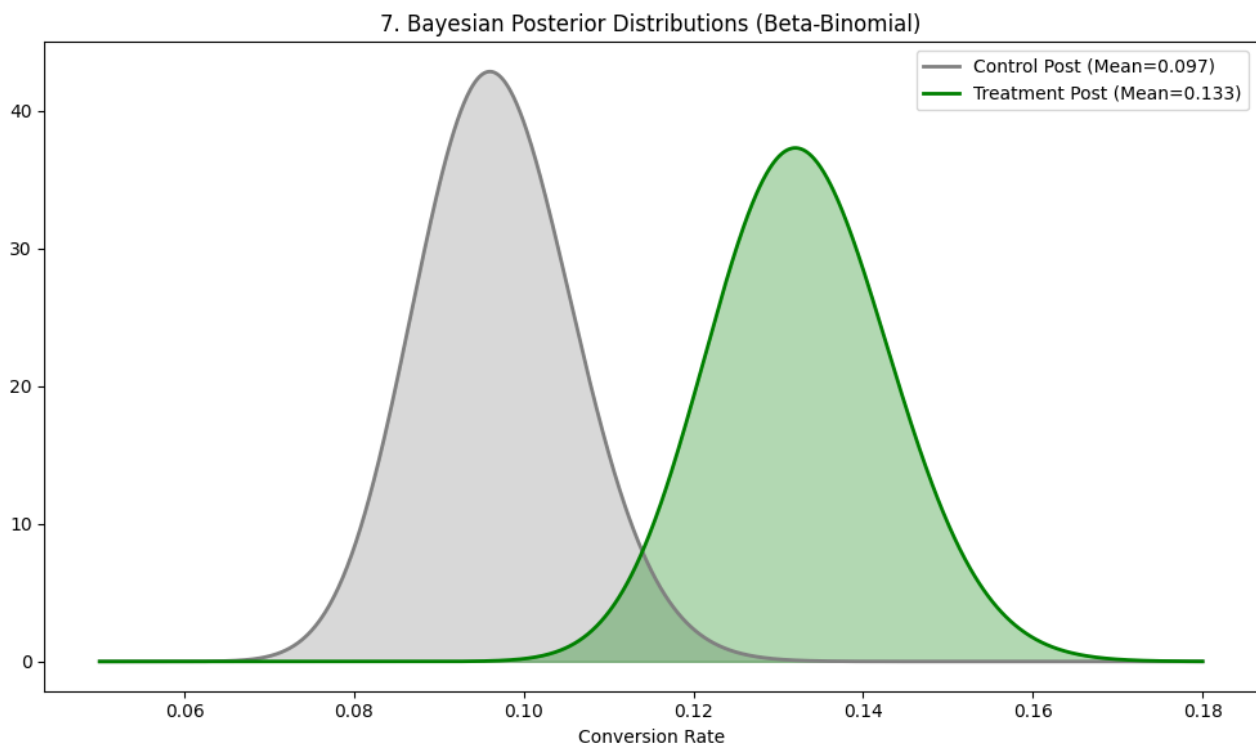


Рисунок 5.9. Апостеріорні щільності розподілу ймовірностей конверсії

5.3.4. Перевірка статистичних гіпотез та мінімізація бізнес-ризиків

Фундаментальним етапом дослідження визначено процедуру формальної перевірки того, чи є зафіксована різниця конверсій наслідком реального впливу нового дизайну, чи ж результатом випадкової флуктуації вибірки. З цією метою було сформульовано дві статистичні гіпотези, серед яких нульова $H_0 : p_c = p_t$ припускає відсутність ефекту, тоді як альтернативна $H_1 : p_c \neq p_t$ стверджує про наявність значущої різниці. В якості порогу прийняття рішення затверджено рівень значущості $\alpha = 0.05$, що відповідає готовності бізнесу прийняти 5% ризик помилки першого роду.

Задля забезпечення надійності висновків верифікація здійснювалася комплексом із трьох асимптотичних критеріїв, які продемонстрували високу узгодженість результатів.

Першочергово було застосовано критерій Вальда, який оцінює відношення різниці оцінок до їх стандартної похибки. Розрахункове значення статистики склало $Z \approx 2.537$ при досягненні рівня значущості $p\text{-value} = 0.0112$, а такий

результат вказує на відхилення емпіричних даних від нульової гіпотези на більш ніж 2.5 стандартних відхилення.

З метою додаткової верифікації результату було використано критерій Рао або Score Test, також відомий як тест Кохрана, котрий базується на оцінці функції правдоподібності в точці нульової гіпотези. Отримана статистика на рівні $Z \approx 2.533$ із показником $p - value = 0.0113$ повністю підтвердила висновки тесту Вальда.

Фінальним етапом став тест відношення правдоподібності, який порівнює узгодженість даних із моделями за умови H_0 та H_1 . Значення статистики $\chi^2 \approx 6.440$ при $p - value = 0.0112$ остаточно підтвердило наявність ефекту.

Враховуючи той факт, що всі отримані p -значення є строго меншими за критичний рівень $\alpha = 0.05$, нульову гіпотезу було статистично відхилено на користь альтернативної.

З метою верифікації стабільності сформульованого статистичного висновку, а також задля детального вивчення параметрів чутливості обраного критерію до варіацій обсягу вибірки було реалізовано процедуру симуляції послідовного аналізу. Методологічний алгоритм формування графічної залежності, представленої на рисунку 5.10, передбачав виконання ітеративних обчислень величини P -value для вибірок, обсяг яких поступово збільшувався у діапазоні від $N = 100$ до $N = 2000$ одиниць із фіксованим кроком приросту у 20 спостережень. У рамках кожного окремого етапу для відповідного поточного зрізу емпіричних даних здійснювався розрахунок статистики Вальда згідно з математичним виразом:

$$Z = \frac{\hat{p}_t - \hat{p}_c}{SE} \quad (5.15)$$

на базі якого надалі встановлювався актуальний рівень значущості.

Завдяки візуалізації траєкторії руху показника P-value у логарифмічній системі координат стає можливою ідентифікація критичної закономірності, згідно з якою на ранніх стадіях експерименту, що відповідають лівій частині графіка, крива характеризується високим рівнем волатильності та наявністю хаотичних стрибків, які неодноразово перетинають червону пунктирну лінію порогового бар'єру $\alpha = 0.05$. Виявлена нестабільність є математичним доказом небезпеки прийняття рішень на основі проміжних результатів, що класифікується як Peeking Problem, адже на обмежених вибірках випадковий шум може призводити до хибних висновків щодо наявності або відсутності ефекту. Втім, паралельно з процесом накопичення додаткових даних спостерігається стабілізація кривої, яка починає демонструвати стійку тенденцію до спадання, впевнено переходячи у зону статистичної значущості. Встановлений факт неповернення лінії до порогового рівня на завершальних стадіях за умови $N = 1000$ виступає гарантією того, що зафіксована відмінність у показниках конверсії є достовірним фактом, а не результатом випадкового збігу обставин на початковому етапі тестування.

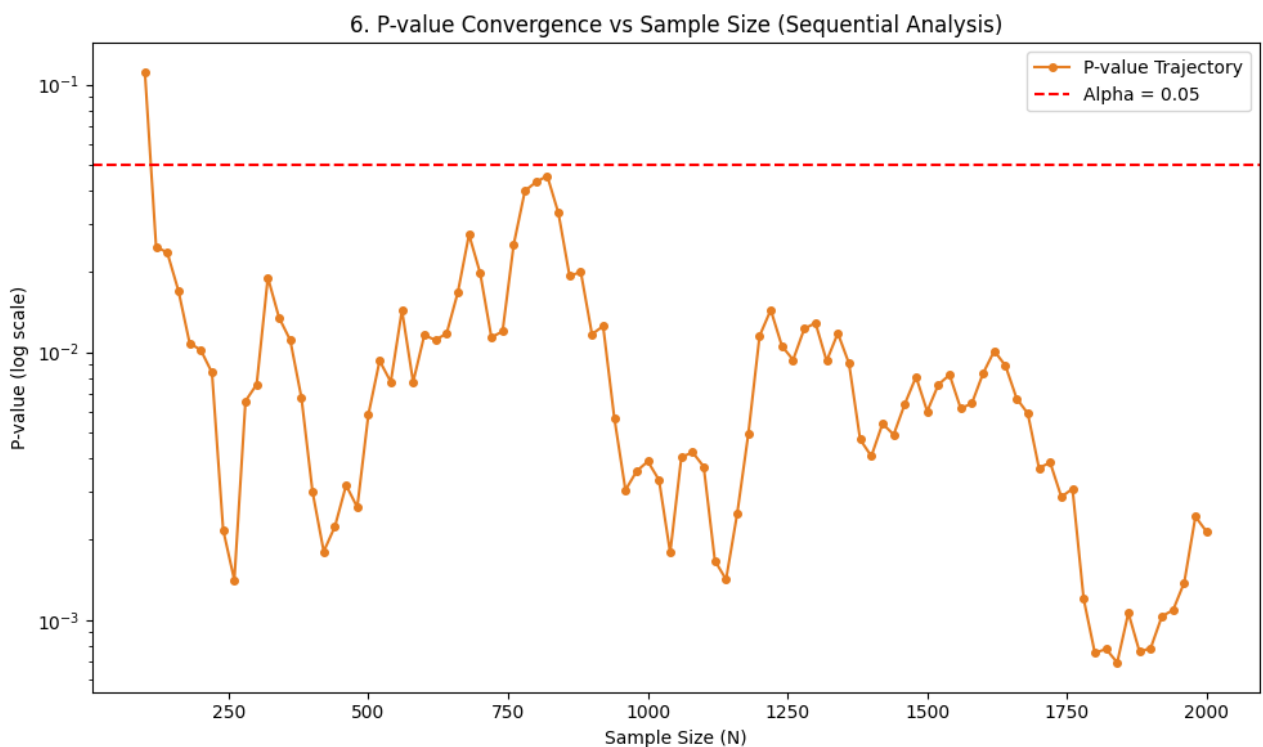


Рисунок 5.10. Траєкторія P-value залежно від обсягу вибірки

5.3.5. Теоретичний аналіз статистичної потужності та верифікація надійності критерію

Перед безпосередньою оцінкою ефективності нового дизайну критично важливим етапом дослідження стала верифікація справності самого вимірювального інструменту, реалізована через симуляцію A/A тестування. У контексті електронної комерції ця процедура моделює ситуацію, коли обом групам користувачів демонструють ідентичний варіант сторінки, що дозволяє перевірити систему на схильність до "галюцинацій" або хибно-позитивних спрацювань. Головна бізнес-мета цього етапу полягає в мінімізації ризику інвестування ресурсів у зміни, які насправді не впливають на фінансові показники, але можуть бути помилково розцінені алгоритмом як успішні.

У рамках процедури технічної верифікації із застосуванням методу Монте-Карло було здійснено генерацію 10 000 незалежних експериментальних циклів за умови дотримання ідентичності параметрів конверсії стосовно обох груп, що формалізується рівністю $p_c = p_t$. За підсумками проведеного моделювання зафіксовано емпіричну частоту прецедентів помилкового виявлення тестом значущої різниці на рівні 0.0475, що у відсотковому еквіваленті становить 4.75%. Зіставлення отриманого індикатора із затвердженим пороговим рівнем значущості $\alpha = 0.05$ засвідчує високий ступінь точності застосованого статистичного критерію. Виявлена близькість величин виступає показником математичної узгодженості методології, оскільки частота помилкових спрацювань інструменту суворо відповідає допускам теорії ймовірностей для стохастичних процесів. Потенційне значне відхилення показників у бік збільшення свідчило б про гіперчутливість алгоритму, створюючи загрозу генерування потоку недостовірних аналітичних висновків, тоді як занижені параметри вказували б на надмірний консерватизм та ймовірність ігнорування фактичних перспектив зростання. З огляду на відповідність фактичного рівня похибок запланованим нормативам, а також враховуючи той факт, що розрахункове зміщення оцінки, визначене як Bias, є знехтувано малим і дорівнює ≈ -0.00001 , використаний математичний апарат слід вважати цілком валідним.

Наступним кроком після процедури підтвердження надійності критерію стало виконання теоретичного аналізу потужності, відомого як Power Analysis, метою якого визначено оцінку ризику виникнення помилки другого роду, що позначається символом β . Фундаментальною математичною основою для проведення обчислювальних операцій обрано метод стандартизації розміру ефекту Cohen's h , специфіка якого полягає у врахуванні нелінійності дисперсії розподілу Бернуллі шляхом застосування арксинус-перетворення згідно з формулою:

$$h = 2 \left(\arcsin \sqrt{p_t} - \arcsin \sqrt{p_c} \right) \quad (5.16)$$

У контексті досліджуваної ситуації реальний ефект від процедури впровадження інтер'єрного дизайну становить +3% до показника конверсії, що характеризується зростанням від рівня 10% до значення 13%. За результатами обчислень розрахункова потужність експерименту за умови наявного обсягу вибірки $N=1000$ склала лише 0.558 або 55.8%, а такий результат дозволяє констатувати, що ймовірність помилки другого роду дорівнює величині:

$$\beta = 1 - 0.558 = 0.442 \quad (5.17)$$

Стосовно бізнес-перспектив зазначений параметр набуває критичного статусу, оскільки він виступає сигналом про наявність 44% шансу того, що застосований статистичний тест виявиться неспроможним зафіксувати реально існуюче зростання ефективності. Фактично існує майже половинна ймовірність реалізації сценарію, за якого компанія ухвалить рішення про відмову від прибуткової стратегії, хибно ідентифікуючи її як неефективну внаслідок фактору недостатньої чутливості інструменту вимірювання.

Процедура графічного аналізу кривої потужності, що відображена на рисунку 5.11, уможлиблює візуалізацію поняття мінімального детектованого ефекту MDE. За умови поточного розміру вибірки $N=1000$ аналітичний тест здатний гарантовано, тобто з імовірністю 80%, ідентифікувати виключно значні зміни конверсії на рівні 4.5% і вище, що відповідає показнику конверсії $\approx 14.5\%$.

З огляду на ту обставину, що в досліджуваному випадку реальний ефект є меншим та становить +3%, він автоматично потрапляє в так звану сіру зону чутливості. Попри той факт, що в межах даного конкретного експерименту вдалося успішно зафіксувати статистичну значущість, для забезпечення системної роботи подібний рівень надійності слід кваліфікувати як недостатній.

З метою підвищення параметрів потужності до загальноприйнятого стандарту 80% найбільш математично обґрунтованою рекомендацією виступає збільшення обсягу вибірки до рівня $N \approx 1500$ користувачів на групу. Фундаментальний фізичний зміст наведеної рекомендації ґрунтується на властивостях дисперсії вибіркової оцінки, котра є обернено пропорційною обсягу вибірки згідно зі співвідношенням $Var \propto 1/n$. Процес збільшення кількості спостережень з 1000 до 1500 неминуче призведе до зменшення стандартної похибки, наслідком чого стане звуження форми розподілів оцінок для контрольної та тестової груп. Мінімізація зони перекриття зазначених розподілів надасть статистичному критерію можливість чіткіше розрізняти корисний сигнал, а саме реальну різницю конверсій, на фоні наявного шуму, що суттєво знизить ризик пропуску корисного для бізнесу впровадження.

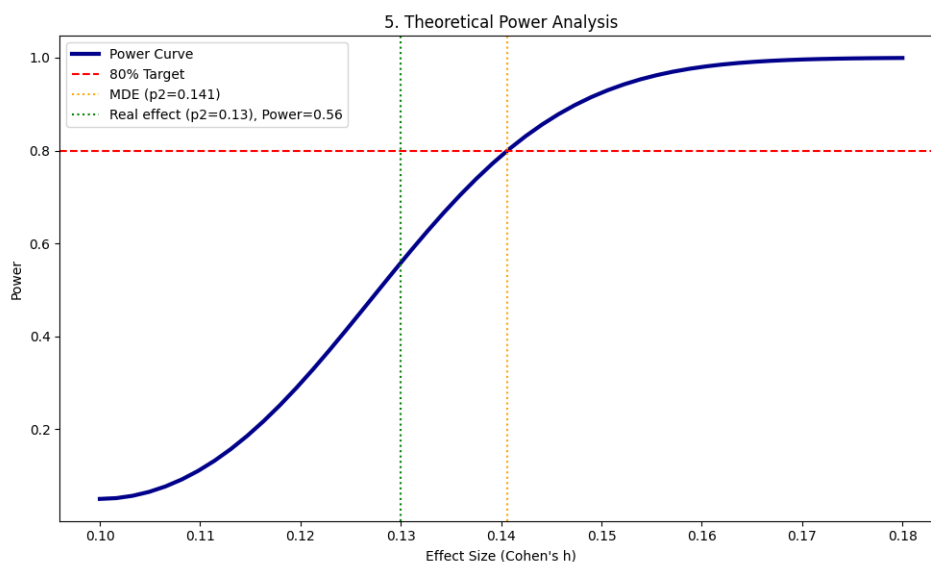


Рисунок 5.11. Крива статистичної потужності: здатність тесту виявляти ефект

ВИСНОВКИ

В результаті виконання дипломної роботи було проведено комплексне дослідження методу максимальної правдоподібності, проаналізовано його теоретичні властивості та реалізовано практичне застосування для дискретних і неперервних розподілів. Отримані результати дозволяють зробити наступні висновки:

1. Проведено теоретичний аналіз властивостей оцінок максимальної правдоподібності, зокрема узгодженості, асимптотичної нормальності та ефективності за критерієм Фішера. Обґрунтовано, що при збільшенні обсягу вибірки оцінки ММП наближаються до істинних значень параметрів, а їх дисперсія досягає мінімальної межі Крамера–Рао.
2. Розроблено програмне забезпечення мовою Python із використанням бібліотек Pandas, NumPy та SciPy, яке дозволяє проводити статистичне оцінювання, моделювання розподілів та візуалізацію результатів. Реалізовано алгоритми для обробки великих масивів даних та генерації синтетичних вибірок.
3. За допомогою методу Монте-Карло емпірично підтверджено властивості асимптотичної незсуненості та узгодженості оцінок. Встановлено, що зі зростанням обсягу вибірки абсолютне зміщення стрімко зменшується від рівня 0.00198 для малих масивів даних до значення 0.000002 при кількості елементів N рівній 10000, а дисперсія оцінки асимптотично наближається до нуля.
4. На прикладі аналізу тривалості поїздок таксі доведено перевагу логнормального розподілу над експоненціальним. Виявлено, що експоненціальна модель не здатна коректно описати унімодальність реальних даних, яка характеризується піковими значеннями в діапазоні від 5 до 10 хвилин, тоді як логнормальний розподіл демонструє високу узгодженість з емпіричною гістограмою та менші значення інформаційних критеріїв.

5. Здійснено практичну перевірку застосування методу максимальної правдоподібності для аналізу бінарних даних за схемою Бернуллі в рамках А/В тестування конверсії веб-користувачів. Результати статистичних тестів, зокрема критерію Вальда та відношення правдоподібності, а також байєсівського аналізу підтвердили статистичну значущість змін, показавши потенційне зростання продажів на 37.5% при впровадженні нового дизайну.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. [1] Fisher R. A. On the mathematical foundations of theoretical statistics // Philosophical Transactions of the Royal Society of London, Series A. – 1922. – Vol. 222. – P. 309–368.
2. [2] Almudevar A. Theory of Statistical Inference. – CRC Press, 2021.
3. [3] Bishop C. M., Bishop H. Deep Learning: Foundations and Concepts. – Springer, 2023.
4. [4] Murphy K. P. Machine Learning: An introduction. – MIT Press, 2022.
5. [5] Lehmann E. L., Casella G. Theory of Point Estimation. 2nd ed. – Springer, 1998.
6. [6] Wasserman L. All of Statistics: A Concise Course in Statistical Inference. – Springer, 2004.
7. [7] Rao C. R. Information and the accuracy attainable in the estimation of statistical parameters // Bulletin of the Calcutta Mathematical Society. – 1945. – Vol. 37. – P. 81–91.
8. [8] Efron B., Hastie T. Computer Age Statistical Inference: Algorithms, Evidence, and Data Science. – Cambridge University Press, 2016.
9. [9] Little R. J. A., Rubin D. B. Statistical Analysis with Missing Data. 3rd ed. – Wiley, 2019.
10. [10] Wainwright M. J. High-Dimensional Statistics: A Non-Asymptotic Viewpoint. – Cambridge University Press, 2019.
11. [11] Goodfellow I., Bengio Y., Courville A. Deep Learning. – MIT Press, 2016.
12. [12] Gretton A., Borgwardt K. M., Rasch M. J., Schölkopf B., Smola A. A Kernel Two-Sample Test // Journal of Machine Learning Research. – 2012. – Vol. 13. – P. 723–773.
13. [13] Gelman A., et al. Bayesian Data Analysis. 3rd ed. – CRC Press, 2013.
14. [14] Документація Scikit-learn [Електронний ресурс]. – Режим доступу до ресурсу: <https://scikit-learn.org/stable/>.

15. [15] Штуль В. С. Неперервний розподіл. GitHub. URL: https://github.com/valentin808/exponential_experiment.git.
16. [16] Штуль В. С. Дискретний розподіл. GitHub. 04.12.2025. URL: https://github.com/valentin808/exponential_experiment.git.
17. [17] Документація Python [Електронний ресурс]. – Режим доступу до ресурсу: <https://docs.python.org/3/>.
18. [18] Murphy K. P. Machine Learning: Advanced Topics. – MIT Press, 2023.
19. [19] New York City Taxi Trip Duration [Електронний ресурс] // Kaggle. – Режим доступу до ресурсу: <https://www.kaggle.com/c/nyc-taxi-trip-duration>.
20. [20] Stephens M. A. EDF Statistics for Goodness of Fit and Some Comparisons // Journal of the American Statistical Association. – 1974. – Vol. 69, No. 347. – P. 730–737.
21. [21] Rubinstein R. Y., Kroese D. P. Simulation and the Monte Carlo Method. 3rd ed. – John Wiley & Sons, 2016.
22. [22] Quantile-Quantile Plot in Python [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.geeksforgeeks.org/machine-learning/qqplot-quantile-quantile-plot-in-python/>.
23. [23] NumPy Random Binomial [Електронний ресурс]. – Режим доступу до ресурсу: <https://numpy.org/doc/stable/reference/random/generated/numpy.random.binomial.html>.
24. [24] Gupta A. K., Nadarajah S. Handbook of Beta Distribution and Its Applications. – CRC Press, 2004.