МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

ТЕОРІЯ ЙМОВІРНОСТЕЙ, ІМОВІРНІСНІ ПРОЦЕСИ ТА МАТЕМАТИЧНА СТАТИСТИКА

КУРС ЛЕКЦІЙ
Частина 2

Навчальний посібник

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня бакалавра
за спеціальностями 126 «Інформаційні системи та технології»,
121 «Інженерія програмного забезпечення»

Укладачі: О.А. Павлов,
О. В. Гавриленко,
О.Г. Жданова

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2022

Рецензент *Клесов О.І.*, д.ф.-м.н., професор, завідувач кафедри математичного аналізу та теорії ймовірностей, фізико-математичний факультет, КПІ ім. Ігоря Сікорського

Відповідальний редактор *Жураковська О.С.*, к.т.н., доцент кафедри інформаційних систем та технологій, факультет інформатики та обчислювальної техніки

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № X від DD.MM.YYYY р.)
за поданням Вченої ради факультету/навчально-наукового інституту
(протокол № 4 від 28.12.2021 р.)

Навчальний посібник призначений для студентів спеціальності 126 «Інформаційні системи та технології» кафедри інформаційних систем та технологій та спеціальності 121 «Інженерія програмного забезпечення» кафедри інформатики та програмної інженерії всіх освітніх програм та всіх форм навчання. У посібнику наведено систематизоване викладення багатьох основних відомостей з теорії ймовірностей, ймовірнісних процесів та математичної статистики з орієнтацією на розв’язання задач з вказаної дисципліни студентами технічних спеціальностей. Метою автора було розвинути навички студентів до розв’язання задач з дисципліни «Теорія ймовірностей та математична статистика» для їх застосування в інженерній практиці.

Реєстр. № НП XX/XX-XXX. Обсяг 3,0 авт. арк.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, м. Київ, 03056
https://kpi.ua
Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

ЗМІСТ

ВСТУП

Дисципліна "Теорія ймовірностей, імовірнісні процеси і математична статистика" має на меті вивчення основ математичної теорії ймовірностей, імовірнісних процесів, вивчення ймовірнісно-статистичного матеріалу, вміння розв'язувати різноманітні задачі згідно класичної схеми, способи перерахунку елементів скінчених множин, вміння застосовувати набуті знання до розв'язку прикладних задач, які виникають на практиці, вміння будувати математичні моделі реальних процесів.

По закінченню вивчення дисципліни студент набуває такі навики та вміння:

- володіння основною термінологією дисципліни, вміння пояснити зміст базових понять і розділів;
- вміння класифікувати стандартні задачі за ознаками, вміння розв'язувати їх;
- володіння навиками роботи з літературою з дисципліни, знання основних підручників, довідників, таблиць. Вміння знайти у літературі необхідну інформацію (спосіб розв'язання задач);
- знання основних прийомів розв'язання стандартних задач теорії ймовірностей (комбінаторні методи, методи, пов'язані з основними теоремами, метод характеристичних функцій, методи розрахунків характеристик дискретних та неперервних розподілів, вміння користуватися біноміальним розподілом та його граничними випадками, нормальним розподілом, законом великих чисел, центральною граничною теоремою тощо)
- вміння проводити стандартні статистичні розрахунки вручну та з використанням комп'ютерних програм, знання основних статистичних процедур: розрахунків середніх, дисперсій, моментів, коефіцієнтів кореляції.

КОМПЕТЕНТНОСТІ

Загальні компетентності:

- ЗК-1. Здатність до абстрактного мислення, аналізу та синтезу.

Спеціальні (фахові, предметні) компетентності

- ФК-13. Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень.

ПРОГРАМНІ РЕЗУЛЬТАТИ НАВЧАННЯ

- ПРН-1. Знати лінійну та векторну алгебру, диференціальне та інтегральне числення, теорію функцій багатьох змінних, теорію рядів, диференціальні рівняння для функції однієї та багатьох змінних, операційне числення, теорію ймовірностей та математичну статистику в обсязі, необхідному для розробки та використання інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації.
- ПРН-2. Застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв’язанні задач проєктування і використання інформаційних систем та технологій.

ПРЕРЕКВІЗИТИ ТА ПОСТРЕКВІЗИТИ ДИСЦИПЛІНИ (МІСЦЕ В СТРУКТУРНО-ЛОГІЧНІЙ СХЕМІ НАВЧАННЯ ЗА ВІДПОВІДНОЮ ОСВІТНЬОЮ ПРОГРАМОЮ)

Пререквізити:

- Спеціальні розділи математики.

Постреквізити:

- Ймовірністні моделі та статистичне оцінювання в інформаційно-управляючих системах.

Навчальний посібник призначений для студентів спеціальності 126 «Інформаційні системи та технології» кафедри інформаційних систем та технологій та спеціальності 121 «Інженерія програмного забезпечення» кафедри інформатики та програмної інженерії всіх освітніх програм та всіх форм навчання.

У посібнику наведено систематизоване викладення багатьох основних відомостей з теорії ймовірностей, ймовірнісних процесів та математичної статистики з орієнтацією на розв’язання задач з вказаної дисципліни студентами технічних спеціальностей. Метою автора було розвинути навички студентів до розв’язання задач з дисципліни «Теорія ймовірностей та математична статистика» для їх застосування в інженерній практиці.

МАТЕМАТИЧНА СТАТИСТИКА

Емпірична функція розподілу

Теоретична функція розподілу $F(x)$ невідома. Над випадковою величиною X проведено n випробувань та фіксуються результати випробувань.

Примітка. в статистиці $x_1 \dots x_n$ – результати випробувань, а n – кількість випробувань.

Ці n чисел звуться *вибіркою* об'ємом n.

По вибірці довжиною n будуємо варіаційний ряд (проведемо сортування членів виборці по збільшенню, де $x_{(1)}$ – найменше та $x_{(n)}$ – найбільше).

Означення. Емпіричною функцією розподілу $W_n(x)$ зветься:

$$W_n(x) = \frac{n_x}{n} = W_n(X < x),$$

де n_x – кількість членів варіаційного ряду, менших за x.

Властивості емпіричної функції розподілу:

1) $\forall x_{(i)} \le x \le x_{(i+1)}$, $i = \overline{1, n-1}$ – емпірична функція розподілу $W_n(x)$ – стала;

2) в точках $x_{(i)}$, $i = \overline{1, n-1}$ емпірична функція розподілу має розрив першого роду на величину $\frac{k}{n}$, де k –кількість членів варіаційного ряду рівних x_i;

3) $\forall x$ при $n \to \infty$ $W_n(x) \to F(x)$ з ймовірністю одиниця. Це означає, що при достатньо великої кількості випробувань nемпірична функція розподілу достатньо точно апроксимує $F(x)$.

Гістограма (емпірична функція щільності)

По варіаційному ряду необхідно побудувати функцію щільності для його оцінки (рисунок 1).

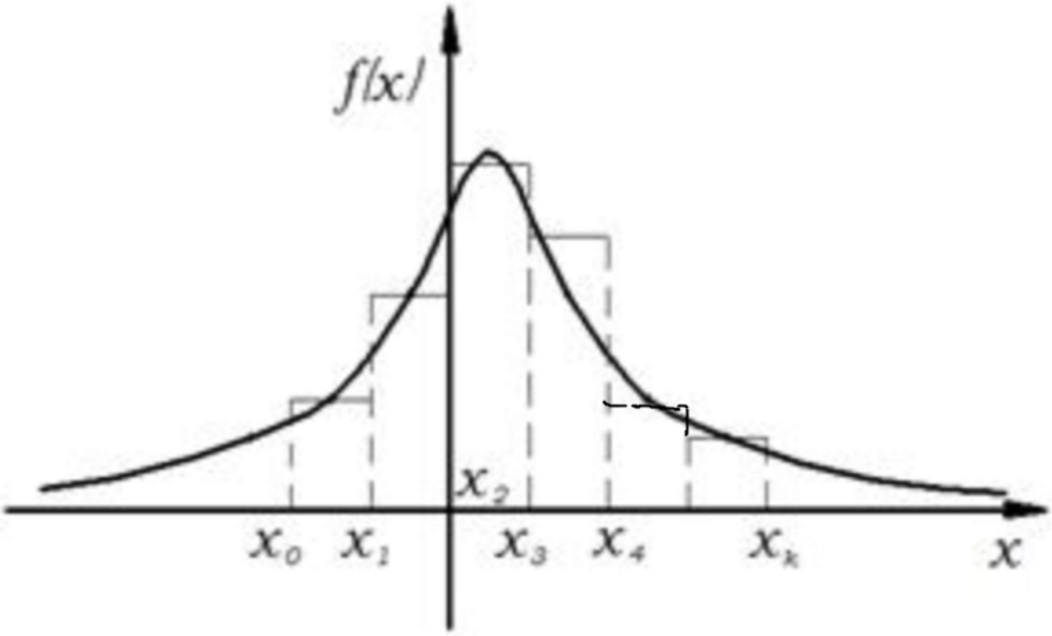


Рисунок 1

Площа кожного прямокутника дорівнює інтегралу від функції щільності по лівому і правому кінцю його основи.

Прямокутна фігура апроксимує функцію тим краще, чим менша довжина відрізку. Проща кожного прямокутника рівна інтегралу від функції щільності по кінцям відрізка, що є основою прямокутника.

Алгоритм побудові гістограми

1) На числовій осі розглядають відрізок $[x(1), x(n)]$.

2) Обираємо число $a \in [8; 12]$.

Примітка. Інтервал знайдено статистично

3) Відрізок $[x(1), x(n)]$ розіб'ємо на відрізки, кожен з котрих містить одне й теж саме число a членів варіаційного ряду та кінці кожного відрізка співпадають з членами варіаційного ряду.

4) На кожному відрізку, як на основі, будуємо прямокутники з однаковою площиною $\frac{a}{n}$.

Побудована фігура і є гістограмою – емпіричною функцією щільності досліджуваної функції щільності випадкової величини X.

Неоднозначністю цього алгоритму є вибір кінців відрізків, на котрих будуються прямокутники. В деяких підручниках рекомендується відрізок $(x(1), x(n))$ розбивати на відрізки рівної довжини (а не по кількості членів варіаційного ряду).

Насправді, гістограма сама по собі нікому не потрібна та використовується тільки для вирішення наступної задачі – І етап розділу мат. статистики «Перевірка складної гіпотези». Суть полягає в наступному: існують статистичні таблиці, де на графіку приведені усі найбільш відомі функції щільності випадкових величин і для кожної функції щільності наведені інженерні задачі, де найчастіше зустрічаються ці функції. Тоді на першому етапі дослідник будує гістограму, порівнює її зі усіма відомими графіками (усіх відомих йому теоретичних функцій щільності), та висуває гіпотезу, що досліджувана випадкова величина X має визначену функцію щільності. А далі необхідно перевірити цю гіпотезу.

ОСНОВИ ТЕОРІЇ ВИБІРКОВОГО МЕТОДУ

Це перший розділ математичної статистики, де вирішується наступна задача: є неперервна випадкова величина X та відома її функція щільності з точністю до числових значень параметрів

Приклад. Нормальний розподіл $n(x, \nu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\nu)^2}{2\sigma^2}}$ має два числових параметра: ν та $\sigma \geq 0$.

Використовуючи вибірку об'єму n (результати n випробувань), необхідно оцінити невідомі значення n параметрів (в нашому прикладі необхідно оцінити математичне сподівання та дисперсію).

Така постановка задачі позникає коли досліджувана випадкова величина задовольняє умовам граничної теоремі (з інженерної точки зору)

Приклад. Виявилось, що інтегральна похибка вимірювань задовольняє загальному випадку центральної граничної теореми (наприклад, є сумою 6-12 незалежних складових, кожна з яких не домінує над іншими). Тоді вважається, що інтегральна похибка вимірювань має нормальний розподіл та по результатам вимірювань необхідно оцінити математичне сподівання і дисперсію.

Існує два способи оцінки невідомих параметрів:

1) точковий – оцінкою параметра $\theta_i,\ i = \overline{1, c}$ є числова скалярна функція, аргументи якої є членами вибірки $\theta_i \approx \widehat{\theta_i}(x_1, \dots, x_n)$;

2) інтервальна оцінка. Вводяться дві функції n аргументів, в які підставляються члени вибірки (ряду): $\underline{\theta_i}(x_1, \dots, x_n)$ та $\overline{\theta_i}(x_1, \dots, x_n)$.

Ці функції задовольняють умові:

з імовірністю $1 - \alpha$ ($\alpha \leq 0.05$; $\alpha \neq 0$) має виконуватися:

$$\underline{\theta_i}(x_1, \dots, x_n) \leq \theta \leq \overline{\theta_i}(x_1, \dots, x_n).$$

Математична статистика побудована з використанням наступного прийому: маємо фізично існуючу випадкову величину X та n результатів випробування на нею ($x_1 \dots x_n$ – вибірка об'єму n). Можна вважати, що ці результати є результатом одного композиційного випробування над віртуальною n – вимірною випадковою величиною $X_1 X_2 \dots X_n$, де $X_i,\ i = \overline{1, n}$ – незалежні віртуальні копії фізично існуючої випадкової величини X (тобто число $x_i,\ i = \overline{1, n}$ – результат випробування над $X_i,\ i = \overline{1, n}$).

<u>Побудова точкових оцінок</u>

Означення. Вибірковою характеристикою, побудованою для оцінки невідомого параметра $\theta_i, i = \overline{1, c}$, зветься числова скалярна функція n випадкових аргументів, що є незалежними копіями фізично існуючої випадкової величини X:

$$\widehat{\theta_i}(X_1, \dots, X_n).$$

Оцінка невідомого параметра береться як реалізація цієї випадкової величини:

$$\theta_i \approx \widehat{\theta_i}(x_1, \dots, x_n),$$

де x_1 – реалізація випадкової величини X_1,…, x_n – реалізація випадкової величини X_n; $x_1, \dots, x_n$ – вибірка, об’ємом n.

Властивості, яким повинна задовольняти вибіркова характеристика, щоб будь-яка її реалізація добре оцінювала параметр $\theta_i, i = \overline{1, c}$:

– незміщеність;

– ефективність;

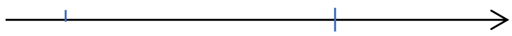
– спроможність.

Незміщеність

$$M\hat{\theta}_i(X_1, \dots, X_n) = \theta_i$$

В цьому випадку результати конкретних випробувань над випадковою величиною $\widehat{\theta_i}(X_1, \dots, X_n)$ будуть групуватись відносного свого математичного сподівання - θ_i.

<u>Обґрунтування:</u>



$M\widehat{\theta_i}(X_1, \dots, X_n)$

Якби $M\widehat{\theta_i}(X_1, \dots, X_n) \neq \theta_i$, то результати конкретних випробувань $\widehat{\theta_i}$ групувалися б не навколо θ_i і були б поганою оцінкою θ_i.

Ефективність

Вибіркова характеристика є ефективною, якщо вона є незміщеною і має найменшу дисперсію з усіх можливих незміщених вибіркових характеристик, побудованих для оцінки θ_i.

<u>Обґрунтування:</u> чим менша дисперсія, тим ближче групуються результати випробувань над випадковою величиною $\widehat{\theta_i}(X_1, \dots, X_n)$ відносно $M\widehat{\theta_i}$.

Спроможність

Вибіркова характеристика $\widehat{\theta}_i(X_1, \dots, X_n)$ зветься спроможною, якщо при $n \to \infty$ ряд вибіркових характеристик збігається до θ_i:

$\lim\limits_{n\to\infty} \widehat{\theta}_i(X_1, \dots, X_n) = \theta_i$ (по відповідному означенню збіжності: з ймовірністю 1, по ймовірності, в середньостатичному).

Приклад

$\nu = MX$ – невідомий параметр.

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$M\bar{X}_n = \nu.$$

σ^2 – невідомий параметр.

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2,$$

$$MS_n^2 = M\frac{1}{n}\sum_{i=1}^{n}(X_i - \nu - (\bar{X}_n - \nu))^2 =$$

$$= M\frac{1}{n}\sum_{i=1}^{n}((X_i - \nu)^2 - 2(X_i - \nu)(\bar{X}_n - \nu) + (\bar{X}_n - \nu)^2) =$$

$$= \left[-2M(\bar{X}_n - \nu)\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(X_i - \nu) = -2M(\bar{X}_n - \nu)^2 = -2D\bar{X}_n\right] =$$

$$= \sigma^2 - 2D\bar{X}_n + D\bar{X}_n = \sigma^2 - D\bar{X}_n = \left[D\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\sigma^2}{n}\right] = \sigma^2 - \frac{\sigma^2}{n}.$$

Вибіркова характеристика S_n^2 є зміщеною вибірковою характеристикою відносно параметра σ^2. Але одночасно вона є асимптотично незміщеною на нескінченності:

$$MS_n^2 \to \sigma^2, \text{ при } n \to \infty.$$

Вибіркова характеристика $\bar{X}_n$ для оцінки вибіркового параметра ν є спроможною відповідно до посиленого закону великих чисел:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \nu, \text{ при } n \to \infty,$$

(з ймовірністю 1).

Задача (приклад на ефективність в класі лінійних оцінок)

Маємо $X_1, \dots, X_n$ незалежних випадкових величини з невідомим, але однаковим математичним сподіванням ν, і різними, але відомими дисперсіями ${\sigma_i}^2$, $i = \overline{1, n}$. Провели випробування, отримали результати випробувань $x_1, \dots, x_n$, де x_i – результат випробування над випадковою величиною X_i. За результатами випробувань треба оцінити ν.

Фізична інтерпретація постановки задачі

Є одна фізична величина, яка вимірюється n приладами, в кожного з яких нульова систематична помилка, а точність оцінюється дисперсіями. Кожним приладом виміряли цю фізичну величину. По результатам випробувань потрібно оцінити цю фізичну величину

Розв'язання. Будемо шукати розв'язок задачі в класі лінійних незміщених оцінок математичного сподівання.

Означення. Вибіркова характеристика зветься лінійною, якщо

$$\nu = M \sum_{i=1}^{n} g_i X_i.$$

З умови незміщеності випливає, що $\sum_{i=1}^{n} g_i = 1$.

Тоді:

$$D\sum_{i=1}^{n} g_i X_i = {g_1}^2{\sigma_1}^2 + {g_2}^2{\sigma_2}^2 + \dots + {g_{n-1}}^2{\sigma_{n-1}}^2 +$$
$$+(1 - g_1 - g_2 - \dots - g_{n-1})^2{\sigma_n}^2$$

З умови ефективності випливає, що цей вираз має бути мінімальним. Оскільки цей вираз є квадратичним, то він має один екстремум. Для його знаходження беремо часткові похідні по g_i і прирівнюємо їх до нуля:

$$2g_i{\sigma_i}^2 = 2g_n{\sigma_n}^2, i = \overline{1, n-1}.$$

Тепер необхідно довести, що ця система має один розв'язок. Тоді будь-який розв'язок системи є єдиним. Для цього необхідно показати, що детермінант рівнянь лінійної системи відносно $g_1, \dots, g_n$не дорівнює нулю. Показати самостійно.

Для знаходження коефіцієнтів $g_1, \dots, g_n$ підставляємо будь-яке значення для g_n і з перших $n-1$ рівнянь знаходимо $g_1, \dots, g_{n-1}$. Тоді:

$$\widehat{g_i} = \frac{g_i}{\sum_{i=1}^{n} g_i}.$$

Показати самим, що $\widehat{g_i}$ задовольняють усім рівняннм системи, тобто є розв'язком.

МЕТОД НАЙБІЛЬШОЇ ПРАВДОПОДІБНОСТІ

Постановка задачі

Неперервний випадок

X – неперервна випадкова величина. Є функція щільності $f(u, \theta_1, \dots, \theta_c)$ з невідомими параметрами $\theta_1, \dots, \theta_c$ та вибірка $x_1, \dots, x_n$. Вважаємо $x_1, \dots, x_n$ результатами випробувань над $X_1 \dots X_n$, де X_i, $i = \overline{1, n}$– незалежні копії X. Знайдемо n-вимірну функцію щільності n-мірної випадкової величини $X_1, \dots, X_n$. Вона дорівнює:

$$f(u_1, \dots, u_n) = \prod_{i=1}^{n} f(u_i, \theta_1, \dots, \theta_c).$$

Означення. Функція найбільшої правдоподібності – числова скалярна функція c-аргументів $\theta_1 \dots \theta_c$, що дорівнює:

$$L(\theta_1, \dots, \theta_c) = \prod_{i=1}^{n} f(x_i, \theta_1, \dots, \theta_c).$$

Примітка. В n-вимірній функції щільності зафіксували значення аргументів: $\mathrm{u_i} = \mathrm{x_i}$, $\mathrm{i} = \overline{1, \mathrm{n}}$, де $\mathrm{x_i}$ – члени вибірки.

Дискретний випадок

$$X \sim \begin{pmatrix} y_1 & \dots & y_s \\ P_1(\theta_1, \dots, \theta_c) & \dots & P_s(\theta_1, \dots, \theta_c) \end{pmatrix}$$

Приклад. Маємо урну з білими та чорними кульками.

$P_1 = \frac{\theta}{N}$, – ймовірність появи білої кульки,

$P_2 = 1 - \frac{\theta}{N}$ – ймовірність появи чорної кульки, де N загальна кількість кульок, θ – кількість білих шариків – невідомий параметр.

Провели n випробувань, в яких y_1 настало m_1 раз, y_2 - m_2 раз, … , y_s - m_s раз ($\sum_{i=1}^{s} m_i = n$).

Означення. Функцією найбільшої правдоподібності називається вираз:

$$L(\theta_1, \dots, \theta_c) = \prod_{i=1}^{n} P_i^{m_i}(\theta_1, \dots, \theta_c).$$

В якості оцінки невідомих параметрів $\theta_1 \dots \theta_c$ Роберт Фішер запропонував брати такі значення $\theta_1 \dots \theta_c$, на котрих досягається максимум функції найбільшої правдоподібності, тобто:

$$\max_{\theta_1 \dots \theta_c} L(\theta_1, \dots, \theta_c).$$

Обґрунтування методу найбільшої правдоподібності

В якості оцінки невідомих параметрів беруться ті, які максимізують ймовірність наставання події, яка настала внаслідок проведених випробувань над випадковою величиною X.

Дискретний випадок

$$P = \prod_{i=1}^{s} P_i^{m_i}(\theta_1, \dots, \theta_c) = L(\theta_1, \dots, \theta_c)$$

(тобто максимізуємо ймовірність того, що настали саме ті результати випробування, які маємо).

Неперервний випадок

Розглянемо ймовірність того, що внаслідок проведених випробувань настане подія:

$$P(x_i \le X_i \le x_i + \Delta x_i) = L(\theta_1, \dots, \theta_c) \prod_{i=1}^{n} \Delta x_i + o\left(\prod_{i=1}^{n} \Delta x_i\right), i = \overline{1, n},$$

де x_i — члени вибірки, а випадкові величини X_i потрапили в відрізки $x_i \le X \le x_i + \Delta x_i$, де Δx_i — достатньо малі числа).

Таким чином максимізація L по параметрам $\theta_1 \dots \theta_c$ еквівалентна максимізації ймовірності потрапляння випадкової величини $X_1 \dots X_n$ в прямокутник, в який вона попала внаслідок проведених випробувань. Отримали задачу на безумовний екстремум:

$$\frac{\partial L(\theta_1,\dots,\theta_c)}{\partial \theta_i} = 0,\ i = \overline{1,n}.$$

Розв'язання. Необхідно знайти усі розв'язки цієї системи рівнянь та обрати той, котрому відповідає максимум функції найбільшої правдоподібності.

Задача спрощується, якщо розглядати

$$\frac{\partial \ln L(\theta_1,\dots,\theta_c)}{\partial \theta_i} = 0,\ i = \overline{1,n},$$

тобто обирають похідні не від функції L, а від її логарифма, оскільки логарифм–монотонно зростаюча функція, тому максимум $\ln L$ не співпаде з максимумом L, але досягається на тих же самих значеннях аргументів. Оскільки L – добуток функцій, то $\ln L$ – сума, тобто система рівнянь спрощується.

Теорема (без доведення)

Оцінки параметрів, отримані методом найбільшої правдоподібності є спроможними та асимптотично незміщеними, ефективними та розподіленими нормально.

Приклад. X має нормальний розподіл, тобто

$$n(x,\nu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\nu)^2}{2\sigma^2}},$$

та є вибірка $x_1, \dots, x_n$. Треба знайти параметри ν та σ^2.

n-вимірна функція щільності:

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \prod_{i=1}^{n} e^{-\frac{(u_i-\nu)^2}{2\sigma^2}} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(u_i-\nu)^2}{2\sigma^2}}.$$

Функція найбільшої правдоподібності має вигляд:

$$L(\theta_1, \dots, \theta_c) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\nu)^2}{2\sigma^2}} = L(\nu, \sigma^2).$$

Якою б не була оцінка σ^2, найбільше значення L досягає при мінімуму $\sum_{i=1}^{n}(x_i - \nu)^2$ по ν, тобто:

$$\min_{\nu} \sum_{i=1}^{n} (x_i - \nu)^2.$$

Звідси випливає, що $\hat{\nu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ – оцінка ν.

Примітка. Отримали реалізацію вибіркової характеристики оцінки математичного сподівання виду:

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(а її реалізація: $\frac{1}{n}\sum_{i=1}^{n} x_i$).

Замість математичного сподівання ν в $L(\nu, \sigma)$ підставляємо $\overline{x_n}$ та знаходимо $\arg\max_{\sigma^2} \ln L(\bar{x}, \sigma^2)$. Отримуємо:

$$\widehat{\sigma^2} = S_1^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x_n})^2.$$

Отримали реалізацію вибіркової характеристики

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X_n})^2$$

ІНТЕРВАЛЬНІ ОЦІНКИ

Розглянемо наступні спеціальні розподіли.

Розподіл χ^2

Означення. Випадкова величина має розподіл $\boldsymbol{\chi^2}$з n-ступенями свободи, якщо її можна представити у вигляді суми $\chi^2 = \sum_{i=1}^{n} X_i^2$, де

$X_1 \ldots X_n$ є незалежними і розподіленими нормовано нормально випадковими величинами. Їх функція щільності дорівнює

$$n(x, 0{,}1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Знайдемо функцію щільності розподілу χ^2. Для цього побудуємо її функцію розподілу:

$$K_n(x) = P(\chi^2 \leq x) = \begin{cases} K_n(x), \\ 0, x < 0 \end{cases} = P\left(\sum_{i=1}^{n} X_i^2 \leq x\right), x \geq 0$$

(функція розподілу, та функція щільності дорівнює 0 для від'ємних значень аргументів, оскільки випадкова величина χ^2 невід'ємно означена). Розглядаємо випадки, коли аргумент є невід'ємним:

$$P\left(\sum_{i=1}^{n} X_i^2 \leq x\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \iint\limits_{\sum_{i=1}^{n} x_i^2 \leq x} e^{-\frac{\sum_{i=1}^{x} x_i{}^2}{2}} dx_1 \ldots dx_n$$

Примітка. x_i - аргументи n-вимірної функції щільності n-вимірної випадкової величини $X_1, \ldots, X_n$.

$$I = \iint\limits_{\sum_{i=1}^{n} x_i^2 \leq x} dx_1 \ldots dx_n$$

Розглянемо цей інтеграл і зробимо заміну змінних:

$$x_i = y_i \sqrt{x}, i = \overline{1, n},$$

$$dx_i = dy_i \sqrt{x},$$

$$I = x^{\frac{n}{2}} \iint\limits_{\sum_{i=1}^{n} y_i^2 \leq 1} dy_1 \ldots dy_n = C_n x^{\frac{n}{2}},$$

$$C_n = \iint\limits_{\sum_{i=1}^{n} x_i^2 \leq x} dx_1 \ldots dx_n,$$

c_n- константа, від *x* не залежить і є об'ємом *n*-вимірної сфери радіусом 1, з центром в початку координат.

$$K_n(x+h) - K_n(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n C_n e^{\frac{-(x+\theta h)}{2}}\left((x+h)^{\frac{n}{2}} - x^{\frac{n}{2}}\right),$$

$$0 < \theta < 1.$$

Примітка. $K_n(x+h) - K_n(x) =$

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n \iint\limits_{x \le \sum_{i=1}^n x_i^2 \le x+h} e^{-\frac{\sum_{i=1}^x {x_i}^2}{2}} dx_1 \dots dx_n.$$

Далі використана **теорема про середнє**:

$$k_n(x) = \lim_{h\to 0} \frac{K_n(x+h) - K_n(x)}{h} = \lim_{h\to 0} \left(\frac{1}{\sqrt{2\pi}}\right)^n \mathrm{C}_n e^{-\frac{x+\theta h}{2}} \left(\frac{(x+h)^{\frac{n}{2}} - x^{\frac{n}{2}}}{h}\right) = T_n x^{\frac{n}{2}-1} e^{-\frac{x}{2}},$$

$$x \ge 0,$$

$$T_n = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{n\mathrm{C}_n}{2}.$$

Знайдемо компактний вигляд для константи T_n:

$$\int_0^\infty T_n x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx = 1,$$

$$T_n = \frac{1}{\int_0^\infty x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx}.$$

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx,$$

$$\Gamma(p) = (p-1)\Gamma(p-1),$$

$$\frac{x}{2} = y,$$

$$x = 2y,$$

$$dx = 2dy,$$

$$T_n = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}.$$

Знаходимо математичне сподівання, дисперсію, характеристичну функцію розподілу χ^2:

$$M(\chi^2) = \int_0^\infty x k_n(x)dx$$

$$M(\chi^2)^2 = \int_0^\infty x^2 k_n(x)dx$$

$$D(\chi^2) = M(\chi^2)^2 - (M\chi^2)^2$$

$$\varphi_{\chi^2}(t) = \int_0^\infty e^{itx} k_n(x)dx$$

Шукаємо перший початковий момент:

$$M(\chi^2) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^\infty x^{\frac{n}{2}} e^{-\frac{x}{2}} dx = \frac{2^{\frac{n}{2}+1}\Gamma\left(\frac{n}{2}+1\right)}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} = n.$$

Шукаємо другий початковий момент:

$$M(\chi^2)^2 = \frac{2^{\frac{n}{2}+2}\Gamma\left(\frac{n}{2}+2\right)}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} = n\left(\frac{n}{2}+1\right) = n^2 + 2n.$$

$$DX = 2n$$

$$\varphi_x(t) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^\infty x^{\frac{n}{2}-1} e^{-\frac{x}{2}(1-it)} dx = \frac{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)(1-2it)^{\frac{n}{2}}} = (1-2it)^{-\frac{n}{2}}$$

Наслідок. Сума двох незалежних випадкових величин, що має розподіл χ^2 з n_1 та n_2 ступенями свободи відповідно має розподіл $\chi_1^2 + \chi_2^2$ сумарної кількості ступенів свободи $\varphi_{(\chi_1^2+\chi_2^2)}(t) = (1-2it)^{-\frac{n_1}{2}}(1-2it)^{-\frac{n_2}{2}} = (1-2it)^{-\frac{n_1+n_2}{2}}$

Доведення завершено, тому що характеристична функція однозначно задає функцію щільності випадкової величини.

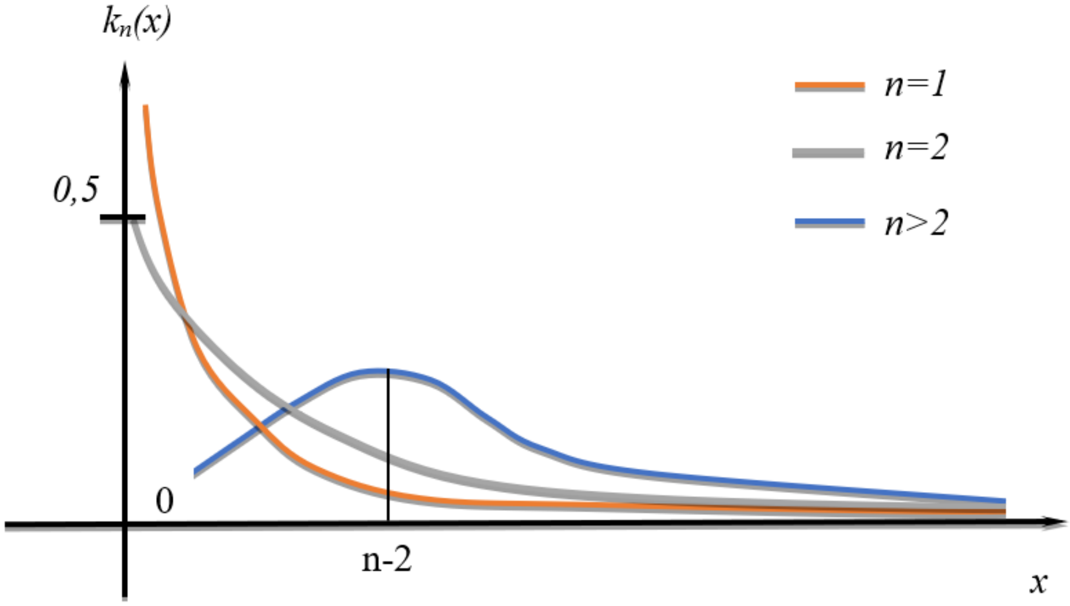


Рисунок 2

Розподіл Ст'юдента

Означення. Випадкова величина має розподіл Ст'юдента з k ступенями свободи, якщо

$$T = \frac{Z\sqrt{k}}{\sqrt{V}},$$

де Z розподілено нормовано нормально, V має розподіл χ^2 з k ступенями свободи.

Знайдемо функцію щільності випадкової величини T.

T приймає значення на всій числовій осі.

$$S_k(x) = P\left(\frac{Z\sqrt{k}}{\sqrt{V}} \le x\right) = P\left(Z \le x\frac{\sqrt{V}}{\sqrt{k}}\right) = P\left(ZV\begin{Bmatrix} 0 \le V < \infty \\ -\infty < Z \le x\dfrac{\sqrt{V}}{\sqrt{k}} \end{Bmatrix}\right)$$

Беремо інтеграл від двовимірної функції щільності по цій області:

$$\frac{1}{\sqrt{2\pi}2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}\iint_{\substack{0\le v<\infty \\ -\infty<z\le x\frac{\sqrt{v}}{\sqrt{k}}}} v^{\frac{k}{2}-1}e^{-\frac{v}{2}}e^{-\frac{z^2}{2}}dzdv =$$

$$= T_k \int_0^{\infty} v^{\frac{k}{2}-1} e^{-\frac{v}{2}} \left[\int_{-\infty}^{x\frac{\sqrt{v}}{\sqrt{k}}} e^{-\frac{z^2}{2}} dz \right]$$

Знаходимо функцію щільності як похідну від функції розподілу:

$$S_k(x) = \frac{dS_k(x)}{dx} = T_k \int_0^{\infty} v^{\frac{k}{2}-1} e^{-\frac{v}{2}} \left[\frac{d}{dx} \int_{-\infty}^{x\frac{\sqrt{v}}{\sqrt{k}}} e^{-\frac{z^2}{2}} dz \right] dv.$$

Треба зробити таку заміну змінних, щоб у внутрішньому інтегралі був тільки x:

$$\begin{pmatrix} z = y\dfrac{\sqrt{v}}{\sqrt{k}} \\ dz = \dfrac{\sqrt{v}}{\sqrt{k}} dy \\ \dfrac{d}{dx} \displaystyle\int_{-\infty}^{\infty} f(u) du = f(x) \end{pmatrix}$$

$$S_K(x) = \frac{T_k}{\sqrt{k}} \int_0^{\infty} v^{\frac{k-1}{2}} e^{-\frac{v}{2}} e^{-\frac{x^2 v}{2k}} dv = \frac{T_k}{\sqrt{k}} \int_0^{\infty} v^{\frac{k-1}{2}} e^{-\frac{v}{2}\left(1+\frac{x^2}{k}\right)} dv$$

$$= \frac{T_k 2^{\frac{k+1}{2}} \Gamma\left(\dfrac{k+1}{2}\right)}{\sqrt{k}(1+\frac{x^2}{k})^{\frac{k+1}{2}}} = D_k \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

При великих k розподіл Ст’юдента практично збігається з нормованим нормальним розподілом (рисунок 3).

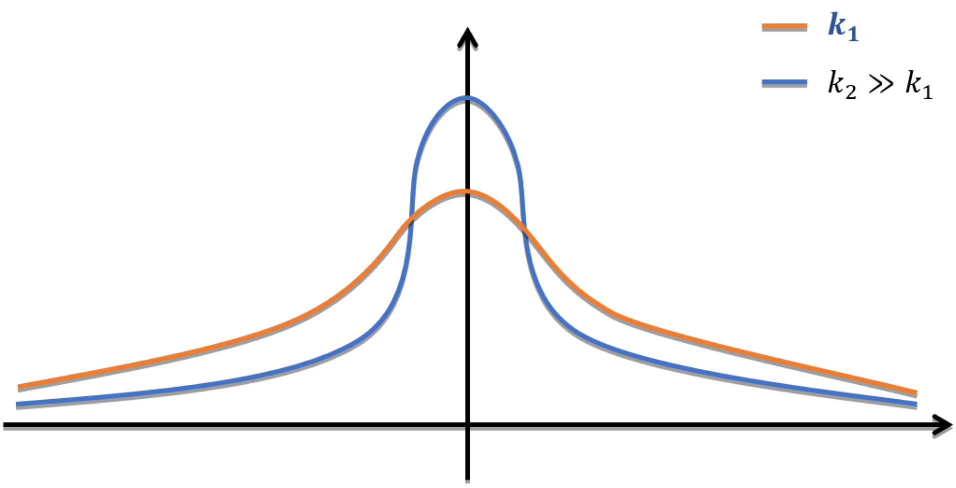


Рисунок 3

Розподіл Фішера

Означення. Випадкова величина F має розподіл Фішера з n_1 і n_2 ступенями свободи якщо:

$$F = \frac{\frac{V_1}{n_1}}{\frac{V_2}{n_2}},$$

де V_1, V_2 - незалежні випадкові величини, розподілені по закону χ^2 з n_1 і n_2 ступенями свободи відповідно.

Функція щільності випадкової величини F:

$$f_F(x) = \begin{cases} \dfrac{Cx^{\frac{n_1-2}{2}}}{(n_2+n_1x)^{\frac{n_1+n_2}{2}}}, x \geq 0 \\ x < 0 \end{cases}$$

Розподіл випадкової величини $\frac{nS^2}{\sigma^2}$

Випадкові величини $X_1, \dots, X_n$ незалежні, однаково розподілені за нормальним законом з довільними параметрами v, σ. Випадкова величина

$$nS^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2,$$

де $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Знайти розподіл випадкової величини $\frac{\mathrm{nS}^2}{\sigma^2}$.

Розв'язання.

Виконаємо заміну: $Z_i = X_i - \bar{X}_n = X_i - \frac{1}{n}\sum_{i=1}^{n} X_i$, $i = \overline{1, n}$. Випадкові величини Z_i мають нормальний розподіл.

Випадкові величини Z_i залежні (зв'язані лінійно), бо $\sum_{i=1}^{n} Z_i = 0$. Тому знаходження розподілу $\frac{\mathrm{nS}^2}{\sigma^2}$ стає нетрівіальним.

Покажемо, що випадкова величина nS^2 не залежить від математичного сподівання X_i. Дійсно, у вираз nS^2 підставимо $X_i + M$, де M – будь-яке дійсне число:

$$nS^2 = \sum_{i=1}^{n}(X_i + M - \frac{1}{n}\sum_{i=1}^{n}(X_i + M))^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

Тому можна вважати, що $MX_i = 0$ (Насправді MX_i може бути не рівним нулю, але це не впливає на значення nS^2).

Розглянемо перетворення випадкового вектора X:

$Y = AX$,

де $X = \begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix}, Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}$.

Розпишемо цю систему скалярно:

$$Y_1 = (X_1 - X_2)\frac{1}{\sqrt{2}}$$

$$Y_2 = (X_1 + X_2 - 2X_3)\frac{1}{\sqrt{2 \cdot 3}}$$

$$\dots$$

$$Y_i = (X_1 + \dots + X_i - iX_{i+1})\frac{1}{\sqrt{i(i+1)}}$$

$$\dots$$

$$Y_{n-1} = (X_1 + \dots + X_{n-1} - (n-1)X_n)\frac{1}{\sqrt{(n-1)n}}$$

$$Y_n = \frac{1}{\sqrt{n}}X_1 + \frac{1}{\sqrt{n}}X_2 + \dots + \frac{1}{\sqrt{n}}X_n = \sqrt{n}\bar{X}_n$$

Запишемо матрицю A:

$$\begin{pmatrix} \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{i(i+1)}} & \cdots & (\frac{1}{\sqrt{i(i+1)}} - \frac{i}{\sqrt{i(i+1)}}) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{(n-1)n}} & \cdots & \cdots & \cdots & \cdots & (\frac{1}{\sqrt{(n-1)n}} - \frac{n-1}{\sqrt{(n-1)n}}) \\ \frac{1}{\sqrt{n}} & \cdots & \cdots & \cdots & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix}$$

Самим довести наступні властивості матриці A:

$$\sum_{j=1}^{n} a_{i_1 j} a_{i_2 j} = 0, \forall i_1, i_2 i_1 \neq i_2$$

$$\sum_{i=1}^{n} a_{i j_1} a_{i j_2} = 0, \forall j_1, j_2 j_1 \neq j_2$$

$$\sum_{j=1}^{n} {a_{ij}}^2 = 1, i = \overline{1, n}$$

$$\sum_{i=1}^{n} {a_{ij}}^2 = 1, j = \overline{1, n}$$

A – ортогональна матриця (модуль детермінанта матриці рівний одиниці), $Y = AX$. З цього випливає, що $\sum_{i=1}^{n} {X_i}^2 = \sum_{i=1}^{n} {Y_i}^2$.

Тобто, лінійне перетворення матриці A є лише поворотом координат відносно початку координат.

Властивості компонент випадкового вектора Y:

$MY_i = 0, DY_i = \sigma^2, i = \overline{1, n}.$

Довести самим, використовуючи властивості матриці A.

$Y_i \dots Y_n$ – n-вимірна випадкова величина, яка має n-вимірний нормальний розподіл, тому що квадратна матриця A є ортогональною, а тому – невиродженою.

Наслідок. Випадкові величини Y_i розподілені нормально з

$MY_i = 0, DY_i = \sigma^2, i = \overline{1,n}$.

Покажемо, що $cov\left(Y_i, Y_j\right) = 0, \forall i \neq j$.

$cov\left(Y_i, Y_j\right) = M\left(Y_i, Y_j\right)$ у нашому випадку.

Далі використовуємо наступні формули:

$M\left(d_1 X_i d_2 X_j\right) = d_1 d_2 \mathrm{cov}(X_i, X_j) = 0$, оскільки випадкові величини X_i, X_j незалежні;

$M(d_1 X_i d_2 X_i) = d_1 d_2 \sigma^2$, бо $M{X_i}^2 = \sigma^2$ $(M(X_i) = 0)$.

Наслідок. Випадкові величини Y_i розподілені нормально за законом $n(x, 0, \sigma)$; є незалежними, бо $cov\left(Y_i, Y_j\right) = 0$.

Таким чином,

$$nS^2 = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 = \sum_{i=1}^{n} {X_i}^2 - 2n{\bar{X}_n}^2 + n{\bar{X}_n}^2 =$$

$$= \sum_{i=1}^{n} {X_i}^2 - n{\bar{X}_n}^2 == \sum_{i=1}^{n} {Y_i}^2 - {\bar{Y}_n}^2 = \sum_{i=1}^{n-1} {Y_i}^2,$$

$$nS^2 = \sum_{i=1}^{n-1} {Y_i}^2$$

$$\frac{nS^2}{\sigma^2} = \sum_{i=1}^{n-1} \left(\frac{Y_i}{\sigma}\right)^2$$

Таким чином, $\frac{nS^2}{\sigma^2}$ – розподіл χ^2 з $n - 1$ степенями свободи.

Примітка. у нас зменшилась на одиницю кількість доданків при переході від X_i до Y_i , бо існує один лінійний зв'язок $\sum_{i=1}^{n}\left(X_i - \overline{X_n}\right) = 0$,

$$\bar{X}_n = \frac{\bar{Y}_n}{\sqrt{n}}.$$

Випадкові величини $\bar{X}_n$, $\frac{nS^2}{\sigma^2}$ – незалежні, тому, що $\bar{X}_n = \frac{\bar{Y}_n}{\sqrt{n}}$, $\frac{nS^2}{\sigma^2} = \sum_{i=1}^{n-1}\left(\frac{Y_i}{\sigma}\right)^2$.

Інтервальні оцінки для параметрів нормального розподілу

Є випадкова величина X, що розподілена нормально. Провели n випробувань, маємо вибірку $x_1, \dots, x_n$, яку вважаємо результатом одного випробування над n-вимірною випадковою величиною $X_1, \dots, X_n$, де X_i, $i = \overline{1, n}$.

$$\mathrm{P}\left(\underline{\theta}_i(x_1 \dots x_n) \le \theta_i \le \overline{\theta}_i(x_1 \dots x_n)\right) = 1 - \alpha, \alpha \le 0.05$$

Розглянемо випадкову величину

$$\frac{\overline{X_n} - \nu}{\frac{\sigma}{\sqrt{n}}}, \text{ де } \overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Випадкова величина $\bar{X}_n$ має нормальний розподіл (дивись наслідок теореми про лінійне перетворення n-вимірного нормально розподіленого випадкового вектору. В нашому випадку $\mathrm{rang}A = 1$)

$$\mathrm{M}\,\overline{X_n} = \nu$$

$$\mathrm{D}\,X_n = \frac{\sigma^2}{n},$$

де $MX = \nu, DX = \sigma^2$.

Таким чином,

$$\frac{\overline{X_n} - \nu}{\frac{\sigma}{\sqrt{n}}}$$

має нормований нормальний розподіл.

Знайдемо таке t_α, для якого виконується наступне рівняння:

$$P\left(\left|\frac{\overline{X_n} - \nu}{\frac{\sigma}{\sqrt{n}}}\right| \le t_\alpha\right) = 1 - \alpha.$$

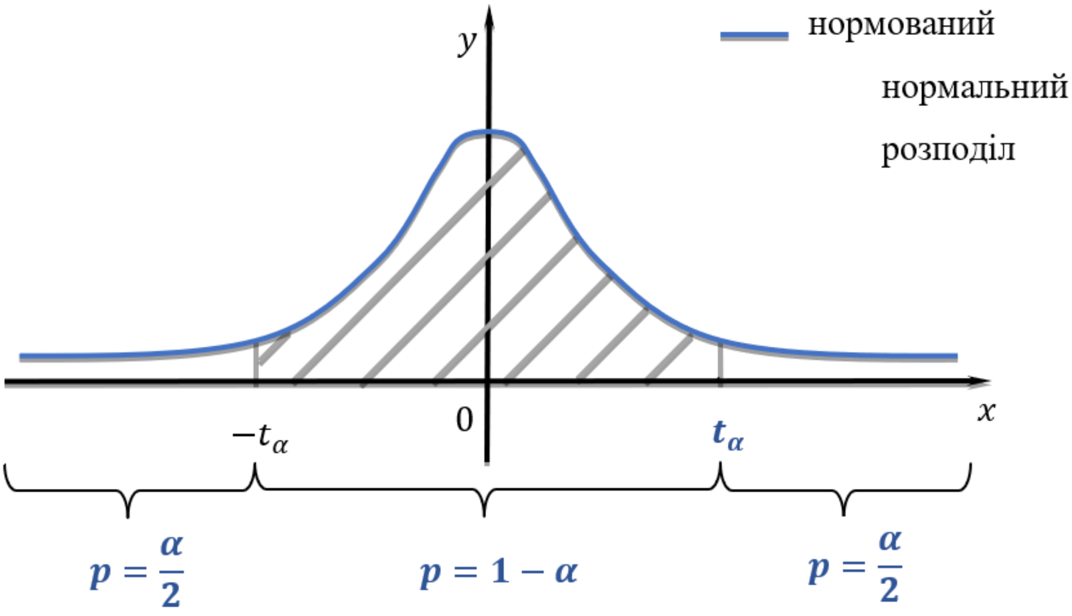


Рисунок 4

$$\Phi_0(t_\alpha) = \frac{1-\alpha}{2}$$

$$t_\alpha = \Phi_0^{-1}\left(\frac{1-\alpha}{2}\right)$$

де Φ_0^{-1} - обернена функція Лапласа від додатного значення аргумента. Знаходимо t_α по таблицям для оберненої функції Лапласа і будуємо інтервальну оцінку.

$$P\left(\overline{X}_n - t_\alpha \frac{\sigma}{\sqrt{n}} \leq \nu \leq \overline{X}_n + t_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$\overline{x}_n - t_\alpha \frac{\sigma}{\sqrt{n}} \leq \nu \leq \overline{x}_n + t_\alpha \frac{\sigma}{\sqrt{n}},$$

де

$$\underline{\theta}_i(X_1 \ldots X_n) = \overline{X}_n - t_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\overline{\theta}_i(X_1 \ldots X_n) = \overline{X}_n + t_\alpha \frac{\sigma}{\sqrt{n}}$$

Наслідок 1. Обов'язково існує ненульова ймовірність помилитись, бо при $\alpha = 0, t_\alpha = +\infty$

Наслідок 2. Це той випадок, коли до випробувань заздалегідь можна знайти кількість випробувань щоб оцінити невідоме мат сподівання із заданою точністю.

Примітка. Інтервальна оцінка має еквівалентну назву **«довірчий інтервал».**

<u>Інтервальна оцінка для математичного сподівання при невідомій дисперсії</u>

$$\frac{\frac{\overline{X}_n - \nu}{\frac{\sigma}{\sqrt{n}}}\sqrt{n-1}}{\sqrt{\frac{nS_n^2}{\sigma^2}}}, \qquad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \;, S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

має розподіл Ст'юдента з (n-1) ступенями свободи.

Пояснення. $\frac{\overline{X}_n - \nu}{\frac{\sigma}{\sqrt{n}}}$ – нормований нормальний розподіл.

$\sqrt{\frac{nS_n^2}{\sigma^2}}$ – розподіл χ^2 з (n-1) ступенями свободи, вони незалежні.

По таблиці для розподілу Ст'юдента знаходимо таке число $t_{\alpha,n-1}$, для якого виконуються:

$$P\left(\left|\frac{\frac{\overline{X}_n - \nu}{\frac{\sigma}{\sqrt{n}}}\sqrt{n-1}}{\sqrt{\frac{nS_n^2}{\sigma^2}}}\right| \le t_{\alpha,n-1}\right) = 1 - \alpha$$

$$P\left(\overline{X}_n - t_{\alpha,n-1}\frac{S_n}{\sqrt{n-1}} \le \nu \le \overline{X}_n + t_{\alpha,n-1}\frac{S_n}{\sqrt{n-1}}\right) = 1 - \alpha$$

Наслідок. Якщо n – невелике, то $t_{\alpha,n-1} > t_\alpha$, тобто довірчий інтеграл більший.

Заздалегідь нам невідома кількість випробувань, щоб отримати оцінку заданої точності.

Інтервальна оцінка для дисперсії

Розглянемо випадковку величину $\frac{nS_n^2}{\sigma^2}$ - має розподіл χ^2 з (n-1) ступенями свободи. За таблицями для розподілу χ^2 знаходимо числа χ_1^2 та χ_2^2, такі що

$$P(\chi^2 \leq \chi_1^2) = \frac{\alpha}{2},$$

$$P(\chi^2 \geq \chi_2^2) = \frac{\alpha}{2}.$$

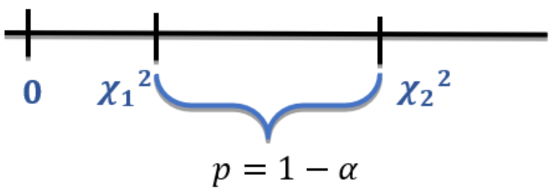


Рисунок 5

Таким чином отримаємо:

$$P(\chi_1^2 \leq \frac{nS_n^2}{\sigma^2} \leq \chi_2^2) = 1 - \alpha \text{ або } P(\frac{nS_n^2}{\chi_2^2} \leq \sigma^2 \leq \frac{nS_n^2}{\chi_1^2}) = 1 - \alpha$$

Тоді з ймовірністю $1 - \alpha$ виконується:

$$\frac{nS_n^2}{\chi_2^2} \leq \sigma^2 \leq \frac{nS_n^2}{\chi_1^2}$$

<u>Інтервальні оцінки для дискретної випадкової величини</u>

Нехай дискретна випадкова величина задається таблицею:

$$X \sim \begin{pmatrix} x_1 & \dots & x_s \\ P_1 & \dots & P_s \end{pmatrix}$$

Проведено n випробувань. При достатньо великому n

$$W_n(X = x_i) = \frac{n_i}{n} \approx P_i,$$

де n_i–кількість випробувань, в кожному з яких настало x_i.n повторів одного випробування будемо вважати як n віртуальних незалежних копій цього випробування. З кожним j-им віртуальним випробуванням зв'яжемо випадкову величину

$$X_j^i = \begin{cases} 1, \text{якщо настало } x_i \\ 0, \text{якщо не настало } x_i \end{cases}.$$

Тоді X_j^i задається таблицею $\begin{Bmatrix} 1 & 0 \\ P_i & 1 - P_i \end{Bmatrix}$.

Випадкова величина $S_n^i = \sum_{j=1}^{n} X_j^i$ має біноміальний розподіл.

При цьому:

$$MS_n^i = np_i,$$

$$DS_n^i = np_i(1 - p_i).$$

Розглянемо випадкову величину $\frac{S_n^i}{n}$ – реалізація котрій є число $\frac{n_i}{n}$:

$$M\frac{S_n^i}{n} = P_{i,}$$

$$D\frac{S_n^i}{n} = \frac{P_i(1 - P_i)}{n}.$$

Тоді для достатньо великих n за теоремою Муавра-Лапласа можна вважати, що випадкова величина:

$$\frac{\frac{S_n^i}{n} - P_i}{\sqrt{\frac{P_i(1 - P_i)}{n}}}$$

має нормований нормальний розподіл.

Тоді по таблицям значень оберненої функції Лапласа знаходимо t_α, для якого виконується:

$$P\left(\left|\frac{\frac{S_n^i}{n} - P_i}{\sqrt{\frac{P_i(1 - P_i)}{n}}}\right| \le t_\alpha\right) = 1 - \alpha$$

$$\frac{S_n^i}{n} - t_\alpha\sqrt{\frac{P_i(1 - P_i)}{n}} \le P_i \le \frac{S_n^i}{n} + t_\alpha\sqrt{\frac{P_i(1 - P_i)}{n}}$$

$$\frac{n_i}{n} - t_\alpha \sqrt{\frac{P_i(1-P_i)}{n}} \le P_i \le \frac{n_i}{n} + t_\alpha \sqrt{\frac{P_i(1-P_i)}{n}}$$

Знайдемо верхню оцінку $P_i(1-P_i)$:

$$\max_{P_i} P_i(1-P_i) = \frac{1}{4}$$

В результаті отримали

$$\frac{n_i}{n} - t_\alpha \frac{1}{2\sqrt{n}} \le P_i \le \frac{n_i}{n} + t_\alpha \frac{1}{2\sqrt{n}}$$

Примітка: можемо заздалегідь знайти необхідну кількість випробувань для знаходження P_i з заданою точністю.

ПРОСТІ, НУЛЬОВІ ТА СКЛАДНІ ГІПОТЕЗИ

Постановка задачі

Не маємо ніякої інформації про розподіл неперервної величини X. Необхідно знайти оцінку її функції щільності.

Методологія розв’язання

1) Проводимо n випробувань та отримуємо вибірку об’єму n.
2) Будуємо гістограму.
3) Порівнюючи гістограму зі всіма відомими графіками функції щільності висуваємо гіпотезу щодо невідомої функції щільності.

Гіпотези можуть бути простими та складними.

Проста гіпотеза однозначно задає невідому функцію $f(x)$.

Проста гіпотеза виникає, коли n – достатньо велике число. В цьому випадку гіпотеза виявляє функцію щільності з точністю до числових значень параметрів. Вважаючи гіпотезу точною, використовуємо метод найбільшої правдоподібності для оцінки невідомих значень числових параметрів. В силу збіжності оцінок, отриманих методом найбільшої правдоподібності до невідомих параметрів, для достатньо великого n

вважаємо їх точними значеннями цих параметрів. Ці оцінки підставляємо в функцію щільності і отримуємо просту гіпотезу.

Складна гіпотеза задає невідому функцію $f(x)$ щільності з точністю до числових значень невідомих параметрів.

Виникає тоді, коли кількість випробувань мале число.

Нульова гіпотеза $\boldsymbol{H_0}$ виникає в наступному випадку: маємо n незалежних випадкових величин, функції щільності яких задані з точністю до значень параметрів (або параметру). Висувається гіпотеза про те, що один з параметрів має одне й те саме значення (наприклад значення математичних сподівань випадкових величин рівні).

МЕТОДОЛОГІЯ ПЕРЕВІРКИ ПРОСТОЇ (НУЛЬОВОЇ) ГІПОТЕЗИ

Над випадковою величною (випадковими величинами при H_0) проводяться випробування та фіксуються результати, котрі вважаються результатом одного випробування над віртуальними незалежними копіями цієї випадкової величини (випадковими величинами при H_0). Кількість віртуальних випадкових величин дорівнює кількості проведених випробувань. Вводиться числова скалярна функція введених віртуальних випадкових величин, котра називається критерієм перевірки гіпотези.

Обов'язкова умова, що накладається на критерій перевірки гіпотези: проста гіпотеза або нульова гіпотеза однозначно задає функцію щільності (або функцію розподілу) критерія перевірки гіпотези. Оскільки критерій перевірки – скалярна випадкова величина (числова скалярна функція випадкових аргументів), то її простір елементарних подій – числова вісь або відрізок числової вісі.

Простір елементарних подій критерія перевірки Ω задається як:

$$D_{\text{доп}} \cup D_{\text{кр}}$$

де $D_{\text{доп}}$ – область допустимих значень, а $D_{\text{кр}}$ – область критичних значень.

При цьому:

$$D_{\text{доп}} \cap D_{\text{кр}} = \{\emptyset\} = V$$

Критичну область беруть із умови:

$$P\big(K(\cdot) \in D_{\text{кр}}\big) = \alpha \leq 0.05$$

Де $K(\cdot)$ - критерій перевірки гіпотези, та $\cdot$ - її випадкові аргументи.

Допустима область – область протилежна критичної області.

Підставив в критерій перевірки гіпотези замість значень віртуальних випадкових аргументів значення, які настали в ході випробування над віртуальними випадковими величинами, отримуємо реалізацію критерія перевірки гіпотези.

Якщо реалізація критерія належіть критичної області, то вважається, що гіпотеза не є вірною (оскільки настала подія з ймовірністю α).

Нехай реалізація критерія належить допустимої області. Тоді продовжуємо дослідження (оскільки гіпотеза може бути невірною, але настали такі події, при яких реалізація критерію попала в допустиму область), тобто статистичні данні не суперечать гіпотезі. Дослідження продовжуються за допомогою інших критеріїв. Статистично доведено, що найбільш «жорстким» критерієм перевірки є критерій χ^2. В процессі перевірки гіпотези можна зробити помилки першого и другого роду. Помилка першого роду – відкинути гіпотезу, що є вірною. Ймовірність помилки першого роду дорівнює α. Помилка другого роду – прийняти гіпотезу, що є невірною. В загальному випадку знайти ймовірність помилки другого роду неможливо. Можна сформулювати методологічний прийом, що зменшує ймовірність помилки другого роду: таким чином обирати критичну область, щоб, чим більше була невірна гіпотеза, яка перевіряється, тим більше була насправді ймовірність попадання критерію в критичну область.

КЛАСИФІКАЦІЯ КРИТИЧНИХ ОБЛАСТЕЙ

Область великих значень критерія задається умовою

$$P(K(\cdot) \geq t_\alpha^1) = \alpha,$$

і використовується тоді, коли відомо, якщо гіпотеза хибна то у критерію є тенденція збільшення своїх значень.

Область малих значень критерія:

$$P(K(\cdot) \leq t_\alpha^2) = \alpha.$$

Коли немає жодної інформації:

А) $P(|K(\cdot)| \geq t_\alpha^{31}) = \alpha$, використовується тоді коли область симетрична відносно осі ординат;

Б) $P\begin{pmatrix} K(\cdot) \leq t_\alpha^{32} \\ K(\cdot) \geq t_\alpha^{33} \end{pmatrix} = \alpha.$

Приклад для перевірки простої гіпотези *H_0*

Маємо дві незалежні випадкові величини X і Y, розподілені нормально:

$$X: n(x, \nu_x, \sigma_x),$$
$$Y: n(y, \nu_y, \sigma_y).$$

Дисперсії – відомі. Математичні сподівання – невідомі. Провели n_1 випробувань над X та n_2 випробування над Y. Отримали наступні вибірки:

$$X: x_1 \ldots x_{n_1},$$
$$Y: y_1 \ldots y_{n_2}.$$

По результатам статистичних даних перевірити нульову гіпотезу H_0, що $\nu_x = \nu_y$

Вважаємо, що отримані вибірки є результатом складного випробування над n_1+n_2-вимірною випадковою величиною, $X_1 \ldots X_{n_1}$, $Y_1 \ldots Y_{n_2}$

Будуємо критерій перевірки простої гіпотези

$$K(X_1 \ldots X_{n_1}, Y_1 \ldots Y_{n_2}) = \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\dfrac{\sigma_x^2}{n_1} + \dfrac{\sigma_y^2}{n_2}}}$$

$$\overline{X}_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$$

$$\overline{Y}_{n_2} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$$

В знаменнику корінь із дисперсії чисельника

В чисельнику випадкова величина, математичне сподівання якої дорівнює v_x-v_y.

Якщо H_0 вірна, то $v_x=v_y$ і критерій перевірки гіпотези має нормований нормальний розподіл.

Вибір критичної області для нормованого нормального розподілу (рисунок 6):

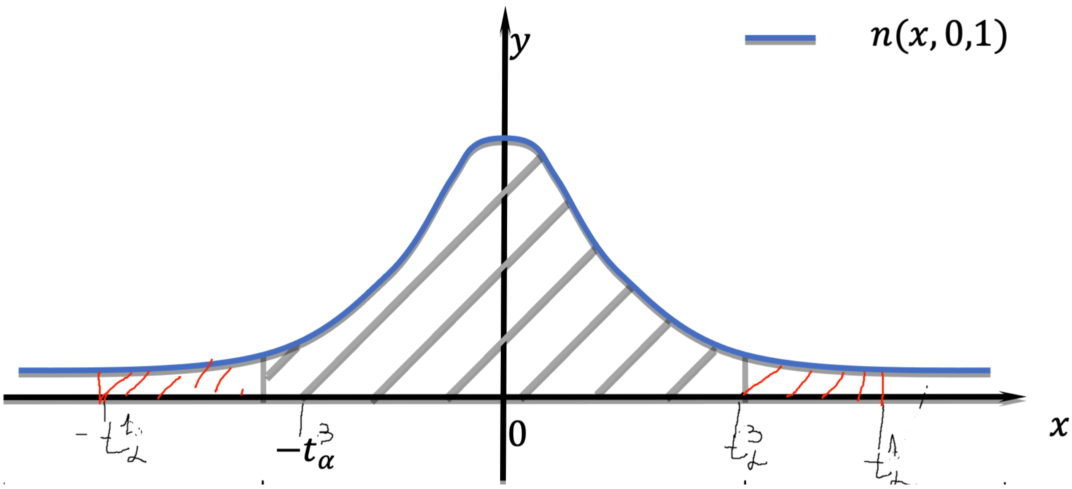


Рисунок 6

$$v_x > v_y$$

$$\Phi_0(t_\alpha^1) = \frac{1}{2} - \alpha$$

$$t_\alpha^1 = \Phi_0^{-1}\left(\frac{1}{2} - \alpha\right)$$

Якщо $v_y > v_x$, то

$$t_\alpha^2 = -t_\alpha^1.$$

Якщо немає жодної інформації, вибирається 3 критична область

$$\Phi_0(t_\alpha^3) = \frac{1-\alpha}{2},$$

$$t_\alpha^3 = {\Phi_0}^{-1}\left(\frac{1-\alpha}{2}\right).$$

КРИТЕРІЙ ЗНАКІВ

Вирішуємо наступну **задачу**: є два незалежних випадкових об'єкти, що діють під впливом спільного фактору. Необхідно перевірити гіпотезу про те, що зміна значення фактора впливає на випадкові події однаково.

Статистична постановка задачі

Маємо $2n$-мірну випадкову величину $\mathrm{X}_1 \dots \mathrm{X}_N X_1^1 \dots X_N^1$, всі компоненти якої незалежні.

Перевіримо нульову гіпотезу про те, що розподіл X_i та X_i^1 при довільному і однаковий.

Результати експериментів:

$$x_1 \ldots x_N,$$

$$x_1^1 \ldots x_N^1.$$

Для спрощення: випадкові величини X_i та X_i^1 неперервні.

Розглянемо випадкову величину $K_N(+)$і її реалізацію $k_N(+)$.

$k_n(+)$ – кількість додатних різниць $x_i - x_i^1$, $i = \overline{1, N}$.

Якщо гіпотеза вірна, то випадкова величина $K_n(+)$має біноміальний розподіл. Це дозволяє сформулювати основну ваду критерію знаків: з того, що додатня і від'ємна різниця має ймовірність наставання 1/2 не випливає, що випадкові величини X_i та X_i^1 мають однаковий розподіл. Майже кожен критерій перевірки гіпотез перевіряє не саму гіпотезу а її наслідок.

Критична область перевірки критерію гіпотези є:

$$P\left(\left|K_N(+)-\frac{N}{2}\right|\geq A\right)=\alpha\leq 0{,}05$$

точне значення числа А знаходиться з розв’язку наступного алгебраїчного рівняння

$$\sum_{m=0}^{\frac{N}{2}-A} C_N^m\left(\frac{1}{2}\right)^N+\sum_{m=\frac{N}{2}+A}^{N} C_N^m\left(\frac{1}{2}\right)^N=\alpha.$$

При достатньо великому N розв’язання задачі можна суттєво спростити:

$$\frac{K_N(+)-\frac{N}{2}}{\sqrt{\frac{N}{4}}},$$

яке наближено має нормований нормальний розподіл (т.Муавра - Лапласа).

$$P\left(\left|\frac{k_N(+)-\frac{N}{2}}{\sqrt{\frac{N}{4}}}\right|>t_\alpha\right)=\alpha$$

$$A=t_\alpha\sqrt{\frac{N}{4}}$$

$$\Phi_0(t_\alpha)=\frac{1-\alpha}{2}$$

$$t_\alpha=\Phi_0^{-1}\left(\frac{1-\alpha}{2}\right)$$

КРИТЕРІЙ χ^2 ДЛЯ ПЕРЕВІРКИ ПРОСТОЇ ГІПОТЕЗИ

Є найбільш «жорсткім» і найбільш використовуваним критерієм для перевірки простої або складної гіпотези.

Постановка задачі. Маємо випадкову величину X, вибірку $x_1, \dots, x_n$ об'єму n, необхідно перевірить просту гіпотезу про функцію щільності $f(x)$.

Алгоритм перевірки гіпотези

Як і при побудові гістограми розглядається відрізок (x_1, x_n), обирається число $a \epsilon [8; 12]$, і цій відрізок б'ється на відрізки Δ_i, в кожний з яких попало a членів варіаційного ряду (члені вибірки).

Знаходимо ймовірність того, що внаслідок випробування:

$$P(X \in \Delta_i) = P_i = \int\limits_{y(i)}^{y(i+1)} f(x)\,dx,$$

де $y(i)$ – лівий кінець i-ого відрізку.

Оскільки X – неперервна, то усі члени ряду – різні, та $x(1) = y(1)$.

$$\int\limits_{y(i)}^{y(i+1)} f(x)\,dx \approx f\left(y_i + \frac{1}{2}\big(y(i+1) - y(i)\big)\right)\big(y(i+1) - y(i)\big).$$

Теоретичний недолік критерію χ^2: критерій перевіряє не вірна або невірна проста гіпотеза, а вірність числа P_i (тобто перевіряє не гіпотезу, а її наслідок).

С кожним Δ_i відрізком зв'язуємо випадкову величину M_i–кількість випробувань з n, в кожному з яких випадкова величина прийняла значення в Δ_i відрізку. Реалізація довільного M_i на кожному відрізку дорівнює a, оскільки так побудовані відрізки Δ_i. Якщо n – достатньо велика кількість випробувань, то випадкові величини M_i мають нормальний розподіл (за теоремою Муавра – Лапласа).

Якщо гіпотеза є вірною (тобто числа P_i є вірними), то критерій χ^2 має розподіл χ^2 з $r-1$ ступенями свободи.

$$\chi^2 = \sum_{i=1}^{r}\left(\frac{M_i - np_i}{\sqrt{np_i}}\right)^2.$$

Примітка. Доведення цього результату наведено у відповідному розділі з *.

Реалізація цього критерію дорівнює:

$$\sum\left(\frac{a - np_i}{\sqrt{np_i}}\right)^2,$$

тобто M_i для кожного i прийняла значення a.

В якості критичної області обирається область великих значень критерію, оскільки якщо числа P_i невірні, то у кожного доданка в сумі у чисельнику вираз є поліномом ступеня n^2, а у знаменнику – ступеня n, то чим більше невірні P_i, тим більше має тенденцію к збільшенню своїх значень критерій. В таблиці для розполілу χ^2 з $r-1$ступеню свободи знаходиться таке α, для якого виконується

$$P(\chi^2 \geq \chi_\alpha{}^2) = \alpha.$$

КРИТЕРІЙ χ^2 ДЛЯ ПЕРЕВІРКИ СКЛАДНОЇ ГІПОТЕЗИ

Постановка задачі. Нехай гіпотеза задає функцію щільності $f(x, \theta_1, \ldots, \theta_c)$, числові значення параметрів $\theta_1, \ldots, \theta_c$ – невідомі.

Вважаючи складну гіпотезу вірною за методом найбільшої правдоподібності знаходимо оцінку невідомих параметрів $\widehat{\theta_i}$ і підставляємо знайдені оцінки в аналітичну функцію щільності, а далі повторюємо послідовність дій, як і для простої гіпотези. Остаточний вираз критерію має вигляд:

$$\chi^2 = \sum_{i=1}^{r}\left(\frac{M_i - n\hat{p}_i}{\sqrt{n\hat{p}_i}}\right).$$

Фішер строго довів, що якщо складна гіпотеза вірна, то критерій χ^2 має $n - c - 1$ ступенів свободи.

КРИТЕРІЙ $n\omega^2$ ДЛЯ ПЕРЕВІРКИ ПРОСТОЇ ГІПОТЕЗИ

Проста гіпотеза задала аналітичний вираз $F(x)$ – функцію розподілу неперервно розподіленої випадкової величини.

Примітка. Насправді гіпотезу можна висунути тільки для функції щільності $f(x)$, далі по ній будується функція $F(x)$ і критерій $n\omega^2$ перевіряє просту гіпотезу для $F(x)$.

$$W_n(x) = \begin{cases} 0, x < x(1) \\ \frac{k}{n}, x(k) < x \leq x(i+1) \\ 1, x > x(n) \end{cases}$$

Примітка. X_i неперервна, тому числа $x(i)$ різні.

При $n \to \infty$, $\forall x$ за посиленим законом великих чисел $W_n(x)$ з ймовірністю одиниця збігається до $F(x)$.

Вводимо випадкову величину ω^2:

$$\omega^2 = \int_{-\infty}^{\infty} \left(F(x) - \omega_n(x)\right)^2 dF(x) = \int_{-\infty}^{x(i)} F(x)^2 dF(x) +$$

$$+ \sum_{k=1}^{n-1} \int_{x(k)}^{x(k+1)} \left[\frac{k}{n} - F(x)\right]^2 dF(x) + \int_{x(n)}^{\infty} \left(1 - F(x)\right)^2 dF(x).$$

Якщо гіпотеза вірна, то інтеграл прямує до 0.

Розглянемо випадкову величину $n\omega^2$.

Теоретично доведено, що при $n \to \infty$ випадкові величини $n\omega^2$ збігаються до випадкової величини з граничним розподілом, що є табульовним.

На практиці цей граничний розподіл використовують при $n \geq 40$.

Критична область – область великих значень критерія. Для заданого α знаходиться за таблицями.

БАГАТОВИМІРНИЙ ДИСПЕРСІЙНИЙ АНАЛІЗ

Багатовимірний дисперсійний аналіз досліджує вплив множини факторів на досліджуваний випадковий об'єкт.

Приклади

Однофакторний дисперсійний аналіз

Є фізичні величини, що вимірюються одним приладом, але різними операторами. Треба дослідити чи впливає 1 фактор – вимірювання різними операторами на систематичну похибку результату вимірювань.

Двофакторний дисперсійний аналіз

Фізичні величини вимірюються різними приладами і різними операторами. Треба дослідити чи впливають два фактори: вимірювання різними приладами та різними операторами на похибку вимірювань.

Однофакторний дисперсійний аналіз

Статистична постановка задачі. Маємо m незалежних випадкових величин $X_1 \dots X_n$ розподілених по нормальному закону з однаковою, але невідомою дисперсією і невідомими математичними сподіваннями. Над кожною випадковою величиною проведено по n випробувань, результати яких представлені у вигляді:

$$\left[\begin{bmatrix} x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \cdots & x_{nn} \end{bmatrix}\right],$$

де x_{ij} – це результат j-го випробування над X_i випадковою величиною.

Практична інтерпретація задачі. Маємо фізичні величини. Один оператор над кожною з них проводить по n вимірювань.

Треба перевірити нульову гіпотезу H_0 про те, що

$$\nu_{\mathrm{X}_1} = \nu_{\mathrm{X}_2} = \cdots = \nu_{\mathrm{X}_n} = \nu$$

(випадкові величини мають однакові математичні сподівання). Фактор, що перевіряється, не впливає на математичне сподівання кожної випадкової величини.

Розв'язання

$\overline{x_i} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$ – середнє арифметичне по i-ій виборці.

$\bar{x} = \frac{1}{n \cdot m}\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}$ – загальне середнє арифметичне.

$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} \overline{x_i}$ – загальне середнє арифметичне є середнім арифметичним середнього арифметичного по виборкам

Має місце тотожність однофакторного дисперсійного аналізу:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_{ij} - \bar{x}\right)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_{ij} - \overline{x_i} + \overline{x_i} - \bar{x}\right)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(\overline{x_i} - \bar{x})^2$$

$$+\sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_{ij} - \bar{x}_i\right)^2;$$

$$Q_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}(\overline{x_i} - \bar{x})^2;$$

$$Q_2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_{ij} - \overline{x_i}\right)^2;$$

$$Q = Q_1 + Q_2.$$

Самим довести цю тотожність.

Якщо гіпотеза, яка перевіряється, є вірною, то $\overline{x_i}$ і $\bar{x}$ оцінюють одне і те саме число v. Якщо перевіряєма гіпотеза є вірною, то Q_1 є невеликою і тим меншою чим більше n.

Показати, що величина Q_2 від перевіряємої гіпотези не залежить (підставити замість x_{ij} числа $x_{ij} + c_i$ і показати, що Q_2 не змінюється).

Методологія перевірки нульової гіпотези ґрунтується на порівнянні величин Q_1 та Q_2.

Використовуємо стандартний прийом математичної статистики і вважаємо, що $m \cdot n$ чисел x_{ij} є результатом 1 віртуального складного випробування над $m \cdot n$ вимірною випадковою величиною $\{X_{ij}\}$ $i = \overline{1, m}$ $j = \overline{1, n}$, де X_{ij} мають нормальний розподіл і незалежні (X_{ij} – є незалежною віртуальною копією X_i)

Використаємо тотожність

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}\right)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{X}_i - \bar{X})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2 .$$

Якщо вірна гіпотеза, то випадкова величина $\frac{Q}{\delta^2}$ має розподіл χ^2 з

$m \cdot n - 1$ ступенями свободи (повна аналогія з $\frac{n \cdot S^2}{\sigma^2}$)

$\frac{Q_1}{\delta^2}$ розподілена по закону χ^2 з m-1 степенями свободи.

Примітка. Випадкові величини $\bar{X}_i$ – незалежні, однаково розподілені за нормальним законом з параметрами: математичне сподівання v, дисперсія $\frac{\sigma^2}{n}$, якщо вірно перевіряємо гіпотези, та

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{X}_i - \bar{X})^2 = n\sum_{i=1}^{m}(\bar{X}_i - \bar{X})^2.$$

$\frac{Q_2}{\delta^2}$ розподілена по закону χ^2 з $m \cdot (n-1)$ ступенями свободи.

Для фіксованого i $\frac{\sum_{j=1}^{m}(X_{ij} - \bar{X}_i)^2}{\sigma^2}$ має розподіл χ^2 з $n-1$ ступенями свободи $\Rightarrow$ $\sum_{i=1}^{m}\left(\frac{\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_i\right)^2}{\sigma^2}\right)$ має розподіл χ^2 з $m \cdot (n-1)$ ступенями свободи, бо для різних i ці випадкові величини незалежні, тому їх сума має розподіл χ^2 з сумарною кількістю степеней свободи.

Покажемо, що випадкові величини Q_1 та Q_2 незалежні. Дійсно, оскільки X_{ij} незалежні та розподілені однаково якщо вірна гіпотеза, що перевіряється, то $\overline{X_1} \dots \overline{X_n}$ незалежні між собою.

Аналогічно незалежні і випадкові величини $n \cdot {S_1}^2, \dots, n \cdot {S_n}^2$, а також $\bar{X}_i \;\; n\bar{S}_j, \;\; \forall i \neq j,$ а також незалежні між собою $n \cdot S_i^2$ та $\overline{X_\iota}, \; i = \overline{1, n}$ (дивись властивості розподілу $\frac{ns^2}{\sigma^2}$).

Q_1та Q_2 є числовими скалярними функціями незалежних аргументів, тому є незалежними (див. відповідну лекцію першої частини курсу).

Критерій перевірки нульової гіпотези є наступним:

$$F = \frac{\frac{Q_1}{\sigma^2} \Big/ m-1}{\frac{Q_2}{\sigma^2} \Big/ m(n-1)}.$$

Це розподіл Фішера з $m-1, m \cdot (n-1)$ ступенями свободи.

Критична область – область великих значень критерію.

МЕТОД НАЙМЕНШИХ КВАДРАТІВ ДЛЯ СКАЛЯРНОГО ДЕТЕРМІНОВАНОГО ВХОДУ

Постановка задачі

Лінія регресії має вигляд:

$$Y(x) = \sum_{j=0}^{r} \theta_j x^j + \varepsilon,$$

де ε – випадкова величина, що має довільний розподіл і її математичне сподівання $M\varepsilon = 0$, дисперсія $\sigma^2 < \infty$ - невідома, $\theta_j, j = \overline{0, r}$ невідомі коефіцієнти, детермінований вхід x.

Примітка. Деякий об’єкт детерміновану вхідну змінну x перетворює за законом:

$$\sum_{j=0}^{r} \theta_j x^j + \varepsilon.$$

На вхід об'єкту подаються значення вхідної змінної $x_i, i = \overline{1,n}$ і вимірюється відповідні значення вихідної змінної $Y(x)$ – числа $y_i,\ i = \overline{1,n}$.

Задача. За результатами випробувань оцінити невідомі значення $\theta_j, j = \overline{0,r}$.

Таким чином, після проведення випробувань отримано $x_i y_i, i = \overline{1,n}$, де $y_i = \sum_{j=0}^{r} \theta_j x_i{}^j + \delta_i$, δ_i – реалізація випадкової величини ε.

Метод найменших квадратів полягає в тому, що знаходження оцінок $\theta_j, j = \overline{0,r}$ зводиться до наступної задачі:

$$\arg\min_{\theta_0 \dots \theta_r} \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{r} \theta_j x^j \right)^2. \quad (1)$$

<u>Якісне пояснення суті методу найменших квадратів</u>

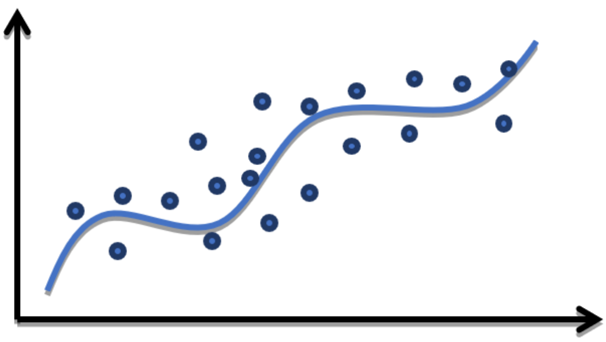


Рисунок 7

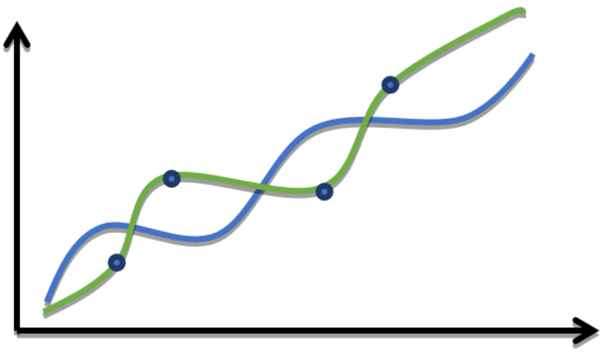


Рисунок 8

Якщо кількість експериментів необмежено велика, то формула (1) повинна дати точні значення коефіцієнтів лінії регресії. В цьому випадку

$$\min_{\theta_0 \dots \theta_r} \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{r} \theta_j x^j \right)^2 = \sum_{i=1}^{n} \delta_i{}^2.$$

Якби в цьому випадку оцінки коефіцієнтів, отримані методом найменших квадратів, не давали точні значення, сума квадратів була б гарантовано більшою, бо на неї накладалася б помилка від того, що лінія регресії знайдена неточно. А ми знаходимо ці оцінки мінімізуючи суму квадратів.

Якщо випробувань небагато, можна отримати погані чи навіть абсурдні результати (рисунок 2). Дійсно, наприклад, якщо кількість випробувань дорівнює кількості невідомих коефіцієнтів, то сума квадратів в (1) дорівнює нулю і лінія регресії, побудована по знайденим коефіцієнтам, проходить через всі точки експерименту.

Розв'язання задачі

$$y_i = \sum_{j=0}^{r} \theta_j x_i{}^j + \delta_i \,, i = \overline{1, n}; \tag{2}$$

$$Y_i = \sum_{j=0}^{r} \theta_j x_i{}^j + \Delta_i \,\,, i = \overline{1, n}. \tag{3}$$

Введемо позначення:

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \theta = \begin{pmatrix} \theta_0 \\ \vdots \\ \theta_r \end{pmatrix}, \delta = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix}, \Delta = \begin{pmatrix} \Delta_1 \\ \vdots \\ \Delta_n \end{pmatrix};$$

$$A = \begin{pmatrix} 1 & x_1 & x_1{}^2 & \dots & x_1{}^r \\ 1 & x_2 & x_2{}^2 & \dots & x_2{}^r \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n{}^2 & \dots & x_n{}^r \end{pmatrix}.$$

Тоді системи (2), (3) можна переписати у векторній формі:

$$y = A\theta + \delta, Y = A\theta + \Delta,$$

а (1):

$$\arg \min_{\theta} (y - A\theta)^T (y - A\theta).$$

Використовуючи рівність $(AB)^T = B^T A^T$ отримуємо:

$$(y - A\theta)^T (y - A\theta) = y^T y - y^T A\theta - \theta^T A^T y + \theta^T A^T A\theta =$$
$$= y^T y - 2\theta^T A^T y + \theta^T A^T A\theta$$

Примітка: $y^T A\theta = \theta^T A^T y$ – бо це скаляри.

Тоді (1): $\arg \min_{\theta} (y^T y - 2y^T A\theta + \theta^T A^T A\theta)$.

Так як оптимальний розв'язок – це внутрішня точка, то необхідно вирішити систему:

$$\frac{\partial (y^T y - 2\theta^T A^T y + \theta^T A^T A\theta)}{\partial \theta_j} = 0, j = \overline{0, r}.$$

Запишемо цю систему рівнянь у векторній формі, використовуючи мнемонічне правило взяття похідної по вектору:

$$-2A^T y + 2A^T A\theta = \bar{0}.$$

Отримали систему рівнянь:

$$A^T y = 2A^T A\theta.$$

Звідки:

$$\hat{\theta} = (A^T A)^{-1} A^T y$$
$$\hat{\theta} = (A^T A)^{-1} A^T Y.$$

В другому вірнянні $\hat{\theta}$ – випадковий вектор, а в першому $\hat{\theta}$ – його реалізація.

ВЛАСТИВОСТІ ОЦІНОК, ОТРИМАНИХ МЕТОДОМ НАЙМЕНШИХ КВАДРАТІВ

1) Незміщеність

Доведення

$$M\Delta = \begin{pmatrix} M\Delta_1 \\ \vdots \\ M\Delta_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Тоді з $Y = A\theta + \Delta$: $MY = A\theta$. Знаходимо математичне сподівання оцінок:

$$M\hat{\theta} = M((A^T A)^{-1} A^T Y) = (A^T A)^{-1} A^T MY = (A^T A)^{-1} A^T A\theta = \theta.$$

Теорема Маркова. Оцінки, отримані методом найменших квадратів, є ефективні в класі всіх лінійних незміщених оцінок.

Доведення

Означення. Оцінка $\theta^* = \begin{pmatrix} {\theta_0}^* \\ \vdots \\ {\theta_r}^* \end{pmatrix}$ є лінійною, якщо

$$\theta^* = UY,$$

де U – довільня числова матриця, розмір U– $(r+1) \times n$

З умови незміщеності:

$$M\theta^* = \theta,$$

чи

$$M(UY) = UMY = UA\theta = \theta.$$

Тоді:

$$UA = E, \tag{4}$$

де E – одинична матриця розміру $(r+1) \times (r+1)$.

Примітка. U і A – прямокутні матриці. З цього випливає, що з матричної рівності (4) матрицю U знайти неможливо.

За умови виконання (4) теорема Маркова означає:

$$D\hat{\theta}_j \le D{\theta^*}_j, j = \overline{0, r}.$$

Доведення

У випадкового вектора Y компоненти – незалежні випадкові величини Y_i з дисперсією $DY_i = \sigma^2, i = \overline{1,n}$. Тоді коваріаційна матриця вектора Y:

$$covY = (cov(Y_i Y_j))^n = \sigma^2 E$$

Знайдемо $cov\theta^*$ - довільна незміщена лінійна векторна оцінка.

$$cov\theta^* = covUY = U\sigma^2 EU^T = \sigma^2 UU^T$$

Примітка. $covAX = AcovXA^T$ (формула виведена в розділі «Лінійні перетворення випадкового вектора компоненти якого мають багатовимірний нормальний розподіл).

Знайдемо:

$$cov\hat{\theta} = cov((A^T A)^{-1} A^T Y) = (A^T A)^{-1} A^T \sigma^2 EA(A^T A)^{-1} =$$
$$= \sigma^2 (A^T A)^{-1} A^T A (A^T A)^{-1} = \sigma^2 (A^T A)^{-1}$$

Примітка. матриця $(A^T A)^{-1}$ є симетричною, тому збігається з $((A^T A)^{-1})^T$.

Для спрощення, далі позначимо : $(A^T A)^{-1} = K$.

Має місце наступна тотожність:

$$UU^T = K + (U - KA^T)(U - KA^T)^T$$

Доведення

$$K + (U - KA^T)(U - KA^T)^T = K + UU^T - UAK - KA^T U^T + KA^T AK =$$
$$= \left| \begin{matrix} UA = E \\ A^T U^T = E \end{matrix} \right| = K + UU^T - K - K + K = UU^T$$

Що й необхідно було довести.

Помножимо вираз на σ^2:

$$\sigma^2 UU^T = \sigma^2 K + \sigma^2 (U - KA^T)(U - KA^T)^T,$$

або

$$cov\theta^* = cov\hat{\theta} + \left| \begin{matrix} \text{Матриця з невід'ємними елементами} \\ \text{на головній діагоналі} \end{matrix} \right|$$

На головній діагоналі матриці $cov\theta^*$ стоять $D\theta^*{}_j$, на головній діагоналі $cov\hat{\theta}$ стоять $D\hat{\theta}_j$. Таким чином: $D\hat{\theta}_j \leq D\theta^*{}_j\ j = \overline{0,r}$, що й необхідно було довести.

Примітка. Отримані результати мають місце також для наступної задачі.

$$Y(\bar{x}) = \theta_0 + \theta_1 x_1 + \cdots + \theta_r x_r + \varepsilon,$$

де $\bar{x} = (x_1, \dots, x_r)^{\mathrm{T}}$ – векторний детермінований вхід, $\varepsilon -$ випадкова величина з нульовим математичним сподіванням та обмеженою дисперсією σ^2. $\theta_0, \dots, \theta_r$ $-$невідомі коефіцієнти.

Результати випробувань мають вигляд:

$$y_i = \theta_0 + \theta_1 x_{1i} + \cdots + \theta_r x_{ri} + \delta_i\ (i = \overline{1,n}).$$

Оцінки невідомих коефіцієнтів знаходяться з виразу

$$\arg\min_{\theta_0 \dots \theta_r} \sum_{i=1}^{n} \left(y_i - \theta_0 - \sum_{j=1}^{r} \theta_j x_{ji}\right)^2.$$

В цьому випадку матриця *A* має вигляд:

$$A = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{r1} \\ 1 & x_{12} & x_{22} & \dots & x_{r2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{rn} \end{pmatrix}.$$

Всі наступні результати аналогічні викладеним вище.

КРИТЕРІЙ Χ-КВАДРАТ ДЛЯ ПЕРЕВІРКИ ІМОВІРНОСТЕЙ

Події $A_1, A_2, \dots A_r$ – несумісні і $\sum_{k=1}^{r} p_k = 1$. До такої постановки приводить, наприклад, перевірка простої гіпотези про закон розподілу випадкової величини X за вибіркою $x_1, x_2, \dots x_n$. Дійсно, розбиваючи область значень випадкової величини X на r відрізків $\bar{\Delta}_i, i = \overline{1,r}$ що не перетинаються, та використовуючи щільність імовірності $\mathrm{f}(x)$, що перевіряємо, знаходимо імовірності p_i наступу події $\mathrm{A_i}$ – попадання у результаті випробування випадкової величини X до відрізку $\bar{\Delta}_i,\ i = \overline{1,r}$.

За вибіркою $x_1, x_2, \dots x_n$ необхідно перевірити гіпотезу про імовірності p_i, $i = \overline{1,r}$.

Для зручності вводимо формально незалежні однаково розподілені величини $\xi_1, \xi_2, \dots \xi_n$, кожна з яких приймає значення $k = \overline{1,r}$ з відповідною імовірністю $p_к$ (з кожним i-м випробуванням зв'язується випадкова величина ξ_i, що приймає значення номера відрізку, в який потрапила випадкова величина X у цьому випробуванні, тобто настала випадкова подія $\mathrm{A_k}$).

Позначимо $\mathrm{v_k}$ (випадкова величина) число величин серед $\xi_1, \dots, . \xi_n$, що приймають значення k.

Очевидно, що $\frac{\mathrm{v}_k}{n}$ є частістю відповідної події A_k (попадання випадкової величини X у k-й інтервал).

Розглянемо випадкову величину

$$\Delta_k = \frac{(\mathrm{v}_k - np_k)}{\sqrt{np_k}} = \sum_{i=1}^{n} \frac{(\xi_{ik} - p_k)}{\sqrt{np_k}},$$

де $\xi_{ik} = 1$, якщо $\xi_i = k$ і $\xi_{ik} = 0$, в іншому випадку ($k = \overline{1,r}$), тобто

$$\mathrm{v}_k = \sum_{i=1}^{n} \xi_{ik}$$

(по усіх випробуваннях для k-го проміжку, $\xi_{ik} = 1$ коли в і-му випробуванні величина X потрапила в k-й інтервал).

Оскільки $\mathrm{v_k}$ розподілена по біноміальному закону $\mathrm{Mv_k = np_k}$, де М - символ мат. сподівання, покажемо, що:

$$M(\mathrm{v}_k - np_k)(\mathrm{v}_l - np_l) = \sum_{i=1}^{n} M\left(\xi_{ik} - p_k\right)\left(\xi_{il} - p_l\right) =$$

$$= n \begin{cases} -p_k p_l, & l \neq k \\ p_k(1-p_k), & l = k \end{cases}.$$

Дійсно:

$$M(\mathrm{v}_k - np_k)(\mathrm{v}_l - np_l) = M\left(\sum_{i=1}^{n}\left(\xi_{ik} - p_k\right)\sum_{j=1}^{n}\left(\xi_{jl} - p_l\right)\right) =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} M\left(\xi_{ik} - p_k\right)\left(\xi_{jl} - p_l\right) = \sum_{i=1}^{n} M\left(\xi_{ik} - p_k\right)\left(\xi_{il} - p_l\right),$$

так як $M\xi_{ik} = 1 \cdot p_k + 0 \cdot (1 - p_k), k = \overline{1,r}$; ξ_{ik}, ξ_{jl} – незалежні при $i \neq j$, отже:

$$M\left(\xi_{ik} - p_k\right)\left(\xi_{jl} - p_l\right) = cov\left(\xi_{ik}, \xi_{jl}\right) = 0 .$$

Нехай $l \neq k$, тоді

$$\sum_{i=1}^{n} M\left(\xi_{ik} - p_k\right)\left(\xi_{il} - p_l\right) = M\xi_{ik}\xi_{il} - p_k p_l - p_k p_l + p_k p_l = -p_k p_l .$$

Оскільки $\xi_{ik}\xi_{jl} = 0$ (в одному випробуванні одночасно події A_k і A_l настати не можуть).

$$M\sum_{i=1}^{n}\left(\xi_{ik} - p_k\right)^2 = \sum_{i=1}^{n} M\left(\xi_{ik} - p_k\right)^2 = \sum_{i=1}^{n}\left[M\xi_{ik}^2 - \left(M\xi_{ik}\right)^2\right] =$$

$$= \sum_{i=1}^{n}(p_k - p_k^2) = np_k(1 - p_k)$$

(тут $M\left(\xi_{ik} - p_k\right)^2 = M\xi_{ik}^2 - (M\xi_{ik})^2$ за властивістю дисперсії).

Розглянемо матрицю коваріації B величин $\Delta_i, \dots, \Delta_r$:

$$B = \left[cov(\Delta_i, \Delta_j)\right]^r$$

При $k \neq 1$:

$$cov(\Delta_k, \Delta_l) = M\big((\Delta_k - M\Delta_k)(\Delta_l - M\Delta_l)\big) = |M\Delta_i = 0, \forall i = \overline{1,r}| = M\Delta_k\Delta_l =$$

$$= M\frac{(\mathrm{v}_k - np_k)(\mathrm{v}_l - np_l)}{n\sqrt{p_k p_l}} = -\sqrt{p_k p_l} ,$$

$$cov(\Delta_k, \Delta_k) \; = M\frac{(\mathrm{v}_k - np_k)^2}{np_k} = q_k \, .$$

А отже матрицю B можна представити у вигляді

$$B = I - \begin{pmatrix} \sqrt{p_1} \\ \sqrt{p_2} \\ \dots \\ \sqrt{p_r} \end{pmatrix} \begin{pmatrix} \sqrt{p_1} & \sqrt{p_2} & \dots & \sqrt{p_r} \end{pmatrix} =$$

$$= \begin{pmatrix} q_1 & -\sqrt{p_1 p_2} & \dots & -\sqrt{p_1 p_r} \\ -\sqrt{p_2 p_1} & q_2 & \dots & -\sqrt{p_2 p_r} \\ \dots & \dots & \dots & \dots \\ -\sqrt{p_r p_1} & -\sqrt{p_r p_2} & \dots & q_r \end{pmatrix},$$

де I – одинична матриця.

Означення. Матриця називається **ортогональною**, якщо сума добутків відповідних елементів будь-яких двох рядків (стовбців) дорівнює нулю. Сума квадратів елементів будь-якого рядка (стовпця) дорівнює одиниці.

Розглянемо ортогональну матрицю $C = \{c_{ij}\}$ розміром $(r \times r)$, перший рядок якої буде $(\sqrt{p_1} \dots \sqrt{p_r})$ (**довести самим, що така матриця існує і вона не єдина**) і лінійне перетворення

$$\eta_j = \sum_{j=1}^{r} c_{jk}\, \Delta_k \, , j = \overline{1, r} \, ;$$

$$\eta = C\Delta \, .$$

При цьому

$$\eta_1 = \sum_{k=1}^{r} \frac{\sqrt{p_k(\upsilon_k - np_k)}}{\sqrt{np_k}} = \frac{\sum_{k=1}^{r} \upsilon_k - n \sum_{k=1}^{r} p_k}{\sqrt{n}} = 0 \, ,$$

оскільки $\sum_{k=1}^{r} \mathrm{v}_k = n, \sum_{k=1}^{r} p_k = 1$.

Легко перевірити (**доводиться від супротивного простою підстановкою)**, що матрицею коваріації випадкових величин $\eta_1, \eta_2, \dots \eta_r$ матриця CBC^{T}, де C^{T} - транспонована матриця і, отже, $C^{\mathrm{T}} = I^4$.

Примітка. Випливає з властивостей матриці. Для ортогональної матриці $C^{\mathrm{T}} = C^{-1}$. Тоді:

$$\mathrm{CBC}^{\mathrm{T}} = I - C\begin{pmatrix}\sqrt{p_1}\\ \cdots \\ \sqrt{p_r}\end{pmatrix}\left(\sqrt{p_1} \dots \sqrt{p_r}\right)C^{\mathrm{T}} = I - \begin{pmatrix}1\\0\\ \dots\\0\end{pmatrix}\begin{pmatrix}1 & 0 & \dots & 0\end{pmatrix} =$$

$$= \begin{pmatrix}0 & 0 & \dots & 0\\ 0 & 1 & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & 0 & 0 & 1\end{pmatrix} \qquad (4)$$

Примітка. На головній діагоналі стоять дисперсії.

Дана рівність випливає з властивостей ортогональної матриці.

Як бачимо

$$M\eta_k\eta_l = \begin{cases} 1, k = l \neq 1 \\ 0, k \neq l, k = l = 1\end{cases}.$$

За властивістю ортогональної матриці C маємо:

$$\sum_{k=1}^{r} \Delta_k^2 = \sum_{k=2}^{r} \eta_k^2 ,$$

оскільки лінійне перетворення відповідає лише повороту осей і, отже, залишає евклідову норму перетвореного вектора без змін.

Таким чином випадкова величина

$$\chi^2 = \sum_{k=1}^{r} \frac{(\mathrm{v}_k - np_k)^2}{np_k} = \sum_{k=2}^{r} \eta_k^2$$

$(\eta_1 = 0)$ є сумою $r - 1$ некорельованих величин $\eta_2, \dots \eta_r$ з нульовим математичним сподіванням ($\forall M\Delta_k = 0$) та одиничною дисперсією.

Випадкові величини η_j можна представити у вигляді нормованої суми незалежних, однаково розподілених доданків

$$\begin{pmatrix} \eta_2 \\ \dots \\ \eta_r \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} \eta_{i2} \\ \dots \\ \eta_{ir} \end{pmatrix},$$

де $\eta_{ij} = \sum_{k=1}^{r} \frac{c_{jk}\left(\xi_{ik} - p_k\right)}{\sqrt{np_k}}$ $(\eta_j = \sum_{k=1}^{r} c_{jk}\Delta_k$, $\Delta_k = \sum_{i=1}^{n} \frac{(\xi_{ik} - p_k)}{\sqrt{np_k}})$.

Далі доводиться (використовується векторний аналог центральної граничної теореми), що компоненти цього вектору мають $n -$ вимірний нормальний розподіл, $cov(\eta_i, \eta_j) = 0$, випадкові величини $\eta_j -$ незалежні і кожна мас одновимірний нормальний розподіл.

Отже, випадкова величина $\chi^2 = \sum_{k=2}^{r} \eta_k^2$ має, у випадку коли $p_1, \dots p_r$ справжні імовірності, розподіл χ - квадрат з $r - 1$ степенем свободи.

Гіпотезу, яку перевіряємо заперечує реалізація критерію, обчислена наоснові вибірки та, яка задовольняє нерівність:

$$\widehat{\chi^2} = \sum \frac{(\widehat{\nu_k} - np_k)^2}{np_k} \geq x_a,$$

де $\widehat{\nu_k}$- реалізація випадкової величини ν_k, обчислена за вибіркою; $x_a -$ відповідна квантиль χ - квадрат розподілу з $r - 1$ степенем свободи, $P(\chi^2 \geq x_a) = a$.

ВИБІР МІЖ ДВОМА КОНКУРУЮЧИМИ ГІПОТЕЗАМИ. ПОСЛІДОВНИЙ АНАЛІЗ.

Розглянемо приклад випробувань Бернуллі відносно ймовірності "успіху", в яких маються 2 гіпотези: $p = p_0$ (гіпотеза H_0) і $p = p_1$ (гіпотеза H_1). Припустимо, H_0 – основна гіпотеза, в той час як H_1 – конкуруюча. Можливу фактичну відмінність в значеннях H_0 і H_1 проілюструємо на задачі виявлення «сигналу» при наявності шумів.

Припустимо, що "сигнал" може бути 2-х типів: $\theta = 1, \theta = 0$ (відсутність сигналу). На кожному кроці $k = \overline{1, n}$ незалежно з ймовірностями δ відбувається вплив на сигнал шумів і "спостерігач" замість справжнього сигналу отримує послідовність, яку можна інтерпретувати як вибірку із послідовності незалежних випадкових величин $\xi_1, \xi_2, \ldots$ з можливими значеннями 0 чи 1 та відповідним розподілом ймовірностей $P(0/H_1) = P(1/H_0) = \delta$, $P\left(\frac{0}{H_1}\right) = P\left(\frac{1}{H_0}\right) = 1 - \delta$, що відповідають вказаним гіпотезам H_0- сигнал відсутній ($\theta = 0$) і H_1- сигнал є ($\theta = 1$). В усіх випробуваннях ($k = 1,2, \ldots$) сигнал або передається, або ні.

Очевидно, що тут імовірністю "успіху" ($\xi_k = 1$) є відповідно $p_0 = \delta$ і $p_1 = 1 - \delta$. Прийняття H_1, якщо вірною є гіпотеза H_0, означає невірний сигнал, і очевидні випадки, коли наслідки цього настільки серйозні, що необхідно максимально зменшити імовірність появи такого випадку. Для цього можна, наприклад, вимагати, щоб імовірність прийняття невірної гіпотези була не більше заданого і достатньо малого α.

Повернемося до загальної схеми випробувань Бернуллі з довільними p_0 і p_1. Визначимо, яке рішення можна прийняти за вибіркою $x_1, \ldots, x_n$. Критерій, що відкидає H_0 (і тим самим приймає альтернативне рішення H_1), сконструювати таким чином, розглядаючи вибірку $(x_1, \ldots, x_n)$ як точку в n -вимірному просторі R^n. Позначимо $P_{\xi_1,\ldots,\xi_n}(\bullet/H)$ розподіл ймовірностей при вказаній гіпотезі $H = H_0, H_1$. Задамо достатньо мале $\alpha > 0$, оберемо критичну область $S \subseteq R^n$, що задовольняє вимогу

$$P_{\xi_1,\dots,\xi_n}(S/H_0) \le \alpha \tag{5}$$

і за отриманою вибіркою $(x_1, \dots, x_n)$ будемо відкидати H_0 лише тоді, коли $(x_1, \dots, x_n) \in S$. При цьому можуть бути помилки. Помилкою першого роду є прийняття H_1, коли вірна H_0, на основі вимоги (1), імовірність цього не більше α; помилкою другого роду є прийняття H_0, коли насправді вірна альтернативно гіпотеза H_1. Очевидно, ймовірність помилки другого роду є

$$P_{\xi_1,\dots,\xi_n}\left(\frac{\bar{S}}{H_1}\right) = 1 - P_{\xi_1,\dots,\xi_n}(S/H_1) = \beta \tag{6}$$

- імовірність того, що вибіркова точка $(x_1, \dots, x_n)$ потрапить в додаток критичної області S в R^n. Бажано обирати критичну область $S \subseteq R^n$ так, щоб при заданій імовірності α відкинути вірну основну гіпотезу, тоді ймовірність β - помилки другого роду, була б мінімальною. Це можна зробити на основі відношення правдоподібності:

$$L(x_1, \dots, x_n) = \frac{P_{\xi_1,\dots\xi_n}\left(x_1, \dots, \frac{x_n}{H_1}\right)}{P_{\xi_1,\dots\xi_n}\left(x_1, \dots, \frac{x_n}{H_0}\right)}, \tag{7}$$

де у чисельнику і знаменнику стоять ймовірності появи вибірки (ці події незалежні), тому логарифм відношення залишається у вигляді

$$log\, L\,(x_1, \dots, x_n) = \sum_{k=1}^{n} log \frac{P(x_k/H_1)}{P(x_k/H_0)}$$

оскільки випадкові величини $\xi_1, \dots, \xi_n$ незалежні і кожна з них приймає значення 0 чи 1 в залежності від того, чи досягнуто "успіху" у випробуванні чи ні.

Таким чином, $log\, L\,(x_1, \dots, x_n)$ представлено у вигляді суми значень незалежних однаково розподілених величин, де $P(\bullet/H)$ - розподіл ймовірностей окремо взятих величин $\xi_1, \dots, \xi_n$при відповідній гіпотезі $H = H_0, H_1$. Іншими словами $P(\xi_i/H), \forall i = \overline{1, n}$ є числові скалярні функції від випадкових величин ξ_i, і, відповідно, є також незалежними випадковими величинами.

Для критичної області виду $S = \{x: log\, L\,(x_1, \dots, x_n) \geq y_\alpha\}$ це дає можливість обрати границю критичної області на основі центральної граничної теореми, взявши замість точного розподілу ймовірностей випадкові величини $log\, L\,(x_1, \dots, x_n) = \sum_{k=1}^{n} log \frac{P(\xi_k/H_1)}{P(\xi_k/H_0)}$ з незалежними $\xi_1, \dots, \xi_n$ відповідний наближений нормальний розподіл. Розглянемо мат. сподівання

$$\alpha = M\, log \frac{P(\xi_k/H_1)}{P(\xi_k/H_0)}$$

величин, значення яких представлені в (7).

Означення. Строго вгнутою називається функція $f(x)$, яка задовольняє умові:

$f(\alpha_1 x_1 + \alpha_2 x_2) > \alpha_1 f(x_1) + \alpha_2 f(x_2), \alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_1 + \alpha_2 = 1.$

Легко довести, що функція $f(x) = log\, x$- строго вгнута.

Тоді при гіпотезі H_0 отримаємо:

$a = p_0\, log \frac{p_1}{p_0} + q_0\, log \frac{q_1}{q_0} < log(p_1 + q_1) = 0$,

де $\alpha_1 = p_0, \alpha_2 = q_0, x_1 = \frac{p_1}{p_0}, x_2 = \frac{q_1}{q_0}$.

Дійсно, випадкова величина $log \frac{P(x_k/H_1)}{P(x_k/H_0)}$, якщо вірна гіпотеза H_0, задається таблицею:

Таблиця 1

$log \frac{P(1/H_1)}{P(1/H_0)} = log \frac{p_1}{p_0}$	$log \frac{P(0/H_1)}{P(0/H_0)} = log \frac{q_1}{q_0}$
p_0	q_0

Аналогічно при гіпотезі H_1 отримаємо:

$$a = p_1\, log \frac{p_1}{p_0} + q_1\, log \frac{q_1}{q_0} = -\left(p_1\, log \frac{p_0}{p_1} + q_1\, log \frac{q_0}{q_1}\right) > 0\ ,$$

оскільки $p_1\, log\frac{p_0}{p_1} + q_1\, log\frac{q_0}{q_1} < log(p_0 + q_0) = 0$.

Отже, на основі закону великих чисел відносно (7) можна стверджувати, що при $n \to \infty$ з імовірністю 1

$$log\, L\,(x_1, \dots, x_n) \to -\infty, \tag{8}$$

якщо вірна гіпотеза H_0, і

$$log\, L\,(x_1, \dots, x_n) \to +\infty, \tag{9}$$

якщо вірна H_1 (зі збіжності $\frac{1}{n} log\, L\,(x_1, \dots, x_n)$ до a випливає збіжність $log\, L\,(x_1, \dots, x_n)$ до na).

При $n \to \infty$ послідовність (7) обмежена зверху з імовірністю 1, якщо вірна гіпотеза H_0.

Тому для монотонно спадних при $A \to \infty$ подій A^c, де A^c означає перший вихід розглядуваної послідовності $log\, L\,(x_1, \dots, x_n)$ вище рівня A: $\lim\limits_{A\to\infty} P(A^c)$.

Звідси випливає існування такої верхньої границі A, що для заданого $\alpha > 0 P(A^c) \le \alpha$.

Аналогічно, якщо вірна гіпотеза H_1, то при $n \to \infty$ послідовність (7) обмежена знизу з імовірністю 1, та існує така нижня границя B, що події B^c- перший вихід цієї послідовності нижче рівня B мають імовірність $P(B^c) \le \beta$ для заданого $\beta > 0$. В той самий час послідовність (7) з імовірністю 1 при гіпотезі H_0 знаходиться нижче рівня B, а при гіпотезі H_1- вище рівня A.

Таким чином, залежно від того, яка гіпотеза дійсно має місце, величина $log\, L\,(x_1, \dots, x_n)$ з імовірністю 1 при деякому n потрапляє вище A або нижче B (при довільних скінчених A і B).

Можна запропонувати такий послідовний критерій вибору альтернативи H_0 чи H_1 у міру послідовного накопичення даних $x_1, \dots, x_n$, ($n = 1,2, \dots$), а саме, рішення приймається, як тільки послідовність (7) виходить за установлені границі, точніше, при першому $n = 1,2,\ldots$ коли $log\, L\,(x_1, \ldots, x_n) \geq A$ чи $log\, L\,(x_1, \ldots, x_n) \leq B$. У першому випадку приймається гіпотеза H_1, а у другому - H_0. Для належних A і B, очевидно, при такому рішенні, помилки першого та другого роду не будуть перевищувати заданих α і β.

При будь-якому $n = 1,2,\ldots$ в усіх точках $(x_1, \ldots, x_n)$, де при першому виході за верхню границю A приймається гіпотеза H_1, маємо:

$$P_{\xi_1,\ldots,\xi_n}\left(x_1, \ldots, \frac{x_n}{H_1}\right) \geq cP_{\xi_1,\ldots,\xi_n}\left(x_1, \ldots, \frac{x_n}{H_0}\right), \tag{10}$$

де $log\, c = A$.

Дійсно з $log\, L\,(x_1, \ldots, x_n) = log\frac{P_{\xi_1,\ldots,\xi_n}(x_1,\ldots,x_n/H_1)}{P_{\xi_1,\ldots,\xi_n}(x_1,\ldots,x_n/H_0)} > A$ випливає (10).

Просумувавши по всім $x_1, \ldots, x_n$ зі змінними $n = 1,2, \ldots$, зліва отримаємо імовірність $1 - \beta$ прийняти вірну гіпотезу H_1, а справа – імовірність α помножену на c, тобто:

$$1 - \beta \geq c\alpha\,,$$

оскільки сумування відбувається по всім вибіркам $x_1, \ldots, x_n$, $(n = 1,2,\ldots)$, для яких $log\, L\,(x_1, \ldots, x_n) \geq A$, і, відповідно, зліва сумуються імовірності появи цих вибірок при гіпотезі H_1 (сумарна імовірність дорівнює $1 - \beta$), а справа – імовірності, помножені на c появи цих вибірок при гіпотезі H_0 (сумарна імовірність дорівнює $c\alpha$).

Отримаємо:

$$A \leq log\frac{1-\beta}{\alpha}. \tag{11}$$

Зокрема, $A \leq log\frac{1}{\alpha} = -\,log\,\alpha$ для будь-якого β, і відповідно, для будь-якої нижньої границі B.

Аналогічно:

$$B \geq log\frac{\beta}{1-\alpha} \geq log\,\beta \tag{12}$$

для будь-якого α (для будь-якої верхньої границі A).

Очевидно, якщо підіймаючи при фіксованому B верхню границю A, ми лише зменшуємо імовірність виходу з неї, тому, замінивши A на $A' = -\log \alpha$, отримаємо відповідне $\alpha' \leq \alpha$. Зрозуміло, що імовірність виходу за B раніше, ніж за A, при заміні A на $A' \geq A$, може тільки збільшуватись. Але, якщо взяти $B' = \log \beta$ замість B, для відповідного β' отримаємо:

$$\log \beta' \leq B', \beta' \leq \beta.$$

Таким чином, границі A' та B' гарантують, що імовірності помилок першого та другого роду будуть:

$$\alpha' \leq \alpha, \beta' \leq \beta.$$

БАЄСІВСЬКИЙ ПІДХІД ДО ПЕРЕВІРКИ ГІПОТЕЗ ТА ОЦІНКИ ПАРАМЕТРІВ

Розглянемо схему кошиків з білими та чорними кулями, будемо вважати число білих куль $M = \theta$ – невідомим параметром з можливими значеннями $\theta = 1, \dots, N-1$ (виключаючи крайні значення $\theta = 0$ і $\theta = N$ – загальна кількість куль у кошику). Нехай з кошика виконується вибірка з поверненням, результати якої представлені як вибірка $x_1, \dots, x_n$ з незалежних однаково розподілених випадкових величин, де $\xi_k = 1$вказує на появу на k-му кроці білої кулі, $\xi_k = 0$ – чорної. При байєсівському підході до питання про невідоме θ вихідним є формальне введення апріорних ймовірностей $\pi(\theta)$ того, що θ – істинне значення параметра (використовується Баєсівський варіант теорії ймовірностей, де числа вважаються випадковими величинами, що мають відповідний розподіл).

Задання ймовірностей $\sum \pi(\theta) = 1$ може бути основане на дуже суб'єктивних міркуваннях.

За вибіркою $x_1, \ldots, x_n$ визначаються відповідні апостеріорні імовірності

$$\pi(\theta/x_1, \ldots, x_n) = \frac{\pi(\theta) P_{\xi_1,\ldots,\xi_n}\left(x_1, \ldots, \frac{x_n}{\theta}\right)}{\sum_\theta \pi(\theta) P_{\xi_1,\ldots,\xi_{ii}}\left(x_1, \ldots, \frac{x_n}{\theta}\right)} - \qquad (13)$$

оцінка імовірності появи білої кулі.

Дійсно, якщо значення θ вважати гіпотезою $B_i, i = \overline{1, N-1}$, а появу вибірки $x_1, \ldots, x_n$ – подією A, то за формулою Байєса отримаємо

$$P(B_i/A) = \frac{P(B_i)P(A/B_i)}{\sum_i P(B_i)P(A/B_i)}$$

В (13) $P_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n/\theta)$ є сукупній розподіл випадкових величин $\xi_1, \ldots, \xi_n$, при значенні параметра θ

У розглядуваній схемі кошиків імовірність успіху (появи білої кулі) $P(\theta) = \frac{\theta}{N}$. Візьмемо в якості оцінки за вибіркою значення

$$\hat{P} = \sum_\theta P(\theta)\pi(\theta/x_1, \ldots, x_n) \qquad (14)$$

Покажемо, що апостеріорні імовірності (13) мають наступну чудову властивість: з імовірністю 1 при $n \to \infty$

$$\pi(\theta/x_1, \ldots, x_n) \to 1 \qquad (15)$$

для істинного значення θ при будь-якій апріорній імовірності $\pi(\theta) > 0$. Для цього зручно позначити θ_0 істинне значення параметру, вважаючи θ вільними змінними. Як відомо відношення правдоподібності

$$L(x_1, \ldots, x_n/\theta) = \frac{P_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n/\theta)}{P_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n/\theta_0)}$$

$$(H_1 = \theta, H_0 = \theta_0)$$

для $\theta \neq \theta_0$таке, що при $n \to \infty$ з імовірністю 1 $L(x_1, \ldots, x_n/\theta) \to 0$, якщо θ_0істинне значення кількості білих куль. ($log\, L\,(x_1, \ldots, x_n/\theta) \to -\infty$, якщо вірна гіпотеза $H_0 = \theta_0$ [див. (8)],) і відповідно,

$$\pi(\theta_0/x_1, \ldots, x_n) =$$

$$= \frac{\pi(\theta_0)P_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n/\theta_0)}{\pi(\theta_0)P_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n/\theta_0) + \sum_{\theta\neq\theta_0}\pi(\theta)P_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n/\theta_0)} =$$

$$= \frac{\pi(\theta_0)\frac{P_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n/\theta_0)}{P_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n/\theta_0)}}{\pi(\theta_0)\frac{P_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n/\theta_0)}{P_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n/\theta_0)} + \sum_{\theta\neq\theta_0}\pi(\theta)\frac{P_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n/\theta)}{P_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n/\theta_0)}} =$$

$$= \frac{\pi(\theta_0)}{\pi(\theta_0) + \sum_{\theta\neq\theta_0}\pi(\theta)L(x_1,\dots,x_n/\theta)} \to \frac{\pi(\theta_0)}{\pi(\theta_0)} = 1$$

що й потрібно було довести.

Як наслідок із співвідношення (15) для запропонованої в (14) оцінки $\hat{P}$ невідомої імовірності $P(\theta_0) = P$ отримаємо, що з імовірністю 1 при $n \to \infty$

$$\hat{P} \to P \tag{16}$$

Візьмемо в якості оцінки $\hat{\theta}$ невідомого параметру таке значення $\theta = \hat{\theta}$, для якого апостеріорна імовірність (13) найбільша:

$$\pi(\hat{\theta}/x_1 \dots x_n) = \max_{\theta} \pi(\theta/x_1,\dots,x_n) \tag{17}$$

Очевидно, $\hat{\theta} \to \theta_0$ при$n \to \infty$, і більше того, з імовірністю 1 для достатньо великих n маємо $\hat{\theta} = \theta_0$ оскільки згідно (15) для $\theta = \theta_0$з імовірністю 1 для достатньо великих n $\pi(\theta_0/x_1,\dots,x_n) > \frac{1}{2}$, а для решти

$\theta \neq \theta_0$ – лише їх скінченне число $\pi(\theta_0/x_1,\dots,x_n) < \frac{1}{2}$ і оцінка найбільшої правдоподібності, що визначається з умови (17) буде $\hat{\theta} = \theta_0$. Зазначена властивість оцінок називається спроможністю.

НАЙБІЛЬШ ПОТУЖНИЙ КРИТЕРІЙ

Нехай $x_1,\dots,x_n$- вибірка із сукупності випадкових величин $\xi_1,\dots,\xi_n$, спільний розподіл імовірності яких залежить від невідомого параметра $\theta \in \Theta$ і є дві гіпотези: основна H_0 про те, $\theta = \theta_0$,та конкуруюча H_1, про те, що $\theta_1 \neq \theta_0$. Для зручності позначень введемо випадкову величину $\xi = (\xi_1,\dots,\xi_n)$ в R^n. Усі наступні доведення будуть основані на співвідношенні правдоподібності $L(x/\theta)$ такій функції змінної $x = (x_1,\dots,x_n) \in R^n$, залежної від параметра $\theta \in \Theta$, що для величини $\phi(\xi)$ - функції від випадкової величини $\xi \in R^n$- при будь-якому θ відповідне математичне сподівання може бути розраховано за формулою

$$M\phi(\xi) = M_0\phi(\xi)L\left(\frac{\xi}{\theta}\right), \tag{18}$$

де M_0- символ математичного сподівання , взятий при $\theta \in \theta_0$.

Приклад 1. Нехай ξ - багатовимірна дискретна випадкова величина з можливими значеннями $x \in X$із деякої зліченої множини $X \subseteq R^n$,і її розподіл ймовірностей

$$P(x/\theta) = P_\xi(x/\theta), x \in X$$

для $\theta \in \theta_0$ строго додатного. Покажемо, що тоді

$$L(x/\theta) = \frac{P_\xi(x/\theta)}{P_\xi(x/\theta_0)}\,. \tag{19}$$

Дійсно, у цьому випадку

$$M_0\phi(\xi)L(\xi/\theta) = \sum_i \phi(x_i)\frac{P_\xi(x_i/\theta)}{P_\xi(x_i/\theta_0)}P_\xi(x_i/\theta_0) =$$

$$= \sum_i \phi(x_i)\,P_\xi(x_i/\theta) = M\phi(\xi)$$

Приклад 2. Нехай ξ - багатовимірна випадкова величина із сукупною щільністю ймовірностей

$$f(x/\theta) = f_\xi(x/\theta), x \in R^n,$$

яка для $\theta = \theta_0$ строго додатня. Покажемо, що

$$L\left(\frac{x}{\theta}\right) = \frac{f_\xi\left(\frac{x}{\theta}\right)}{f_\xi\left(\frac{x}{\theta_0}\right)}\,. \tag{20}$$

Дійсно, у цьому випадку

$$M_0\phi(\xi)L\left(\frac{\xi}{\theta}\right) = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\phi(x)\frac{f_\xi\left(\frac{x}{\theta}\right)}{f_\xi\left(\frac{x}{\theta_0}\right)}f_\xi\left(\frac{x}{\theta_0}\right)dx_1\ldots dx_n = M\phi(\xi).$$

Припустимо, вибір між основною гіпотезою H_0 та конкуруючою простою $(\theta = \theta_1)$ H_1 здійснюється на основі критерію з так званою критичною областю $S \subseteq R^n$. Основна гіпотеза H_0 відкидається лише тоді, коли вибіркова точка $x = (x_1, \dots, x_n)$ потрапляє в S (при попаданні xв доповнення $S^c = R^n \setminus S$ гіпотеза H_0 приймається). У такому рішенні, звичайно, можуть бути помилки. Наприклад, помилка першого роду – відкинути гіпотезу H_0 при умові, коли вона вірна, має імовірність

$$a = M_0 1_S(\xi),$$

де $1_s(x)$ означає індикаторну функцію, $1_s(x) = 1$ при $x \in S$ і $1_S(x) = 0$ в іншому випадку. Дійсно

$$a = M_0 1_S(\xi) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} 1_S\,(x) f_\xi\left(\frac{x}{\theta_0}\right) dx_1 \dots dx_n =$$

$$= \int \dots \int f_\xi(x/\theta_0) dx_1 \dots dx_n = P(\xi \in S/\theta = \theta_0)$$

(викладки для дискретного випадку проводяться аналогічно).

Помилка другого роду – прийнята H_0 при умові, коли насправді вірна гіпотеза H_1 має імовірність

$$\beta = P\left(\xi \in \frac{S^c}{\theta} = \theta_1\right) = M 1_s(\xi) = M[1_{R^n}(\xi) - 1_S(\xi)] = 1 - M 1_S(\xi) =$$

$$= 1 - M_0 1_S(\xi) L(\xi/\theta_1)$$

на основі (1.15). Тут M математичне сподівання, яке береться при $\theta = \theta_1$. При заданому a бажано мати критерій, що забезпечує найменше β. Величину $1 - \beta$ називають потужністю критерію. Покажемо, як можна обрати найбільш потужний критерій. Візьмемо критичну область виду

$$S_0 = \{x : L(x/\theta_1) > c\},$$

тобто S_0, містить усі x, для яких $L(x/\theta_1) > c$, де c підібрана так, щоб $M_0 1_{s_0}(\xi) = a_0$ для належного a_0.

При використанні цього критерію на практиці необхідно для кожного конкретного вигляду розподілу, в залежності від параметра θ_1, знайти аналітично (або чисельними методами) числове значення константи c.

Візьмемо також будь-який інший критерій з критичною областю S, для якої помилка першого роду $a \leq a_0$. Нехай β і β_0 – імовірності помилок другого роду у критеріях S та S_0. Тоді

$$\beta_0 = 1 - M_0 1_{S_0}(\xi) L\left(\frac{\xi}{\theta_1}\right), \beta = 1 - M_0 1_S(\xi) L\left(\frac{\xi}{\theta_1}\right);$$

$$\begin{aligned}
&\beta - \beta_0 = M_0 1_{S_0}(\xi) L(\xi/\theta_1) - M_0 1_S(\xi) L(\xi/\theta_1) = \\
&= \int_{S_0}\ldots\int 1_{S_0}(x) L(x/\theta_1) f_\xi(x/\theta_0) dx_1 \ldots dx_n - \int_{S}\ldots\int 1_S(x) L(x/\theta_1) f_\xi(x/\theta_0) dx_1 \ldots dx_n = \\
&= \int_{S_0\setminus\Delta}\ldots\int 1(x) L(x/\theta_1) f_\xi(x/\theta_0) dx_1 \ldots dx_n - \int_{S\setminus\Delta}\ldots\int 1(x) L(x/\theta_1) f_\xi(x/\theta_0) dx_1 \ldots dx_n = \\
&= M_0 1_{S_0\setminus\Delta}(\xi) L(\xi/\theta_1) - M_0 1_{S\setminus\Delta}(\xi) L(\xi/\theta_1)
\end{aligned}$$

де $\Delta = S \cap S_0$ - спільна частина областей S та S_0, $S_0\setminus\Delta, S_0\setminus\Delta$ - доповнення до неї. Для дискретного випадку викладки проводяться аналогічно.

За умовою, помилка першого роду для S є $a \leq a_0, a = M_0 1_S(\xi)$. Тому

$$M_0 1_{S\setminus\Delta}(\xi) = M_0 1_S(\xi) - M_0 1_\Delta(\xi) = a - M_0 1_\Delta(\xi) \leq$$

$$\leq a_0 - M_0 1_\Delta(\xi) = M_0 1_{S_0}(\xi) - M_0 1_\Delta(\xi) = M_0 1_{S_0\setminus\Delta}(\xi),$$

тобто $M_0 1_{S\setminus\Delta}(\xi) \leq M_0 1_{S_0\setminus\Delta}(\xi)$.

Використовуючи ту обставину, що $L(x/\theta_1) \leq c$ при $x \in S\setminus\Delta$ (тобто $x \notin S_0$) і $L(x/\theta_1) > c$ при $x \in S_0\setminus\Delta, (x \notin S)$, отримаємо нерівності

$$M_0 1_{S\setminus\Delta}(\xi) L\left(\frac{\xi}{\theta_1}\right) \leq c M_0 1_{S\setminus\Delta}(\xi) \leq M_0 1_{S_0\setminus\Delta}(\xi) \leq M_0 1_{S_0\setminus\Delta}(\xi) L\left(\frac{\xi}{\theta_1}\right),$$

звідки $\beta - \beta_0 \geq 0$. Отже потужність критерію є найбільшою серед усіх критеріїв S, у яких помилка першого роду не перевищує a_0. Оптимальний у цьому розумінні критерій S_0 називають критерієм Неймана-Пірсона.

ІНФОРМАЦІЯ ФІШЕРА І НЕРІВНІСТЬ ДЛЯ СЕРЕДНЬОКВАДРАТИЧНИХ ПОМИЛОК

Розглянемо задачу оцінювання невідомого параметра $\theta \in \Theta$ в розподілі ймовірностей випадкової величини $\xi = (\xi_1, \dots, \xi_n)$ за її спостерігаючим значенням – вибірковій точці $x = (x_1, \dots, x_n) \in R^n$. Оцінкою θ будемо називати дійсну функцію $\varphi = \varphi(x)$ від вибіркової точки $x \in R^n$, розглядаючи одночасно $\varphi = \varphi(\xi)$ як випадкову величину. Вкажемо нижню границю для дисперсії помилок в оцінці $\hat{\theta} = \varphi(x)$ за вибіркою $x = (x_1, \dots, x_n)$. Для дійсного $\theta \in \Theta$ відповідно (18):

$$M\varphi(\xi) = M_0\varphi(\xi)L(\xi/\theta) = a(\theta)$$

(вважається, що $\varphi(x)$ може бути зміщеною оцінкою, тобто не завжди $M\varphi(\xi) = \theta$). Зрозуміло, що $M1 = M_0L(\xi/\theta) = 1$.

Дійсно, математичне сподівання константи дорівнює самій константі.

Припустимо, що ці вирази як функції розглядуваної дійсної компоненти θ можна продиференціювати під знаком математичного сподівання (операції математичного сподівання та диференціювання по θ можна поміняти місцями як лінійні операції диференціювання та інтегрування). Тоді при відповідному θ:

$$\frac{\partial}{\partial\theta}M_0\varphi(\xi)L(\xi/\theta) = M_0\varphi(\xi)\frac{\partial L(\xi/\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}a(\theta), \tag{21}$$

$$\frac{\partial}{\partial\theta}M_1 = \frac{\partial}{\partial\theta}M_0L(\xi/\theta) = M_0\frac{\partial L(\xi/\theta)}{\partial\theta} = 0$$

Домноживши друге з цих рівнянь на $a(\theta)$ та віднявши його потім від першого, отримаємо

$$M_0[\varphi(\xi) - a(\theta)]\frac{\partial}{\partial\theta}L(\xi/\theta) = \frac{\partial}{\partial\theta}a(\theta),$$

$$a(\theta)M_0\frac{\partial}{\partial\theta}L(\xi/\theta) = M_0a(\theta)\frac{\partial}{\partial\theta}L(\xi/\theta),$$

оскільки $a(\theta)$ для операції математичного сподівання при $\theta = \theta_0$ є константою.

Припустимо тепер, що похідна

$$\frac{\partial}{\partial\theta}\log L(x/\theta) = \frac{1}{L(x/\theta)}\frac{\partial}{\partial\theta}L(x/\theta)$$

дозволяє отримати величину, що інтегрується у квадраті, тобто скінчені вирази

$$M\left[\frac{\partial}{\partial\theta}\log L(\xi/\theta)\right]^2, \forall\theta \in \Theta.$$

Нехай:

$$\eta_1 = [\varphi(\xi) - a(\theta)]\sqrt{L(\xi/\theta)},$$

$$\eta_2 = \frac{\partial}{\partial\theta}\log L(\xi/\theta)\sqrt{L(\xi/\theta)}.$$

Згідно нерівності Коші-Буняковського до величин η_1 і η_2 можна записати

$$|M_0\eta_1\eta_2| \leq \sqrt{M_0{\eta_1}^2 M_0{\eta_2}^2}.$$

Тоді:

$$M_0\eta_1\eta_2 = M_0[\varphi(\xi) - a(\theta)]\frac{L(\xi/\theta)}{L(\xi/\theta)}\frac{\partial L(\xi/\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}a(\theta),$$

$$M_0{\eta_1}^2 M_0{\eta_2}^2 \geq \left[\frac{\partial}{\partial\theta}a(\theta)\right]^2.$$

Згідно (18):

$$M_0{\eta_1}^2 = M_0[\varphi(\xi) - a(\theta)]^2 L(\xi/\theta) = M[\varphi(\xi) - a(\theta)]^2 = D\varphi(\xi)$$

(тут ми припускаємо, що θ – істинне значення невідомого параметра),

$$M_0{\eta_2}^2 = M_0\left[\frac{\partial}{\partial\theta}\log L(\xi/\theta)\right]^2 L(\xi/\theta) = M\left[\frac{\partial}{\partial\theta}\log L(\xi/\theta)\right]^2.$$

Тоді ми перейдемо до наступної нерівності Рао-Крамера для дисперсії $D\varphi(\xi)$ оцінки φ:

$$D\varphi(\xi) \geq \left[\frac{\partial}{\partial\theta}a(\theta)\right]^2 I(\theta)^{-1}, \qquad (22)$$

в якій $I(\theta) = M\left[\frac{\partial}{\partial\theta}\log L(\xi/\theta)\right]^2$ – називають інформацією Фішера.

Так для незміщених оцінок $M\varphi(\xi) = a(\theta) = \theta$ нерівність (22) дає $D\varphi(\xi) \geq I(\theta)^{-1}$.

Оцінку називають **ефективною**, якщо $D\varphi(\xi)$ досягає вказаної в (22) нижньої границі $D\varphi(\xi) = I(\theta)^{-1}$.

Покажемо, що $I(\theta)$ не залежить від θ_0. Розглянемо $I(\theta)$ для неперервного випадку (для дискретного розглянуті самостійно).

$$I(\theta) = M\left[\frac{\partial}{\partial\theta}\log L(\xi/\theta)\right]^2 =$$

$$= \int_{-\infty}^{\infty} \dots \int \frac{1}{L^2(x/\theta)}\left[\frac{\partial}{\partial\theta}L(x/\theta)\right]^2 f_\xi\,(x/\theta)dx_1 \dots dx_n =$$

$$= \int_{-\infty}^{\infty} \dots \int \frac{f_\xi^2(x/\theta_0)}{{f_\xi}^2(x/\theta)} \frac{\left[\frac{\partial}{\partial\theta}f_\xi(x/\theta)\right]^2}{f_\xi^2(x/\theta_0)} f_\xi\,(x/\theta)dx_1 \dots dx_n =$$

$$= \int_{-\infty}^{\infty} \dots \int \frac{\left[\frac{\partial}{\partial\theta}f_\xi(x/\theta)\right]^2}{f_\xi(x/\theta_0)} dx_1 \dots dx_n$$

від θ_0 не залежить ($f_\xi(x/\theta_0)$ не залежить від θ, а тому винесено з-під знаку диференціювання). Згідно другій рівності (21):

$$M\frac{\partial}{\partial\theta}\log L(\xi/\theta) = M_0\frac{\partial}{\partial\theta}\log L(\xi/\theta)L(\xi/\theta) = M_0\frac{\partial}{\partial\theta}L(\xi/\theta) = 0,$$

і, відповідно, $I(\theta) = D\frac{\partial}{\partial\theta}\log L(\xi/\theta)$, дисперсія випадкової величини $\frac{\partial}{\partial\theta}\log L(\xi/\theta)$.

Приклад 1. Дискретний випадок

$$I(\theta) = M\left[\frac{\partial}{\partial\theta}\log L(\xi/\theta)\right]^2 = M\left[\frac{\partial}{\partial\theta}\big(\log P_\xi(x/\theta) - \log P_\xi(x/\theta_0)\big)\right]^2 =$$

$$= M\left[\frac{\partial}{\partial\theta}\log P_\xi(x/\theta)\right]^2$$

Приклад 2. Неперервний випадок

$$I(\theta) = M\left[\frac{\partial}{\partial\theta}\log L(\xi/\theta)\right]^2 = M\left[\frac{\partial}{\partial\theta}\big(\log f_\xi(x/\theta) - \log f_\xi(x/\theta_0)\big)\right]^2$$

$$= M\left[\frac{\partial}{\partial\theta}\log f_\xi(x/\theta)\right]^2$$

ЛІТЕРАТУРА

1. Гихман И.И., Скороход А. В., Ядренко М.И. Теория вероятностей и математическая статистика. –К.: Вища школа, 1979. -408с.
2. Розанов Ю.А. Теория вероятностей, случайные процессы и математическая статистика. –М.: Наука, 1985. -320с.
3. Смирнов Н.В., Дунин – Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. –М.: Наука, 1965. -512с.
4. Худсон Д. Статистика для физиков. –М.: Мир, 1967. -242с.
5. Крамер Г. Математические методы статистики. –М.: Мир, 1975. -648с.
6. Андерсон Т. Введение в многомерный статистический анализ. –М.: Физматгиз, 1963. -500с.
7. Коваленко И.Н., Филиппова А.А. Теория вероятностей и математическая статистика. –М.: Высшая школа, 1982. -256с.
8. Вентцель Е.С., Овчаров Л.А. Теория слаучайных процессов и её инженерные приложения. –М.: Наука, 1991. -384с.
9. Свешников А.А. Прикладные методы теории случайных функция. –М.:наука, 1968. -463с.
10. Анисимов В.В., Закусило О.К., Донченко В.С. Элементы теории саммового обслуживания и асимптотического анализа систем. –К.: Вища школа, 1987. -248с.
11. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA в среде WINDOWS. –М.: Финансы и статистика, 1999. -384с.
12. Конышева Л.К., Назаров Д.М. Основы теории нечётких множеств. –СПб.: Питер, 2011. -192с.

13. Павлов О.А. Лекції по курсу “Математична статистика та потоки подій”. – К.: НТУУ “КПІ”, 2000. – 39 с.
14. Згуровський М.З. , Павлов А.А. Принятие решений в сетевых системах с ограничеными ресурсами. – К.: Наукова думка, 2010. – 574 с.
15. Королюк В.С., Турбин А.Ф. Полумарковские процессы и их приложения. –К.: Наукова думка, 1976. – 184 с.
16. Павлов А.А. и др.. Основы системного анализа и проектирования АСУ. _К.: Выща школа, 1991 с. – 367 с.
17. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Изд. Дом “Вильямс”, 2007. – 912 с.
18. Павлов О.А., Гавриленко О.В., Рибачук Л.В. Навчальний посібник з дисципліни «Теорія ймовірностей, імовірнісні процеси та математична статистика». Курс лекцій. Частина 1 для студентів спеціальності 126 «Інформаційні системи та технології» [Електронний ресурс]:навчальний посібник. – Київ: КПІ, 2021. – 154 с.