Перевірка статистичних гіпотез

Лекція № 9

9.1. Гіпотези та критерії.
9.2. Підходи до порівняння критеріїв

Вступ

Маючи вибірку, ми можемо висунути кілька взаємовиключаючих гіпотез по теоретичному розподілу, одній з яких слід віддати перевагу над іншими. Завдання пов'язані з вибором однієї з кількох гіпотез вирішується побудовою статистичного критерію. Як правило, за вибіркою кінцевого обсягу безпомилкових висновків про розподіл зроблено бути не може, тому завжди є небезпека вибрати невірну гіпотезу.

Вступ

Так, кидаючи монету, можна висувати припущення про справжню ймовірність того, що випаде герб. Припустимо, є дві гіпотези: ймовірність або знаходиться в межах 0,45-0,55, або ні. Отримавши після ста кидків рівно 51 герб, ми напевно виберемо першу гіпотезу. Однак є ненульові шанси на те, що і при $p = 0{,}3$ випаде 51 герб: вибираючи першу гіпотезу, ми можемо помилитися.
Навпаки, отримавши 33 герба, ми швидше за все віддамо перевагу другій гіпотезі. І знову не виключена можливість, що настільки далеке від половини число гербів є просто результат випадковості, а монета насправді симетрична.

Нехай дана вибірка $\xi_1, \xi_2, ..., \xi_n$ з розподілу $\mathfrak{F}$. Ми будемо вважати вибірку набором незалежних випадкових величин з одним і тим же розподілом, хоча в ряді завдань і ці припущення потребують перевірки. Тоді однакова розподіленість або незалежність спостережень не передбачається.

Означення 1

Гіпотезою H називається будь-яке припущення про розподіл спостережень:

$$H = \{\mathfrak{F} = \mathfrak{F}_1\} \quad \text{або} \quad \{\mathfrak{F} = \mathbb{F}\},$$

де $\mathbb{F}$-деяка підмножина в множині всіх розподілів. Гіпотеза H називається простою, якщо вона вказує на єдине розподіл: $\mathfrak{F} = \mathfrak{F}_1$. Інакше H називається складною: $\mathfrak{F} = \mathbb{F}$.
Якщо гіпотез всього дві, то одну з них прийнято називати основною, а іншу – альтернативою, або відхиленням від основної гіпотези.

9.1. Гіпотези та критерії

Перерахуємо типові завдання перевірки гіпотез.

- Вибір з декількох простих гіпотез: є $H_1 = \{\mathfrak{F} = \mathfrak{F}_1\}, \ldots, H_k = \{\mathfrak{F} = \mathfrak{F}_k\}$, і інші припущення неможливі.
- Проста основна гіпотеза і складна альтернатива: $H_1 = \{\mathfrak{F} = \mathfrak{F}_1\}, H_2 = \{\mathfrak{F} \neq \mathfrak{F}_1\}$.
- Складна основна гіпотеза і складна альтернатива: $H_1 = \{F \in \mathbb{F}\}, H_2 = \{F \notin \mathbb{F}\}$. Наприклад, гіпотеза про нормальність: $H_1 = \{\text{розподіл } \mathfrak{F} \text{ є нормальним}\}$ при альтернативній $H_2 = \{H_1 \text{ є невірною}\}$.
- Гіпотеза однорідності: є кілька вибірок; основна гіпотеза полягає в тому, що ці вибірки витягнуті з одного розподілу.

- Гіпотеза незалежності: по вибірці $(X_1, Y_1), ..., (X_n, Y_n)$ з n незалежних спостережень пари випадкових величин перевіряється гіпотеза $H_1 = \{X_i \text{ і } Y_i \text{ незалежні}\}$ при альтернативі $H_2 = \{H_1 \text{ є невірною}\}$. Обидві гіпотези є складними.
- Гіпотеза випадковості. В експерименті спостерігаються n випадкових величин $X_1, ..., X_n$ і перевіряється складна гіпотеза $H_1 = \{X_1, ..., X_n \text{ незалежні і однаково розподілені}\}$. Це завдання ставлять, наприклад, при перевірці якості генератора випадкових чисел.

Нехай дана вибірка $X_1, ..., X_n$, щодо розподілу якоївисунуті гіпотези $H_1, ..., H_k$.

Означення 2

Критерієм $\sigma = \sigma(X_1, ..., X_n)$ називається вимірне відображення

$$\sigma : R_n \to \{H_1, ..., H_k\}$$

з множини всіх можливих значень вибірки у множині гіпотез. Вимірність розуміється в звичайному сенсі: $\{\omega \mid \sigma(X_1, ..., X_n) = H_i\}$ є подією при будь-якому $i = 1, ..., k$.

Означення 3

Кажуть, що сталася помилка i-го роду крітерію σ, якщо критерій відкинув вірну гіпотезу H_i. Ймовірністю помилки i-го роду критерію σ називається число

$$\alpha_i(\sigma) = P_{H_i}(\sigma(\overline{X}) \neq H_i).$$

Розглянемо детально випадок, коли є дві прості гіпотези про розподіл спостережень $H_1 = \{\mathfrak{F} = \mathfrak{F}_1\}$ та $H_2 = \{\mathfrak{F} = \mathfrak{F}_2\}$ Тоді будь-який критерій $\sigma(\overline{X})$ приймає не більше двох значень.
Це означає, що область $\mathbb{R}^n$ ділиться на дві частини $\mathbb{R}^n = S \bigcup (\mathbb{R}^2 \setminus S)$ так, що

$$\sigma(\overline{X}) = \begin{cases} H_1, & \overline{X} \in \mathbb{R}^2 \setminus S; \\ H_2, & \overline{X} \in S, \end{cases}$$

Область S, в якій приймається друга (альтернативна) гіпотеза, називається критичною областю.

Означення 4

Імовірність помилки першого роду $\alpha_1 = \alpha_1(\sigma)$ інакше називають розміром або критичним рівнем критерію σ:

$$\alpha_1 = \alpha_1(\sigma) = P_{H_1}(\sigma(\overline{X}) \neq= H_1) = P_{H_1}(\sigma(\overline{X}) = H_2) = P_{H_1}(\overline{X} \in S).$$

Потужністю критерію σ називають величину $1 - \alpha_2$, де $\alpha_2 = \alpha_2(\sigma)$ - ймовірність помилки другого роду критерію σ. Потужність критерію дорівнює

$$1 - \alpha_2(\sigma) = 1 - P_{H_2}(\sigma(\overline{X}) \neq H_2) = P_{H_2}(\sigma(\overline{X}) = H_2) = P_{H_2}(\overline{X} \in S).$$

Зауважимо, що ймовірності помилок першого і другого роду обчислюються при різних припущеннях про розподіл (вірна H_1 або вірна H_2),тому ніякими фіксованими співвідношеннями виду $\alpha_1 \equiv 1 - \alpha_2$ ці помилки не пов'язані.

Як порівнювати критерії?

Зрозуміло, критерій тим краще, ніж меньше ймовірності його помилок. Але якщо порівнювати критерії за двома ймовірностями помилок одночасно, щоб

$$\alpha_i(\sigma_1) \leq \alpha_i(\sigma_2)) \quad \text{при } i = 1, 2,$$

то занадто багато критерії виявляться непорівнянними.

Наприклад, розглянемо два крайніх випадки, коли критерій, незалежно від вибірки, завжди приймає одну і ту ж гіпотезу.

Приклад 1

Нехай критерій $\sigma(\overline{X}) \equiv H_1$ завжди вибирає першу гіпотезу. Тоді

$$\alpha_1 = P_{H_1}(\sigma = H_2) = 0, \;\; \alpha_2 = P_{H_2}(\sigma = H_1) = 1.$$

Навпаки: нехай критерій $\sigma(\overline{X}) \equiv H_2$ завжди вибирає другу гіпотезу.

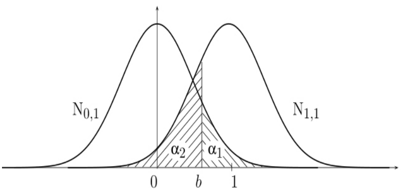
Приклад 2

Задано вибірку обсягу $n = 1$ з нормального розподілу $N(a, 1)$ і дві прості гіпотези $H_1 = a = 0$ і $H_2 = a = 1$. Розглянемо при деякому $b \in \mathbb{R}$ наступний критерій:

$$\sigma(X_1) = \begin{cases} H_1, & X_1 \le b; \\ H_2, & X_1 > b. \end{cases}$$

Зобразимо на графіку відповідні гіпотезам щільності та розподілимо ймовірності помилок першого і другого роду критерію σ

$$\alpha_1 = P_{H_1}(X_1 > b), \;\; \alpha_2 = P_{H_2}(X_1 \le b).$$



Бачимо, що з ростом числа b ймовірність помилки першого роду α_1 зменшується, але ймовірність помилки другого роду α_2 зростає.

9.2. Підходи до порівняння критеріїв

Отже, приклади 1 і 2 показують загальну тенденцію:

при спробі зменшити одну з ймовірностей помилок інша, як правило, збільшується.

Так, якщо зменшувати $\alpha_1 = P_{H_1}(\overline{X} \in S)$ за рахунок звуження критичної області S, то одночасно буде рости ймовірність помилки другого роду й зменшуватися потужність критерію

$$1 - \alpha_2 = P_{H_2}(\overline{X} \in S).$$

9.2. Підходи до порівняння критеріїв.

Перерахуємо загальноприйняті підходи до порівняння критеріїв. Обмежимося для простоти завданням перевірки двох простих гипотез. Нехай маємо критерії σ і ρ з вірогідністю помилок першого і другого роду $\alpha_1(\sigma), \alpha_1(\sigma)$ і $\alpha_1(\rho), \alpha_2(\rho)$.

Мінімакстний підхід. Кажуть, що критерій σ не гірше критерію ρ в сенсі минимаксного підходу, якщо

$$\max\{\alpha_1(\sigma), \alpha_2(\sigma)\} \le \max\{\alpha_1(\rho), \alpha_2(\rho)\}.$$

Означення 5

Критерій σ називається мінімаксним, якщо він не гірше за всіх інших критеріїв в сенсі минимаксного підходу. Інакше кажучи, мінімаксний критерій має найменшу «найбільшу помилку» $\max\{\alpha_1(\sigma), \alpha_2(\sigma)\}$ серед всіх інших критеріїв.

9.2. Підходи до порівняння критеріїв

Байєсівський підхід. Цей підхід застосовують у наступних випадках:

- якщо відомо апріорі, що з імовірністю r вірна гіпотеза H_1, а з ймовірністю $s = 1 - r$ – гіпотеза H_2;
- якщо задана лінійна «функція втрат»: втрати від помилкового рішення на рівні r, якщо відбувається помилка першого роду, і рівні s, якщо другого. Тут $r + s$ вже не обов'язково дорівнює 1, але втрати можна звести до одиниці нормування $r' = \frac{r}{(r+s)}4$ і $s' = \frac{s}{(r+s)}$.

Нехай апріорні ймовірності або втрати r і s задані. Кажуть що критерій σ не гірше критерію ρ в сенсі байєсівського підходу, якщо

$$r\alpha_1(\sigma) + s\alpha 2(\sigma) \leq r\alpha_1(\rho) + s\alpha_1(\rho).$$

Означення 6

Критерій σ називають Байєсовим, якщо він не гірше за всіх інших критеріїв в сенсі байєсівського підходу.

Вибір найбільш потужного критерію. Помилки першого і другого роду зазвичай нерівноправні. Тому виникає бажання контролювати одну з помилок.
Наприклад, зафіксувати ймовірність помилки першого роде на досить низькому (безпечному) рівні і розглядати тільки критерії з такою ж або ще меншою ймовірністю цієї помилки. Серед них найкращим слід визнати критерій с найменшою вірогідністю помилки другого роду.
Введемо при $\varepsilon \in [0, 1]$ клас критеріїв $K_\varepsilon = \left\{\sigma(\overline{X}) | \alpha_1(\sigma) \leq \varepsilon\right\}$.

Означення 7

Критерій $\sigma \in K_\varepsilon$ називають найбільш потужним критерієм (НПК) розміру ε, якщо $\alpha_2(\sigma_0) \leq \alpha_2(\sigma)$ для будь-якого іншого критерію $\sigma \in K_\varepsilon$.

Дякуємо за увагу!